

# Package ‘astrodatR’

February 19, 2015

**Type** Package

**Title** Astronomical Data

**Version** 0.1

**Date** 2014-08-08

**Author** Eric D. Feigelson

**Maintainer** “Eric D., Feigelson” <edf@astro.psu.edu>

**Description** A collection of 19 datasets from contemporary astronomical research. They are described in the textbook ‘Modern Statistical Methods for Astronomy with R Applications’ by Eric D. Feigelson and G. Jogesh Babu (Cambridge University Press, 2012, Appendix C) or on the website of Penn State’s Center for Astrostatistics (<http://astrostatistics.psu.edu/datasets>). These datasets can be used to exercise methodology involving: density estimation; heteroscedastic measurement errors; contingency tables; two-sample hypothesis tests; spatial point processes; nonlinear regression; mixture models; censoring and truncation; multivariate analysis; classification and clustering; inhomogeneous Poisson processes; periodic and stochastic time series analysis.

**License** GPL

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2014-08-12 22:20:23

## R topics documented:

asteroid_dens . . . . .	2
sensor_Be . . . . .	3
COMBO17_lowz . . . . .	3
COUP_var . . . . .	4
ell_gal_profile . . . . .	5
exoplanet_RV . . . . .	6
GlobClus_mag . . . . .	6
GlobClus_prop . . . . .	7
GX . . . . .	8
HIP . . . . .	9

LMC_dist . . . . .	9
plan_neb_LF . . . . .	10
protostellar_disks . . . . .	11
protostellar_jets . . . . .	12
SDSS_ptsrc_test . . . . .	12
SDSS_ptsrc_train . . . . .	13
SDSS_QSO . . . . .	14
Shapley_galaxy . . . . .	15
Sun_spot_num . . . . .	15

<b>Index</b>	<b>17</b>
--------------	-----------

---

asteroid_dens	<i>Densities of asteroids</i>
---------------	-------------------------------

---

### Description

This data set gives measured values for the physical density (in grams/cm<sup>3</sup>) of 26 asteroids in the main Asteroid Belt of our Solar System, along with their measurement errors. This dataset is useful for analysis involving heteroscedastic measurement errors in 1-dimension, such as weighted density estimation and regression. Further description and references for the dataset are given in Appendix C.1 of Feigelson & Babu (2012).

### Usage

```
asteroid_dens
```

### Format

A table containing 26 rows and 3 columns with header row.

### Source

Feigelson and Babu (2012)

### References

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

---

censor\_Be

*Stellar abundances and planets*

---

### Description

This data set allows investigation of whether the presence or absence of a giant planet is correlated with the level of lithium and beryllium in the photosphere of the host star. The dataset consists of 3 measured variables for 68 solar-type stars, roughly half with and half without detected planets (Santos et al. 2002). Some missing data are present. This dataset is useful for survival analysis involving correlation of singly- and doubly-censored data. Note that astronomical data are typically left-censored. Further description and references for the dataset are given in Appendix C.4 of Feigelson & Babu (2012).

### Usage

censor\_Be

### Format

A table containing 68 rows and 8 columns with header row. Column 2 is a sample indicator (1 = star with planet; 2 = star without planet). Columns 4 and 7 are censoring indicators (1 = detected; 0 = undetected, left-censored).

### Source

Feigelson and Babu (2012), Santos et al. (2002)

### References

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

Santos, N. C., Garcia Lopez, R. J., Israelian, G., Mayor, M., Rebolo, R., Garcia-Gil, A., Perez de Taoro, M. R. & Randich, S. (2002) Beryllium abundances in stars hosting giant planets, *Astropnomy & Astrophysics*, 386, 1028-1-38

---

COMB017\_lowz

*Galaxy color-magnitude diagram*

---

### Description

This data set gives two photometric quantities for 572 low redshift ( $z < 0.3$ ) galaxies from the COMBO-17 survey (Classifying Objects by Medium-Band Observations in 17 Filters). The quantities are: absolute magnitude in the blue band,  $M(B)$ ; and ultraviolet-to-blue color index,  $M(280)-M(B)$ . This dataset is useful for 2-dimensional density estimation, clustering, and mixture models. Further description is given in Appendix C.10 of Feigelson & Babu (2012). The full COMBO-17 dataset can be accessed at <http://vizier.u-strasbg.fr/viz-bin/VizieR?-source=II%2F253A>.

