Ashish Mahabal
California Institute of Technology

R - IV

Jet Propulsion Laboratory
Keck Institute for Space Studies
Sponsored by the Jet Propulsion Laboratory and by the Keck Institute for Space Studies, California Institute of Technology
Caltech

astRowRap

(With Tejas Kale)

- An R package aimed at astronomers to:
  - introduce them to various statistical tests that can be widely used in astronomy but are not (as of yet).
  - give access to interesting historical and modern astronomical data-sets for quick tests.
  - provide worked out examples with the data-sets to aid understanding and provide context.
  - http://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaKDDguide Section 7

Using astRowRap

- In the R console, load it just like any other package

```
> library(astRowRap)
```

- Search for all available tests using

```
> ??astrowrap
```

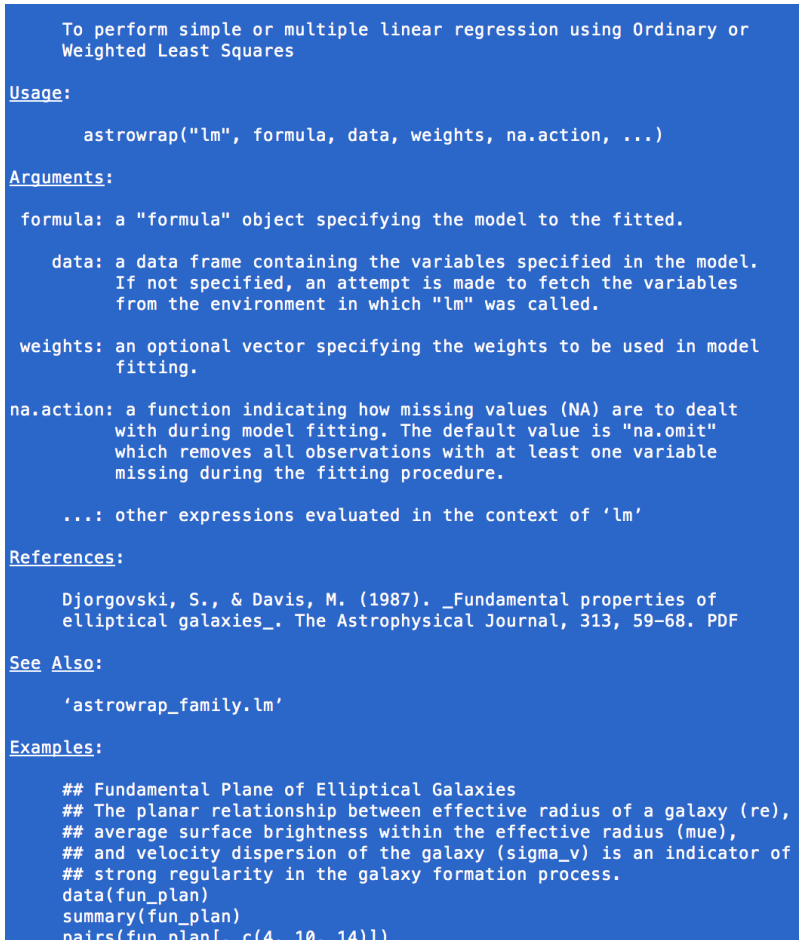
```
Help files with alias or concept or title matching 'astrowrap_' using
regular expression matching:


astRowRap::astrowrap_agnes
                        Agglomerative Clustering
astRowRap::astrowrap_anova
                        Analysis of Variance for model fits
astRowRap::astrowrap_biplot.princomp
                        PCA Biplot
astRowRap::astrowrap_boot
                        Bootstrap resampling
astRowRap::astrowrap_boot.array
                        Bootstrap Samples
astRowRap::astrowrap_boot.ci
                        Confidence Intervals of Bootstrap Parameters
astRowRap::astrowrap_cophenetic
                        Cophenetic Distance
astRowRap::astrowrap_family.lm
                        Elements of a Model Fit Object
astRowRap::astrowrap_fanny
                        Fuzzy Clustering
astRowRap::astrowrap_glm
                        Fit Generalized Linear Models
astRowRap::astrowrap_glm.diag
                        Parameters for assessing GLM fits
astRowRap::astrowrap_glm.diag.plots
                        Diagnostic plots for GLM fits
astRowRap::astrowrap_hclust
                        Agglomerative Hierarchical Clustering
astRowRap::astrowrap_kmeans
                        k-Means clustering
astRowRap::astrowrap_lda
                        Linear Discriminant Analysis
astRowRap::astrowrap_lm
                        Linear Models
astRowRap::astrowrap_lm.fit
                        Basic Engine for Straight Line Fitting
astRowRap::astrowrap_prcomp
                        Principal Component Analysis
astRowRap::astrowrap_princomp
                        Principal Component Analysis
astRowRap::astrowrap_summary
                        Summary of an object


:
```

