Package ‘Lock5Data’

October 12, 2022

Title  Datasets for “Statistics: UnLocking the Power of Data"
Version 3.0.0
Maintainer Robin Lock <rlock@stlawu.edu>
Description Datasets for the third edition of “Statistics: Unlocking the Power of Data” by Lock^5
Includes version of datasets from earlier editions.
Depends R (>= 3.5.0)
License GPL-2
Encoding UTF-8
LazyData true
RoxygenNote 7.1.1
NeedsCompilation no
Author Robin Lock [aut, cre]
Repository CRAN
Date/Publication 2021-07-22 22:40:10 UTC

R topics documented:

Lock5Data-package ................................................................. 5
ACS ......................................................................................... 6
ACS2010 .................................................................................. 7
AllCountries ........................................................................... 8
AllCountries1e ........................................................................ 9
AllCountries2e ........................................................................ 10
APMultipleChoice ................................................................. 11
April14Temps ........................................................................ 11
April14Temps1e ...................................................................... 12
April14Temps2e ...................................................................... 12
BaseballHits1e ....................................................................... 13
BaseballHits2014 ................................................................... 14
BaseballHits2019 ................................................................... 14
BaseballSalaries2015 ............................................................. 15
BaseballSalaries2019 ............................................................. 16

1
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>BaseballTimes</td>
<td>16</td>
</tr>
<tr>
<td>Benford</td>
<td>17</td>
</tr>
<tr>
<td>BikeCommute</td>
<td>18</td>
</tr>
<tr>
<td>BodyFat</td>
<td>18</td>
</tr>
<tr>
<td>BodyTemp50</td>
<td>19</td>
</tr>
<tr>
<td>BootAtlantaCorr</td>
<td>20</td>
</tr>
<tr>
<td>CaffeineTaps</td>
<td>20</td>
</tr>
<tr>
<td>CAOSEExam</td>
<td>21</td>
</tr>
<tr>
<td>CarbonDioxide</td>
<td>21</td>
</tr>
<tr>
<td>CarbonDioxide2e</td>
<td>22</td>
</tr>
<tr>
<td>CarDepreciation</td>
<td>23</td>
</tr>
<tr>
<td>Cars2015</td>
<td>23</td>
</tr>
<tr>
<td>Cars2020</td>
<td>24</td>
</tr>
<tr>
<td>Cereal</td>
<td>25</td>
</tr>
<tr>
<td>CityTemps</td>
<td>26</td>
</tr>
<tr>
<td>CityTemps2e</td>
<td>26</td>
</tr>
<tr>
<td>CocaineTreatment</td>
<td>27</td>
</tr>
<tr>
<td>ColaCalcium</td>
<td>28</td>
</tr>
<tr>
<td>CollegeScores</td>
<td>28</td>
</tr>
<tr>
<td>CollegeScores2yr</td>
<td>30</td>
</tr>
<tr>
<td>CollegeScores4yr</td>
<td>31</td>
</tr>
<tr>
<td>CommuteAtlanta</td>
<td>33</td>
</tr>
<tr>
<td>CommuteStLouis</td>
<td>33</td>
</tr>
<tr>
<td>CompassionateRats</td>
<td>34</td>
</tr>
<tr>
<td>CricketChirps</td>
<td>35</td>
</tr>
<tr>
<td>DDS</td>
<td>35</td>
</tr>
<tr>
<td>DecemberFlights</td>
<td>36</td>
</tr>
<tr>
<td>DecemberFlights2e</td>
<td>36</td>
</tr>
<tr>
<td>DietDepression</td>
<td>37</td>
</tr>
<tr>
<td>Digits</td>
<td>38</td>
</tr>
<tr>
<td>DogOwner</td>
<td>38</td>
</tr>
<tr>
<td>DrugResistance</td>
<td>39</td>
</tr>
<tr>
<td>EducationLiteracy</td>
<td>40</td>
</tr>
<tr>
<td>EducationLiteracy2e</td>
<td>40</td>
</tr>
<tr>
<td>ElectionMargin</td>
<td>41</td>
</tr>
<tr>
<td>EmployedACS</td>
<td>41</td>
</tr>
<tr>
<td>EmployedACS2010</td>
<td>42</td>
</tr>
<tr>
<td>ExerciseHours</td>
<td>43</td>
</tr>
<tr>
<td>FacebookFriends</td>
<td>44</td>
</tr>
<tr>
<td>FatMice18</td>
<td>44</td>
</tr>
<tr>
<td>FireAnts</td>
<td>45</td>
</tr>
<tr>
<td>FisherIris</td>
<td>46</td>
</tr>
<tr>
<td>FishGills12</td>
<td>46</td>
</tr>
<tr>
<td>FishGills3</td>
<td>47</td>
</tr>
<tr>
<td>Flight179</td>
<td>47</td>
</tr>
<tr>
<td>Flight433</td>
<td>48</td>
</tr>
<tr>
<td>Flight433_2e</td>
<td>48</td>
</tr>
<tr>
<td>FloridaLakes</td>
<td>49</td>
</tr>
</tbody>
</table>
topics documented:

- FootballBrain .................................................. 50
- ForestFires ...................................................... 50
- GeneticDiversity ................................................. 51
- GlobalInternet2010 ............................................ 52
- GlobalInternet2019 ............................................. 53
- GolfRound ......................................................... 53
- GPAbySex .......................................................... 54
- GSWarriors2016 .................................................. 54
- GSWarriors2019 .................................................. 56
- HappyPlanetIndex ............................................... 57
- HeatCognition ...................................................... 58
- HeightData ........................................................ 59
- HockeyPenalties2011 ......................................... 60
- HockeyPenalties2019 ......................................... 61
- HollywoodMovies ................................................. 61
- HollywoodMovies2011 ......................................... 62
- HollywoodMovies2013 ......................................... 63
- HomesForSale ..................................................... 64
- HomesForSale2e .................................................. 65
- HomesForSaleCA ................................................. 65
- HomesForSaleCA2e ............................................. 66
- HomesForSaleCanton .......................................... 66
- HomesForSaleCanton2e ....................................... 67
- HomesForSaleNY ................................................ 67
- HomesForSaleNY2e ............................................. 68
- HomingPigeons .................................................... 69
- Honeybee ............................................................ 69
- HoneybeeCircuits ................................................ 70
- HoneybeeWaggle ................................................ 70
- HotDogs1e .......................................................... 71
- HotDogs2015 ....................................................... 72
- HotDogs2019 ....................................................... 72
- HouseStarts2015 ................................................ 73
- HouseStarts2018 ................................................ 73
- HumanTears25 ..................................................... 74
- HumanTears50 ..................................................... 75
- Hurricanes2014 .................................................. 75
- Hurricanes2018 .................................................. 76
- ICUAdmissions ..................................................... 76
- ImmuneTea ........................................................ 77
- InkjetPrinters ...................................................... 78
- LifeExpectancyVehicles ....................................... 78
- LifeExpectancyVehicles1e ..................................... 79
- LifeExpectancyVehicles2e ..................................... 80
- LightatNight ....................................................... 80
- LightatNight4Weeks ............................................. 81
- LightatNight8Weeks ............................................. 82
- MalevolentUniformsNFL ....................................... 83
R topics documented:

- MalevolentUniformsNHL
- MammalLongevity
- ManhattanApartments
- ManhattanApartments2011
- MarriageAges
- MastersGolf
- MateChoice
- MentalMuscle
- MiamiHeat
- MindsetMatters
- MustangPrice
- NBAPlayers2011
- NBAPlayers2015
- NBAPlayers2019
- NBAStandings2011
- NBAStandings2016
- NBAStandings2019
- NFLContracts2015
- NFLContracts2019
- NFLPreSeason2014
- NFLPreseason2019
- NFLScores2011
- NFLScores2018
- NHANES
- NutritionStudy
- OlympicMarathon2008
- OlympicMarathon2012
- OlympicMarathon2016
- OrganicEffect
- OttawaSenators
- OttawaSenators2010
- OttawaSenators2019
- PASeniors
- PizzaGirl
- PumpkinBeer
- QuizPulse10
- RandomP50N200
- RestaurantTips
- RetailSales
- RetailSales2011
- RockandRoll2012
- RockandRoll2015
- RockandRoll2019
- SalaryGender
- SampColleges
- SampColleges2yr
- SampColleges4yr
- SampCountries
<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>SampCountries1e</td>
<td>119</td>
</tr>
<tr>
<td>SampCountries2e</td>
<td>120</td>
</tr>
<tr>
<td>SandP500</td>
<td>121</td>
</tr>
<tr>
<td>SandP5001e</td>
<td>122</td>
</tr>
<tr>
<td>SandP5002e</td>
<td>122</td>
</tr>
<tr>
<td>SandwichAnts</td>
<td>123</td>
</tr>
<tr>
<td>SandwichAnts2</td>
<td>124</td>
</tr>
<tr>
<td>SkateboardPrices</td>
<td>124</td>
</tr>
<tr>
<td>SleepCaffeine</td>
<td>125</td>
</tr>
<tr>
<td>SleepStudy</td>
<td>125</td>
</tr>
<tr>
<td>Smiles</td>
<td>127</td>
</tr>
<tr>
<td>SpeedDating</td>
<td>127</td>
</tr>
<tr>
<td>SplitBill</td>
<td>128</td>
</tr>
<tr>
<td>StatGrades</td>
<td>129</td>
</tr>
<tr>
<td>StockChanges</td>
<td>129</td>
</tr>
<tr>
<td>StorySpoilers</td>
<td>131</td>
</tr>
<tr>
<td>StressedMice</td>
<td>132</td>
</tr>
<tr>
<td>StudentSurvey</td>
<td>132</td>
</tr>
<tr>
<td>SynchronizedMovement</td>
<td>133</td>
</tr>
<tr>
<td>TenCountries</td>
<td>134</td>
</tr>
<tr>
<td>TenCountries1e</td>
<td>135</td>
</tr>
<tr>
<td>TenCountries2e</td>
<td>135</td>
</tr>
<tr>
<td>TextbookCosts</td>
<td>136</td>
</tr>
<tr>
<td>ToenailArsenic</td>
<td>136</td>
</tr>
<tr>
<td>TrafficFlow</td>
<td>137</td>
</tr>
<tr>
<td>USStates</td>
<td>137</td>
</tr>
<tr>
<td>USStates1e</td>
<td>139</td>
</tr>
<tr>
<td>USStates2e</td>
<td>140</td>
</tr>
<tr>
<td>WaterStriders</td>
<td>141</td>
</tr>
<tr>
<td>WaterTaste</td>
<td>141</td>
</tr>
<tr>
<td>Wetsuits</td>
<td>142</td>
</tr>
<tr>
<td>YoungBlood</td>
<td>143</td>
</tr>
</tbody>
</table>

Index 144

Description

Datasets for first, second, and third editions of Statistics: Unlocking the Power of Data by Lock^5
Details

Package: Lock5Data
Type: Package
Version: 3.0.0
Date: 2021-07-22
License: GPL-2
LazyLoad: yes

Author(s)

Robin Lock
Maintainer: Robin Lock <rlock@stlawu.edu>

Description

Data from a sample of individuals in the American Community Survey

Format

A data frame with 2000 observations on the following 9 variables.

Sex 0=female and 1=male
Age Age (years)
Married 0=not married and 1=married
Income Wages and salary for the past 12 months (in $1,000's)
HoursWk Hours of work per week
Race asian, black, other, or white
USCitizen 1=citizen and 0=noncitizen
HealthInsurance 1=have health insurance and 0=no health insurance
Language 1=English spoken at home and 0=other

Details

The American Community Survey, administered by the US Census Bureau, is given every year to a random sample of about 3.5 million households (about 3% of all US households). Data on a random sample of 1% of all US residents are made public (after ensuring anonymity), and we have selected a random sub-sample of n = 2000 from the 2017 data for this dataset.

** Updated for 3e (earlier version is ACS2010). **
Source

The full public dataset can be downloaded at https://www.census.gov/programs-surveys/acs/microdata.html, and the full list of variables are at https://www.census.gov/programs-surveys/acs/microdata/documentation.html

ACS2010 American Community Survey - 2010

Description

Data from a sample of individuals in the 2010 American Community Survey

Format

A dataset with 1000 observations on the following 9 variables.

- **Sex**: 0=female and 1=male
- **Age**: Age (years)
- **Married**: 0=not married and 1=married
- **Income**: Wages and salary for the past 12 months (in $1,000’s)
- **HoursWk**: Hours of work per week
- **Race**: asian, black, white, or other
- **USCitizen**: 1=citizen and 0=noncitizen
- **HealthInsurance**: 1=have health insurance and 0=no health insurance
- **Language**: 1=native English speaker and 0=other

Details

The American Community Survey, administered by the US Census Bureau, is given every year to a random sample of about 3.5 million households (about 3% of all US households). Data on a random sample of 1% of all US residents are made public (after ensuring anonymity), and we have selected a random sub-sample of n = 1000 from the 2010 data for this dataset.

** From 2e - dataset has been updated for 3e **

Source

Description

Data on the countries of the world

Format

A data frame with 217 observations on the following 26 variables.

Country  Country name
Code  Three-letter code for country
LandArea  Size in 1000 sq. km.
Population  Population in millions
Density  Number of people per square kilometer
GDP  Gross Domestic Product (in $US) per capita
Rural  Percentage of population living in rural areas
CO2  CO2 emissions (metric tons per capita)
PumpPrice  Price for a liter of gasoline ($US)
Military  Percentage of government expenditures directed toward the military
Health  Percentage of government expenditures directed towards healthcare
ArmedForces  Number of active duty military personnel (in 1,000's)
Internet  Percentage of the population with access to the internet
Cell  Cell phone subscriptions (per 100 people)
HIV  Percentage of the population with HIV
Hunger  Percent of the population considered undernourished
Diabetes  Percent of the population diagnosed with diabetes
BirthRate  Births per 1000 people
DeathRate  Deaths per 1000 people
ElderlyPop  Percentage of the population at least 65 years old
LifeExpectancy  Average life expectancy (years)
FemaleLabor  Percent of females 15 - 64 in the labor force
Unemployment  Percent of labor force unemployed
Energy  Kilotons of oil equivalent
Electricity  Electric power consumption (kWh per capita)
Developed  Categories for kilowatt hours per capita, 1= under 2500, 2=2500 to 5000, 3=over 5000
Details

Data for each variable were collected for 2018 (or most recently available year). Within a variable all country measurements are from the same year, but the year may vary between different variables depending on availability.

** This dataset is updated from an earlier versions (now Allcountries1e and AllCountries2e) **

Source


---

<table>
<thead>
<tr>
<th>AllCountries1e</th>
<th>AllCountries - 1e</th>
</tr>
</thead>
</table>

**Description**

Data on the countries of the world

**Format**

A dataset with 213 observations on the following 18 variables.

- **Country** Name of the country
- **Code** Three letter country code
- **LandArea** Size in sq. kilometers
- **Population** Population in millions
- **Energy** Energy usage (kilotons of oil)
- **Rural** Percentage of population living in rural areas
- **Military** Percentage of government expenditures directed toward the military
- **Health** Percentage of government expenditures directed towards healthcare
- **HIV** Percentage of the population with HIV
- **Internet** Percentage of the population with access to the internet
- **Developed** Categories for kilowatt hours per capita, 1= under 2500, 2=2500 to 5000, 3=over 5000
- **BirthRate** Births per 1000 people
- **ElderlyPop** Percentage of the population at least 65 years old
- **LifeExpectancy** Average life expectancy (years)
- **CO2** CO2 emissions (metric tons per capita)
- **GDP** Gross Domestic Product (per capita)
- **Cell** Cell phone subscriptions (per 100 people)
- **Electricity** Electric power consumption (kWh per capita)

**Details**

Most data from 2008 to avoid many missing values in more recent years.

** From 1e - dataset has been updated for 2e **
AllCountries2e

Source

Data collected from the World Bank website, worldbank.org.

AllCountries2e

Description

Data on the countries of the world

Format

A dataset with 215 observations on the following 25 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Name of the country</td>
</tr>
<tr>
<td>LandArea</td>
<td>Size in 1000 sq. kilometers</td>
</tr>
<tr>
<td>Population</td>
<td>Population in millions</td>
</tr>
<tr>
<td>Density</td>
<td>Number of people per square kilometer</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product (in $US) per capita</td>
</tr>
<tr>
<td>Rural</td>
<td>Percentage of population living in rural areas</td>
</tr>
<tr>
<td>CO2</td>
<td>CO2 emissions (metric tons per capita)</td>
</tr>
<tr>
<td>PumpPrice</td>
<td>Price for a liter of gasoline ($US)</td>
</tr>
<tr>
<td>Military</td>
<td>Percentage of government expenditures directed toward the military</td>
</tr>
<tr>
<td>Health</td>
<td>Percentage of government expenditures directed towards healthcare</td>
</tr>
<tr>
<td>ArmedForces</td>
<td>Number of active duty military personnel (in 1,000’s)</td>
</tr>
<tr>
<td>Internet</td>
<td>Percentage of the population with access to the internet</td>
</tr>
<tr>
<td>Cell</td>
<td>Cell phone subscriptions (per 100 people)</td>
</tr>
<tr>
<td>HIV</td>
<td>Percentage of the population with HIV</td>
</tr>
<tr>
<td>Hunger</td>
<td>Percent of the population considered undernourished</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Percent of the population diagnosed with diabetes</td>
</tr>
<tr>
<td>BirthRate</td>
<td>Births per 1000 people</td>
</tr>
<tr>
<td>DeathRate</td>
<td>Deaths per 1000 people</td>
</tr>
<tr>
<td>ElderlyPop</td>
<td>Percentage of the population at least 65 years old</td>
</tr>
<tr>
<td>LifeExpectancy</td>
<td>Average life expectancy (years)</td>
</tr>
<tr>
<td>FemaleLabor</td>
<td>Percent of females 15 - 64 in the labor force</td>
</tr>
<tr>
<td>Unemployment</td>
<td>Percent of labor force unemployed</td>
</tr>
<tr>
<td>Energy</td>
<td>Energy usage (kilotons of oil equivalent)</td>
</tr>
<tr>
<td>Electricity</td>
<td>Electric power consumption (kWh per capita)</td>
</tr>
<tr>
<td>Developed</td>
<td>Categories for kilowatt hours per capita, 1= under 2500, 2=2500 to 5000, 3=over 5000</td>
</tr>
</tbody>
</table>

Details

Data for each variable were collected for years between 2012 and 2014. Within a variable all country measurements are from the same year, but the year may vary between different variables depending on availability.

** From 2e - dataset has been updated for 3e **
APMultipleChoice  
AP Multiple Choice

Description
Correct responses on Advanced Placement multiple choice exams

Format
A dataset with 400 observations on the following variable.

Answer  Correct response: A, B, C, D, or E

Details
Correct responses from multiple choice sections for a sample of released Advanced Placement exams

Source
Sample exams from several disciplines at http://apcentral.collegeboard.com

April14Temps  
April 14th Temperatures

Description
Temperatures in Des Moines, IA and San Francisco, CA on April 14th

Format
A data frame with 25 observations on the following 3 variables.

Year 1995 to 2019
DesMoines Temperature in Des Moines (degrees F)
SanFrancisco Temperature in San Francisco (degrees F)

Details
Average temperature for the day of April 14th in each of 25 years from 1995-2019
** Data set updated for 3e (earlier versions are now April14Temps1e and April14Temps2e) **
**Source**

The University of Dayton Average Daily Temperature Archive at [https://academic.udayton.edu/kissock/http/Weather/citylistUS.htm](https://academic.udayton.edu/kissock/http/Weather/citylistUS.htm)

---

**April14Temps1e**  
**April 14th Temperatures -1e**

**Description**

Temperatures in Des Moines, IA and San Francisco, CA on April 14th

**Format**

A dataset with 16 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Year</th>
<th>1995-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>DesMoines</td>
<td>Temperature in Des Moines (degrees F)</td>
</tr>
<tr>
<td>SanFrancisco</td>
<td>Temperature in San Francisco (degrees F)</td>
</tr>
</tbody>
</table>

**Details**

Average temperature for the day of April 14th in each of 16 years from 1995-2010  
**From 1e - dataset has been updated for 2e**

**Source**

The University of Dayton Average Daily Temperature Archive at [http://academic.udayton.edu/kissock/http/Weather/citylistUS.htm](http://academic.udayton.edu/kissock/http/Weather/citylistUS.htm)

---

**April14Temps2e**  
**April 14th Temperatures - 2e**

**Description**

Temperatures in Des Moines, IA and San Francisco, CA on April 14th

**Format**

A dataset with 21 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Year</th>
<th>1995 to 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>DesMoines</td>
<td>Temperature in Des Moines (degrees F)</td>
</tr>
<tr>
<td>SanFrancisco</td>
<td>Temperature in San Francisco (degrees F)</td>
</tr>
</tbody>
</table>
Baseball Hits

Details

Average temperature for the day of April 14th in each of 21 years from 1995-2015
** From 2e - dataset has been updated for 3e **

Source

The University of Dayton Average Daily Temperature Archive at
http://academic.udayton.edu/kissock/http/Weather/citylistUS.htm

Baseball Hits

Description

Number of hits, wins, and other stats for MLB teams - 2011

Format

A dataset with 30 observations on the following 14 variables.

<table>
<thead>
<tr>
<th>Team</th>
<th>Name of baseball team</th>
</tr>
</thead>
<tbody>
<tr>
<td>League</td>
<td>Either American AL or National NL League</td>
</tr>
<tr>
<td>Wins</td>
<td>Number of wins for the season</td>
</tr>
<tr>
<td>Runs</td>
<td>Number of runs scored</td>
</tr>
<tr>
<td>Hits</td>
<td>Number of hits</td>
</tr>
<tr>
<td>Doubles</td>
<td>Number of doubles</td>
</tr>
<tr>
<td>Triples</td>
<td>Number of triples</td>
</tr>
<tr>
<td>HomeRuns</td>
<td>Number of home runs</td>
</tr>
<tr>
<td>RBI</td>
<td>Number of runs batted in</td>
</tr>
<tr>
<td>StolenBases</td>
<td>Number of stolen bases</td>
</tr>
<tr>
<td>CaughtStealing</td>
<td>Number of times caught stealing</td>
</tr>
<tr>
<td>Walks</td>
<td>Number of walks</td>
</tr>
<tr>
<td>Strikeouts</td>
<td>Number of strikeouts</td>
</tr>
<tr>
<td>BattingAvg</td>
<td>Team batting average</td>
</tr>
</tbody>
</table>

Details

Data from the 2010 Major League Baseball regular season.
** From 1e - dataset has been updated for 2e **

Source

**Baseball Hits - 2014**

**Description**

Number of hits, wins, and other stats for MLB teams - 2014

**Format**

A dataset with 30 observations on the following 14 variables.

- **Team**: Name of baseball team (3-character code)
- **League**: Either AL or NL
- **Wins**: Number of wins for the season
- **Runs**: Number of runs scored
- **Hits**: Number of hits
- **Doubles**: Number of doubles
- **Triples**: Number of triples
- **HomeRuns**: Number of home runs
- **RBI**: Number of runs batted in
- **StolenBases**: Number of stolen bases
- **CaughtStealing**: Number of times caught stealing
- **Walks**: Number of walks
- **Strikeouts**: Number of strikeouts
- **BattingAvg**: Team batting average

**Details**

Data from the 2014 Major League Baseball regular season.

**From 2e - dataset has been updated for 3e**

**Source**


---

**Baseball Team Statistics (2019)**

**Description**

Number of hits, wins, and other stats for MLB teams in 2019
BaseballSalaries2015

Format
A data frame with 30 observations on the following 14 variables.