**Usage**

COMB017\_lowz

**Format**

A table containing 572 rows and 2 columns with header row.

**Source**

Wolf et al. (2003), Feigelson and Babu (2012)

**References**

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

Wolf, C., Meisenheimer, K., Rix, H.-W., Borch, A., Dye, S. and Kleinheinrich, M. (2003) The COMBO-17 survey: Evolution of the galaxy luminosity function from 25 000 galaxies with  $0.2 < z < 1.2$ , *Astronomy and Astrophysics*, 401, 73-98

---

COUP\_var

*COUP: X-ray source variability*


---

**Description**

This dataset consists of three time series representing (in)homogeneous Poisson processes. They are tables of arrival times of individual X-ray photons from magnetically flaring young stars in the Orion Nebula Cluster, obtained from the Chandra Orion Ultradeep Project (COUP; Getman et al. 2005). NASA's Chandra X-ray Observatory is described at [http://en.wikipedia.org/wiki/Chandra\\_X-ray\\_Observatory](http://en.wikipedia.org/wiki/Chandra_X-ray_Observatory). COUP detected 1616 X-ray sources, mostly flaring young stars. The first source, COUP 263, has 209 photon arrival times that appear to represent a constant X-ray intensity. The second source, COUP 551, has 678 photons and exhibits at least two flares. The third source, COUP 554, is much stronger with 14,258 photons; it exhibits very high-amplitude, nearly continuous flaring. The dataset has three columns: the source identifier, the photon arrival time, and the photon X-ray energy. Gaps in the data streams due to the satellite observatory orbit have been removed. Photon energy may also represent a time dependent process as flares typically have higher average energy X-rays than quiescent periods. The dataset is discussed at [http://astrostatistics.psu.edu/datasets/Chandra\\_flares.html](http://astrostatistics.psu.edu/datasets/Chandra_flares.html). This dataset is useful for statistical methods treating event data and inhomogeneous Poisson processes.

**Usage**

COUP\_var

**Format**

A table containing 15,145 rows and 3 columns

**Source**

Getman et al. (2005)

**References**

Getman, K. V., Flaccomio, E., Broos, P. S., Grosso, N., Tsujimoto, M., Townsley, L., Garmire, G. P., Kastner, J., Li, J., Harnden, F. R., Jr., Wolk, S., Murray, S. S., Lada, C. J., Muench, A. A., McCaughrean, M. J., Meeus, G., Damiani, F., Micela, G., Sciortino, S., Bally, J., Hillenbrand, L. A., Herbst, W., Preibisch, T., Feigelson, E. D. (2005), Chandra Orion Ultradeep Project: Observations and source lists, *Astrophysical Journal Supplements*, 160, 319-352 <http://adsabs.harvard.edu/abs/2005ApJS...160..319G>

---

ell_gal_profile	<i>Elliptical galaxy radial profiles</i>
-----------------	--

---

**Description**

This dataset gives the radial surface brightness profiles for three elliptical galaxies in the nearby Virgo galaxy cluster (Kormendy et al. 2009). They are useful for exercising nonlinear regression in 2-dimensions; the Sersic (1968) profile function is commonly used. The samples are described in Appendix C.11 of Feigelson & Babu (2012) with regression illustrated in Chapter 7.

**Usage**

ell\_gal\_profile

**Format**

A table containing 150 rows and 3 columns with header row. It includes 58 entries for NGC 4472, 52 rows for NGC 4406, and 40 rows for NGC 4551.