Using astRowRap

- View help for any test by simply keying in something like

```
> ?astrowrap_lm
```

- Documentation for a test includes a detailed example using a relevant astronomical data-set and an explanation of the typical workflow involved in running the analysis



To perform simple or multiple linear regression using Ordinary or Weighted Least Squares

Usage:

```
astrowrap("lm", formula, data, weights, na.action, ...)
```

Arguments:

formula	a "formula" object specifying the model to the fitted.
data	a data frame containing the variables specified in the model. If not specified, an attempt is made to fetch the variables from the environment in which "lm" was called.
weights	an optional vector specifying the weights to be used in model fitting.
na.action	a function indicating how missing values (NA) are to dealt with during model fitting. The default value is "na.omit" which removes all observations with at least one variable missing during the fitting procedure.
...	other expressions evaluated in the context of 'lm'

References:

Djorgovski, S., & Davis, M. (1987). _Fundamental properties of elliptical galaxies_. The Astrophysical Journal, 313, 59-68. PDF

See Also:

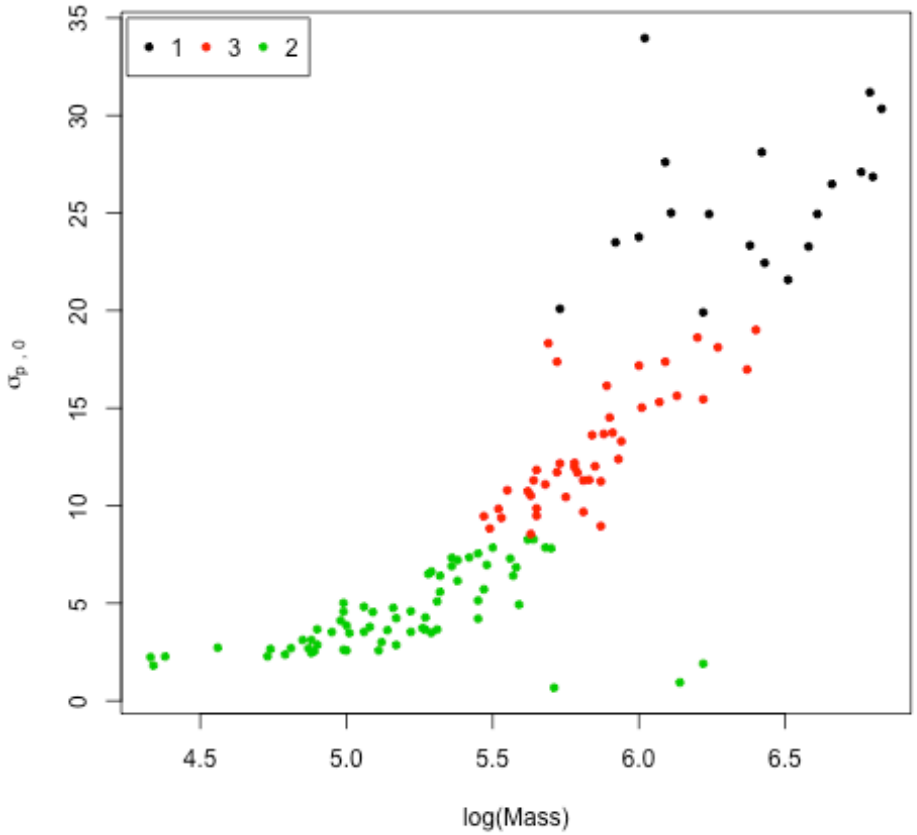
'astrowrap_family.lm'