Team  Name of baseball team (3-character code)
League Either AL or NL
Wins  Number of wins for the season
Runs  Number of runs scored
Hits  Number of hits
Doubles  Number of doubles
Triples  Number of triples
HomeRuns  Number of home runs
RBI  Number of runs batted in
StolenBases  Number of stolen bases
CaughtStealing  Number of times caught stealing
Walks  Number of walks
Strikeouts  Number of strikeouts
BattingAvg  Team batting average

Details
Offensive team statistics for the 2019 Major League Baseball regular season.
** Updated for 3e (earlier versions are now BaseballHits2014 and BaseballHits1e)

Source

BaseballSalaries2015  MLB Player Salaries in 2015

Description
Opening Day salaries for all Major League Baseball players in 2015

Format
A dataset with 868 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Name</th>
<th>Player's name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salary</td>
<td>2015 season salary (in millions)</td>
</tr>
<tr>
<td>Team</td>
<td>Abbreviated team name</td>
</tr>
<tr>
<td>Position</td>
<td>Code for player's main position</td>
</tr>
</tbody>
</table>
Details
Yearly salary (in millions of dollars) for all players on the rosters of Major League Baseball teams at the start of the 2015 season.
** From 2e - dataset has been updated for 3e **

Source
http://www.usatoday.com/sports/mlb/salaries

BaseballSalaries2019  MLB Player Salaries in 2019

Description
Opening Day salaries for all Major League Baseball players in 2019

Format
A data frame with 877 observations on the following 4 variables.

Name  Player’s name
Salary  2019 season salary (in millions)
Team  Abbreviated team name
POS  Code for player’s main position

Details
Yearly salary (in millions of dollars) for all players on the rosters of Major League Baseball teams at the start of the 2019 season.
** Updated for 3e (earlier version for 2015 is at BaseballSalaries2015). **

Source
https://databases.usatoday.com/mlb-salaries/

BaseballTimes  Baseball Game Times

Description
Information for a sample of 30 Major League Baseball games played during the 2011 season

Format
A dataset with 30 observations on the following 9 variables.
Details

Data from a sample of boxscores for Major League Baseball games played in August 2011.

Source


Benford

<table>
<thead>
<tr>
<th>Benford data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
</tr>
<tr>
<td>Two examples to test Benford’s Law</td>
</tr>
<tr>
<td>Format</td>
</tr>
<tr>
<td>A dataset with 9 observations on the following 4 variables.</td>
</tr>
<tr>
<td>Digit</td>
</tr>
<tr>
<td>BenfordP</td>
</tr>
<tr>
<td>Address</td>
</tr>
<tr>
<td>Invoices</td>
</tr>
<tr>
<td>Details</td>
</tr>
<tr>
<td>Leading digits from 1188 addresses sampled from a phone book and 7273 amounts from invoices sampled at a company.</td>
</tr>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Thanks to Prof. Richard Cleary for providing the data</td>
</tr>
</tbody>
</table>
BikeCommute  
_Bike Commute_

**Description**
Commute times for two kinds of bicycle

**Format**
A dataset with 56 observations on the following 9 variables.

- **Bike** Type of material Carbon or Steel
- **Date** Date of the bike commute
- **Distance** Length of commute (in miles)
- **Time** Total commute time (hours:minutes:seconds)
- **Minutes** Time converted to minutes
- **AvgSpeed** Average speed during the ride (miles per hour)
- **TopSpeed** Maximum speed (miles per hour)
- **Seconds** Time converted to seconds
- **Month** Categories: 1 Jan 2 Feb 3 Mar 4 Apr 5 May 6 June 7 July

**Details**
Data from a personal experiment to compare commuting time based on a randomized selection between two bicycles made of different materials.

**Source**
Thanks to Dr. Groves for providing his data.

**References**

BodyFat  
_Body Measurements_

**Description**
Percent fat and other body measurements for a sample of men

**Format**
A dataset with 100 observations on the following 10 variables.
Details

This is a subset of a larger sample of men who each had a percent body fat estimated by an underwater weighing technique. Other measurements were taken to see how they might be used to predict the body fat percentage.

Source

These data were contributed by Roger Johnson, then at Carleton University, to the Datasets Archive at the Journal of Statistics Education.

https://ww2.amstat.org/publications/jse/v4n1/datasets.johnson.html

The data were originally supplied by Dr. A. Garth Fisher, Human Performance Research Center, Brigham Young University, Provo, Utah 84602.

---

BodyTemp50

<table>
<thead>
<tr>
<th>BodyFat</th>
<th>Percent body fat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Age in years</td>
</tr>
<tr>
<td>Weight</td>
<td>Weight in pounds</td>
</tr>
<tr>
<td>Height</td>
<td>Height in inches</td>
</tr>
<tr>
<td>Neck</td>
<td>Neck circumference in cm.</td>
</tr>
<tr>
<td>Chest</td>
<td>Chest circumference in cm.</td>
</tr>
<tr>
<td>Abdomen</td>
<td>Abdomen circumference in cm.</td>
</tr>
<tr>
<td>Ankle</td>
<td>Ankle circumference in cm.</td>
</tr>
<tr>
<td>Biceps</td>
<td>Extended biceps circumference in cm.</td>
</tr>
<tr>
<td>Wrist</td>
<td>Wrist circumference in cm.</td>
</tr>
</tbody>
</table>

---

Description

Sample of 50 body temperatures

Format

A data frame with 50 observations on the following 3 variables.

BodyTemp  Body temperature in degrees F
Pulse     Pulse rates (beat per minute)
Sex       F=Female, M=Male

Details

Body temperatures and pulse rates for a sample of 50 healthy adults. Note the Sex variable was labeled as Gender in earlier versions of this dataset. We acknowledge that this binary dichotomization is not a complete or inclusive representation of reality.
Source

http://jse.amstat.org/v4n2/datasets.shoemaker.html

---

BootAtlantaCorr  Bootstrap Correlations for Atlanta Commutes

Description

Bootstrap correlations between Time and Distance for 500 commuters in Atlanta

Format

A dataset with 1000 observations on the following variable.

CorrTimeDist  Correlation between Time and Distance for a bootstrap sample of Atlanta commuters

Details

Correlations for bootstrap samples of Time vs. Distance for the data on Atlanta commuters in CommuteAtlanta.

Source

Computer simulation

---

CaffeineTaps  Caffeine Taps

Description

Finger tap rates with and without caffeine

Format

A dataset with 20 observations on the following 2 variables.

Taps  Number of finger taps in one minute
Group  Treatment with levels Caffeine NoCaffeine
**Details**

Results from a double-blind experiment where a sample of male college students were asked to tap their fingers at a rapid rate. The sample was then divided at random into two groups of ten students each. Each student drank the equivalent of about two cups of coffee, which included about 200 mg of caffeine for the students in one group but was decaffeinated coffee for the second group. After a two hour period, each student was tested to measure finger tapping rate (taps per minute). The goal of the experiment was to determine whether caffeine produces an increase in the average tap rate.

**Source**


---

**CAOSExam**

**CAOS Exam Scores**

**Description**

Scores on a pre-test and post-test of basic statistics concepts

**Format**

A dataset with 10 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Student</th>
<th>ID code for student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>CAOS Pretest score</td>
</tr>
<tr>
<td>Posttest</td>
<td>CAOS Posttest score</td>
</tr>
</tbody>
</table>

**Details**

The CAOS (Comprehensive Assessment of Outcomes in First Statistics Course) exam is designed to measure comprehension of basic statistical ideas in an introductory statistics course. This dataset has scores for ten students who took the CAOS pre-test at the start of a course and the post-test during the course itself. Each exam consists of 40 multiple choice questions and the score is the percentage correct.

**Source**

A sample of 10 students from an introductory statistics course. Find out more about the CAOS exam at [http://app.gen.umn.edu/artist/caos.html](http://app.gen.umn.edu/artist/caos.html)

---

**CarbonDioxide**

**Carbon Dioxide Levels**
Description

Atmospheric carbon dioxide levels by year

Format

A data frame with 12 observations on the following 2 variables.

Year  Every five years from 1960 to 2015
C02   Carbon dioxide level in parts per million

Details

Carbon dioxide levels in the atmosphere over a 55 year span from 1960-2015.
** Updated for 3e (earlier version is now CarbonDioxide2e) **

Source

Dr. Pieter Tans, NOAA/ESRL. Values recorded at the Mauna Loa Observatory in Hawaii. https://gml.noaa.gov/ccgg/trends/

---

Description

Atmospheric carbon dioxide levels by year

Format

A dataset with 11 observations on the following 2 variables.

Year  Every five years from 1960 to 2010
C02   Carbon dioxide level in parts per million

Details

Carbon dioxide levels in the atmosphere over a 50 year span from 1960-2010.
** From 2e - dataset has been updated for 3e **

Source

Dr. Pieter Tans, NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/). Values recorded at the Mauna Loa Observatory in Hawaii.
**CarDepreciation**  
*Car Depreciation*

**Description**  
Depreciation for 20 car models.

**Format**  
A dataset with 20 observations on the following 4 variables.

- **Car**  Name of the car model  
- **New**  Price of a new car  
- **Used**  Value after new car leaves the lot after purchase  
- **Depreciation**  Drop in value when a new car is driven away

**Details**  
Twenty car models were selected at random from kellybluebook.com. Original price (in dollars) and value after the car has been driven 10 miles were recorded for each model. The depreciation is the difference (New-Used).

**Source**  
New and used automobile costs determined using 2015 models selected from kellybluebook.com.

---

**Cars2015**  
*2020 Car Models*

**Description**  
Information about new car models in 2020

**Format**  
A dataset with 110 observations on the following 24 variables.

- **Make**  Manufacturer (e.g. Chevrolet, Toyota, etc.)  
- **Model**  Car model (e.g. Impala, Prius, ...)
- **Type**  Vehicle category (Small, Hatchback, Sedan, Sporty, Wagon, SUV, 7Pass)  
- **LowPrice**  Lowest MSRP (in $1,000)  
- **HighPrice**  Highest MSRP (in $1,000)  
- **Drive**  Type of drive (FWD, RWD, AWD)  
- **CityMPG**  City miles per gallon (EPA)  
- **HwyMPG**  Highway miles per gallon (EPA)
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FuelCap</td>
<td>Fuel capacity (in gallons)</td>
</tr>
<tr>
<td>Length</td>
<td>Length (in inches)</td>
</tr>
<tr>
<td>Width</td>
<td>Width (in inches)</td>
</tr>
<tr>
<td>Height</td>
<td>Height (in inches)</td>
</tr>
<tr>
<td>Wheelbase</td>
<td>Wheelbase (in inches)</td>
</tr>
<tr>
<td>UTurn</td>
<td>Diameter (in feet) needed for a U-turn</td>
</tr>
<tr>
<td>Weight</td>
<td>Curb weight (in pounds)</td>
</tr>
<tr>
<td>Acc030</td>
<td>Time (in seconds) to go from 0 to 30 mph</td>
</tr>
<tr>
<td>Acc060</td>
<td>Time (in seconds) to go from 0 to 60 mph</td>
</tr>
<tr>
<td>QtrMile</td>
<td>Time (in seconds) to go ¼ mile</td>
</tr>
<tr>
<td>Page Num</td>
<td>Page number in the Consumer Reports New Car Buying Guide</td>
</tr>
<tr>
<td>Size</td>
<td>Small, Midsized, or Large</td>
</tr>
</tbody>
</table>

**Details**

Data for a set of 110 new car models in 2015 based on information in the Consumer Reports.

**From 2e - dataset has been updated for 3e**

**Source**

Data on new car models in 2020 accessed from Consumer Reports website. [https://www.consumerreports.org/cars/](https://www.consumerreports.org/cars/)

---

**Description**

Information about new car models in 2020

**Format**

A data frame with 110 observations on the following 21 variables.

- **Make**  Manufacturer (e.g. Chevrolet, Toyota, etc.)
- **Model** Car model (e.g. Impala, Highlander, ...)
- **Type**  Vehicle category (Hatchback, Minivan, Sedan, Sporty, SUV, or Wagon)
- **LowPrice**  Lowest MSRP (in $1,000)
- **HighPrice** Highest MSRP (in $1,000)
- **CityMPG**  City miles per gallon (EPA)
- **HwyMPG**  Highway miles per gallon (EPA)
- **Seating**  Seating capacity
- **Drive**  Type of drive (AWD, FWD, or RWD)
- **Acc030**  Time (in seconds) to go from 0 to 30 mph
Acc060  Time (in seconds) to go from 0 to 60 mph
QtrMile  Time (in seconds) to go ¼ mile
Braking  Distance to stop from 60 mph (dry pavement)
FuelCap  Fuel capacity (in gallons)
Length   Length (in inches)
Width    Width (in inches)
Height   Height (in inches)
Wheelbase Wheelbase (in inches)
UTurn    Diameter (in feet) needed for a U-turn
Weight   Curb weight (in pounds)
Size     Large, Midsized, or Small

Details
Data for a set of 110 new car models in 2020 based on information in the Consumer Reports.
** Updated for 3e (an earlier version from 2015 is at Cars2015). **

Source
Data on new car models in 2020 accessed from Consumer Reports website. https://www.consumerreports.org/cars/

Cereal  

Description
Nutrition information for a sample of 30 breakfast cereals

Format
A dataset with 30 observations on the following 10 variables.

<table>
<thead>
<tr>
<th>Name</th>
<th>Brand name of cereal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>Manufacturer coded as G=General Mills, K=Kellog’s or Q=Quaker</td>
</tr>
<tr>
<td>Serving</td>
<td>Serving size (in cups)</td>
</tr>
<tr>
<td>Calories</td>
<td>Calories (per cup)</td>
</tr>
<tr>
<td>Fat</td>
<td>Fat (grams per cup)</td>
</tr>
<tr>
<td>Sodium</td>
<td>Sodium (mg per cup)</td>
</tr>
<tr>
<td>Carbs</td>
<td>Carbohydrates (grams per cup)</td>
</tr>
<tr>
<td>Fiber</td>
<td>Dietary Fiber (grams per cup)</td>
</tr>
<tr>
<td>Sugars</td>
<td>Sugars (grams per cup)</td>
</tr>
<tr>
<td>Protein</td>
<td>Protein (grams per cup)</td>
</tr>
</tbody>
</table>
Details

Nutrition contents for a sample of breakfast cereals, derived from nutrition labels. Values are per cup of cereal (rather than per serving).

Source

Cereal data obtained from nutrition labels at http://www.nutritionresource.com/foodcomp2.cfm?id=0800

<table>
<thead>
<tr>
<th>CityTemps</th>
<th>City Temperatures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

Mean monthly temperature in Moscow, Melbourne, and San Francisco for 2017 and 2018

Format

A data frame with 24 observations on the following 5 variables.

- **Year** 2017 or 2018
- **Month** 1=January through 12=December
- **Moscow** Monthly temperatures in Moscow (Russia)
- **Melbourne** Monthly temperatures in Melbourne (Australia)
- **San.Francisco** Monthly temperatures in San Francisco (United States)

Details

Mean monthly temperatures in degrees C for the years 2017 and 2018 in each of three cities.

** Updated for 3e (an earlier version for 2014 and 2015 is at CityTemps2e). **

Source

Source: KNMI Climate Explorer at https://climexp.knmi.nl/selectstation.cgi?id=someone@somewhere Use station codes 94866 (Melbourne), 72494 (San Francisco), 27612 (Moscow).

<table>
<thead>
<tr>
<th>CityTemps2e</th>
<th>City Temperatures - 2e</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

Mean monthly temperature in Moscow, Melbourne, and San Francisco for 2014 and 2015

Format

A dataset with 24 observations on the following 5 variables.
Moscow
Monthly temperatures in Moscow (Russia)

Melbourne
Monthly temperatures in Melbourne (Australia)

SanFrancisco
Monthly temperatures in San Francisco (United States)

Details
Mean monthly temperatures in degrees Celsius for the years 2014 and 2015 in each of three cities.
** From 2e - dataset has been updated for 3e **

Source
KNMI Climate Explorer at https://climexp.knmi.nl/selectstation.cgi?id=someone@somewhere

<table>
<thead>
<tr>
<th>CocaineTreatment</th>
<th>Cocaine Treatment</th>
</tr>
</thead>
</table>

Description
Relapse/no relapse responses to three different treatments for cocaine addiction

Format
A dataset with 72 observations on the following 2 variables.

| Treatment drug: Desipramine, Lithium, or Placebo |
| Did the patient relapse? no or yes |

Details
Data from an experiment to investigate the effectiveness of the two drugs, desipramine and lithium, in the treatment of cocaine addiction. Subjects (cocaine addicts seeking treatment) were randomly assigned to take one of the treatment drugs or a placebo. The response variable is whether or not the subject relapsed (went back to using cocaine) after the treatment.

Source
**ColaCalcium**  
*Cola Calcium*

**Description**
Calcium excretion with diet cola and water

**Format**
A dataset with 16 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Drink</th>
<th>Type of drink: Diet cola or Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcium</td>
<td>Amount of calcium excreted (in mg.)</td>
</tr>
</tbody>
</table>

**Details**
A sample of 16 healthy women aged 18 - 40 were randomly assigned to drink 24 ounces of either diet cola or water. Their urine was collected for three hours after ingestion of the beverage and calcium excretion (in mg.) was measured. The researchers were investigating whether diet cola leaches calcium out of the system, which would increase the amount of calcium in the urine for diet cola drinkers.

**Source**
Larson, Amin, Olsen, and Poth, Effect of Diet Cola on Urine Calcium Excretion, Endocrine Reviews, 31[3]: S1070, June 2010. These data are recreated from the published summary statistics, and are estimates of the actual data.

---

**CollegeScores**  
*College Scorecard*

**Description**
Information on all US post-secondary schools collected by the Department of Education for the College Scorecard

**Format**
A data frame with 6141 observations on the following 37 variables.

<table>
<thead>
<tr>
<th>Name</th>
<th>Name of the school</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>State where school is located</td>
</tr>
<tr>
<td>ID</td>
<td>ID number for school</td>
</tr>
<tr>
<td>Main</td>
<td>Main campus? (1=yes, 0=branch campus)</td>
</tr>
</tbody>
</table>
Accred  Accreditation agency
MainDegree  Predominant undergrad degree (0=not classified, 1=certificate, 2=associate, 3=bachelors, 4=only graduate)
HighDegree  Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)
Control  Control of school (Private, Profit, Public)
Region  Region of country (Midwest, Northeast, Southeast, Territory, West)
Locale  Locale (City, Rural, Suburb, Town)
Latitude  Latitude
Longitude  Longitude
AdmitRate  Admission rate
MidACT  Median of ACT scores
AvgSAT  Average combined SAT scores
Online  Only online (distance) programs
Enrollment  Undergraduate enrollment
White  Percent of undergraduates who report being white
Black  Percent of undergraduates who report being black
Hispanic  Percent of undergraduates who report being Hispanic
Asian  Percent of undergraduates who report being Asian
Other  Percent of undergraduates who don’t report one of the above
PartTime  Percent of undergraduates who are part-time students
NetPrice  Average net price (cost minus aid)
Cost  Average total cost for tuition, room, board, etc.
TuitionIn  In-state tuition and fees
TuitionOut  Out-of-state tuition and fees
TuitionFTE  Net Tuition revenue per FTE student
InstructFTE  Instructional spending per FTE student
FacSalary  Average monthly salary for full-time faculty
FullTimeFac  Percent of faculty that are full-time
Pell  Percent of students receiving Pell grants
CompRate  Completion rate (percent who finish program within 150% of normal time)
Debt  Average debt for students who complete program
Female  Percent of female students
FirstGen  Percent of first-generation students
MedIncome  Median family income (in $1,000)

Details
The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains a small subsets of the variables in the full College Scorecard.
Source


CollegeScores2yr College Scorecard - Two Year

Description

Information on all US colleges and universities that primarily grant associate’s degrees, collected by the Department of Education for the College Scoreboard.

Format

A data frame with 1141 observations on the following 37 variables.

Name Name of the school
State State where school is located
ID ID number for school
Main Main campus? (1=yes, 0=branch campus)
Accred Accreditation agency
MainDegree Predominant undergrad degree (2=associate)
HighDegree Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)
Control Control of school (Private, Profit, Public)
Region Region of country (Midwest, Northeast, Southeast, Territory, West)
Locale Locale (City, Rural, Suburb, Town)
Latitude Latitude
Longitude Longitude
AdmitRate Admission rate
MidACT Median of ACT scores
AvgSAT Average combined SAT scores
Online Only online (distance) programs
Enrollment Undergraduate enrollment
White Percent of undergraduates who report being white
Black Percent of undergraduates who report being black
Hispanic Percent of undergraduates who report being Hispanic
Asian Percent of undergraduates who report being Asian
Other Percent of undergraduates who don’t report one of the above
PartTime Percent of undergraduates who are part-time students
NetPrice Average net price (cost minus aid)
Cost Average total cost for tuition, room, board, etc.
TuitionIn In-state tuition and fees
TuitionOut Out-of-state tuition and fees
TuitionFTE Net Tuition revenue per FTE student
InstructFTE Instructional spending per FTE student
FacSalary Average monthly salary for full-time faculty
FullTimeFac Percent of faculty that are full-time
Pell Percent of students receiving Pell grants
CompRate Completion rate (percent who finish program within 150% of normal time)
Debt Average debt for students who complete program
Female Percent of female students
FirstGen Percent of first-generation students
MedIncome Median family income (in $1,000)

Details
The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains a small subset of the variables in the full College Scorecard and only the schools that primarily grant associate’s degrees (MainDegree=2). The CollegeScores dataset contains these and other schools with other degree types.

Source

CollegeScores4yr College Scorecard - Four Year

Description
Information on all US colleges and universities that primarily grant bachelor’s degrees, collected by the Department of Education for the College Scoreboard

Format
A data frame with 2012 observations on the following 37 variables.

Name Name of the school
State State where school is located
ID ID number for school
Main Main campus? (1=yes, 0=branch campus)
Accred  Accreditation agency
MainDegree  Predominant undergrad degree (3=bachelors)
HighDegree  Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)
Control  Control of school (Private, Profit, Public)
Region  Region of country (Midwest, Northeast, Southeast, Territory, West)
Locale  Locale (City, Rural, Suburb, Town)
Latitude  Latitude
Longitude  Longitude
AdmitRate  Admission rate
MidACT  Median of ACT scores
AvgSAT  Average combined SAT scores
Online  Only online (distance) programs
Enrollment  Undergraduate enrollment
White  Percent of undergraduates who report being white
Black  Percent of undergraduates who report being black
Hispanic  Percent of undergraduates who report being Hispanic
Asian  Percent of undergraduates who report being Asian
Other  Percent of undergraduates who don’t report one of the above
PartTime  Percent of undergraduates who are part-time students
NetPrice  Average net price (cost minus aid)
Cost  Average total cost for tuition, room, board, etc.
TuitionIn  In-state tuition and fees
TuitionOut  Out-of-state tuition and fees
TuitionFTE  Net Tuition revenue per FTE student
InstructFTE  Instructional spending per FTE student
FacSalary  Average monthly salary for full-time faculty
FullTimeFac  Percent of faculty that are full-time
Pell  Percent of students receiving Pell grants
CompRate  Completion rate (percent who finish program within 150% of normal time)
Debt  Average debt for students who complete program
Female  Percent of female students
FirstGen  Percent of first-generation students
MedIncome  Median family income (in $1,000)

Details
The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains a small subset of the variables in the full College Scorecard and only the schools that primarily grant bachelor’s degrees (MainDegree=3). The CollegeScores dataset contains these and other schools with other degree types.
CommuteStLouis

Source


CommuteAtlanta  Commute Atlanta

Description

Commute times and distances for a sample of 500 people in Atlanta

Format

A data frame with 500 observations on the following 5 variables.

<table>
<thead>
<tr>
<th>City</th>
<th>Atlanta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Age of the respondent (in years)</td>
</tr>
<tr>
<td>Distance</td>
<td>Commute distance (in miles)</td>
</tr>
<tr>
<td>Time</td>
<td>Commute time (in minutes)</td>
</tr>
<tr>
<td>Sex</td>
<td>F or M</td>
</tr>
</tbody>
</table>

Details

Data from the US Census Bureau’s American Housing Survey (AHS) which contains information about housing and living conditions for samples from certain metropolitan areas. These data were extracted from respondents in the Atlanta metropolitan area. They include only cases where the respondent worked somewhere other than home. Values show the time (in minutes) and distance (in miles) that respondents typically traveled on their commute to work each day as well as age and sex.