**Source**

Feigelson and Babu (2012), Kormendy et al. (2009), Sersic (1968)

**References**

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK: Cambridge University Press

Kormendy, J., Fisher, D. B., Cornell, M. E. & Bender, R. (2009) Structure and formation of elliptical and spheroidal galaxies, *Astrophysical Journal Supplements*, 182, 216-309 <http://adsabs.harvard.edu/abs/2009ApJS...182..216K>

Sersic, J. L. (1968) *Atlas de Galaxias Australes*, Obs. Astronomico de Cordoba <http://adsabs.harvard.edu/abs/1968adga.book.....S>

---

 exoplanet\_RV

*Exoplanet radial velocities*


---

### Description

This dataset gives radial velocity time series for three stars hosting massive planets. The technique of detecting extrasolar planets by periodic Doppler variations in the stellar radial velocity is summarized at [http://en.wikipedia.org/wiki/Radial\\_velocity](http://en.wikipedia.org/wiki/Radial_velocity). Data are provided here for three stars: HD 88133 (17 measurements), HD 37124 (52 measurements), and HD 3651 (138 measurements). The four columns give the star identifier, observation times (in modified Julian days), heliocentric radial velocities in meters/sec, and standard deviation of the radial velocities. References and discussion can be found at [http://astrostatistics.psu.edu/datasets/exoplanet\\_Doppler.html](http://astrostatistics.psu.edu/datasets/exoplanet_Doppler.html). This dataset is useful for statistical methods treating detection of non-sinusoidal periodicities in sparse irregularly-spaced time series with heteroscedastic measurement errors.

### Usage

exoplanet\_RV

### Format

A table containing 27 rows and 4 columns

---

 GlobClus\_mag

*Globular cluster magnitudes*


---

### Description

This data set gives near-infrared K-band magnitudes for 81 globular clusters in the Milky Way Galaxy (MWG) and 360 globular clusters in the nearby Andromeda Galaxy (M31). The data are obtained from Nantais et al. (2006) and the samples are more fully described in Appendix C.3 of Feigelson & Babu (2012). This dataset is useful for univariate two-sample tests, parametric estimation, and treatment of truncation.

### Usage

GlobClus\_mag

### Format

A table containing 441 rows and 3 columns with header row. The first column is a subsample indicator.

### Source

Nantais et al. (2006), Feigelson and Babu (2012)

## References

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

Nantais, J. B., Huchra, J. P., Barmby, P., Olsen, K. A. & Jarrett, T. H. (2006) Nearby spiral globular cluster systems. I. Luminosity functions, *Astronomical Journal*, 131, 1416-1425 <http://adsabs.harvard.edu/abs/2006AJ...131.1416N>

---

GlobClus\_prop

*Galactic globular cluster properties*

---

## Description

This dataset provides astronomical and astrophysical properties of 147 globular star clusters in the Milky Way Galaxy obtained from the catalog of Webbink (1985). Properties include Galactic location, integrated stellar luminosity, metallicity, ellipticity, central surface brightness, color, and seven measures of dynamical state (core and tidal radius, concentration, central star density and relaxation time, central velocity dispersion and escape velocity). The sample is described in Appendix C.7 of Feigelson & Babu (2012) with statistical analysis in Chapter 8. It is useful to exercise techniques of multivariate analysis and regression.

## Usage

GlobClus\_prop

## Format

A table containing 147 rows and 20 columns with header row. Several columns have 'NA' entries.

## Source

Feigelson and Babu (2012), Webbink (1985)

## References

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

Webbink, R. F., Structure parameters of galactic globular clusters, in *Dynamics of Star Clusters*, Dordrecht NL:Reidel, 541-577 <http://adsabs.harvard.edu/abs/1985IAUS..113..541W>

---

GX

*GX 5-1: X-ray source variability*

---

### Description

This dataset is a time series of the rapid changes in brightness seen in the Galactic X-ray binary star GX 5-1 lying near the Galactic Center. X-ray binaries produce prodigious high energy radiation as gas from a companion star is accreted onto a compact companion (white dwarf, neutron star or black hold); see [http://en.wikipedia.org/wiki/X-ray\\_binary](http://en.wikipedia.org/wiki/X-ray_binary). While some systems show periodic variations from stellar rotation and orbits, others show stochastic quasi-periodic variations (QPOs) and red noise (1/f-type noise) from the accretion disk around the compact star; see [http://en.wikipedia.org/wiki/Quasi-periodic\\_oscillations](http://en.wikipedia.org/wiki/Quasi-periodic_oscillations). Over a thousand studies of the QPO phenomenon have been published.