Examples:

```
## Fundamental Plane of Elliptical Galaxies
## The planar relationship between effective radius of a galaxy (re),
## average surface brightness within the effective radius (mue),
## and velocity dispersion of the galaxy (sigma_v) is an indicator of
## strong regularity in the galaxy formation process.
data(fun_plan)
summary(fun_plan)
pairs(fun_plan[, c(4, 10, 14)])
```

Tests covered

- **Regression Analysis**
  - **Simple Linear Model**
  - **Generalized Linear Model**
  - **ANOVA**
- **Clustering**
  - **Hierarchical Clustering**
  - **k-Means Clustering**
- **Dimensionality Reduction**
  - **Principal Component Analysis**
  - **Linear Discriminant Analysis**
  - **Biplot**
- **Bootstrap**



Data included

- CRTS Lightcurves
- Faber-Jackson: Absolute magnitude and velocity dispersion of E and S0 type galaxies (to illustrate simple linear modelling)
- Colour-magnitude data of COMBO-17 galaxies (to demonstrate H-clustering)
- Properties of globular clusters from NGC5128 galaxy (to show k-Means clustering)
- Magnitudes of quasars in different optical wave-bands (to demonstrate generalised linear modelling)

Kmeans example

- Using data from Globular Clusters in NGC5128
  (Chattopadhyay, A. K et al., ApJ, 2009, 705, 1533)
- data(ngc5128)
- ngc5128_pca <- ngc5128[, c("log_rh", "sigma_p0", "mu_vh")]
- astrowrap("summary", astrowrap("princomp", ngc5128_pca))
- ngc5128_clus <- ngc5128[-c(which(ngc5128$NAME %in% c("C156", "C159", "FIGC15"), arr.ind=TRUE)), -1]
- ngc5128_kmeans <- astrowrap("kmeans", ngc5128_pca, centers=3)

{swirl}

Learn R, in R.

- Stands for “statistics with interactive R learning”.
- A software package built for the R environment that creates an interactive learning environment within the R console.
- Intended to allow educators to create concise lessons for introducing students to R. Can be take further to explain sophisticated statistical concepts to a mature audience.
- Original idea by Nick Carchedi, John Hopkins University.
- For more information, visit http://www.swirlstats.com
- We will include swirl modules in astRowRap. Currently alpha-.

More about Swirl

- Lessons available on Swirl's GitHub repository on introductory R programming, Data analysis, Regression models, etc.

- A supporting package called 'swirlify' allows anyone to create their own lesson which can be shared with the world using email, GitHub links, and variety of other ways