Source


CommuteStLouis  Commute Times in St. Louis

Description

Commute times and distances for a sample of 500 people in St. Louis

Format

A dataset with 500 observations on the following 5 variables.
Details

Data from the US Census Bureau’s American Housing Survey (AHS) which contains information about housing and living conditions for samples from certain metropolitan areas. These data were extracted from respondents in the St. Louis metropolitan area. They include only cases where the respondent worked somewhere other than home. Values show the time (in minutes) and distance (in miles) that respondents typically traveled on their commute to work each day as well as age and sex.

Source


Description

Would a rat attempt to free a trapped rat?

Format

A dataset with 30 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Sex of the rat: coded as F or M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empathy</td>
<td>Freed the trapped rat? no or yes</td>
</tr>
</tbody>
</table>

Details

In a recent study, some rats showed compassion by freeing another trapped rat, even when chocolate served as a distraction and even when the rats would then have to share the chocolate with their freed companion.

Source

**Description**

Cricket chirp rate and temperature

**Format**

A dataset with 7 observations on the following 2 variables.

| Temperature | Air temperature in degrees F |
| Chirps      | Cricket chirp rate (chirps per minute) |

**Details**

The data were collected by E.A. Bessey and C.A. Bessey who measured chirp rates for crickets and temperatures during the summer of 1898.

**Source**


---

**Description**

Funding for individuals by the California Department of Developmental Services (DDS),

**Format**

A dataset with 1000 observations on the following 6 variables.

<table>
<thead>
<tr>
<th>ID</th>
<th>ID code for subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgeCohort</td>
<td>Age group (0–5, 6–12, 13–17, 18–21, 22–50, 50+)</td>
</tr>
<tr>
<td>Age</td>
<td>Age in years</td>
</tr>
<tr>
<td>Expenditures</td>
<td>Annual expenditures in dollars</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Ethnic group</td>
</tr>
</tbody>
</table>
Details

The California Department of Developmental Services (DDS) allocates funds to support developmentally disabled California residents (such as those with autism, cerebral palsy, or intellectual disabilities) and their families. We refer to those supported by DDS as DDS consumers. The dataset DDS includes data on annual expenditure (in $), ethnicity, age, and gender for 1000 DDS consumers.

Source


Description

Difference between actual and scheduled arrival for United and Delta flights in December 2018.

Format

A data frame with 2000 observations on the following 2 variables.

Airline Delta or United
Difference Actual - Scheduled arrival times (in minutes)

Details

For a sample of 1000 December flights (in 2018) from each airline, we find the difference between actual and scheduled arrival times. A negative value indicates the flight arrived early.

** Updated for 3e (earlier version from 2014 is in DecemberFlights2e.)

Source

Downloaded from the Bureau of Transportation Statistics (https://www.transtats.bts.gov/).

Description

Difference between actual and scheduled arrival for a sample of United and Delta flights in December 2014.

Format

A dataset with 2000 observations on the following 2 variables.
DietDepression

Airline Delta or United
Difference Difference (Actual - Scheduled arrival times)

Details
For a sample of 1000 December flights (in 2014) from each airline, we find the difference between actual and scheduled arrival times. A negative value indicates the flight arrived early.
** From 2e - dataset has been updated for 3e **

Source
Downloaded from the Bureau of Transportation Statistics (https://www.bts.gov/). More specific URL is https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time.

---

DietDepression Diet and Depression

Description
Results from a study of a short-term diet intervention on depression.

Format
A data frame with 75 observations on the following 10 variables.

Group Control or Diet

CESD1 CESD depression score on Day 1
CESD21 CESD depression score on Day 21
CESDdiff Change in CESD depression score
DASS1 DASS depression score on Day 1
DASS21 DASS depression score on Day 21
DASSdiff Change in DASS depression score
BM1 BMI Body Mass Index on Day 1
BM21 Body Mass Index on Day 21
BMIdiff Change in Body Mass Index

Details
A group of researchers in Australia conducted a short (three-week) dietary intervention in a randomized controlled experiment. In the study, 75 college-age students with elevated depression symptoms and relatively poor diet habits were randomly assigned to either a healthy diet intervention group or a control group. The researchers recorded the change over the three-week period on two different numeric scales of depression (the CESD scale and the DASS scale). The CESD (Centre for Epidemiological Studies Depression) score is based more on clinical observations, while the DASS (Depression, Anxiety, and Stress Scale) depends more on self-reported information. They also recorded body mass index (BMI) at the start and end of the 21 day period.
Source

---

### Digits

<table>
<thead>
<tr>
<th>Digits</th>
<th>Digit Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description
Digits from social security numbers and student selected "random numbers"

Format
A dataset with 150 observations on the following 7 variables.

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>Four digit random numbers given by a sample of students</td>
</tr>
<tr>
<td>RND1</td>
<td>First digit</td>
</tr>
<tr>
<td>RND2</td>
<td>Second digit</td>
</tr>
<tr>
<td>RND3</td>
<td>Third digit</td>
</tr>
<tr>
<td>RND4</td>
<td>Fourth digit</td>
</tr>
<tr>
<td>SSN8</td>
<td>Eighth digit of social security number</td>
</tr>
<tr>
<td>SSN9</td>
<td>Last digit of social security number</td>
</tr>
</tbody>
</table>

Details
A sample of students were asked to give a random four digit number. The numbers are given in the dataset, along with separate columns for each of the four digits. The data also show the last two digits of each student’s social security number (SSN).

Source
In-class student surveys from several classes.

---

### DogOwner

<table>
<thead>
<tr>
<th>DogOwner</th>
<th>Dog/Owner matches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description
Experiment to match dogs with owners

Format
A dataset with 25 observations on the following variable.

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match</td>
<td>Was the dog correctly paired with it’s owner? no or yes</td>
</tr>
</tbody>
</table>
Details

Pictures were taken of 25 owners and their purebred dogs, selected from dog parks. Study participants were shown a picture of an owner together with pictures of two dogs (the owner’s dog and another random dog from the study) and asked to choose which dog most resembled the owner. Each dog-owner pair was viewed by 28 naive undergraduate judges, and the pairing was deemed "correct" (yes) if the majority of judges (more than 14) chose the correct dog to go with the owner.

** In first edition, but not as dataset in 2e **

Source


Description

Effect on drug resistance by level of treatment in mice.

Format

A dataset with 72 observations on the following 5 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Untreated, Light, Moderate, or Aggressive</td>
</tr>
<tr>
<td>Weight</td>
<td>Mouse weight in grams</td>
</tr>
<tr>
<td>RBC</td>
<td>Red blood cell density</td>
</tr>
<tr>
<td>ResistantDensity</td>
<td>Density of resistant parasites</td>
</tr>
<tr>
<td>DaysInfectious</td>
<td>Days infectious with resistant parasites</td>
</tr>
</tbody>
</table>

Details

In an experiment to study drug resistance in mice, groups of 18 mice were injected with a mixture of drug-resistant and drug-susceptible malaria parasites. One group received no treatment while the others got limited, moderate, or aggressive amounts of anti-malarial treatment. The weight and red blood cell density reflect the initial health of the mice. Density of resistant parasites and number of days infectious measure the effectiveness of the treatment.

Source


http://dx.doi.org/10.1371/journal.ppat.1003578

Description

Education spending and literacy rates for countries.

Format

A data frame with 170 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Country</th>
<th>Name of country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>Three-letter code for country</td>
</tr>
<tr>
<td>Education</td>
<td>Education spending (as a percentage of GDP)</td>
</tr>
<tr>
<td>Literacy</td>
<td>Literacy rate</td>
</tr>
</tbody>
</table>

Details

For each country, we have public spending on education (as a percentage of GDP) and literacy rate (percentage of the population who can read and write).

** Updated for 3e (an earlier version is at EducationLiteracy2e). **

Source

Most recent data (as of 2019) for each country obtained from https://www.worldbank.org/en/home.

Description

Education spending and literacy rates for countries.

Format

A dataset with 188 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Country</th>
<th>Name of country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>Education spending (as a percentage of GDP)</td>
</tr>
<tr>
<td>Literacy</td>
<td>Literacy rate</td>
</tr>
</tbody>
</table>
Details

For each country, we have public spending on education (as a percentage of GDP) and literacy rate (percentage of the population who can read and write).

** From 2e - dataset has been updated for 3e **

Source

Most recent data (as of 2015) for each country obtained from worldbank.org and http://www.knoema.com

<table>
<thead>
<tr>
<th>ElectionMargin</th>
<th>Election Margin</th>
</tr>
</thead>
</table>

Description

Approval rating and election margin for recent presidential elections

Format

A dataset with 12 observations on the following 5 variables.

<table>
<thead>
<tr>
<th>Year</th>
<th>Certain election years from 1940-2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate</td>
<td>Incumbent US president</td>
</tr>
<tr>
<td>Approval</td>
<td>Presidential approval rating at time of election</td>
</tr>
<tr>
<td>Margin</td>
<td>Margin of victory/defeat (as a percentage)</td>
</tr>
<tr>
<td>Result</td>
<td>Outcome of the election for the incumbent: Lost or Won</td>
</tr>
</tbody>
</table>

Details

Data include US Presidential elections since 1940 in which an incumbent was running for president. The approval rating for the sitting president is compared to the margin of victory/defeat in the election.

** Updated for 2e (original is now ElectionMargin1e) **

Source


EmployedACS

Description

Employed individuals from the American Community Survey (ACS) dataset
Format

A data frame with 1287 observations on the following 9 variables.

- **Sex**: 0=female and 1=male
- **Age**: Age (years)
- **Married**: 0=not married and 1=married
- **Income**: Wages and salary for the past 12 months (in $1,000’s)
- **HoursWk**: Hours of work per week
- **Race**: asian, black, other, white
- **USCitizen**: 1=citizen and 0=noncitizen
- **HealthInsurance**: 1=have health insurance and 0= no health insurance
- **Language**: 1=native English speaker and 0=other

Details

This is a subset of the ACS dataset including only 1287 individuals who were employed. (HoursWk>0)

**Updated for 3e (an earlier version is at EmployedACS2010).**

Source

The full public dataset can be downloaded at [https://www.census.gov/programs-surveys/acs/microdata/access.html](https://www.census.gov/programs-surveys/acs/microdata/access.html), and the full list of variables is at [https://www.census.gov/programs-surveys/acs/microdata.html](https://www.census.gov/programs-surveys/acs/microdata.html)
Details

This is a subset of the ACS dataset including only 431 individuals who were employed.
** From 2e - dataset has been updated for 3e **

Source

The full public dataset can be downloaded at
http://www.census.gov/acs/www/data documentation/pums data/, and the full list of variables are at

<table>
<thead>
<tr>
<th>Exercise Hours</th>
<th>Exercise Hours</th>
</tr>
</thead>
</table>

Description

Amount of exercise per week for students (and other variables)

Format

A data frame with 50 observations on the following 7 variables.

Year Year in school (1=First year,..., 4=Senior)
Sex  F or M
Hand Left (l) or Right (r) handed?
Exercise Hours of exercise per week
TV Hours of TV viewing per week
Pulse Resting pulse rate (beats per minute)
Pierces Number of body piercings

Details

Data from an in-class survey of statistics students asking about amount of exercise, TV viewing, handedness, sex, pulse rate, and number of body piercings. Note the Sex variable was labeled as Gender in earlier versions of this dataset. We acknowledge that this binary dichotomization is not a complete or inclusive representation of reality.

Source

In-class student survey.
Facebook Friends

Description
Data on number of Facebook friends and grey matter density in brain regions related to social perception and associative memory.

Format
A dataset with 40 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMdensity</td>
<td>Normalized z-scores of grey matter density in certain brain regions</td>
</tr>
<tr>
<td>FBfriends</td>
<td>Number of friends on Facebook</td>
</tr>
</tbody>
</table>

Details
A recent study in Great Britain examines the relationship between the number of friends an individual has on Facebook and grey matter density in the areas of the brain associated with social perception and associative memory. The study included 40 students at City University London.

Source

Fat Mice 18

Description
Weight gain for mice with different nighttime light conditions

Format
A dataset with 18 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light</td>
<td>Light treatment: LD= normal light/dark cycle OR LL=bright light at night</td>
</tr>
<tr>
<td>WgtGain4</td>
<td>Weight gain (grams over a four week period)</td>
</tr>
</tbody>
</table>
Details

This is a subset of the LightatNight dataset, showing body mass gain in mice after 4 weeks for two of the treatment conditions: a normal light/dark cycle (LD) or a bright light on at night (LL).

** In first edition, but not 2e **

Source

Fonken, L., et. al., "Light at night increases body mass by shifting time of food intake," Proceedings of the National Academy of Sciences, October 26, 2010; 107(43): 18664-18669.

<table>
<thead>
<tr>
<th>FireAnts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fire Ants</td>
</tr>
</tbody>
</table>

Description

Reactions of lizards to the presence of fire ants.

Format

A dataset with 80 observations on the following 3 variables.

- Invasion: Coded as Uninvaded or Invaded, depending on if the lizard comes from a region with fire ants
- Twitches: Number of twitches the lizard makes when encountering fire ants
- Flee: Time for the lizard to flee in seconds (more than one minute is recorded as 61).

Details

The red imported fire ant, *Solenopsis invicta*, is native to South America, but has an expansive invasive range, including much of the southern United States (invasion of this ant is predicted to go global). In the United States, these ants occupy similar habitats as fence lizards. The ants eat the lizards and the lizards eat the ants, and in either scenario the venom from the fire ant can be fatal to the lizard. The study explored the question of whether lizards learn to adapt their behavior if their environment has been invaded by fire ants by taking lizards from an uninvaded habitat (eastern Arkansas) and lizards from an invaded habitat (southern Alabama, which has been invaded for more than 70 years), exposing them to fire ants, and measuring how long it takes each lizard to flee and the number of twitches each lizard does.

Source

FisherIris  
*Fisher’s Iris Data*

**Description**

Measurements of three iris species

**Format**

A dataset with 150 observations on the following 5 variables.

<table>
<thead>
<tr>
<th>Type</th>
<th>Species of iris, Setosa, Virginica, or Versicolor</th>
</tr>
</thead>
<tbody>
<tr>
<td>PetalLength</td>
<td>Petal length in mm.</td>
</tr>
<tr>
<td>PetalWidth</td>
<td>Petal width in mm.</td>
</tr>
<tr>
<td>SepalLength</td>
<td>Sepal length in mm.</td>
</tr>
<tr>
<td>SepalWidth</td>
<td>Sepal width in mm.</td>
</tr>
</tbody>
</table>

**Details**

Data used in Fisher’s 1936 paper, this famous dataset looks at measurements for samples of three different species of iris. The petal is part of the flower itself and the sepals are green leaves, directly under the petals, providing support.

**Source**


FishGills12  
*Fish Respiration and Calcium - Full Data*

**Description**

An experiment to look at fish respiration rates in water with different levels of calcium.

**Format**

A dataset with 360 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Calcium</th>
<th>Amount of calcium in the water (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GillRate</td>
<td>Respiration rate (beats per minute)</td>
</tr>
</tbody>
</table>
**Details**

Fish were randomly assigned to twelve tanks with different levels (measured in mg/L) of calcium. Respiration rate was measured as number of gill beats per minute.

**Source**

Thanks to Prof. Brad Baldwin for supplying the data.

---

**FishGills3**  
*Fish Respiration and Calcium*

<table>
<thead>
<tr>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiration rate for fish in three levels of calcium.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>A dataset with 90 observations on the following 2 variables.</td>
</tr>
<tr>
<td>Calcium Level of calcium Low 0.71 mg/L, Medium 5.24 mg/L, or High 18.24 mg/L</td>
</tr>
<tr>
<td>GillRate Respiration rate (beats per minute)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish were randomly assigned to three tanks with different levels (low, medium and high) of calcium. Respiration rate was measured as number of gill beats per minute.</td>
</tr>
</tbody>
</table>

**Source**

Thanks to Prof. Brad Baldwin for supplying the data.

---

**Flight179**  
*Flight times*

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flight times for Flight 179 (Boston-SF) and Flight 180 (SF-Boston).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>A dataset with 36 observations on the following 3 variables.</td>
</tr>
<tr>
<td>Date Date of the flight (5th, 15th and 25th of each month in 2010</td>
</tr>
<tr>
<td>Flight179 Flying time (Boston-SF) in minutes</td>
</tr>
<tr>
<td>Flight180 Flying time (SF-Boston) in minutes</td>
</tr>
</tbody>
</table>
Details

United Airlines Flight 179 was a daily flight from Boston to San Francisco. Flight 180 goes in the other direction (SF to Boston). The data show the airborne flying times for each flight on the three dates each month (5th, 15th and 25th) in 2010.

** In first edition, but not in 2e - replaced by Flight433 **

Source

Data collected from the Bureau of Transportation Statistics website at
http://www.bts.gov/xml/ontimesummarystatistics/src/dstat/OnetimeSummaryAirtime.xml

---

Flight433  
Flight 433

Description

Flight times for Flight 433 (Boston-SF) in January 2019.

Format

A data frame with 28 observations on the following variable.

AirTime  Airborne flying time (in minutes) for Flight 433, Boston to San Francisco

Details

United Airlines Flight 433 was a daily flight from Boston to San Francisco. The data show the airborne flying times for the flight on each day of January 2019.

**Updated for 3e (earlier version from 2016 is in Flight433_2e) **

Source

Data collected from the Bureau of Transportation Statistics website at https://www.transtats.bts.gov/

---

Flight433_2e  
Flight 433 - 2e

Description

Flight times for Flight 433 (Boston-SF) in January 2016.

Format

A dataset with 31 observations on the following 1 variable.
Airtime  Airborne flying time (in minutes) for Flight 433, Boston to San Francisco

Details

United Airlines Flight 433 was a daily flight from Boston to San Francisco. The data show the airborne flying times for the flight on each day of January 2016.

** From 2e - dataset has been updated for 3e **

Source

Data collected from the Bureau of Transportation Statistics website at

http://www.bts.gov/xml/ontimesummarystatistics/src/dstat/OntimeSummaryAirtime.xml

FloridaLakes

Description

Water quality measurements for a sample of lakes in Florida

Format

A dataset with 53 observations on the following 12 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>An identifying number for each lake</td>
</tr>
<tr>
<td>Lake</td>
<td>Name of the lake</td>
</tr>
<tr>
<td>Alkalinity</td>
<td>Concentration of calcium carbonate (in mg/L)</td>
</tr>
<tr>
<td>pH</td>
<td>Acidity</td>
</tr>
<tr>
<td>Calcium</td>
<td>Amount of calcium in water</td>
</tr>
<tr>
<td>Chlorophyll</td>
<td>Amount of chlorophyll in water</td>
</tr>
<tr>
<td>AvgMercury</td>
<td>Average mercury level for a sample of fish (large mouth bass) from each lake</td>
</tr>
<tr>
<td>NumSamples</td>
<td>Number of fish sampled at each lake</td>
</tr>
<tr>
<td>MinMercury</td>
<td>Minimum mercury level in a sampled fish</td>
</tr>
<tr>
<td>MaxMercury</td>
<td>Maximum mercury level in a sampled fish</td>
</tr>
<tr>
<td>ThreeYrStdMercury</td>
<td>Adjusted mercury level to account for the age of the fish</td>
</tr>
<tr>
<td>AgeData</td>
<td>Mean age of fish in each sample</td>
</tr>
</tbody>
</table>

Details

This dataset describes characteristics of water and fish samples from 53 Florida lakes. Some variables (e.g. Alkalinity, pH, and Calcium) reflect the chemistry of the water samples. Mercury levels were recorded for a sample of large mouth bass selected at each lake.

Source

Description

Brain measurements for non-football players, football players with no concussion history, and football players with a concussion history.

Format

A dataset with 75 observations on the following 5 variables.

- **Group**: Control=no football, FBNoConcuss=football player but no concussions, or FBConcuss=football player with concussion history
- **Hipp**: Total hippocampus volume, in microL
- **LeftHipp**: Left hippocampus volume, in microL
- **Years**: Number of years playing football
- **Cognition**: Cognitive testing composite reaction time score, given as a percentile

Details

The study included 3 groups, with 25 cases in each group. The control group consisted of healthy individuals with no history of brain trauma who were comparable to the other groups in age, sex, and education. The second group consisted of NCAA Division 1 college football players with no history of concussion, while the third group consisted of NCAA Division 1 college football players with a history of concussion. High resolution MRI was used to collect brain hippocampus volume. Data were collected between June 2011 and August 2013. The data values given here are estimated from information given in the paper.

Source


Description

Characteristics of forest fires in Montesinho park (Portugal)
Genetic Diversity

Format
A data frame with 517 observations on the following 13 variables.

X  West to east coordinates for the site (1=farthest west to 9=farthest east)
Y  North to south coordinates for the site (1=farthest north to 9=farthest south)
Month  Month of the year (jan to dec)
Day  Day of the week (sun to sat)
FFMC  Fine fuel moisture code
DMC  Duff moisture code
DC  Drought code
ISI  Initial spread index
Temp  Outside temperature (in celsius)
RH  Relative humidity (in %)
Wind  Wind speed (in km/h)
Rain  Rain in past 30 minutes (in mm/sq-m)
Area  Total burned area (in hectares)

Details
Data were recorded for fires in the Montesinho natural park in Portugal between January 2000 and December 2003. A map of the park (see the pdf linked below) is divided into 9x9 grid sections (given by the x,y-coordinates in the first two columns of the dataset). There are four components of a Fire Weather Index that rate how weather conditions might increase fire danger. FFMC, DMC, and DC reflect various measures of moisture content, while the ISI score indicated how fast a fire might spread (for example, by wind). For all four measures larger values are associated with more fire danger. Fires that are less than 100 square meters in size (0.01 hectares) are recorded as Area=0.

Source

Genetic Diversity

Description
Genetic diversity for different populations are compared to the distance from East Africa.

Format
A dataset with 52 observations on the following 5 variables.
Details

The data give a measure of genetic diversity for different populations and the geographic distance of each population from East Africa (Addis Ababa, Ethiopia), as one would travel over the surface of the earth by land (migration long ago is thought to have happened by land).

Source


GlobalInternet2010

Global Internet Usage - 2010

Description

Internet usage for several countries

Format

A dataset with 9 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Country</th>
<th>Name of country</th>
</tr>
</thead>
<tbody>
<tr>
<td>PercentFastConnection</td>
<td>Percent of internet users with a fast connection</td>
</tr>
<tr>
<td>HoursOnline</td>
<td>Average number of hours online in February 2011</td>
</tr>
</tbody>
</table>

Details

The Nielsen Company measured connection speeds on home computers in nine different countries. Variables include the percent of internet users with a fast connection (defined as 2Mb/sec or faster) and the average amount of time spent online, defined as total hours connected to the web from a home computer during the month of February 2011. **From 2e - dataset has been updated for 3e**

Source

NielsenWire, "Swiss Lead in Speed: Comparing Global Internet Connections", April 1, 2011
**Global Internet Usage**

### Description

Internet usage for several countries

### Format

A data frame with 9 observations on the following 3 variables.

- **Country**: Name of country
- **InternetSpeed**: Average download speed (in Mb)
- **HoursOnline**: Average hours online per day

### Details

The Worldwide Broadband Speed League tests internet speeds at millions of access points around the world. The average download speed for each country is derived from those data. The DataReport site provides summaries of country level data on internet usage obtained from various sources. The average number of hours spent online for each country is based on survey data reported at that site.