This dataset consists of 65,536 evenly spaced measurements of X-ray counts detected in 1/128 second intervals by the Japanese Ginga satellite (Norris et al. 1990). While the photon counts constitute a Poisson process, the values are sufficiently high (~70 counts/interval) that a Gaussian process assumption is reasonable. The dataset is described in Appendix C.12 of Feigelson & Babu (2012) and Hertz and Feigelson (1995). It can be used to study stochastic temporal behaviors in evenly-spaced time series.

### Usage

GX

### Format

A stream of 65,536 2-digit integers that can be read into R using ‘scan(GX)’

### Source

Feigelson and Babu (2012), Hertz & Feigelson (1995), Norris et al. (1990)

### References

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

Hertz, P. and Feigelson, E. D. (1995) A sample of astronomical time series, in *applications of Time Series Analysis in Astronomy and Meteorology*. T. SubbaRao (ed.), Chapman & Hall, 340-356

Norris, J. P., Hertz, P., Wood, K. S., Vaughan, B., A., Michelson, P. F., Mitsuda, K. & Dotani, T. (1990) Independence of short time scale fluctuations of quasi-periodic oscillations and low frequency noise in GX5-1, *Astrophysical Journal*, 361, 514-526. <http://adsabs.harvard.edu/abs/1990ApJ...361..514N>



---

HIP	<i>Hipparchos stars</i>
-----	-------------------------

---

### Description

This dataset is a subset of the large catalog of stellar positions and parallaxes (inverse of distances) obtained by the European Space Agency's Hipparchos satellite (Perryman et al. 1997). The 2719 stars are chosen from the Hipparchos Input Catalogue (HIP) with the criterion that the parallax lies between 20 and 25 milliarcsecond (corresponding to distance between 40 and 50 parsecs). This subset includes most of the nearby Hyades star cluster. The dataset also includes celestial location, proper motion in right ascension and declination, visual magnitude, and  $B-V$  color index. The sample is described in Appendix C.6 of Feigelson & Babu (2012). This dataset is useful for 2- and 3-dimensional clustering, mixture models, regression, and spatial point processes.

### Usage

HIP

### Format

A table containing 2719 rows and 9 columns with header row. The last column has a few 'NA' entries.

### Source

Feigelson and Babu (2012), Perryman et al. (1997)

### References

- Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press
- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., Hoeg, E., Bastian, U., Bernacca, P. L., Creze, M., Donati, F., Grenon, M., Grewing, M., van Leeuwen, F., van der Marel, H., Mignard, F., Murray, C. A., Le Poole, R. S., Schrijver, H., Turon, C., Arenou, F., Froeschle, M., and Petersen, C. S. (1997) The HIPPARCOS Catalogue, *Astronomy and Astrophysics*, 323, L49-L52.

---

LMC_dist	<i>Distance to the Large Magellanic Cloud</i>
----------	---

---

**Description**

This dataset is a collection of 25 measurements of the distance to the Large Magellanic Cloud (LMC), the largest galaxy now orbiting the Milky Way Galaxy (Clementini 2003). The measurements are made by different research groups using different methods: Cepheid and RR Lyrae luminosities, red clump giants, Supernova 1987a expansion, and so forth. Each distance modulus (DM) measurement is accompanied by its standard deviation. The dataset is discussed at [http://astrostatistics.psu.edu/datasets/LMC\\_distance.html](http://astrostatistics.psu.edu/datasets/LMC_distance.html). This dataset is useful for statistical methods treating heteroscedastic measurement errors.