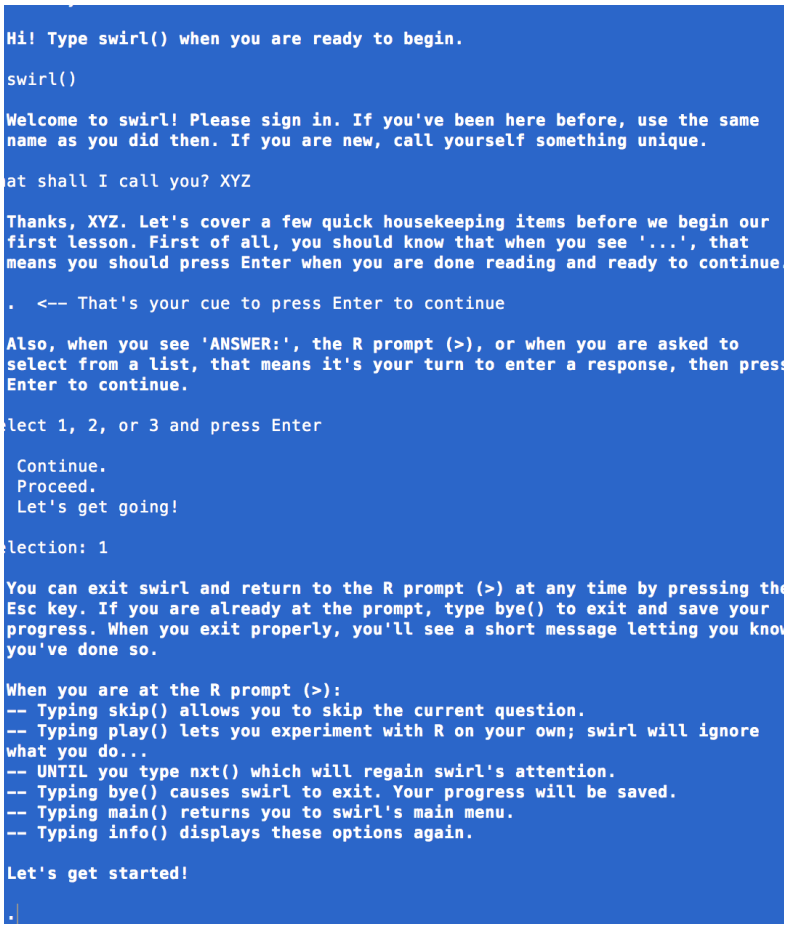
Using Swirl

- Load the Swirl package

```
> library(swirl)
```

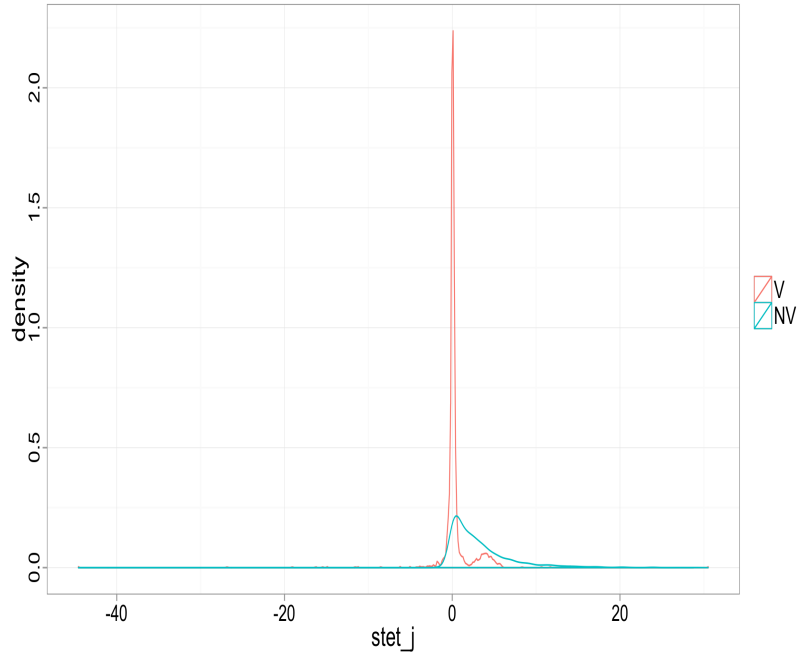
- Enter the interactive Swirl environment with the following command following which you will meet a welcome like one on the right

```
> swirl()
```



astRowRap + Swirl

- We are creating Swirl lessons using astRowRap documentation, primarily for beginners to the R environment.
- In the lesson on Linear Discriminant Analysis (LDA), we explain the algorithm using the light-curve data from Catalina Real-time Transient Survey (CRTS).



Next time …

- Classes (S3, S4, …)