**Updated for 3e (earlier version from 2011 is at GlobalInternet2011).**

### Source

- Internet speeds for 2019 downloaded from [https://www.cable.co.uk/broadband/speed/worldwide-speed-league/](https://www.cable.co.uk/broadband/speed/worldwide-speed-league/)
- Online hours for 2019 downloaded from [https://datareportal.com/library](https://datareportal.com/library)

---

**Golf Round**

### Description

Scorecard for 18 holes of golf

### Format

A data frame with 18 observations on the following 4 variables.

- **Hole**: Hole number (1 to 18)
- **Distance**: Length of the hole (in yards)
- **Par**: Par for the hole
- **Score**: Actual number of stokes needed in this round
Details
Data come from a scorecard for one round of golf at the Potsdam Country Club. Par is the expected number of strokes a good golfer should need to complete the hole.

Source
Personal file

GPAbySex

GPA by Sex

Description
Data from a survey of introductory statistics students.

Format
A dataset with 343 observations on the following 6 variables.

Exercise  Hours of exercise (per week)
SAT       Combined SAT scores (out of 1600)
GPA       Grade Point Average (0.00-4.00 scale)
Pulse     Pulse rate (beats per minute)
Piercings Number of body piercings
CodedSex  0=female or 1=male

Details
This is a subset of the StudentSurvey dataset where cases with missing values have been dropped and sex is coded as a 0/1 indicator variable.

Source
A first day survey over several different introductory statistics classes.

GSWarriors2016

Golden State Warriors Basketball - 2016

Description
Game log data for the Golden State Warriors basketball team in 2015-2016

Format
A dataset with 82 observations on the following 33 variables.
<table>
<thead>
<tr>
<th>Game ID number for each game</th>
<th>Date</th>
<th>Location</th>
<th>Away or Home</th>
<th>Opp</th>
<th>Opponent team</th>
<th>Win</th>
<th>Game result: L or W</th>
<th>FG</th>
<th>Field goals made</th>
<th>FGA</th>
<th>Field goals attempted</th>
<th>FG3</th>
<th>Three-point field goals made</th>
<th>FG3A</th>
<th>Three-point field goals attempted</th>
<th>FT</th>
<th>Free throws made</th>
<th>FTA</th>
<th>Free throws attempted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rebounds</td>
<td>Total rebounds</td>
<td>OffReb</td>
<td>Offensive rebounds</td>
<td>Assists</td>
<td>Number of assists</td>
<td>Steals</td>
<td>Number of steals</td>
<td>Blocks</td>
<td>Number of shots blocked</td>
<td>Turnovers</td>
<td>Number of turnovers</td>
<td>Fouls</td>
<td>Number of fouls</td>
<td>OppFG</td>
<td>Opponent’s field goals made</td>
<td>OppFGA</td>
<td>Opponent’s Field goals attempted</td>
<td>OppFG3</td>
<td>Opponent’s Three-point field goals made</td>
</tr>
<tr>
<td>OppRebounds</td>
<td>Opponent’s Total rebounds</td>
<td>OppOffReb</td>
<td>Opponent’s Offensive rebounds</td>
<td>OppAssists</td>
<td>Opponent’s assists</td>
<td>OppSteals</td>
<td>Opponent’s steals</td>
<td>OppBlocks</td>
<td>Opponent’s shots blocked</td>
<td>OppTurnovers</td>
<td>Opponent’s turnovers</td>
<td>OppFouls</td>
<td>Opponent’s fouls</td>
<td>OppPoints</td>
<td>Opponent’s points scored</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Details**

Information from online boxscores for all 82 regular season games played by the Golden State Warriors basketball team during the 2015-2016 season.

**From 2e - dataset has been updated for 3e **

**Source**

Data for the 2015-2016 Golden State games downloaded from

Description

Game log data for the Golden State Warriors basketball team in 2018-2019

Format

A data frame with 82 observations on the following 33 variables.

- **Game**  ID number for each game
- **Date**  Date the game was played (mm/dd/yyyy)
- **Location**  Away or Home
- **Opp**  Opponent team
- **Win**  Game result: L or W
- **Points**  Number of points scored
- **FG**  Field goals made
- **FGA**  Field goals attempted
- **FG3**  Three-point field goals made
- **FG3A**  Three-point field goals attempted
- **FT**  Free throws made
- **FTA**  Free throws attempted
- **Rebounds**  Total rebounds
- **OffReb**  Offensive rebounds
- **Assists**  Number of assists
- **Steals**  Number of steals
- **Blocks**  Number of shots blocked
- **Turnovers**  Number of turnovers
- **Fouls**  Number of fouls
- **OppPoints**  Opponent’s points scored
- **OppFG**  Opponent’s field goals made
- **OppFGA**  Opponent’s field goals attempted
- **OppFG3**  Opponent’s three-point field goals made
- **OppFG3A**  Opponent’s three-point field goals attempted
- **OppFT**  Opponent’s free throws made
- **OppFTA**  Opponent’s free throws attempted
- **OppRebounds**  Opponent’s total rebounds
- **OppOffReb**  Opponent’s offensive rebounds
OppAssists  Opponent’s assists
OppSteals  Opponent’s steals
OppBlocks  Opponent’s shots blocked
OppTurnovers  Opponent’s turnovers
OppFouls  Opponent’s fouls

**Details**

Information from online boxscores for all 82 regular season games played by the Golden State Warriors basketball team during the 2018-2019 season.

** Source**


**Description**

Measurements related to happiness and well-being for 143 countries.

**Format**

A dataset with 143 observations on the following 11 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>Score on a 0-10 scale for average level of happiness (10 is happiest)</td>
</tr>
<tr>
<td>LifeExpectancy</td>
<td>Average life expectancy (in years)</td>
</tr>
<tr>
<td>Footprint</td>
<td>Ecological footprint - a measure of the (per capita) ecological impact</td>
</tr>
<tr>
<td>HLY</td>
<td>Happy Life Years - combines life expectancy with well-being</td>
</tr>
<tr>
<td>HPI</td>
<td>Happy Planet Index (0-100 scale)</td>
</tr>
<tr>
<td>HPIRank</td>
<td>HPI rank for the country</td>
</tr>
<tr>
<td>GDPperCapita</td>
<td>Gross Domestic Product (per capita)</td>
</tr>
<tr>
<td>HDI</td>
<td>Human Development Index</td>
</tr>
<tr>
<td>Population</td>
<td>Population (in millions)</td>
</tr>
</tbody>
</table>
### Details

Data for 143 countries from the Happy Planet Index Project that works to quantify indicators of happiness, well-being, and ecological footprint at a country level.

### Source

Data downloaded from http://www.happyplanetindex.org/data/

---

#### HeatCognition

#### Heat and Cognition

### Description

Effect of heat on cognitive ability

### Format

A data frame with 46 observations on the following 3 variables.

- **AC**  Whether the student had air conditioning on in the room, No or Yes
- **MathZRT**  Z-score of reaction time solving math problems
- **ColorsZRT**  Z-score of reaction time solving STROOP color problems

### Details

Forty-six college students were asked to solve cognitive problems first thing in the morning during a heat wave in their Northeastern city. Twenty of the students had air-conditioning in their rooms and twenty-six did not. Z-scores of reaction times are given for math problems and for color dissonance problems.

### Source

Description

Heights measured for the same 94 children over 18 years.

Format

A dataset with 94 observations on the following 33 variables.
ID Identification number
Sex M or F
Year_1 Height (in cm.) at age 1 year
Year_1.25 Height (in cm.) at age 1.25 years
Year_1.5 Height (in cm.) at age 1.5 years
Year_1.75 Height (in cm.) at age 1.75 years
Year_2 Height (in cm.) at age 2 years
Year_3 Height (in cm.) at age 3 years
Year_4 Height (in cm.) at age 4 years
Year_5 Height (in cm.) at age 5 years
See below for full list of years...
Year_17.5 Height (in cm.) at age 17.5 years
Year_18 Height (in cm.) at age 18 years

Details
In the 1940’s and 1950’s, the heights of 39 boys and 54 girls, in centimeters, were measured at 30 different time points between the ages of 1 and 18 years as part of the University of California Berkeley growth study. Ages for measurement are 1, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 7, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 15.5, 16, 16.5, 17, 17.5, 18.

Source

---

**Description**
Penalty minutes (per game) for NHL teams in 2010-11

**Format**
A dataset with 30 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Team</th>
<th>Name of the team</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIMperG</td>
<td>Average penalty minutes per game</td>
</tr>
</tbody>
</table>

**Details**
Data give the average number of penalty minutes for each of the 30 National Hockey League (NHL) teams during the 2010-11 regular season.

** From 2e - dataset has been updated for 3e **
**Source**

Data obtained online at www.nhl.com

---

**HockeyPenalties2019**

**Hockey Penalties (2019)**

**Description**

Penalty minutes (per game) for NHL teams in 2018-2019

**Format**

A data frame with 30 observations on the following 4 variables.

- **Team**: Name of the team
- **PIM**: Average penalty minutes per game
- **OppPIM**: Average opponent's penalty minutes per game
- **Playoff**: Did the team make the playoffs? (N or Y)

**Details**

Data give the average number of penalty minutes for each of the 30 National Hockey League (NHL) teams (and their opponents) during the 2018-2019 regular season.

**Updated for 3e (earlier version from 2010-11 is at HockeyPenalties2011).**

**Source**

Data obtained online at https://www.hockey-reference.com/leagues/NHL_2019.html#all_stats

---

**HollywoodMovies**

**Hollywood Movies**

**Description**

Data on movies released in Hollywood between 2012 and 2018
**Format**

A data frame with 1295 observations on the following 15 variables.

- **Movie Title of the movie**
- **LeadStudio** Primary U.S. distributor of the movie
- **RottenTomatoes** Rotten Tomatoes rating (critics)
- **AudienceScore** Audience rating (via Rotten Tomatoes)
- **Genre** One of Action Adventure, Black Comedy, Comedy, Concert, Documentary, Drama, Horror, Musical, Romantic Comedy, Thriller, or Western
- **TheatersOpenWeek** Number of screens for opening weekend
- **OpeningWeekend** Opening weekend gross (in millions)
- **BOAvgOpenWeekend** Average box office income per theater, opening weekend
- **Budget** Production budget (in millions)
- **DomesticGross** Gross income for domestic (U.S.) viewers (in millions)
- **WorldGross** Gross income for all viewers (in millions)
- **ForeignGross** Gross income for foreign viewers (in millions)
- **Profitability** WorldGross as a percentage of Budget
- **OpenProfit** Percentage of budget recovered on opening weekend
- **Year** Year the movie was released

**Details**

Information from 1295 movies released from Hollywood between 2012 and 2018. **Updated for 3e (earlier versions are HollywoodMovies2013 and HollywoodMovies2011).**

**Source**

Movie data obtained from

[https://www.boxofficemojo.com/](https://www.boxofficemojo.com/)
[https://www.the-numbers.com/](https://www.the-numbers.com/)
[https://www.rottentomatoes.com/](https://www.rottentomatoes.com/)

---

**HollywoodMovies2011 Hollywood Movies in 2011**

**Description**

Data on movies released in Hollywood in 2011

**Format**

A dataset with 136 observations on the following 14 variables.
Movie | Title of movie
--- | ---
LeadStudio | Studio that released the movie
RottenTomatoes | Rotten Tomatoes rating (reviewers)
AudienceScore | Audience rating (via Rotten Tomatoes)
Story | General theme - one of 21 themes
Genre | Action Adventure Animation Comedy Drama Fantasy Horror Romance Thriller
TheatersOpenWeek | Number of screens for opening weekend
BOAverageOpenWeek | Average opening week box office income (per theater)
DomesticGross | Gross income for domestic viewers (in $ millions)
ForeignGross | Gross income for foreign viewers (in $ millions)
WorldGross | Gross income for all viewers (in $ millions)
Budget | Production budget (in $ millions)
Profitability | WorldGross as a percentage of Budget
OpeningWeekend | Opening weekend gross (in $ millions)

Details


**This dataset has been updated for 2e with more years of data (in HollywoodMovies)**

Source


HollywoodMovies2013 Hollywood Movies - 2013

Description

Data on movies released in Hollywood between 2007 and 2013

Format

A dataset with 970 observations on the following 16 variables.

Movie | Title of movie
--- | ---
LeadStudio | Studio that released the movie
RottenTomatoes | Rotten Tomatoes rating (reviewers)
AudienceScore | Audience rating (via Rotten Tomatoes)
Story | General theme - one of 21 themes
Genre | One of 14 possible genres
TheatersOpenWeek | Number of screens for opening weekend
OpeningWeekend | Opening weekend gross (in $ millions)
BOAverageOpenWeek | Average opening week box office income (per theater)
Details


** From 2e - dataset has been updated for 3e **

Source


---

HomesForSale  Homes For Sale (2019)

Description

Data on homes for sale in four states in 2019

Format

A data frame with 120 observations on the following 5 variables.

State  Location of the home (CA, NJ, NY, or PA)
Price  Asking price (in $1,000’s)
Size   Area of all rooms (in 1,000’s sq. ft.)
Beds   Number of bedrooms
Baths  Number of bathrooms

Details

Data for samples of homes for sale in each state, selected from zillow.com.

** Updated for 3e (earlier version from 2010 is in HomesForSale2e). **

Source

HomesForSaleCA

HomesForSale2e  Home for Sale - 2e

Description
Data on homes for sale in four states

Format
A dataset with 120 observations on the following 5 variables.

- **State**: Location of the home: CA NJ NY PA
- **Price**: Asking price (in $1,000's)
- **Size**: Area of all rooms (in 1,000’s sq. ft.)
- **Beds**: Number of bedrooms
- **Baths**: Number of bathrooms

Details
Data for samples of homes for sale in each state, selected from zillow.com.

** From 2e - dataset has been updated for 3e **

Source

HomesForSaleCA  Homes For Sale in California (2019)

Description
Data for a sample of homes offered for sale in California

Format
A data frame with 30 observations on the following 5 variables.

- **State**: Location of the home (CA)
- **Price**: Asking price (in $1,000's)
- **Size**: Area of all rooms (in 1,000’s sq. ft.)
- **Beds**: Number of bedrooms
- **Baths**: Number of bathrooms
**Details**

Data for a sample of homes for sale in California, selected from zillow.com. This is a subset of the HomesForSale dataset.

**Updated for 3e (earlier version from 2010 is in HomesForSaleCA2e).**

**Source**


---

**HomesForSaleCA2e**

*Home for Sale in California -2e*

---

**Description**

Data for a sample of homes offered for sale in California

**Format**

A dataset with 30 observations on the following 5 variables.

<table>
<thead>
<tr>
<th>State</th>
<th>Location of the home: CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Asking price (in $1,000's)</td>
</tr>
<tr>
<td>Size</td>
<td>Area of all rooms (in 1,000's sq. ft.)</td>
</tr>
<tr>
<td>Beds</td>
<td>Number of bedrooms</td>
</tr>
<tr>
<td>Baths</td>
<td>Number of bathrooms</td>
</tr>
</tbody>
</table>

---

**Details**

Data for samples of homes for sale in California, selected from zillow.com.

**From 2e - dataset has been updated for 3e **

**Source**


---

**HomesForSaleCanton**

*Homes For Sale in Canton, NY (2019)*

---

**Description**

Data for a sample of homes offered for sale in Canton, NY
HomesForSaleNY

Format
A data frame with 30 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Asking price (in $1,000's)</td>
</tr>
<tr>
<td>Size</td>
<td>Area of all rooms (in 1,000’s sq. ft.)</td>
</tr>
<tr>
<td>Beds</td>
<td>Number of bedrooms</td>
</tr>
<tr>
<td>Baths</td>
<td>Number of bathrooms</td>
</tr>
</tbody>
</table>

Details
Data for a sample of homes for sale in Canton, NY, selected from zillow.com.
** Updated for 3e (earlier version from 2010 is in HomesForSaleCanton2e). **

Source

HomesForSaleCanton2e Homes for sale in Canton, NY - 2e

Description
Prices of homes for sale in Canton, NY

Format
A dataset with 10 observations on the following variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Asking price for the home (in $1,000’s)</td>
</tr>
</tbody>
</table>

Details
Data for samples of homes for sale in Canton, NY, selected from zillow.com.
** From 2e - dataset has been updated for 3e **

Source

HomesForSaleNY Homes For Sale in New York (2019)

Description
Data for a sample of homes offered for sale in New York (state)
Format

A data frame with 30 observations on the following 5 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>Location of the home (NY)</td>
</tr>
<tr>
<td>Price</td>
<td>Asking price (in $1,000's)</td>
</tr>
<tr>
<td>Size</td>
<td>Area of all rooms (in 1,000's sq. ft.)</td>
</tr>
<tr>
<td>Beds</td>
<td>Number of bedrooms</td>
</tr>
<tr>
<td>Baths</td>
<td>Number of bathrooms</td>
</tr>
</tbody>
</table>

Details

Data for a sample of homes for sale in New York, selected from zillow.com. This is a subset of the HomesForSale dataset.

** Updated for 3e (earlier version from 2010 is in HomesForSaleNY2e). **

Source

**Homing Pigeons**

**Description**

Results from the 2019 Midwest Classic Homing Pigeon race

**Format**

A data frame with 1412 observations on the following 5 variables.

- **Position** Finishing position in the race
- **Loft** Name of the pigeon’s home loft
- **Sex** C=cock (male) or H=hen (female)
- **Distance** Distance (in miles) from release point to home loft
- **Speed** Speed (in yards per minute)

**Details**

Finishing results from 1412 pigeons completing the 2019 Midwest Classic race for homing pigeons on June 30, 2019. Each loft may enter multiple pigeons.

**Source**

Final race report from the Midwest Homing Pigeon Association, downloaded from [http://www.midwesthpa.com/MIDFinalReports.htm](http://www.midwesthpa.com/MIDFinalReports.htm)

---

**Honeybee Colonies**

**Description**

Number of honeybee colonies (1995-2012)

**Format**

A dataset with 18 observations on the following 2 variables.

- **Year**
- **Colonies** Estimated number of honeybee colonies in the US (in thousands)
Details

Data collected from the USDA on the estimated number of honeybee colonies in the US for the years 1995 through 2012.

Source


HoneybeeCircuits  Honeybee Circuits

Description

Number of circuits for honeybee dances and nest quality

Format

A dataset with 78 observations on the following 2 variables.

Circuits  Number of waggle dance circuits for a returning scout bee
Quality   Quality of the nest site: High or Low

Details

When honeybees are looking for a new home, they send out scouts to explore options. When a scout returns, she does a "waggle dance" with multiple circuit repetitions to tell the swarm about the option she found. The bees then decide between the options and pick the best one. Scientists wanted to find out how honeybees decide which is the best option, so they took a swarm of honeybees to an island with only two possible options for new homes: one of very high honeybee quality and one of low quality. They then kept track of the scouts who visited each option and counted the number of waggle dance circuits each scout bee did when describing the option.

Source


HoneybeeWaggle  Honeybee Waggle

Description

Honeybee dance duration and distance to nesting site
Format

A dataset with 7 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Distance</th>
<th>Distance to the potential nest site (in meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>Duration of the waggle dance (in seconds)</td>
</tr>
</tbody>
</table>

Details

When honeybee scouts find a food source or a nice site for a new home, they communicate the location to the rest of the swarm by doing a "waggle dance." They point in the direction of the site and dance longer for sites farther away. The rest of the bees use the duration of the dance to predict distance to the site.

Source


---

HotDogs1e                  Hot Dog Eating Contest

Description

Winning number of hot dogs consumed in an eating contest

Format

A dataset with 10 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Year</th>
<th>Year of the contest: 2002-2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>HotDogs</td>
<td>Winning number of hot dogs consumed</td>
</tr>
</tbody>
</table>

Details

Every Fourth of July, Nathan’s Famous in New York City holds a hot dog eating contest, in which contestants try to eat as many hot dogs (with buns) as possible in ten minutes. The winning number of hot dogs are given for each year from 2002-2011.

** From 1e - dataset has been updated for 2e **

Source

Downloaded from https://en.wikipedia.org/wiki/Nathan’s_Hot_Dog_Eating_Contest
**HotDogs2015**  
*Hot Dog Eating Contest - 2015*

**Description**
Winning number of hot dogs consumed in an eating contest

**Format**
A dataset with 14 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Year</th>
<th>Year of the contest: 2002-2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>HotDogs</td>
<td>Winning number of hot dogs consumed</td>
</tr>
</tbody>
</table>

**Details**
Every Fourth of July, Nathan’s Famous in New York City holds a hot dog eating contest, in which contestants try to eat as many hot dogs (with buns) as possible in ten minutes. The winning number of hot dogs are given for each year from 2002-2015.

**From 2e - dataset has been updated for 3e**

**Source**
Downloaded from [https://en.wikipedia.org/wiki/Nathan’s_Hot_Dog_Eating_Contest](https://en.wikipedia.org/wiki/Nathan’s_Hot_Dog_Eating_Contest)

---

**HotDogs2019**  
*Hot Dog Eating Contest*

**Description**
Winning number of hot dogs consumed in an eating contest (2002-2019)

**Format**
A data frame with 18 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Year</th>
<th>Year of the contest: 2002 to 2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>HotDogs</td>
<td>Winning number of hot dogs consumed</td>
</tr>
</tbody>
</table>

**Details**
Every Fourth of July, Nathan’s Famous in New York City holds a hot dog eating contest, in which contestants try to eat as many hot dogs (with buns) as possible in ten minutes. The winning number of hot dogs are given for each year from 2002-2019.

**Data set updated for 3e (earlier versions are HotDogs2015 and HotDogs1e)**
**HouseStarts2015**  
*Housing Starts - 2015*

**Description**
Quarterly housing starts in the United States from 2000-2015

**Format**
A dataset with 64 observations on the following 3 variables.

- **Year**: Year (2000 to 2015)
- **Quarter**: Q1=Jan-Mar, Q2=Apr-June, Q3=July-Sept, Q4=Oct-Dec
- **Houses**: New US residential house construction starts (in thousands)

**Details**
Number of new homes started in the US for each quarter from 2000-2015.

**From 2e - dataset has been updated for 3e**

**Source**
Census.gov website https://www.census.gov/econ/currentdata/
https://www.census.gov/econ/currentdata/dbsearch?program=RESCONST&startYear=2000&endYear=2016&categories=5

---

**HouseStarts2018**  
*Housing Starts (2000-2018)*

**Description**
Quarterly housing starts in the United States from 2000-2018

**Format**
A data frame with 76 observations on the following 3 variables.

- **Year**: Year (2000 to 2018)
- **Quarter**: Q1=Jan-Mar, Q2=Apr-June, Q3=July-Sept, Q4=Oct-Dec
- **Houses**: New US residential house construction starts (in thousands)
HumanTears25

Details

Number of new homes started in the US for each quarter from 2000-2018. Updated for 3e (earlier version is in HouseStarts2015)

Source


| HumanTears25 | Human Tears - Sadness and Sexual Arousal |

Description

Differences in sadness and sexual arousal ratings for 25 men sniffing female tears or a placebo in a matched pairs experiment.

Format

A data frame with 25 observations on the following 2 variables.

SexDiff Difference in sexual arousal rating (placebo rating - tears rating)
SadDiff Difference in sadness rating (placebo rating - tears rating)

Details

Twenty-five men had a pad attached to their upper lip that contained either female tears collected from women who watched a sad film or a salt solution (as a placebo) that had been trickled down the same women’s faces. The data were collected following a double-blind matched pairs design, where the order was randomized. The men were shown pictures of female faces and asked "To what extent is this face sad?" or "To what extent is this face sexually arousing?" Men’s answers were input using a Visual Analog Scale, which were then converted to a scale with results between about 200 and 800. The data show the difference in rating (placebo rating minus sadness rating) for each man for the sad question (SadDiff) or the sexual arousal question (SexDiff). Data are approximated from information given in the article.

Source

**HumanTears50**  
*Human Tears - Testosterone*

**Description**
Differences in testosterone levels for 50 men in a matched pairs experiment, where the differences are between sniffing female tears and sniffing a placebo.