**Usage**

LMC\_dist

**Format**

A table containing 25 rows and 3 columns

**Source**

Clementini et al. (2003)

**References**

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

Clementini, G., Gratton, R., Bragaglia, A., Carretta, E., Di Fabrizio, L. and Maio, M. (2003) Distance to the Large Magellanic Cloud: The RR Lyrae Stars, *Astronomical Journal*, 125, 1309-1329

---

plan\_neb\_LF

*Planetary nebula luminosity function*

---

**Description**

This dataset contains the visual magnitudes of 531 planetary nebulae in five nearby galaxies: M 31 (Andromeda), M 81, NGC 3379, NGC 4494 and NGC 4382 (Ciardullo et al. 2002 and references therein). If the distribution of planetary nebula luminosities (its ‘luminosity function’) is universal, then the offsets between the distributions can be used to estimate galaxy distances. However, the samples are truncated at different levels for each galaxy. The dataset is discussed at [http://astrostatistics.psu.edu/datasets/plan\\_neb.html](http://astrostatistics.psu.edu/datasets/plan_neb.html) (with references) and can be used to exercise statistical methods of density estimation and regression in the presence of truncation.

**Usage**

plan\_neb\_LF

**Format**

A table containing 531 rows and 2 columns with header row

**Source**

Ciardullo et al. (2002)

**References**

Ciardullo, R., Feldmeier, J. J., Jacoby, G. H., Kuzio de Naray, R., Laychak, M. B. and Durrell, P. R. (2002) Planetary nebulae as standard candles. XII. Connecting the Population I and Population II distance scales, *Astrophysical Journal*, 577, 31-50

---

protostellar\_disks      *Protostellar disks*

---

**Description**

This data set gives rxc contingency table of the observed population of 5 evolutionary stages of pre-main sequence stars in four nearby star forming clouds. Clouds are listed in order of increasing age, and stellar evolutionary stages are based on the properties of their infrared-emitting circumstellar disks. Sample sizes in a cell range from 0 to 179 stars. This dataset is useful for analysis involving  $r \times c$  contingency tables with small samples. Further description and references for the dataset are given in Appendix C.2 of Feigelson & Babu (2012).

**Usage**

protostellar\_disks

**Format**

A table containing 4 rows and 5 columns with header row and column.

**Source**

Feigelson and Babu (2012)

**References**

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

---

protostellar\_jets      *Protostellar jets*

---

### Description

This data set gives a 2x2 contingency table of the presence of jets produced by protostars in single and multiple systems. The sample has 21 total systems (Reipurth et al. 2004). This dataset is useful for analysis involving 2x2 contingency tables of small samples. Further description and references for the dataset are given in Appendix C.2 of Feigelson & Babu (2012).

### Usage

protostellar\_jets

### Format

A table containing 2 rows and 2 columns with header row.

### Source

Reipurth et al. (2004), Feigelson and Babu (2012)

### References

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

Reipurth, B., Rodriquez, L. F., Anglada, G. & Bally, J. (2004) Radio continuum jets from protostellar objects, *Astronomical Journal*, 127, 1736-1746 <http://adsabs.harvard.edu/abs/2004AJ...127.1736R>

---

SDSS\_ptsrc\_test      *Sloan Digital Sky Survey point source photometry: Test sample*

---

### Description

Together with SDSS\_ptsrc\_train, these two datasets are designed to exercise methods of supervised classification in low dimensions. The datasets are extracted from the very large (~300 million) Sloan Digital Sky Survey in five photometric bands (ugriz). See York et al. (2000) and [http://en.wikipedia.org/wiki/Sloan\\_Digital\\_Sky\\_Survey](http://en.wikipedia.org/wiki/Sloan_Digital_Sky_Survey) for background on the survey. The SDSS\_test sample has four color indices (u-g, g-r, r-i, i-z) for 12884 unclassified point sources from the 5th Data Release. The SDSS\_train sample has the four color indices together with known classes for 9000 well-characterized Sloan point sources: 2000 quasars (Class 1), 5000 main sequence and giant stars (Class 2), and 2000 white dwarfs (Class 3). The references, extraction and pre-processing necessary to obtain these samples are described in Appendix C.9 of Feigelson & Babu (2012). Application of statistical and machine learning classification methods area given in Chapter 9.