**Format**
A data frame with 50 observations on the following 3 variables.

- **Placebo**  Testosterone level after sniffing a placebo
- **Tears**  Testosterone level after sniffing female tears
- **Difference**  Difference in testosterone level (Placebo - Tears)

**Details**
Fifty men had a pad attached to their upper lip that contained either female tears collected from women who watched a sad film or a salt solution (as a placebo) that had been trickled down the same women’s faces. The data were collected following a double-blind matched pairs design, where the order was randomized and the data were collected on consecutive days. After sniffing each substance (placebo or tears), men had their salivary testosterone levels measured, in pg/ml. Data are approximated from information given in the article.

**Source**

---

**Hurricanes2014**  
*Hurricanes - 2014*

**Description**
Hurricanes making landfall on the US east coast each year (1914-2014)

**Format**
A dataset with 64 observations on the following 3 variables.

- **Year**  Year (1914 to 2014)
- **Hurricanes**  Number of hurricanes making landfall on US East coast
**ICUAdmissions**

**Details**

Number of hurricanes making landfall on the East coast of the United States - yearly 1914-2014.
**From 2e - dataset has been updated for 3e**

**Source**


---

**Hurricanes2018**  *Hurricanes (1914 to 2018)*

**Description**

Hurricanes in the North Atlantic each year (1914-2018)

**Format**

A data frame with 105 observations on the following 2 variables.

- **Year** Year (1914 to 2018)
- **Hurricanes** Number of North Atlantic hurricanes

**Details**

Number of North Atlantic hurricanes - yearly 1914-2018.
**Updated for 3e (earlier version through 2014 is in Hurricanes2014).**

**Source**

Weather Underground website at [https://www.wunderground.com/hurricane/archive](https://www.wunderground.com/hurricane/archive)

---

**ICUAdmissions**  *Intensive Care Unit Admissions*

**Description**

Data from patients admitted to an intensive care unit
Format

A dataset with 200 observations on the following 21 variables.

- **ID**: Patient ID number
- **Status**: Patient status: 0=lived or 1=died
- **Age**: Patient’s age (in years)
- **Sex**: 0=male or 1=female
- **Race**: Patient’s race: 1=white, 2=black, or 3=other
- **Service**: Type of service: 0=medical or 1=surgical
- **Cancer**: Is cancer involved? 0=no or 1=yes
- **Renal**: Is chronic renal failure involved? 0=no or 1=yes
- **Infection**: Is infection involved? 0=no or 1=yes
- **CPR**: Patient gets CPR prior to admission? 0=no or 1=yes
- **Systolic**: Systolic blood pressure (in mm of Hg)
- **HeartRate**: Pulse rate (beats per minute)
- **Previous**: Previous admission to ICU within 6 months? 0=no or 1=yes
- **Type**: Admission type: 0=e elective or 1=emergency
- **Fracture**: Fractured bone involved? 0=no or 1=yes
- **PO2**: Partial oxygen level from blood gases under 60? 0=no or 1=yes
- **PH**: pH from blood gas under 7.25? 0=no or 1=yes
- **PCO2**: Partial carbon dioxide level from blood gas over 45? 0=no or 1=yes
- **Bicarbonate**: Bicarbonate from blood gas under 18? 0=no or 1=yes
- **Creatinine**: Creatinine from blood gas over 2.0? 0=no or 1=yes
- **Consciousness**: Level: 0=conscious, 1=deep stupor, or 2=coma

Details

Data from a sample of 200 patients following admission to an adult intensive care unit (ICU).

Source

DASL dataset downloaded from [http://lib.stat.cmu.edu/DASL/Datafiles/ICU.html](http://lib.stat.cmu.edu/DASL/Datafiles/ICU.html)

---

**ImmuneTea**  
**Immune Tea**

Description

Interferon gamma production and tea drinking

Format

A dataset with 21 observations on the following 2 variables.

- **InterferonGamma**  
  Measure of interferon gamma production
- **Drink**  
  Type of drink: Coffee or Tea
Details

Eleven healthy non-tea-drinking individuals were asked to drink five or six cups of tea a day, while ten healthy non-tea and non-coffee-drinkers were asked to drink the same amount of coffee, which has caffeine but not the L-theanine that is in tea. The groups were randomly assigned. After two weeks, blood samples were exposed to an antigen and production of interferon gamma was measured.

Source


---

InkjetPrinters

Description

Data from online reviews of inkjet printers

Format

A dataset with 20 observations on the following 6 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Model name of printer</td>
</tr>
<tr>
<td>PPM</td>
<td>Printing rate (pages per minute) for a benchmark set of print jobs</td>
</tr>
<tr>
<td>PhotoTime</td>
<td>Time (in seconds) to print 4x6 color photos</td>
</tr>
<tr>
<td>Price</td>
<td>Typical retail price (in dollars)</td>
</tr>
<tr>
<td>CostBW</td>
<td>Cost per page (in cents) for printing in black &amp; white</td>
</tr>
<tr>
<td>CostColor</td>
<td>Cost per page (in cents) for printing in color</td>
</tr>
</tbody>
</table>

Details

Information from reviews of inkjet printers at PCMag.com in August 2011.

Source

Inkjet printer reviews found at http://www.pcmag.com/reviews/printers, August 2011.

---

LifeExpectancyVehicles

Life Expectancy and Vehicle Registrations (2017)
Life Expectancy and Vehicle Registrations - 1e

Description


Format

A dataset with 40 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Year</th>
<th>LifeExpectancy</th>
<th>Vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year (1970 to 2009)</td>
<td>Average life expectancy (in years) for babies born in the year</td>
<td>Number of motor vehicles registered in the US (in millions)</td>
</tr>
</tbody>
</table>

Details

Life expectancy (in years for babies born each year) and number of vehicles registered in the US for each year from 1970 to 2009.

** Updated for 3e (earlier versions are LifeExpectancyVehicles2e and LifeExpectancyVehicles1e) **

Source


Lifetime data from the Centers for Disease Control and Prevention, National Center for Health Statistics https://www.cdc.gov/nchs/hus/contents2019.htm?search=Life_expectancy,
Details

Life expectancy (in years for babies born each year) and number of vehicles registered in the US for each year from 1970 to 2009.
** From 1e - dataset has been updated for 2e **

Source

Vehicle registrations from US Census Bureau, http://www.census.gov/compendia/statab/cats/transportation.html
Lifetime data from the Centers for Disease Control and Prevention, National Center for Health Statistics, Health Data Interactive, www.cdc.gov/nchs/hdi.htm

---

LifeExpectancyVehicles2e

*Life Expectancy and Vehicle Registrations - 2e*

Description

Yearly US life expectancy and number of registered vehicles (1970-2013)

Format

A dataset with 44 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Year</th>
<th>LifeExpectancy</th>
<th>Year</th>
<th>Vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average life expectancy (in years) for babies born in the year</td>
<td></td>
<td>Number of motor vehicles registered in the US (in millions)</td>
</tr>
</tbody>
</table>

Details

Life expectancy (in years for babies born each year) and number of vehicles registered in the US for each year from 1970 to 2013.
** From 2e - dataset has been updated for 3e **

Source

Vehicle registrations from US Census Bureau, http://www.census.gov/compendia/statab/cats/transportation.html
Lifetime data from the Centers for Disease Control and Prevention, National Center for Health Statistics, Health Data Interactive, www.cdc.gov/nchs/hdi.htm

---

LightatNight

*Light at Night for Mice*

Description

Data on body mass gain from an experiment with mice having different nighttime light conditions
**Format**

A dataset with 18 observations on the following 2 variables.

- **Group** Light=dim light at night or Dark=dark at night
- **BMGain** Body mass gain (in grams over a three week period)

**Details**

In this study, 18 mice were randomly split into two groups. One group was on a normal light/dark cycle (Dark) and the other group had light during the day and dim light at night (Light). The dim light was equivalent to having a television set on in a room. The mice in darkness ate most of their food during their active (nighttime) period, matching the behavior of mice in the wild. The mice with dim light at night, however, consumed much of their food during the well-lit rest period, when most mice are usually sleeping. The change in body mass was recorded after three weeks.

**See also LightatNight4Weeks or LightatNight8Weeks for more variables measured at other points in the same experiment, with a third experimental condition which had 9 additional mice with a bright light on all the time.**

**Source**

Fonken, L., et. al., "Light at night increases body mass by shifting time of food intake,” Proceedings of the National Academy of Sciences, October 26, 2010; 107(43): 18664-18669.

---

**LightatNight4Weeks**

**Light at Night for Mice - After 4 Weeks**

**Description**

Data from an experiment with mice having different nighttime light conditions

**Format**

A dataset with 27 observations on the following 9 variables.

- **Light** DM=dim light at night, LD=dark at night, or LL=bright light at night
- **BMGain** Body mass gain (in grams over a four week period)
- **Corticosterone** Blood corticosterone level (a measure of stress)
- **DayPct** Percent of calories eaten during the day
- **Consumption** Daily food consumption (grams)
- **GlucoseInt** Glucose intolerant? No or Yes
- **GTT15** Glucose level in the blood 15 minutes after a glucose injection
- **GTT120** Glucose level in the blood 120 minutes after a glucose injection
- **Activity** A measure of physical activity level
Details

In this study, 27 mice were randomly split into three groups. One group was on a normal light/dark cycle (LD), one group had bright light on all the time (LL), and one group had light during the day and dim light at night (DM). The dim light was equivalent to having a television set on in a room. The mice in darkness ate most of their food during their active (nighttime) period, matching the behavior of mice in the wild. The mice in both dim light and bright light, however, consumed more than half of their food during the well-lit rest period, when most mice are sleeping. Values in this dataset are recorded after four weeks in the experimental condition.

** This dataset was named LightatNight in the first edition **

** See also LightatNight8Weeks for the same data after 8 weeks or LightatNight with just BMGain after 3 weeks for the DM and LD groups. **

Source

Fonken, L., et. al., "Light at night increases body mass by shifting time of food intake," Proceedings of the National Academy of Sciences, October 26, 2010; 107(43): 18664-18669.

---

** LightatNight8Weeks **

*** Light at Night for Mice - After 8 Weeks ***

Description

Data from an experiment with mice having different nighttime light conditions

Format

A dataset with 27 observations on the following 9 variables.

- Light
  - DM = dim light at night, LD = dark at night, or LL = bright light at night
- BMGain
  - Body mass gain (in grams over an eight week period)
- Corticosterone
  - Blood corticosterone level (a measure of stress)
- DayPct
  - Percent of calories eaten during the day
- Consumption
  - Daily food consumption (grams)
- GlucoseInt
  - Glucose intolerant? No or Yes
- GTT15
  - Glucose level in the blood 15 minutes after a glucose injection
- GTT120
  - Glucose level in the blood 120 minutes after a glucose injection
- Activity
  - A measure of physical activity level

Details

In this study, 27 mice were randomly split into three groups. One group was on a normal light/dark cycle (LD), one group had bright light on all the time (LL), and one group had light during the day and dim light at night (DM). The dim light was equivalent to having a television set on in a room. The mice in darkness ate most of their food during their active (nighttime) period, matching the behavior of mice in the wild. The mice in both dim light and bright light, however, consumed more than half of their food during the well-lit rest period, when most mice are sleeping. Values in this
dataset are recorded after eight weeks in the experimental condition. ** See also LightatNight4Weeks for the same data after 4 weeks or LightatNight with just BMGain after 3 weeks for just the DM and LD groups. **

**Source**


---

**Description**

Perceived malevolence of uniforms and penalties for National Football League (NFL) teams

**Format**

A dataset with 28 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>NFLTeam</th>
<th>Team name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFL_Malevolence</td>
<td>Score reflecting the &quot;malevolence&quot; of a team’s uniform</td>
</tr>
<tr>
<td>ZPenYds</td>
<td>Z-score for penalty yards</td>
</tr>
</tbody>
</table>

**Details**

Participants with no knowledge of the teams rated the jerseys on characteristics such as timid/aggressive, nice/mean and good/bad. The averages of these responses produced a "malevolence" index with higher scores signifying impressions of more malevolent uniforms. To measure aggressiveness, the authors used the amount of penalty yards converted to z-scores and averaged for each team over the seasons from 1970-1986.

**Source**


---

**Description**

Perceived malevolence of uniforms and penalties for National Hockey League (NHL) teams
Format

A dataset with 28 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>NHLTeam</th>
<th>Team name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHL_Malevolence</td>
<td>Score reflecting the &quot;malevolence&quot; of a team's uniform</td>
</tr>
<tr>
<td>ZPenMin</td>
<td>Z-score for penalty minutes</td>
</tr>
</tbody>
</table>

Details

Participants with no knowledge of the teams rated the jerseys on characteristics such as timid/aggressive, nice/mean and good/bad. The averages of these responses produced a "malevolence" index with higher scores signifying impressions of more malevolent uniforms. To measure aggressiveness, the authors used the amount of penalty minutes converted to z-scores and averaged for each team over the seasons from 1970-1986.

Source


Description

Longevity and gestation period for mammals

Format

A dataset with 40 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Animal</th>
<th>Species of mammal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestation</td>
<td>Time from fertilization until birth (in days)</td>
</tr>
<tr>
<td>Longevity</td>
<td>Average lifespan (in years)</td>
</tr>
</tbody>
</table>

Details

Dataset with average lifespan (in years) and typical gestation period (in days) for 40 different species of mammals.

Source

Description

Apt price for sale in Manhattan in 2019

Format

A data frame with 20 observations on the following variable.

Rent Monthly rent (in dollars)

Details

Monthly rents for a sample of 20 one-bedroom apartments in Manhattan, NY that were advertised on Craig’s List in November, 2019.

Source


Description

Monthly rent for one-bedroom apartments in Manhattan, NY

Format

A dataset with 20 observations on the following variable.

Rent Monthly rent in dollars

Details

Monthly rents for a sample of 20 one-bedroom apartments in Manhattan, NY that were advertised on Craig’s List in July, 2011.

** From 2e - dataset has been updated for 3e **
Source
Apartments advertised on Craig’s List at newyork.craigslist.org, July 5, 2011.

<table>
<thead>
<tr>
<th>MarriageAges</th>
<th>Marriage Ages</th>
</tr>
</thead>
</table>

Description
Ages for husbands and wives from marriage licenses

Format
A dataset with 100 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Husband</th>
<th>Age of husband at marriage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wife</td>
<td>Age of wife at marriage</td>
</tr>
</tbody>
</table>

Details
Data from a sample of 100 marriage licenses in St. Lawrence County, NY gives the ages of husbands and wives for newly married couples.

Source
Thanks to Linda Casserly, St. Lawrence County Clerk’s Office

<table>
<thead>
<tr>
<th>MastersGolf</th>
<th>Masters Golf Scores</th>
</tr>
</thead>
</table>

Description
Scores from the 2011 Masters golf tournament

Format
A dataset with 20 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>First</th>
<th>First round score (in relation to par)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final</td>
<td>Final four round score (in relation to par)</td>
</tr>
</tbody>
</table>

Details
Data for a random sample of 20 golfers who made the cut at the 2011 Masters golf tournament.
Source


MateChoice  Fruitfly Survival - by Mate Choice

Description

Number of fruitflies surviving depending on number of mating choices.

Format

A dataset with 50 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Choice</th>
<th>Number of surviving larvae (out of 200) when female had a choice of mates</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoChoice</td>
<td>Number of surviving larvae (out of 200) when female had only one choice for a mate</td>
</tr>
<tr>
<td>Difference</td>
<td>Choice - NoChoice</td>
</tr>
</tbody>
</table>

Details

In an experiment, two hundred larvae from female fruitflies that were exposed to many male fruitflies were tracked to see how many survived. This was compared to a different set of 200 larvae from females that were exposed to only one male each. Values in the dataset give how many of the 200 larvae survived. This process was replicated 50 times, so each row of the dataset corresponds to the survival counts (and difference) for one run, starting with 200 larvae of each type.

Source


MentalMuscle  Mental Muscle

Description

Comparing actual movements to mental imaging movements

Format

A dataset with 32 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Action</th>
<th>Treatment: Actual motions or Mental imaging motions</th>
</tr>
</thead>
<tbody>
<tr>
<td>PreFatigue</td>
<td>Time (in seconds) to complete motions before fatigue</td>
</tr>
<tr>
<td>PostFatigue</td>
<td>Time (in seconds) to complete motions after fatigue</td>
</tr>
</tbody>
</table>
Details

In this study, participants were asked to either perform actual arm pointing motions or to mentally imagine equivalent arm pointing motions. Participants then developed muscle fatigue by holding a heavy weight out horizontally as long as they could. After becoming fatigued, they were asked to repeat the previous mental or actual motions. Eight participants were assigned to each group, and the time in seconds to complete the motions was measured before and after fatigue.

Source


| MiamiHeat | Miami Heat Basketball |

Description

Game log data for the Miami Heat basketball team in 2010-11

Format

A dataset with 82 observations on the following 33 variables.

<table>
<thead>
<tr>
<th>Game</th>
<th>Date</th>
<th>Location</th>
<th>Opp</th>
<th>Win</th>
<th>FG</th>
<th>FGA</th>
<th>FG3</th>
<th>FG3A</th>
<th>FT</th>
<th>FTA</th>
<th>Rebounds</th>
<th>OffReb</th>
<th>Assists</th>
<th>Steals</th>
<th>Blocks</th>
<th>Turnovers</th>
<th>Fouls</th>
<th>Points</th>
<th>OppFG</th>
<th>OppFGA</th>
<th>OppFG3</th>
<th>OppFG3A</th>
</tr>
</thead>
</table>
OppFT  Opponent’s Free throws made
OppFTA  Opponent’s Free throws attempted
OppOffReb  Opponent’s Offensive rebounds
OppRebounds  Opponent’s Total rebounds
OppAssists  Opponent’s assists
OppSteals  Opponent’s steals
OppBlocks  Opponent’s shots blocked
OppTurnovers  Opponent’s turnovers
OppFouls  Opponent’s fouls
OppPoints  Opponent’s points scored

Details

Information from online boxscores for all 82 regular season games payed by the Miami Heat basketball team during the 2010-11 season.

** This is from the first edition, updated in second edition to GSWarriors dataset **

Source

Data for the 2010-11 Miami games downloaded from

<table>
<thead>
<tr>
<th>MindsetMatters</th>
<th>Mindset Matters</th>
</tr>
</thead>
</table>

Description

Data from a study of perceived exercise with maids

Format

A dataset with 75 observations on the following 14 variables.

Cond  Treatment condition: 0=uninformed or 1=informed
Age  Age (in years)
Wt  Original weight (in pounds)
Wt2  Weight after 4 weeks (in pounds)
BMI  Original body mass index
BMI2  Body mass index after 4 weeks
Fat  Original body fat percentage
Fat2  Body fat percentage after 4 weeks
WHR  Original waist to hip ratio
WHR2  Waist to hip ratio after 4 weeks
Syst  Original systolic blood pressure
Syst2  Systolic blood pressure after 4 weeks
Diast  Original diastolic blood pressure
Diast2  Diastolic blood pressure after 4 weeks
In 2007 a Harvard psychologist recruited 75 female maids working in different hotels to participate in a study. She informed 41 maids (randomly chosen) that the work they do satisfies the Surgeon General’s recommendations for an active lifestyle (which is true), giving them examples for how and why their work is good exercise. The other 34 maids were told nothing (uninformed). Various characteristics (weight, body mass index, ...) were recorded for each subject at the start of the experiment and again four weeks later. Maids with missing values for weight change have been removed.

Source


### MustangPrice

**Mustang Prices**

**Description**

Price, age, and mileage for used Mustang cars at an internet website

**Format**

A dataset with 25 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Age</th>
<th>Miles</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of the car (in years)</td>
<td>Mileage on the car (in 1,000’s)</td>
<td>Asking price (in $1,000’s)</td>
</tr>
</tbody>
</table>

**Details**

A statistics student, Gabe McBride, was interested in prices for used Mustang cars being offered for sale on an internet site. He sampled 25 cars from the website and recorded the age (in years), mileage (in thousands of miles) and asking price (in $1,000’s) for each car in his sample.

**Source**

Student project with data collected from autotrader.com in 2008.

### NBAPlayers2011

**NBA Players Data for 2010-11 Season**

**Description**

Data from the 2010-2011 regular season for 176 NBA basketball players.
Format

A dataset with 176 observations on the following 25 variables.

- **Player**: Name of player
- **Age**: Age (in years)
- **Team**: Team name
- **Games**: Games played (out of 82)
- **Starts**: Games started
- **Mins**: Minutes played
- **MinPerGame**: Minutes per game
- **FGMade**: Field goals made
- **FGAttempt**: Field goals attempted
- **FGPct**: Field goal percentage
- **FG3Made**: Three-point field goals made
- **FG3Attempt**: Three-point field goals attempted
- **FG3Pct**: Three-point field goal percentage
- **FTMade**: Free throws made
- **FTAttempt**: Free throws attempted
- **FTPct**: Free throw percentage
- **OffRebound**: Offensive rebounds
- **DefRebound**: Defensive rebounds
- **Rebounds**: Total rebounds
- **Assists**: Number of assists
- **Steals**: Number of steals
- **Blocks**: Number of blocked shots
- **Turnovers**: Number of turnovers
- **Fouls**: Number of personal fouls
- **Points**: Number of points scored

Details

Data for 176 NBA basketball players from the 2010-2011 regular season. Includes all players who averaged more than 24 minutes per game.

** From 1e - dataset has been updated (in NBAPlayers2015) for 2e **

Source


---

**NBAPlayers2015**  
**NBA Players Data for 2014-15 Season**

Description

Data from the 2014-2015 regular season for 182 NBA basketball players.
Format

A dataset with 182 observations on the following 25 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player</td>
<td>Name of player</td>
</tr>
<tr>
<td>Position</td>
<td>PG=point guard, SG=shooting guard, PF=power forward, SF=small forward, C=center</td>
</tr>
<tr>
<td>Age</td>
<td>Age (in years)</td>
</tr>
<tr>
<td>Team</td>
<td>Team name</td>
</tr>
<tr>
<td>Games</td>
<td>Games played (out of 82)</td>
</tr>
<tr>
<td>Starts</td>
<td>Games started</td>
</tr>
<tr>
<td>Mins</td>
<td>Minutes played</td>
</tr>
<tr>
<td>MinPerGame</td>
<td>Minutes per game</td>
</tr>
<tr>
<td>FGMade</td>
<td>Field goals made</td>
</tr>
<tr>
<td>FGAttempt</td>
<td>Field goals attempted</td>
</tr>
<tr>
<td>FGpct</td>
<td>Field goal percentage</td>
</tr>
<tr>
<td>FG3Made</td>
<td>Three-point field goals made</td>
</tr>
<tr>
<td>FG3Attempt</td>
<td>Three-point field goals attempted</td>
</tr>
<tr>
<td>FG3Pct</td>
<td>Three-point field goal percentage</td>
</tr>
<tr>
<td>FTMade</td>
<td>Free throws made</td>
</tr>
<tr>
<td>FTAtempt</td>
<td>Free throws attempted</td>
</tr>
<tr>
<td>FTPct</td>
<td>Free throw percentage</td>
</tr>
<tr>
<td>OffRebound</td>
<td>Offensive rebounds</td>
</tr>
<tr>
<td>DefRebound</td>
<td>Defensive rebounds</td>
</tr>
<tr>
<td>Rebounds</td>
<td>Total rebounds</td>
</tr>
<tr>
<td>Assists</td>
<td>Number of assists</td>
</tr>
<tr>
<td>Steals</td>
<td>Number of steals</td>
</tr>
<tr>
<td>Blocks</td>
<td>Number of blocked shots</td>
</tr>
<tr>
<td>Turnovers</td>
<td>Number of turnovers</td>
</tr>
<tr>
<td>Fouls</td>
<td>Number of personal fouls</td>
</tr>
<tr>
<td>Points</td>
<td>Number of points scored</td>
</tr>
</tbody>
</table>

Details

Data for 182 NBA basketball players from the 2014-2015 regular season. Includes all players who averaged more than 24 minutes per game that season.