**Usage**

SDSS\_ptsrc\_test

**Format**

Table SDSS\_ptsrc\_test contains 12,884 rows and 4 columns with header row. No missing or censored data are present.

**Source**

Feigelson and Babu (2012), York et al. (2000)

**References**

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

York, D. G., Adelman, J., Anderson, J. E., Anderson, S. F., Annis, J., and 140 others (2000) The Sloan Digital Sky Survey: Technical Summary, *Astronomical Journal*, 120, 1579-1587 <http://adsabs.harvard.edu/abs/2000AJ...120.1579Y>

---

SDSS\_ptsrc\_train*Sloan Digital Sky Survey point source photometry: Training sample*

---

**Description**

Together with SDSS\_ptsrc\_teset, these two datasets are designed to exercise methods of supervised classification in low dimensions. The datasets are extracted from the very large (~300 million) Sloan Digital Sky Survey in five photometric bands (ugriz). See York et al. (2000) and [http://en.wikipedia.org/wiki/Sloan\\_Digital\\_Sky\\_Survey](http://en.wikipedia.org/wiki/Sloan_Digital_Sky_Survey) for background on the survey. The SDSS\_test sample has four color indices (u-g, g-r, r-i, i-z) for 12884 unclassified point sources from the 5th Data Release. The SDSS\_train sample has the four color indices together with known classes for 9000 well-characterized Sloan point sources: 2000 quasars (Class 1), 5000 main sequence and giant stars (Class 2), and 2000 white dwarfs (Class 3). The references, extraction and pre-processing necessary to obtain these samples are described in Appendix C.9 of Feigelson & Babu (2012). Application of statistical and machine learning classification methods are given in Chapter 9.

**Usage**

SDSS\_ptsrc\_train

**Format**

Table SDSS\_ptsrc\_train contains 9000 rows and 5 columns with header row. No missing or censored data are present.

**Source**

Feigelson and Babu (2012), York et al. (2000)

**References**

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

York, D. G., Adelman, J., Anderson, J. E., Anderson, S. F., Annis, J., and 140 others (2000) The Sloan Digital Sky Survey: Technical Summary, *Astronomical Journal*, 120, 1579-1587 <http://adsabs.harvard.edu/abs/2000AJ....120.1579Y>

---

SDSS\_QSO

*Sloan Digital Sky Survey quasars*

---

**Description**

This large dataset is the catalog of 77,429 quasars (QSOs) produced from the 5th Data Release of the Sloan Digital Sky Survey (SDSS; York et al. 2000, Schneider et al 2007). It provides the SDSS designation (which gives the celestial location), redshift (a measure of distance), magnitudes in five photometric bands (with heteroscedastic measurement errors), measures or indicators of radio and X-ray emission, and absolute magnitude (a measure of luminosity). The sample is described in Appendix C.8 of Feigelson & Babu (2012) which defines the variables including the radio and X-ray censoring indicators. The dataset is useful for multivariate analysis and regression including left-censoring.

**Usage**

SDSS\_QSO

**Format**

A gzipped table containing 77429 rows and 15 columns with header row. No 'NA' entries are present.

**Source**

Feigelson and Babu (2012), Schneider et al. (2007), York et al. (2000)

**References**

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

Schneider, D. P., Hall, P. B., Richards, G. T., Strauss, M. A., Vanden Berk, D. E. and 39 others (2007) The Sloan Digital Sky Survey Quasar Catalog. IV. Fifth Data Release, *Astronomical Journal*, 134, 102-117 <http://adsabs.harvard.edu/abs/2007AJ....134..102S>

York, D. G., Adelman, J., Anderson, J. E., Anderson, S. F., Annis, J., and 140 others (2000) The Sloan Digital Sky Survey: Technical Summary, *Astronomical Journal*, 120, 1579-1587 <http://adsabs.harvard.edu/abs/2000AJ....120.1579Y>