** From 2e - dataset has been updated for 3e **

Source


---

NBAPlayers2019

NBA Players Data for 2018-19 Season

Description

Data from the 2018-2019 regular season for 193 NBA basketball players.
Format

A data frame with 193 observations on the following 26 variables.

- **Player**: Name of player
- **Pos**: PG=point guard, SG=shooting guard, PF=power forward, SF=small forward, C=center
- **Age**: Age (in years)
- **Team**: Team name
- **Games**: Games played (out of 82)
- **Starts**: Games started
- **Mins**: Minutes played
- **MinPerGame**: Minutes per game
- **FGMade**: Field goals made
- **FGAttempt**: Field goals attempted
- **FGPct**: Field goal percentage
- **FG3Made**: Three-point field goals made
- **FG3Attempt**: Three-point field goals attempted
- **FG3Pct**: Three-point field goal percentage
- **FTMade**: Free throws made
- **FTAttempt**: Free throws attempted
- **FTPct**: Free throw percentage
- **OffRebound**: Offensive rebounds
- **DefRebound**: Defensive rebounds
- **Rebounds**: Total rebounds
- **Assists**: Number of assists
- **Steals**: Number of steals
- **Blocks**: Number of blocked shots
- **Turnovers**: Number of turnovers
- **Fouls**: Number of personal fouls
- **Points**: Number of points scored

Details

Data for 193 NBA basketball players from the 2018-2019 regular season. Includes all players who averaged more than 24 minutes per game that season.

** Data set updated for 3e (earlier versions are NBAPlayers2015 and NBAPlayers2011). **

Source

NBAStandings2011  NBA 2010-11 Regular Season Standings

Description

Won-Loss record and statistics for NBA Teams in 2010-2011

Format

A dataset with 30 observations on the following 6 variables.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Team</td>
<td>Team name</td>
</tr>
<tr>
<td>Wins</td>
<td>Number of wins in an 82 game regular season</td>
</tr>
<tr>
<td>Losses</td>
<td>Number of losses</td>
</tr>
<tr>
<td>WinPct</td>
<td>Proportion of games won</td>
</tr>
<tr>
<td>PtsFor</td>
<td>Average points scored per game</td>
</tr>
<tr>
<td>PtsAgainst</td>
<td>Average points allowed per game</td>
</tr>
</tbody>
</table>

Details

Won-Loss record and regular season statistics for 30 teams in the National Basketball Association for the 2010-2011 season.

** From 1e - dataset has been updated for 2e and 3e**

Source


NBAStandings2016  NBA 2015-2016 Regular Season Standings

Description

Won-Loss record and statistics for NBA Teams in 2015-2016

Format

A dataset with 30 observations on the following 6 variables.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Team</td>
<td>Team name</td>
</tr>
<tr>
<td>Wins</td>
<td>Number of wins in an 82 game regular season</td>
</tr>
<tr>
<td>Losses</td>
<td>Number of losses</td>
</tr>
<tr>
<td>WinPct</td>
<td>Proportion of games won</td>
</tr>
<tr>
<td>PtsFor</td>
<td>Average points scored per game</td>
</tr>
<tr>
<td>PtsAgainst</td>
<td>Average points allowed per game</td>
</tr>
</tbody>
</table>
Details

Won-Loss record and regular season statistics for 30 teams in the National Basketball Association for the 2015-2016 season.

** From 2e - dataset has been updated for 3e **

Source


<table>
<thead>
<tr>
<th>NBAStandings2019</th>
<th>NBA 2018-2019 Regular Season Standings</th>
</tr>
</thead>
</table>

Description

Won-Loss record and statistics for NBA Teams in 2018-2019

Format

A data frame with 30 observations on the following 6 variables.

Team   Team name
 Wins   Number of wins in an 82 game regular season
 Losses Number of losses
 WinPct Proportion of games won
 PtsFor Average points scored per game
 PtsAgainst Average points allowed per game

Details

Won-Loss record and regular season statistics for 30 teams in the National Basketball Association for the 2018-2019 season.

** Data set updated for 3e (earlier version are NBAStandings2016 and NBAStandings1e) **

Source


<table>
<thead>
<tr>
<th>NFLContracts2015</th>
<th>NFL Contracts in 2015</th>
</tr>
</thead>
</table>

Description

Dollar size of contracts for all NFL players in 2015

Format

A dataset with 2099 observations on the following 5 variables.
<table>
<thead>
<tr>
<th>Player</th>
<th>Player’s name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Code for the primary position of the player (QB=quarterback, etc.)</td>
</tr>
<tr>
<td>Team</td>
<td>Nickname of the team</td>
</tr>
<tr>
<td>TotalMoney</td>
<td>Total value of the contract (in millions of dollars)</td>
</tr>
<tr>
<td>YearlySalary</td>
<td>Salary (in millions of dollars) for the 2015 season</td>
</tr>
</tbody>
</table>

**Details**

This dataset contains salary information for all National Football League (NFL) players under contract for the 2015 season. Many contracts extend over multiple years, so TotalMoney gives the overall size of the contract and YearlySalary indicates how much of that is to be paid for the 2015 season. All amounts are in millions of dollars.

**Updated for 3e**

**Source**


---

**NFLContracts2019**

**NFL Contracts in 2019**

**Description**

Dollar size of contracts for all NFL players in 2019

**Format**

A data frame with 1988 observations on the following 5 variables.

- **Player** Player’s name
- **Position** Code for the primary position of the player (QB=quarterback, etc.)
- **Team** Nickname of the team
- **TotalMoney** Total value of the contract (in millions of dollars)
- **YearlySalary** Salary (in millions of dollars) for the 2019 season

**Details**

This dataset contains salary information for all National Football League (NFL) players under contract for the 2019 season. Many contracts extend over multiple years, so TotalMoney gives the overall size of the contract and YearlySalary indicates how much of that is to be paid for the 2019 season. All amounts are in millions of dollars.

**Updated for 3e (earlier version is NFLContracts2015).**

**Source**

Wins for NFL Teams (2005-2014)

Description

Number of preseason and regular season wins for NFL teams, each year from 2005 to 2014.

Format

A dataset with 320 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Team</th>
<th>Code for one of 32 NFL teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Season</td>
<td>Year between 2005 and 2014</td>
</tr>
<tr>
<td>Preseason</td>
<td>Number of preseason wins (out of 4 games)</td>
</tr>
<tr>
<td>RegularWins</td>
<td>Number of regular season wins (out of 16 games)</td>
</tr>
</tbody>
</table>

Details

Number of wins in the preseason (out of 4 preseason games) and regular season (out of 16 regular season games) for each of the 32 National Football (NFL) teams over a ten year period from 2005 to 2014.

** From 2e - dataset has been updated for 3e **

Source


Description

Number of preseason and regular season wins for NFL teams, each year from 2005 to 2019.

Format

A data frame with 480 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Team</th>
<th>Code for one of 32 NFL teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Season</td>
<td>Year between 2005 and 2019</td>
</tr>
<tr>
<td>Preseason</td>
<td>Number of preseason wins (out of 4 games)</td>
</tr>
<tr>
<td>RegularWins</td>
<td>Number of regular season wins (out of 16 games)</td>
</tr>
</tbody>
</table>
Details
Number of wins in the preseason (out of 4 preseason games) and regular season (out of 16 regular season games) for each of the 32 National Football (NFL) teams over a fifteen year period from 2005 to 2019.

** Updated for 3e (earlier version is now NFLPreseason2014). **

Source

NFLScores2011 NFL Game Scores in 2011

Description
Results for all NFL games for the 2011 regular season

Format
A dataset with 256 observations on the following 11 variables.

<table>
<thead>
<tr>
<th>Week</th>
<th>HomeTeam</th>
<th>AwayTeam</th>
<th>HomeScore</th>
<th>AwayScore</th>
<th>HomeYards</th>
<th>AwayYards</th>
<th>HomeTO</th>
<th>AwayTO</th>
<th>Date</th>
<th>Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week of the season (1 through 17)</td>
<td>Home team name</td>
<td>Visiting team name</td>
<td>Points scored by the home team</td>
<td>Points scored by the visiting team</td>
<td>Yards gained by the home team</td>
<td>Yards gained by the visiting team</td>
<td>Turnovers lost by the home team</td>
<td>Turnovers lost by the visiting team</td>
<td>Date of the game</td>
<td>Day of the week: Mon, Sat, Sun, or Thu</td>
</tr>
</tbody>
</table>

Details
Data for all 256 regular season games in the National Football League (NFL) for the 2011 season.

** From 2e - dataset has been updated for 3e **

Source
Description

Results for all NFL games for the 2018 regular season

Format

A data frame with 256 observations on the following 11 variables.

- **Week**  Week of the season (1 through 17)
- **HomeTeam**  Home team name
- **AwayTeam**  Visiting team name
- **HomeScore**  Points scored by the home team
- **AwayScore**  Points scored by the visiting team
- **HomeYards**  Yards gained by the home team
- **AwayYards**  Yards gained by the visiting team
- **HomeTO**  Turnovers lost by the home team
- **AwayTO**  Turnovers lost by the visiting team
- **Date**  Date of the game
- **Day**  Day of the week (Mon, Sat, Sun, or Thu)

Details

Data for all 256 regular season games in the National Football League (NFL) for the 2018 season.

** Updated for 3e (earlier version is NFLScores2011). **

Source

**NHANES**

*National Health and Nutrition Examination Survey (NHANES) Subset*

**Description**

A subset of the 2009-2010 National Health and Nutrition Examination Survey (NHANES).

**Format**

A data frame with 4716 observations on the following 5 variables.

- **Case**: Case ID number
- **Organic**: Buy any food labeled organic (past 30 days)? (No or Yes)
- **Health**: Self-rating of health (Excellent, Very good, Fair, Good, or Poor)
- **HealthBinary**: Health with two categories: Poor / Fair / Good or Very good / Excellent
- **Income**: Monthly income? (dollars)

**Details**

This dataset is a subset of the 2009-2010 National Health and Nutrition Examination Survey (NHANES). NHANES is a national survey conducted by the Centers for Disease Control and Prevention (CDC) on a random sample of Americans. This subset contains data on select variables for the subset of people with responses to the questions about buying organic food and self-reported health status.

**Source**

The data were downloaded from [https://www.cdc.gov/nchs/nhanes/index.htm](https://www.cdc.gov/nchs/nhanes/index.htm).

---

**NutritionStudy**

*Nutrition Study*

**Description**

Variables related to nutrition and health for 315 individuals

**Format**

A dataset with 315 observations on the following 17 variables.

- **ID**: ID number for each subject in this sample
- **Age**: Subject's age (in years)
- **Smoke**: Smoker? coded as No or Yes
- **Quetelet**: Weight/(Height^2)
- **Vitamin**: Vitamin use: coded as 1=Regularly, 2=Occasionally, or 3=No
- **Calories**: Number of calories consumed per day
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat</td>
<td>Grams of fat consumed per day</td>
</tr>
<tr>
<td>Fiber</td>
<td>Grams of fiber consumed per day</td>
</tr>
<tr>
<td>Alcohol</td>
<td>Number of alcoholic drinks consumed per week</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>Cholesterol consumed (mg per day)</td>
</tr>
<tr>
<td>BetaDiet</td>
<td>Dietary beta-carotene consumed (mcg per day)</td>
</tr>
<tr>
<td>RetinolDiet</td>
<td>Dietary retinol consumed (mcg per day)</td>
</tr>
<tr>
<td>BetaPlasma</td>
<td>Plasma beta-carotene (ng/ml)</td>
</tr>
<tr>
<td>RetinolPlasma</td>
<td>Plasma retinol (ng/ml)</td>
</tr>
<tr>
<td>Sex</td>
<td>Coded as Female or Male</td>
</tr>
<tr>
<td>VitaminUse</td>
<td>Coded as No Occasional Regular</td>
</tr>
<tr>
<td>PriorSmoke</td>
<td>Smoking status: coded as 1=Never, 2=Former, or 3=Current</td>
</tr>
</tbody>
</table>

**Details**

Data from a cross-sectional study to investigate the relationship between personal characteristics and dietary factors, and plasma concentrations of retinol, beta-carotene and other carotenoids. Study subjects were patients who had an elective surgical procedure during a three-year period to biopsy or remove a lesion of the lung, colon, breast, skin, ovary or uterus that was found to be non-cancerous.

**Source**


---

**OlympicMarathon2008 2008 Olympic Men’s Marathon**

**Description**

Times for all finishers in the men’s marathon at the 2008 Olympics

**Format**

A data frame with 76 observations on the following 5 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>Order of finish</td>
</tr>
<tr>
<td>Athlete</td>
<td>Name of marathoner</td>
</tr>
<tr>
<td>Nationality</td>
<td>Country of marathoner</td>
</tr>
<tr>
<td>Time</td>
<td>Time as H:MM:SS</td>
</tr>
<tr>
<td>Minutes</td>
<td>Time in minutes</td>
</tr>
</tbody>
</table>

**Details**

Results for all finishers in the 2008 Men’s Olympic marathon in Beijing, China.

** This 1e version has been updated for 2e and 3e**
Source


OlympicMarathon2012  2012 Olympic Men’s Marathon

Description

Times for all finishers in the men’s marathon at the 2012 Olympics

Format

A data frame with 85 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athlete</td>
<td>Name of marathoner</td>
</tr>
<tr>
<td>Country</td>
<td>Nationality of marathoner (3 letter country code)</td>
</tr>
<tr>
<td>Time</td>
<td>Time as H:MM:SS</td>
</tr>
<tr>
<td>Minutes</td>
<td>Time in minutes</td>
</tr>
</tbody>
</table>

Details


** From 2e - dataset has been updated for 3e **

Source


OlympicMarathon2016  2016 Olympic Men’s Marathon

Description

Times for all finishers in the men’s marathon at the 2016 Olympics

Format

A data frame with 140 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athlete</td>
<td>Name of marathoner</td>
</tr>
<tr>
<td>Country</td>
<td>Nationality of marathoner (3 letter country code)</td>
</tr>
<tr>
<td>Time</td>
<td>Time as H:MM:SS</td>
</tr>
<tr>
<td>Minutes</td>
<td>Time in minutes</td>
</tr>
</tbody>
</table>
Details

Results for all finishers in the 2016 Men’s Olympic marathon in Rio de Janeiro, Brazil.
** Updated for 3e (earlier versions are now in OlympicMarathon2012 and OlympicMarathon2008)**

Source


OrganicEffect Eating Organic Foods

Description

Data comparing pesticide levels in family members when eating non-organic vs organic food

Format

A dataset with 160 observations on the following 6 variables.

Person Code for family member, Father, Mother, GirlA, GirlB, Boy
Pesticide One of eight different pesticides measured
Day Day of the measurement (Day1, Day3, Day4, or Day6)
NonOrganic Level of the pesticide after eating a non-organic diet
Organic Level of the pesticide after eating an organic diet
Diff Difference = NonOrganic - Organic

Details

A study looked at a Swedish family that ate a conventional diet (non-organic), and then had them eat only organic for two weeks. Pesticide concentrations for several different pesticides were measured in micrograms/g creatinine by testing morning urine. Multiple measurements were taken for each person before the switch to organic foods, and then again after participants had been eating organic for at least one week.

Source

Description

Data for 24 players on the 2014-2015 Ottawa Senators NHL team

Format

A dataset with 24 observations on the following 10 variables.

- **Player**  Players name
- **Position**  D=defense, C=center, RW=right wing, LW=left wing
- **Age**  Age (in years)
- **Games**  Games played in the 2014-15 NHL season (out of 82)
- **Goals**  Goals
- **Assists**  Assists
- **Points**  Goals + Assists
- **PlusMinus**  Difference between (even strength) goals for and against while on ice
- **PenMins**  Number of penalty minutes
- **MinPerGame**  Average minutes on the ice per game

Details

Data for all players (except goalies) who played at least 10 games with the Ottawa Senators hockey team in the 2014-15 NHL season.

**This is an updated version (previous version is now in OttawaSenators1e)**

Source


---

Description

Data for 24 players on the 2009-10 Ottawa Senators

Format

A dataset with 24 observations on the following 2 variables.

- **Points**  Number of points (goals + assists) scored
- **PenMins**  Number of penalty minutes
Details

Points scored and penalty minutes for 24 players (excluding goalies) playing ice hockey for the Ottawa Senators during the 2009-10 NHL regular season.

** From 1e - dataset has been updated for 2e and 3e **

Source


Description

Data for 26 players on the 2018-2019 Ottawa Senators NHL team

Format

A data frame with 26 observations on the following 10 variables.

- **Player** Players name
- **Position** D=defense, C=center, RW=right wing, LW=left wing
- **Age** Age (in years)
- **Games** Games played in the 2018-19 NHL season (out of 82)
- **Goals** Goals
- **Assists** Assists
- **Points** Goals + Assists
- **PlusMinus** Difference between (even strength) goals for and against while on ice
- **PenMins** Number of penalty minutes
- **MinPerGame** Average minutes on the ice per game

Details

Data for all players (except goalies) who played at least 10 games with the Ottawa Senators hockey team in the 2018-2019 NHL season.

** Updated for 3e (previous versions are now OttawaSenators2015 and OttawaSenators1e) **

Source

Description

Information on a sample of high school seniors from the state of Pennsylvania between 2010 and 2019.

Format

A data frame with 457 observations on the following 36 variables.

- **Year**  Year student submitted data
- **Gender**  Female or Male
- **Age**  Age (in years)
- **Hand**  Dominant hand (Left, Right, or Both)
- **Height**  Height (in cm)
- **Foot**  Foot length (in cm)
- **Armspan**  Armspan (in cm)
- **Languages**  Languages spoken
- **GetToSchool**  Main mode of transportation to school (Bus, Car, or Walk - Walk includes bicycle)
- **TravelTime**  Travel time to school (in minutes)
- **ReactionTime**  Time (in seconds) to click when a color changes
- **MemoryScore**  Score in an online memory game
- **Activity**  Favorite physical activity
- **Music**  Favorite genre of music
- **BirthMonth**  Birth month
- **Season**  Favorite season
- **Allergies**  Have allergies? (No or Yes)
- **Vegetarian**  Vegetarian? (No or Yes)
- **FavFood**  Favorite food
- **Drink**  Beverage used most often during the day
- **FavSubject**  Favorite subject in school
- **Sleep1**  Typical hours of sleep on a school night
- **Sleep2**  Typical hours of sleep on a non-school night
- **Occupants**  Number of occupants at home
- **Communicate**  Most often method to communicate with friends
- **TextsSent**  Number of texts sent (previous day)
- **HangHours**  Hours last week spent hanging out with friends
Hours  Hours last week spent doing homework
SportsHours  Hours last week spent playing sports or outdoor activities
VideoGameHours  Hours last week spent playing computer/video games
ComputerHours  Hours last week spent using a computer
TVHours  Hours last week spent watching TV
WorkHours  Hours last week spent working at a paid job
SchoolPressure  Amount of pressure due to schoolwork
Superpower  Most desired superpower (Fly, Freeze time, Invisibility, Super strength, or Telepathy)
Preference  Prefers to be Famous, Happy, Healthy, or Rich

Details
The dataset gives responses for a random sample of high school seniors in Pennsylvania who participated in the Census at Schools project.

Source
Data from U.S. Census at School (https://ww2.amstat.org/censusatschool/) downloaded and used with the permission of the American Statistical Association.

---

PizzaGirl  Pizza Girl Tips

Description
Data on tips for pizza deliveries

Format
A dataset with 24 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Tip</th>
<th>Amount of tip (in dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shift</td>
<td>Data collected over three different shifts</td>
</tr>
</tbody>
</table>

Details
"Pizza Girl" collected data on her deliveries and tips over three different evening shifts.

Source
**Pumpkin Beer**

**Description**

Ratings of different kinds of pumpkin beer by a wife and husband

**Format**

A data frame with 18 observations on the following 8 variables.

- **Name**: Name of pumpkin beer
- **Brewer**: Name of brewery that produced the beer
- **WifeRating**: Rating on a 0-10 scale by the wife
- **HusbandRating**: Rating on a 0-10 scale by the husband
- **WifeComments**: Text of comments by the wife
- **HusbandComments**: Text of comments by the husband
- **Average**: Average of the two ratings (wife and husband)
- **Year**: Year the ratings were done (2011 to 2019)

**Details**

A Lock wife and husband are fans of pumpkin flavored beer, so they have each rated a variety of different brands of pumpkin beer over the years.

**Source**

Personal records

---

**Quiz vs Lecture Pulse Rates**

**Description**

Paired data with pulse rates in a lecture and during a quiz for 10 students

**Format**

A dataset with 10 observations on the following 3 variables.

- **Student**: ID number for the student
- **Quiz**: Pulse rate (beats per minute) during a quiz
- **Lecture**: Pulse rate (beats per minute) during a lecture
Details

Ten students in an introductory statistics class measured their pulse rate (beats per minute) in two settings: first, in the middle of a regular class lecture and second, while taking an in-class quiz.

Source

In-class data collection

Description

Counts and proportions for 5000 simulated samples with n=200 and p=0.50

Format

A dataset with 5000 observations on the following two variables:

- Count: Number of simulated "yes" responses in 200 trials
- Phat: Sample proportion (Count/200)

Details

Results from 5000 simulations of samples of size n=200 from a population with proportion of "yes" responses at p=0.50.

Source

Computer simulation

Description

Tip data from the First Crush Bistro

Format

A dataset with 157 observations on the following 7 variables:

- Bill: Size of the bill (in dollars)
Details

The owner of a bistro called First Crush in Potsdam, NY was interested in studying the tipping patterns of his customers. He collected restaurant bills over a two week period that he believes provide a good sample of his customers. The data recorded from 157 bills include the amount of the bill, size of the tip, percentage tip, number of customers in the group, whether or not a credit card was used, day of the week, and a coded identity of the server.

Source

Thanks to Tom DeRosa at First Crush for providing the tipping data.

---

### RetailSales

**Retail Sales (2009-2019)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tip</td>
<td>Size of the tip (in dollars)</td>
</tr>
<tr>
<td>Credit</td>
<td>Paid with a credit card? n or y</td>
</tr>
<tr>
<td>Guests</td>
<td>Number of people in the group</td>
</tr>
<tr>
<td>Day</td>
<td>Day of the week: m=Monday, t=Tuesday, w=Wednesday, th=Thursday, or f=Friday</td>
</tr>
<tr>
<td>Server</td>
<td>Code for specific waiter/waitress: A, B, or C</td>
</tr>
<tr>
<td>PctTip</td>
<td>Tip as a percentage of the bill</td>
</tr>
</tbody>
</table>

**Description**

Monthly U.S. Retail Sales from 2009 to 2019

**Format**

A data frame with 129 observations on the following 3 variables.

- **Month**  Month (Jan through Dec)  
- **Year**  Years from 2009 to 2019  
- **Sales**  Monthly U.S. retail sales (in billions of dollars)

**Details**

Data show the monthly retail sales (in billions) for the U.S. economy in each month from January 2009 through September 2019.

** Updated for 3e (earlier versions are RetailSales2e and RetailSales1e). **

**Source**

Data downloaded from [https://www.census.gov/retail/](https://www.census.gov/retail/).
### RetailSales2011  
**Retail Sales (2000-2011)**

**Description**

Monthly U.S. Retail Sales

**Format**

A dataset with 136 observations on the following 3 variables.

<table>
<thead>
<tr>
<th>Month</th>
<th>Month of the year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Years from 2000 to 2011</td>
</tr>
<tr>
<td>Sales</td>
<td>U.S. retail sales (in billions of dollars)</td>
</tr>
</tbody>
</table>

**Details**

Data show the monthly retail sales (in billions) for the U.S. economy in each month from January 2000 through April 2011.

**From 1e - dataset has been updated for 2e and 3e**

**Source**

Data downloaded from [http://www.census.gov/retail/](http://www.census.gov/retail/)

---

### RockandRoll2012  
**Rock & Roll Hall of Fame (2012)**

**Description**

Groups and Individuals in the Rock and Roll Hall of Fame (2012)

**Format**

A dataset with 273 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Inductee</th>
<th>Name of the group or individual</th>
</tr>
</thead>
<tbody>
<tr>
<td>FemaleMembers</td>
<td>Yes if individual or member of the group is female, otherwise No</td>
</tr>
<tr>
<td>Category</td>
<td>Type of individual or group: Performer, Non-performer, Early Influence, Lifetime Achievement, Sideman</td>
</tr>
<tr>
<td>People</td>
<td>Number of people in the group</td>
</tr>
</tbody>
</table>
Details

All inductees of the Rock & Roll Hall of Fame as of 2012.
** From 1e - dataset has been updated for 2e and 3e **

Source


---

RockandRoll2015

Rock & Roll Hall of Fame (2015)

Description

Groups and Individuals in the Rock and Roll Hall of Fame (2015)

Format

A dataset with 303 observations on the following 4 variables.

- **Inductee**: Name of the group or individual
- **FemaleMembers**: Yes if individual or member of the group is female, otherwise No
- **Category**: Type of individual or group: Performer, Non-performer, Early Influence, Lifetime Achievement, Sideman
- **People**: Number of people in the group

Details

All inductees of the Rock & Roll Hall of Fame as of 2015.
** From 2e - dataset has been updated for 3e **

Source


---

RockandRoll2019

Rock & Roll Hall of Fame (2019)

Description

Groups and Individuals in the Rock and Roll Hall of Fame as of 2019
**SalaryGender**

**Format**

A data frame with 329 observations on the following 4 variables.

- **Inductee**: Name of the group or individual
- **FemaleMembers**: Yes if individual or member of the group is female, otherwise No
- **Category**: Type of individual or group: Early Influence, Lifetime Achievement, Non-performer, Performer, or Sideman
- **People**: Number of people in the group

**Details**

All inductees of the Rock & Roll Hall of Fame as of 2019.

**Updated for 3e (earlier versions are now RockandRoll2015 and RockandRoll1e)**

**Source**

Rock & Roll Hall of Fame website, [https://www.rockhall.com/inductees/a-z](https://www.rockhall.com/inductees/a-z)

---

**Salary and Gender**

**Description**

Salaries for college teachers

**Format**

A dataset with 100 observations on the following 4 variables.

- **Salary**: Annual salary in $1,000’s
- **Gender**: 0=female or 1=male
- **Age**: Age in years
- **PhD**: 1=have PhD or 0=no PhD

**Details**

A random sample of college teachers taken from the 2010 American Community Survey (ACS) 1-year Public Use Microdata Sample (PUMS).

**Source**

Downloaded from [https://www.census.gov/programs-surveys/acs/data/pums.html](https://www.census.gov/programs-surveys/acs/data/pums.html)
SampColleges

Sample of US Post-secondary Schools

Description

Information for a sample of 50 US post-secondary schools from the Department of Education’s College Scorecard

Format

A data frame with 50 observations on the following 37 variables.

Name  Name of the school
State  State where school is located
ID    ID number for school
Main  Main campus? (1=yes, 0=branch campus)
Accred Accreditation agency
MainDegree Predominant undergrad degree (0=not classified, 1=certificate, 2=associate, 3=bachelors, 4=only graduate)
HighDegree Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4=graduate)
Control Control of school (Private, Profit, Public)
Region Region of country (Midwest, Northeast, Southeast, Territory, West)
Locale Locale (City, Rural, Suburb, Town)
Latitude Latitude
Longitude Longitude
AdmitRate Admission rate
MidACT Median of ACT scores
AvgSAT Average combined SAT scores
Online Only online (distance) programs
Enrollment Undergraduate enrollment
White Percent of undergraduates who report being white
Black Percent of undergraduates who report being black
Hispanic Percent of undergraduates who report being Hispanic
Asian Percent of undergraduates who report being Asian
Other Percent of undergraduates who don’t report one of the above
PartTime Percent of undergraduates who are part-time students
NetPrice Average net price (cost minus aid)
Cost Average total cost for tuition, room, board, etc.
TuitionIn In-state tuition and fees
TuitionOut  Out-of-state tuition and fees
TuitionFTE  Net tuition revenue per FTE student
InstructFTE  Instructional spending per FTE student
FacSalary  Average monthly salary for full-time faculty
FullTimeFac  Percent of faculty that are full-time
Pell  Percent of students receiving Pell grants
CompRate  Completion rate (percent who finish program within 150% of normal time)
Debt  Average debt for students who complete program
Female  Percent of female students
FirstGen  Percent of first-generation students
MedIncome  Median family income (in $1,000)

Details

The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains information from a sample of the 50 schools selected from CollegeScores.

Source


SampColleges2yr  Sample of College Scorecard - Two Year

Description

Information for a sample of 50 US post-secondary schools that primarily grant associate’s degrees, from the Department of Education’s College Scorecard

Format

A data frame with 50 observations on the following 31 variables.

Name  Name of the school
State  State where school is located
ID  ID number for school
Main  Main campus? (1=yes, 0=branch campus)
Accred  Accreditation agency
MainDegree  Predominant undergrad degree (0=not classified, 1=certificate, 2=associate, 3=bachelors, 4=only graduate)
HighDegree  Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)
Control  Control of school (Private, Profit, Public)
Region  Region of country (Midwest, Northeast, Southeast, Territory, West)
Locale  Locale (City, Rural, Suburb, Town)
Enrollment  Undergraduate enrollment
White  Percent of undergraduates who report being white
Black  Percent of undergraduates who report being black
Hispanic  Percent of undergraduates who report being Hispanic
Asian  Percent of undergraduates who report being Asian
Other  Percent of undergraduates who don’t report one of the above
PartTime  Percent of undergraduates who are part-time students
NetPrice  Average net price (cost minus aid)
Cost  Average total cost for tuition, room, board, etc.
TuitionIn  In-state tuition and fees
TuitionOut  Out-of-state tuition and fees
TuitionFTE  Net Tuition revenue per FTE student
InstructFTE  Instructional spending per FTE student
FacSalary  Average monthly salary for full-time faculty
FullTimeFac  Percent of faculty that are full-time
Pell  Percent of students receiving Pell grants
CompRate  Completion rate (percent who finish program within 150% of normal time)
Debt  Average debt for students who complete program
Female  Percent of female students
FirstGen  Percent of first-generation students
MedIncome  Median family income (in $1,000)

Details
Details The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains information from a sample of the two-year colleges selected from all two-year colleges in CollegeScores2yr.

Source
Sample of College Scorecard - Four Year

Description

Information on a sample of 50 US four-year colleges and universities from the Department of Education’s College Scoreboard.

Format

A data frame with 50 observations on the following 37 variables.

- **Name**: Name of the school
- **State**: State where school is located
- **ID**: ID number for school
- **Main**: Main campus? (1=yes, 0=branch campus)
- **Accred**: Accreditation agency
- **MainDegree**: Predominant undergrad degree (3=bachelors)
- **HighDegree**: Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4=graduate)
- **Control**: Control of school (Private, Profit, Public)
- **Region**: Region of country (Midwest, Northeast, Southeast, Territory, West)
- **Locale**: Locale (City, Rural, Suburb, Town)
- **Latitude**: Latitude
- **Longitude**: Longitude
- **AdmitRate**: Admission rate
- **MidACT**: Median of ACT scores
- **AvgSAT**: Average combined SAT scores
- **Online**: Only online (distance) programs
- **Enrollment**: Undergraduate enrollment
- **White**: Percent of undergraduates who report being white
- **Black**: Percent of undergraduates who report being black
- **Hispanic**: Percent of undergraduates who report being Hispanic
- **Asian**: Percent of undergraduates who report being Asian
- **Other**: Percent of undergraduates who don’t report one of the above
- **PartTime**: Percent of undergraduates who are part-time students
- **NetPrice**: Average net price (cost minus aid)
- **Cost**: Average total cost for tuition, room, board, etc.
- **TuitionIn**: In-state tuition and fees
- **TuitionOut**: Out-of-state tuition and fees
SampCountries

TuitionFTE  Net Tuition revenue per FTE student
InstructFTE  Instructional spending per FTE student
FacSalary  Average monthly salary for full-time faculty
FullTimeFac  Percent of faculty that are full-time
Pell  Percent of students receiving Pell grants
CompRate  Completion rate (percent who finish program within 150% of normal time)
Debt  Average debt for students who complete program
Female  Percent of female students
FirstGen  Percent of first-generation students
MedIncome  Median family income (in $1,000)

Details
The US Department of Education maintains a database through its College Scorecard project of
demographic information from all active postsecondary educational institutions that participate in
Title IV. This dataset contains information from a sample of the four-year colleges and universities
selected from all four-year colleges in CollegeScores4yr.

Source
Data downloaded from the US Department of Education’s College Scorecard at https://collegescorecard.
ed.gov/data/ (November 2019)

SampCountries | Sample of Countries

Description
Data on a sample of fifty countries of the world (2018)

Format
A data frame with 50 observations on the following 25 variables.

<table>
<thead>
<tr>
<th>Country</th>
<th>Country name</th>
</tr>
</thead>
<tbody>
<tr>
<td>LandArea</td>
<td>Size in 1000 sq. km.</td>
</tr>
<tr>
<td>Population</td>
<td>Population in millions</td>
</tr>
<tr>
<td>Density</td>
<td>Number of people per square kilometer</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product (in $US) per capita</td>
</tr>
<tr>
<td>Rural</td>
<td>Percentage of population living in rural areas</td>
</tr>
<tr>
<td>CO2</td>
<td>CO2 emissions (metric tons per capita)</td>
</tr>
<tr>
<td>PumpPrice</td>
<td>Price for a liter of gasoline ($US)</td>
</tr>
</tbody>
</table>
Military  Percentage of government expenditures directed toward the military  
Health  Percentage of government expenditures directed towards healthcare  
ArmedForces  Number of active duty military personnel (in 1,000’s)  
Internet  Percentage of the population with access to the internet  
Cell  Cell phone subscriptions (per 100 people)  
HIV  Percentage of the population with HIV  
Hunger  Percent of the population considered undernourished  
Diabetes  Percent of the population diagnosed with diabetes  
BirthRate  Births per 1000 people  
DeathRate  Deaths per 1000 people  
ElderlyPop  Percentage of the population at least 65 years old  
LifeExpectancy  Average life expectancy (years)  
FemaleLabor  Percent of females 15 - 64 in the labor force  
Unemployment  Percent of labor force unemployed  
EnergyUse  Kilotons of oil equivalent  
Electricity  Electric power consumption (kWh per capita)  
Developed  Categories for kilowatt hours per capita, 1= under 2500, 2=2500 to 5000, 3=over 5000

Details
Data from AllCountries for a random sample of 50 countries. Data for 2016-2018 to avoid many missing values in more recent years.
** Updated for 3e (earlier versions are now SampCountries2e and SampCountries1e). **

Source
**Format**

A dataset with 50 observations on the following 13 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Name of the country</td>
</tr>
<tr>
<td>LandArea</td>
<td>Size in sq. kilometers</td>
</tr>
<tr>
<td>Population</td>
<td>Population in millions</td>
</tr>
<tr>
<td>Energy</td>
<td>Energy usage (kilotons of oil)</td>
</tr>
<tr>
<td>Rural</td>
<td>Percentage of population living in rural areas</td>
</tr>
<tr>
<td>Military</td>
<td>Percentage of government expenditures directed toward the military</td>
</tr>
<tr>
<td>Health</td>
<td>Percentage of government expenditures directed towards healthcare</td>
</tr>
<tr>
<td>HIV</td>
<td>Percentage of the population with HIV</td>
</tr>
<tr>
<td>Internet</td>
<td>Percentage of the population with access to the internet</td>
</tr>
<tr>
<td>Developed</td>
<td>Categories for kilowatt hours per capita: 1= under 2500, 2=2500 to 5000, 3=over 5000</td>
</tr>
<tr>
<td>BirthRate</td>
<td>Births per 1000 people</td>
</tr>
<tr>
<td>ElderlyPop</td>
<td>Percentage of the population at least 65 years old</td>
</tr>
<tr>
<td>LifeExpectancy</td>
<td>Average life expectancy (in years)</td>
</tr>
</tbody>
</table>

**Details**

A subset of data from AllCountries for a random sample of 50 countries in 2008.

**From 1e - dataset has been updated for 2e and 3e**

**Source**


**Description**

Data on a sample of fifty countries of the world (2014)

**Format**

A dataset with 50 observations on the following 25 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Name of the country</td>
</tr>
<tr>
<td>LandArea</td>
<td>Size in 1000 sq. kilometers</td>
</tr>
<tr>
<td>Population</td>
<td>Population in millions</td>
</tr>
<tr>
<td>Density</td>
<td>Number of people per square kilometer</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product (in $US) per capita</td>
</tr>
<tr>
<td>Rural</td>
<td>Percentage of population living in rural areas</td>
</tr>
<tr>
<td>CO2</td>
<td>CO2 emissions (metric tons per capita)</td>
</tr>
<tr>
<td>PumpPrice</td>
<td>Price for a liter of gasoline ($US)</td>
</tr>
<tr>
<td>Military</td>
<td>Percentage of government expenditures directed toward the military</td>
</tr>
</tbody>
</table>
### Details

Data from AllCountries for a random sample of 50 countries. Data for 2012-2014 to avoid many missing values in more recent years.

**From 2e - dataset has been updated for 3e**

### Source


---

#### SandP500

**S&P 500 Prices**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily data for S&amp;P 500 Stock Index</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>A data frame with 251 observations on the following 6 variables.</td>
</tr>
</tbody>
</table>

- **Date**  Trading date (mm/dd/yyyy)
- **Open**  Opening value
- **High**  High point for the day
- **Low**  Low point for the day
- **Close**  Closing value
- **Volume**  Shares traded (in millions)
Details

** Updated for 3e (earlier versions are SandP5002e from 2014 and SandP5001e from 2010). **

Source

Downloaded from https://finance.yahoo.com/quote/^GSPC/history?ltr=1

<table>
<thead>
<tr>
<th>SandP5001e</th>
<th>S&amp;P 500 Prices</th>
</tr>
</thead>
</table>

Description

Daily data for S&P 500 Stock Index

Format

A dataset with 252 observations on the following 6 variables.

- Date  Trading date
- Open  Opening value
- High  High point for the day
- Low   Low point for the day
- Close Closing value
- Volume Shares traded (in millions)

Details

Daily prices for the S&P 500 Stock Index for trading days in 2010.
** From 1e - dataset has been updated for 2e and 3e **

Source

Downloaded from http://finance.yahoo.com/q/hp?s=^GSPC+Historical+Prices

<table>
<thead>
<tr>
<th>SandP5002e</th>
<th>S&amp;P 500 Prices - 2e</th>
</tr>
</thead>
</table>

Description

Daily data for S&P 500 Stock Index

Format

A dataset with 252 observations on the following 6 variables.
Details

** From 2e - dataset has been updated for 3e **

Source

Downloaded from http://finance.yahoo.com/q/hp?s=^GSPC+Historical+Prices

SandwichAnts

<table>
<thead>
<tr>
<th>Date</th>
<th>Trading date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open</td>
<td>Opening value</td>
</tr>
<tr>
<td>High</td>
<td>High point for the day</td>
</tr>
<tr>
<td>Low</td>
<td>Low point for the day</td>
</tr>
<tr>
<td>Close</td>
<td>Closing value</td>
</tr>
<tr>
<td>Volume</td>
<td>Shares traded (in millions)</td>
</tr>
</tbody>
</table>

Description

Ant counts on samples of different sandwiches

Format

A dataset with 24 observations on the following 5 variables.

- Butter: Butter on the sandwich? no (Cases with Butter=yes are in SandwichAnts2)
- Filling: Type of filling: Ham & Pickles, Peanut Butter, or Vegemite
- Bread: Type of bread: Multigrain, Rye, White, or Wholemeal
- Ants: Number of ants on the sandwich
- Order: Trial number

Details

As young students, Dominic Kelly and his friends enjoyed watching ants gather on pieces of sandwiches. Later, as a university student, Dominic decided to study this with a more formal experiment. He chose three types of sandwich fillings (vegemite, peanut butter, and ham & pickles), four types of bread (multigrain, rye, white, and wholemeal), and put butter on some of the sandwiches. To conduct the experiment he randomly chose a sandwich, broke off a piece, and left it on the ground near an ant hill. After several minutes he placed a jar over the sandwich bit and counted the number of ants. He repeated the process, allowing time for ants to return to the hill after each trial, until he had two samples for each combination of the factors. This dataset has only sandwiches with no butter. The data in SandwichAnts2 adds information for samples with butter.
Source
http://www.amstat.org/publications/jse/v2n1/mackisack supp.html

---

SandwichAnts2  Sandwich Ants - Part 2

Description
Ant counts on samples of different sandwiches

Format
A dataset with 48 observations on the following 5 variables.

- **Butter**: Butter on the sandwich? no or yes
- **Filling**: Type of filling: Ham & Pickles, Peanut Butter, or Vegemite
- **Bread**: Type of bread: Multigrain, Rye, White, or Wholemeal
- **Ants**: Number of ants on the sandwich
- **Order**: Trial number

Details
As young students, Dominic Kelly and his friends enjoyed watching ants gather on pieces of sandwiches. Later, as a university student, Dominic decided to study this with a more formal experiment. He chose three types of sandwich fillings (vegemite, peanut butter, and ham & pickles), four types of bread (multigrain, rye, white, and wholemeal), and put butter on some of the sandwiches. To conduct the experiment he randomly chose a sandwich, broke off a piece, and left it on the ground near an ant hill. After several minutes he placed a jar over the sandwich bit and counted the number of ants. He repeated the process, allowing time for ants to return to the hill after each trial, until he had two samples for each combination of the three factors.

Source
http://www.amstat.org/publications/jse/v2n1/mackisack supp.html

---

SkateboardPrices  Skateboard Prices

Description
Prices of skateboards for sale online
**SleepStudy**

**Format**

A dataset with 20 observations on the following variable.

| Price | Selling price in dollars |

**Details**

Prices for skateboards offered for sale on eBay.

**Source**

Random sample taken from all skateboards available for sale on eBay on February 12, 2012.

---

**SleepCaffeine**  

**Sleep Caffeine**

**Description**

Experiment to compare word recall after sleep or caffeine.

**Format**

A dataset with 24 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment: Caffeine or Sleep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words</td>
<td>Number of words recalled</td>
</tr>
</tbody>
</table>

**Details**

A random sample of 24 adults were divided equally into two groups and given a list of 24 words to memorize. During a break, one group takes a 90 minute nap while another group is given a caffeine pill. The response variable is the number of words participants are able to recall following the break.

**Source**

Mednick, Cai, Kanady, and Drummond, "Comparing the benefits of caffeine, naps and placebo on verbal, motor and perceptual memory", Behavioural Brain Research, 193 (2008), 79-86.

---

**SleepStudy**  

**Sleep Study**
Description

Data from a study of sleep patterns for college students.

Format

A dataset with 253 observations on the following 27 variables.

- **Gender**: 1=male, 0=female
- **ClassYear**: Year in school, 1=first year, ..., 4=senior
- **LarkOwl**: Early riser or night owl? Lark, Neither, or Owl
- **NumEarlyClass**: Number of classes per week before 9 am
- **EarlyClass**: Indicator for any early classes
- **GPA**: Grade point average (0-4 scale)
- **ClassesMissed**: Number of classes missed in a semester
- **CognitionZscore**: Z-score on a test of cognitive skills
- **PoorSleepQuality**: Measure of sleep quality (higher values are poorer sleep)
- **DepressionScore**: Measure of degree of depression
- **AnxietyScore**: Measure of amount of anxiety
- **StressScore**: Measure of amount of stress
- **DepressionStatus**: Coded depression score: normal, moderate, or severe
- **AnxietyStatus**: Coded anxiety score: normal, moderate, or severe
- **Stress**: Coded stress score: normal or high
- **DASScore**: Combined score for depression, anxiety and stress
- **Happiness**: Measure of degree of happiness
- **AlcoholUse**: Self-reported: Abstain, Light, Moderate, or Heavy
- **Drinks**: Number of alcoholic drinks per week
- **WeekdayBed**: Average weekday bedtime (24.0=midnight)
- **WeekdayRise**: Average weekday rise time (8.0=8 am)
- **WeekdaySleep**: Average hours of sleep on weekdays
- **WeekendBed**: Average weekend bedtime (24.0=midnight)
- **WeekendRise**: Average weekend rise time (8.0=8 am)
- **WeekendSleep**: Average weekend bedtime (24.0=midnight)
- **AverageSleep**: Average hours of sleep for all days
- **AllNighter**: Had an all-nighter this semester? 1=yes, 0=no

Details

The data were obtained from a sample of students who did skills tests to measure cognitive function, completed a survey that asked many questions about attitudes and habits, and kept a sleep diary to record time and quality of sleep over a two week period.

Source

Onyper, S., Thacher, P., Gilbert, J., Gradess, S., "Class Start Times, Sleep, and Academic Performance in College: A Path Analysis," April 2012; 29(3): 318-335. Thanks to the authors for supplying the data.
**SpeedDating**

**Description**

Experiment to study effect of smiling on leniency in judicial matters

**Format**

A dataset with 68 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leniency</td>
<td>Score assigned by a judgment panel (higher is more lenient)</td>
</tr>
<tr>
<td>Group</td>
<td>Treatment group: neutral or smile</td>
</tr>
</tbody>
</table>

**Details**

Hecht and LeFrance conducted a study examining the effect of a smile on the leniency of disciplinary action for wrongdoers. Participants in the experiment took on the role of members of a college disciplinary panel judging students accused of cheating. For each suspect, along with a description of the offense, a picture was provided with either a smile or neutral facial expression. A leniency score was calculated based on the disciplinary decisions made by the participants.

**Source**


---

**SpeedDating**

**Description**

Data from a sample of four minute speed dates.

**Format**

A dataset with 276 observations on the following 22 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DecisionM</td>
<td>Would the male like another date? 1=yes 0=no</td>
</tr>
<tr>
<td>DecisionF</td>
<td>Would the female like another date? 1=yes 0=no</td>
</tr>
<tr>
<td>LikeM</td>
<td>How much the male likes his partner (1-10 scale)</td>
</tr>
<tr>
<td>LikeF</td>
<td>How much the female likes her partner (1-10 scale)</td>
</tr>
<tr>
<td>PartnerYesM</td>
<td>Male’s estimate of chance the female wants another date (1-10 scale)</td>
</tr>
<tr>
<td>PartnerYesF</td>
<td>Female’s estimate of chance the male wants another date (1-10 scale)</td>
</tr>
<tr>
<td>AgeM</td>
<td>Male’s age (in years)</td>
</tr>
</tbody>
</table>
Participants were students at Columbia's graduate and professional schools, recruited by mass email, posted fliers, and fliers handed out by research assistants. Each participant attended one speed dating session, in which they met with each participant of the opposite sex for four minutes. Order and session assignments were randomly determined. After each four minute "speed date," participants filled out a form rating their date on a scale of 1-10 on various attributes. Only data from the first date in each session is recorded here.

Details

Subjects were 48 Israeli students who were randomly assigned to eat in groups of six (three males and three females) at a restaurant. Half the groups were told that they would pay for meals individually and half were told that the group would split the bill equally. The number of items ordered and cost (in Israeli new shekels) was recorded for each individual.

Source


StatGrades

Statistics Exam Grades

Description

Grades on statistics exams

Format

A dataset with 50 observations on the following 3 variables.

Exam1  Score (out of 100 points) on the first exam  
Exam2  Score (out of 100 points) on the second exam  
Final  Score (out of 100 points) on the final exam  

Details

Exam scores for a sample of students who completed a course using Statistics: Unlocking the Power of Data as a text. The dataset contains scores on Exam1 (Chapters 1 to 4), Exam2 (Chapters 5 to 8), and the Final exam (entire book).