---

 Shapley\_galaxy

*Shapley Concentration of galaxies redshift survey*


---

**Description**

This dataset gives results from a galaxy redshift survey of the Shapley Concentration (SCI 124), the richest nearby supercluster of galaxies centered around (RA,Dec) = (13h25m, -30d). It includes ~40 Abell rich clusters as well as regions with low galaxy densities (voids). See [http://en.wikipedia.org/wiki/Shapley\\_Supercluster](http://en.wikipedia.org/wiki/Shapley_Supercluster). The dataset, based on the redshift survey of Drinkwater et al. (2004), has 4215 galaxy redshifts with uncertainties in addition to celestial locations and visible magnitudes. The sample is described in Appendix C.5 of Feigelson & Babu (2012). It is useful for study of 3-dimensional spatial point processes with high-amplitude, anisotropic spatial clustering.

**Usage**

Shapley\_galaxy

**Format**

A table containing 4215 rows and 5 columns with header row. Missing data is denoted by zeros.

**Source**

Drinkwater et al. 2004, Feigelson and Babu (2012)

**References**

Drinkwater, M. J., Parker, Q. A., Proust, D., Slezak, E. & Quintana, H. (2004) The large scale distribution of galaxies in the Shapley Supercluster, *Publ Astron. Soc. Australia*, 21, 89-96

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press

---

 Sun\_spot\_num

*Sunspot numbers*


---

**Description**

This dataset provides measurements of the number of sunspots, magnetic active regions, seen each month on the Sun from 1749 until May 2014. It shows the well-known 11-year solar magnetic cycle as well as other variations. The dataset is described in Appendix C.13 of Feigelson & Babu (2012) and is obtained from the Solar Physics research group at NASA's Marshall Space Flight Center [solarscience.msfc.nasa.gov/SunspotCycle.shtml](http://solarscience.msfc.nasa.gov/SunspotCycle.shtml). The dataset has four columns: Year, Month, Monthly average sunspot number (SSN), and its standard deviation from the International Sunspot Number compiled by the Solar Influences Data Analysis Center in Belgium <http://sidc.oma.be/>. Users are encouraged to retrieve updated versions prior to study. The dataset is useful for the analysis of periodic and aperiodic time series.

**Usage**

Sun\_spot\_num

**Format**

A table containing 3185 rows and 4 columns.

**Source**

Feigelson and Babu (2012)

**References**

Feigelson, E. D. and Babu, G. J. (2012) *Modern Statistical Methods for Astronomy with R Applications*, Cambridge UK:Cambridge University Press



# Index

## \*Topic **datasets**

- asteroid\_dens, [2](#)
  - [censor\\_Be, 3](#)
  - [COMB017\\_lowz, 3](#)
  - [COUP\\_var, 4](#)
  - [ell\\_gal\\_profile, 5](#)
  - [exoplanet\\_RV, 6](#)
  - [GlobClus\\_mag, 6](#)
  - [GlobClus\\_prop, 7](#)
  - [GX, 8](#)
  - [HIP, 9](#)
  - [LMC\\_dist, 9](#)
  - [plan\\_neb\\_LF, 10](#)
  - [protostellar\\_disks, 11](#)
  - [protostellar\\_jets, 12](#)
  - [SDSS\\_ptsrc\\_test, 12](#)
  - [SDSS\\_ptsrc\\_train, 13](#)
  - [SDSS\\_QSO, 14](#)
  - [Shapley\\_galaxy, 15](#)
  - [Sun\\_spot\\_num, 15](#)
- 
- [asteroid\\_dens, 2](#)
  
  - [censor\\_Be, 3](#)
  - [COMB017\\_lowz, 3](#)
  - [COUP\\_var, 4](#)
  
  - [ell\\_gal\\_profile, 5](#)
  - [exoplanet\\_RV, 6](#)
  
  - [GlobClus\\_mag, 6](#)
  - [GlobClus\\_prop, 7](#)
  - [GX, 8](#)
  
  - [HIP, 9](#)
  
  - [LMC\\_dist, 9](#)
  
  - [plan\\_neb\\_LF, 10](#)
  - [protostellar\\_disks, 11](#)
  - [protostellar\\_jets, 12](#)