Source

Random selection of students in an introductory statistics course.

StockChanges

Stock Changes

Description

Stock price change for a sample of stocks from the S&P 500 (August 2-6, 2010)
**Format**

A dataset with 50 observations on the following variable.
Details

A random sample of 50 companies from Standard & Poor’s index of 500 companies was selected. The change in the price of the stock (in dollars) over the 5-day period from August 2 - 6, 2010 was recorded for each company in the sample.

Source

Data obtained from http://money.cnn.com/data/markets/sandp/

---

StorySpoilers

<table>
<thead>
<tr>
<th>StorySpoilers</th>
<th>Story Spoilers</th>
</tr>
</thead>
</table>

Description

Ratings for stories with and without spoilers

Format

A dataset with 12 observations on the following 3 variables.

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story</td>
<td>ID for story</td>
</tr>
<tr>
<td>Spoiler</td>
<td>Average (0-10) rating for spoiler version</td>
</tr>
<tr>
<td>Original</td>
<td>Average (0-10) rating for original version</td>
</tr>
</tbody>
</table>

Details

This study investigated whether a story spoiler that gives away the ending early diminishes suspense and hurts enjoyment. For twelve different short stories, the study's authors created a second version in which a spoiler paragraph at the beginning discussed the story and revealed the outcome. Each version of the twelve stories was read by at least 30 people and rated on a 1 to 10 scale to create an overall rating for the story, with higher ratings indicating greater enjoyment of the story. Stories 1 to 4 were ironic twist stories, stories 5 to 8 were mysteries, and stories 9 to 12 were literary stories.

Source

**Stressed Mice**

**Description**

Time in darkness for mice in different environments.

**Format**

A dataset with 14 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Time</th>
<th>Time spent in darkness (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>Type of environment: Enriched or Standard</td>
</tr>
</tbody>
</table>

**Details**

In the study, mice were randomly assigned to either an enriched environment where there was an exercise wheel available, or a standard environment with no exercise options. After three weeks in the specified environment, for five minutes a day for two weeks, the mice were each exposed to a "mouse bully" - a mouse who was very strong, aggressive, and territorial. One measure of mouse anxiety is amount of time hiding in a dark compartment, with mice who are more anxious spending more time in darkness. The amount of time spent in darkness is recorded for each of the mice.

**Source**


---

**Student Survey Data**

**Description**

Data from a survey of students in introductory statistics courses.

**Format**

A data frame with 362 observations on the following 17 variables.

- Year: Year in school
- Sex: code F=female or M= male
- Smoke: Smoker? No or Yes
- Award: Preferred award: Academy, Nobel, or Olympic
HigherSAT  Which SAT is higher? Math or Verbal
Exercise  Hours of exercise per week
TV  Hours of TV viewing per week
Height  Height (in inches)
Weight  Weight (in pounds)
Siblings  Number of siblings
BirthOrder  Birth order, 1=oldest
VerbalSAT  Verbal SAT score
MathSAT  Math SAT score
SAT  Combined Verbal + Math SAT
GPA  College grade point average
Pulse  Pulse rate (beats per minute)
Piercings  Number of body piercings

Details
Data from an in-class survey given to introductory statistics students over several years. Note the Sex variable was labeled as Gender in earlier versions of this dataset. We acknowledge that this binary dichotomization is not a complete or inclusive representation of reality.

Source
In-class student survey

SynchronizedMovement  Synchronized Movement

Description
Effects of synchronized movement activities

Format
A dataset with 264 observations on the following 11 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>f = female or m = male</td>
</tr>
<tr>
<td>Group</td>
<td>Type of activity. Coded as HS+HE, HS+LE, LS+HE, or LS+LE for High/Low Synchronization + High/Low Exertion</td>
</tr>
<tr>
<td>Synch</td>
<td>Synchronized activity? yes or no</td>
</tr>
<tr>
<td>Exertion</td>
<td>Exertion level: high or low</td>
</tr>
<tr>
<td>PainToleranceBefore</td>
<td>Measure of pain tolerance (mm Hg) before activity</td>
</tr>
<tr>
<td>PainTolerance</td>
<td>Measure of pain tolerance (mm Hg) after activity</td>
</tr>
<tr>
<td>PainTolDiff</td>
<td>Difference (after - before) in pain tolerance</td>
</tr>
<tr>
<td>MaxPressure</td>
<td>Reached the maximum pressure (300 mm Hg) when testing pain tolerance (after)</td>
</tr>
</tbody>
</table>
Details

From a study of 264 high school students in Brazil to examine the effect of doing synchronized movements (such as marching in step or doing synchronized dance steps) and the effect of exertion on variables, such as pain tolerance and attitudes towards others. Students were randomly assigned to activities that involved synchronized or non-synchronized movements involving high or low levels of exertion. Pain tolerance was measured with a blood pressure cuff, going to a maximum possible reading of 300 mmHg.

Source


---

**TenCountries**

<table>
<thead>
<tr>
<th>CloseBefore</th>
<th>Rating of closeness to the group before activity (1=least close to 7=most close)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CloseAfter</td>
<td>Rating of closeness to the group after activity (1=least close to 7=most close)</td>
</tr>
<tr>
<td>CloseDiff</td>
<td>Change on closeness rating (after - before)</td>
</tr>
</tbody>
</table>

**Description**

A subset of the AllCountries data for a random sample of ten countries

**Format**

A data frame with 10 observations on the following 4 variables.

- **Country**  Country name
- **Code**     Three-letter country code
- **Area**     Size in 1000 sq. kilometers
- **PctRural** Percentage of population living in rural areas

**Details**

Area and percent rural for a sample of ten countries from AllCountries dataset.

**Updated for 3e (earlier versions are now TenCountries2e and TenCountries1e)**

**Source**

Description

A subset of the AllCountries data for a random sample of ten countries

Format

A dataset with 10 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Country name</td>
</tr>
<tr>
<td>Code</td>
<td>Three-letter country code</td>
</tr>
<tr>
<td>Area</td>
<td>Size in 1000 sq. kilometers</td>
</tr>
<tr>
<td>PctRural</td>
<td>Percentage of population living in rural areas</td>
</tr>
</tbody>
</table>

Details

Area and percent rural for a sample of ten countries from AllCountries dataset.

** From 1e - dataset has been updated for 2e and 3e **

Source


Details
Area and percent rural for a sample of ten countries from AllCountries dataset.
*From 2e - dataset has been updated for 3e*

Source

---

Textbook Costs

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>General discipline of the course: Arts, Humanities, Natural Science, or Social Science</td>
<td></td>
</tr>
<tr>
<td>Books</td>
<td>Number of books required</td>
</tr>
<tr>
<td>Cost</td>
<td>Total cost (in dollars) for required books</td>
</tr>
</tbody>
</table>

Description
Prices for textbooks for different courses

Format
A data frame with 40 observations on the following 3 variables.

Details
Data are from samples of ten courses in each of four disciplines at a liberal arts college. For each course the bookstore’s website lists the required texts(s) and costs. Data were collected for the Fall 2011 semester.

Source
Bookstore online site

---

Toenail Arsenic

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arsenic</td>
<td>Level of arsenic found in toenails (ppm)</td>
</tr>
</tbody>
</table>

Description
Arsenic in toenails of 19 people using private wells in New Hampshire

Format
A dataset with 19 observations on the following variable.

Arsenic  Level of arsenic found in toenails (ppm)
USStates

Details

Level of arsenic was measured in toenails of 19 subjects from New Hampshire, all with private wells as their main water source.

Source

Adapted from Karagas, et.al., "Toenail Samples as an Indicator of Drinking Water Arsenic Exposure", Cancer Epidemiology, Biomarkers and Prevention 1996;5:849-852.

TrafficFlow

<table>
<thead>
<tr>
<th>Traffic Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>Traffic flow times from a simulation with timed and flexible traffic lights</td>
</tr>
</tbody>
</table>

| **Format** |
| A dataset with 24 observations on the following 3 variables. |
| **Timed** | Delay time (in minutes) for fixed timed lights |
| **Flexible** | Delay time (in minutes) for flexible communicating lights |
| **Difference** | Difference (Timed-Flexible) for each simulation |

Details

Engineers in Dresden, Germany were looking at ways to improve traffic flow by enabling traffic lights to communicate information about traffic flow with nearby traffic lights. The data show results of one experiment where they simulated buses moving along a street and recorded the delay time (in seconds) for both a fixed time and a flexible system of lights. The process was repeated under both conditions for a sample of 24 simulated scenarios.

Source


USStates

<table>
<thead>
<tr>
<th>US State Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>Various data for all 50 US States.</td>
</tr>
</tbody>
</table>
Format

A data frame with 50 observations on the following 22 variables.

State  State name
HouseholdIncome  Median household income (in $1,000’s)
Region  MW=Midwest, NE=Northeast, S=South, W=West
Population  Number of residents (in millions for 2014)
EighthGradeMath  Average score NAEP mathematics for 8th-grade students
HighSchool  % of residents (ages 25-34) who are high school graduates
College  % of residents (ages 25-34) who are college graduates
IQ  Estimated mean IQ score of residents
GSP  Gross state product (in $1,000’s per capita)
Vegetables  % of residents eating vegetables at least once per day
Fruit  % of residents eating fruit at least once per day
Smokers  % of residents who smoke
PhysicalActivity  % who do 150+ minutes of aerobic physical activity per week
Obese  % obese residents (BMI 30+)
NonWhite  % nonwhite residents
HeavyDrinkers  % heavy drinkers (men: 14+ drinks/week, women 7+ drinks/week)
Electoral  Number of state votes in the presidential electoral college
ClintonVote  Proportion of votes for Democrat Clinton in 2016 presidential election
Elect2016  State winner in 2016 presidential election (D=Clinton, R=Trump)
TwoParents  % of children living in two-parent households
StudentSpending  School spending (in $1,000 per pupil)
Insured  % of adults (ages 19-64) who have any kind of health coverage

Details

Information from each of the 50 states of the United States. Years vary from 2013 to 2018 depending on data availability.

** Updated for 3e (earlier versions are now USStates2e and USStates1e) **

Source

U.S. Census Bureau, 2013-2017 5-Year American Community Survey

http://factfinder.census.gov/faces/tables Services/jsf/pages/productview.xhtml?pid=ACS_17_5YR_DP03&src=pt

http://factfinder.census.gov/faces/tables Services/jsf/pages/productview.xhtml?pid=ACS_17_5YR_S1501&src=pt
Description

Various data for all 50 US States

Format

A dataset with 50 observations on the following 17 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>Name of state</td>
</tr>
<tr>
<td>HouseholdIncome</td>
<td>Mean household income (in dollars)</td>
</tr>
<tr>
<td>IQ</td>
<td>Mean IQ score of residents</td>
</tr>
<tr>
<td>McCainVote</td>
<td>Percentage of votes for John McCain in 2008 Presidential election</td>
</tr>
<tr>
<td>Region</td>
<td>Area of the country: MW=Midwest, NE=Northeast, S=South, or W=West</td>
</tr>
<tr>
<td>ObamaMcCain</td>
<td>Which 2008 Presidential candidate won state? M=McCain or O=Obama</td>
</tr>
<tr>
<td>Population</td>
<td>Number of residents (in millions)</td>
</tr>
<tr>
<td>EighthGradeMath</td>
<td>Average score NAEP mathematics for 8th-grade students</td>
</tr>
<tr>
<td>HighSchool</td>
<td>Percentage of high school graduates</td>
</tr>
<tr>
<td>GSP</td>
<td>Gross State Product (dollars per capita)</td>
</tr>
<tr>
<td>FiveVegetables</td>
<td>Percentage of residents who eat at least five servings of fruits/vegetables per day</td>
</tr>
<tr>
<td>Smokers</td>
<td>Percentage of residents who smoke</td>
</tr>
<tr>
<td>PhysicalActivity</td>
<td>Percentage of residents who have competed in a physical activity in past month</td>
</tr>
<tr>
<td>Obese</td>
<td>Percentage of residents classified as obese</td>
</tr>
<tr>
<td>College</td>
<td>Percentage of residents with college degrees</td>
</tr>
<tr>
<td>NonWhite</td>
<td>Percentage of residents who are not white</td>
</tr>
<tr>
<td>HeavyDrinkers</td>
<td>Percentage of residents who drink heavily</td>
</tr>
</tbody>
</table>

Details

Information from each of the 50 states of the United States.

** From 1e - dataset has been updated for 2e and 3e **

Source

Various online sources, mostly at www.census.gov
USStates2e

US State Data - 2e

Description

Various data for all 50 US States in 2014.

Format

A dataset with 50 observations on the following 22 variables.

<table>
<thead>
<tr>
<th>State</th>
<th>State name</th>
</tr>
</thead>
<tbody>
<tr>
<td>HouseholdIncome</td>
<td>Median household income (in $1,000’s)</td>
</tr>
<tr>
<td>Region</td>
<td>MW=Midwest, NE=Northeast, S=South, W=West</td>
</tr>
<tr>
<td>Population</td>
<td>Number of residents (in millions for 2014)</td>
</tr>
<tr>
<td>EighthGradeMath</td>
<td>Average score NAEP mathematics for 8th-grade students (2013)</td>
</tr>
<tr>
<td>HighSchool</td>
<td>Percent of residents (ages 25-34) who are high school graduates</td>
</tr>
<tr>
<td>College</td>
<td>Percent of residents (ages 25-34) who are college graduates</td>
</tr>
<tr>
<td>IQ</td>
<td>Estimated mean IQ score of residents</td>
</tr>
<tr>
<td>GSP</td>
<td>Gross state product (in $1,000’s per capita in 2013)</td>
</tr>
<tr>
<td>Vegetables</td>
<td>Percent of residents eating vegetables at least once per day</td>
</tr>
<tr>
<td>Fruit</td>
<td>Percent of residents eating fruit at least once per day</td>
</tr>
<tr>
<td>Smokers</td>
<td>Percent of residents who smoke</td>
</tr>
<tr>
<td>PhysicalActivity</td>
<td>Percent who do 150+ minutes of aerobic physical activity per week</td>
</tr>
<tr>
<td>Obese</td>
<td>Percent obese residents (BMI 30+)</td>
</tr>
<tr>
<td>NonWhite</td>
<td>Percent nonwhite residents (in 2013)</td>
</tr>
<tr>
<td>HeavyDrinkers</td>
<td>Percent heavy drinkers (men: 3+ drinks/day, women 2+ drinks/day)</td>
</tr>
<tr>
<td>Electoral</td>
<td>Number of state votes in the presidential electoral college</td>
</tr>
<tr>
<td>ObamaVote</td>
<td>Proportion of votes for Obama in 2012 presidential election</td>
</tr>
<tr>
<td>ObamaRomney</td>
<td>State winner in 2012 presidential election (O=Obama, R=Romney)</td>
</tr>
<tr>
<td>TwoParents</td>
<td>Percent of children living in two-parent households</td>
</tr>
<tr>
<td>StudentSpending</td>
<td>School spending (in $1,000 per pupil in 2013)</td>
</tr>
<tr>
<td>Insured</td>
<td>Percent of adults (ages 18-64) who have any kind of health coverage</td>
</tr>
</tbody>
</table>

Details

Information from each of the 50 states of the United States (from 2013 or 2014).

** From 2e - dataset has been updated for 3e **

Source

U.S. Census Bureau, 2009-2013 5-Year American Community Survey
http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_5YR_DP03&src=pt
http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_5YR_S1501&src=pt
Water Striders

Description

Mating activity for water striders

Format

A dataset with 10 observations on the following 3 variables.

- AggressiveMale: Hyper-aggressive male in group? No or Yes
- FemalesHiding: Proportion of time the female water striders were in hiding
- MatingActivity: Measure of mean mating activity (higher numbers meaning more mating)

Details

Water striders are common bugs that skate across the surface of water. Water striders have different personalities and some of the males are hyper-aggressive, meaning they jump on and wrestle with any other water strider near them. Individually, because hyper-aggressive males are much more active, they tend to have better mating success than more inactive striders. This study examined the effect they have on a group. Four males and three females were put in each of ten pools of water. Half of the groups had a hyper-aggressive male as one of the males and half did not. The proportion of time females are in hiding was measured for each of the 10 groups, and a measure of mean mating activity was also measured with higher numbers meaning more mating.

Source


Water Taste

Description

Blind taste test to compare brands of bottled water

Format

A dataset with 100 observations on the following 10 variables.
Gender | Gender of respondent: F=Female M=Male
---|---
Age | Age (in years)
Class | Year in school F=First year J=Junior O=Other P=Sophomore SR=Senior
UsuallyDrink | Usual source of drinking water: Bottled, Filtered, or Tap
FavBotWatBrand | Favorite brand of bottled water
Preference | Order of preference: A=Sams Choice, B=Aquafina, C=Fiji, and D=Tap water
First | Top choice among Aquafina, Fiji, SamsChoice, or Tap
Second | Second choice
Third | Third choice
Fourth | Fourth choice

Details

Result from a blind taste test comparing four different types of water (Sam’s Choice, Aquafina, Fiji, and tap water). Participants rank ordered waters when presented in a random order.

Source

"Water Taste Test Data" by M. Leigh Lunsford and Alix D. Dowling Finch in the Journal of Statistics Education (Vol 18, No, 1) 2010

---

Wetsuits

Description

Swim velocity (for 1500 meters) with and without wearing a wetsuit

Format

A dataset with 12 observations on the following 4 variables.

<table>
<thead>
<tr>
<th>Wetsuit</th>
<th>Maximum swim velocity (m/sec) when wearing a wetsuit</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoWetsuit</td>
<td>Maximum swim velocity (m/sec) when wearing a regular bathing suit</td>
</tr>
<tr>
<td>Gender</td>
<td>Gender of swimmer: F or M</td>
</tr>
<tr>
<td>Type</td>
<td>Type of athlete: swimmer or triathlete</td>
</tr>
</tbody>
</table>

Details

A study tested whether wearing wetsuits influences swimming velocity. Twelve competitive swimmers and triathletes swam 1500m at maximum speed twice each; once wearing a wetsuit and once wearing a regular bathing suit. The order of the trials was randomized. Each time, the maximum velocity in meters/sec of the swimmer was recorded.
Source


Description

Effects of transfusions of young blood on exercise endurance in mice

Format

A dataset with 30 observations on the following 2 variables.

<table>
<thead>
<tr>
<th>Plasma</th>
<th>Whether the blood came from a Young or Old mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Runtime</td>
<td>Maximum treadmill run time (in minutes) in a 90-minute window</td>
</tr>
</tbody>
</table>

Details

The data come from a study to see if transfusions of blood plasma from young mice (equivalent to about a 25-year-old person) can counteract or reverse brain aging in old mice (equivalent to about a 70-year-old person.) Old mice were randomly assigned to receive plasma from either a young mouse or another old mouse, and exercise endurance was measured.

Source

Data come from two references, and are estimated from summary statistics and graphs.


Index

* datasets

ACS, 6
ACS2010, 7
AllCountries, 8
AllCountries1e, 9
AllCountries2e, 10
APMultipleChoice, 11
April14Temps, 11
April14Temps1e, 12
April14Temps2e, 12
BaseballHits1e, 13
BaseballHits2014, 14
BaseballHits2019, 14
BaseballSalaries2015, 15
BaseballSalaries2019, 16
BaseballTimes, 16
Benford, 17
BikeCommute, 18
BodyFat, 18
BodyTemp50, 19
BootAtlantaCorr, 20
CaffeineTaps, 20
CAOSExam, 21
CarbonDioxide, 21
CarbonDioxide2e, 22
CarDepreciation, 23
Cars2015, 23
Cars2020, 24
Cereal, 25
CityTemps, 26
CityTemps2e, 26
CocaineTreatment, 27
ColaCalcium, 28
CollegeScores, 28
CollegeScores2yr, 30
CollegeScores4yr, 31
CommuteAtlanta, 33
CommuteStLouis, 33
CompassionateRats, 34
CricketChirps, 35
DDS, 35
DecemberFlights, 36
DecemberFlights2e, 36
DietDepression, 37
Digits, 38
DogOwner, 38
DrugResistance, 39
EducationLiteracy, 40
EducationLiteracy2e, 40
ElectionMargin, 41
EmployedACS, 41
EmployedACS2010, 42
ExerciseHours, 43
FacebookFriends, 44
FatMice18, 44
FireAnts, 45
FisherIris, 46
FishGills12, 46
FishGills3, 47
Flight179, 47
Flight433, 48
Flight433_2e, 48
FloridaLakes, 49
FootballBrain, 50
ForestFires, 50
GeneticDiversity, 51
GlobalInternet2010, 52
GlobalInternet2019, 53
GolfRound, 53
GPAbySex, 54
GSWarriors2016, 54
GSWarriors2019, 56
HappyPlanetIndex, 57
HeatCognition, 58
HeightData, 59
HockeyPenalties2011, 60
HockeyPenalties2019, 61
HollywoodMovies, 61
StorySpoilers, 131
StressedMice, 132
StudentSurvey, 132
SynchronizedMovement, 133
TenCountries, 134
TenCountries1e, 135
TenCountries2e, 135
TextbookCosts, 136
ToenailArsenic, 136
TrafficFlow, 137
USStates, 137
USStates1e, 139
USStates2e, 140
WaterStriders, 141
WaterTaste, 141
Wetsuits, 142
YoungBlood, 143

* package
  Lock5Data-package, 5

ACS, 6
ACS2010, 7
AllCountries, 8
AllCountries1e, 9
AllCountries2e, 10
APMultipleChoice, 11
April14Temps, 11
April14Temps1e, 12
April14Temps2e, 12

BaseballHits1e, 13
BaseballHits2014, 14
BaseballHits2019, 14
BaseballSalaries2015, 15
BaseballSalaries2019, 16
BaseballTimes, 16
Benford, 17
BikeCommute, 18
BodyFat, 18
BodyTemp50, 19
BootAtlantaCorr, 20

CaffeineTaps, 20
CAOExam, 21
CarbonDioxide, 21
CarbonDioxide2e, 22
CarDepreciation, 23
Cars2015, 23
Cars2020, 24

Cereal, 25
CityTemps, 26
CityTemps2e, 26
CocaineTreatment, 27
ColaCalcium, 28
CollegeScores, 28
CollegeScores2yr, 30
CollegeScores4yr, 31
CommuteAtlanta, 33
CommuteStLouis, 33
CompassionateRats, 34
CricketChirps, 35

DDS, 35
DecemberFlights, 36
DecemberFlights2e, 36
DietDepression, 37
Digits, 38
DogOwner, 38
DrugResistance, 39

EducationLiteracy, 40
EducationLiteracy2e, 40
ElectionMargin, 41
EmployedACS, 41
EmployedACS2010, 42
ExerciseHours, 43

FacebookFriends, 44
FatMice18, 44
FireAnts, 45
FisherIris, 46
FishGills12, 46
FishGills3, 47
Flight179, 47
Flight433, 48
Flight433_2e, 48
FloridaLakes, 49
FootballBrain, 50
ForestFires, 50

GeneticDiversity, 51
GlobalInternet2010, 52
GlobalInternet2019, 53
GolfRound, 53
GPAbySex, 54
GSWarriors2016, 54
GSWarriors2019, 56

HappyPlanetIndex, 57
SandP500, 121
SandP500le, 122
SandP5002e, 122
SandwichAnts, 123
SandwichAnts2, 124
SkateboardPrices, 124
SleepCaffeine, 125
SleepStudy, 125
Smiles, 127
SpeedDating, 127
SplitBill, 128
StatGrades, 129
StockChanges, 129
StorySpoilers, 131
StressedMice, 132
StudentSurvey, 132
SynchronizedMovement, 133

TenCountries, 134
TenCountriesle, 135
TenCountries2e, 135
TextbookCosts, 136
ToenailArsenic, 136
TrafficFlow, 137

USStates, 137
USStatesle, 139
USStates2e, 140

WaterStriders, 141
WaterTaste, 141
Wetsuits, 142

YoungBlood, 143