“R” 101

Samantha Ahern

What is “R”?

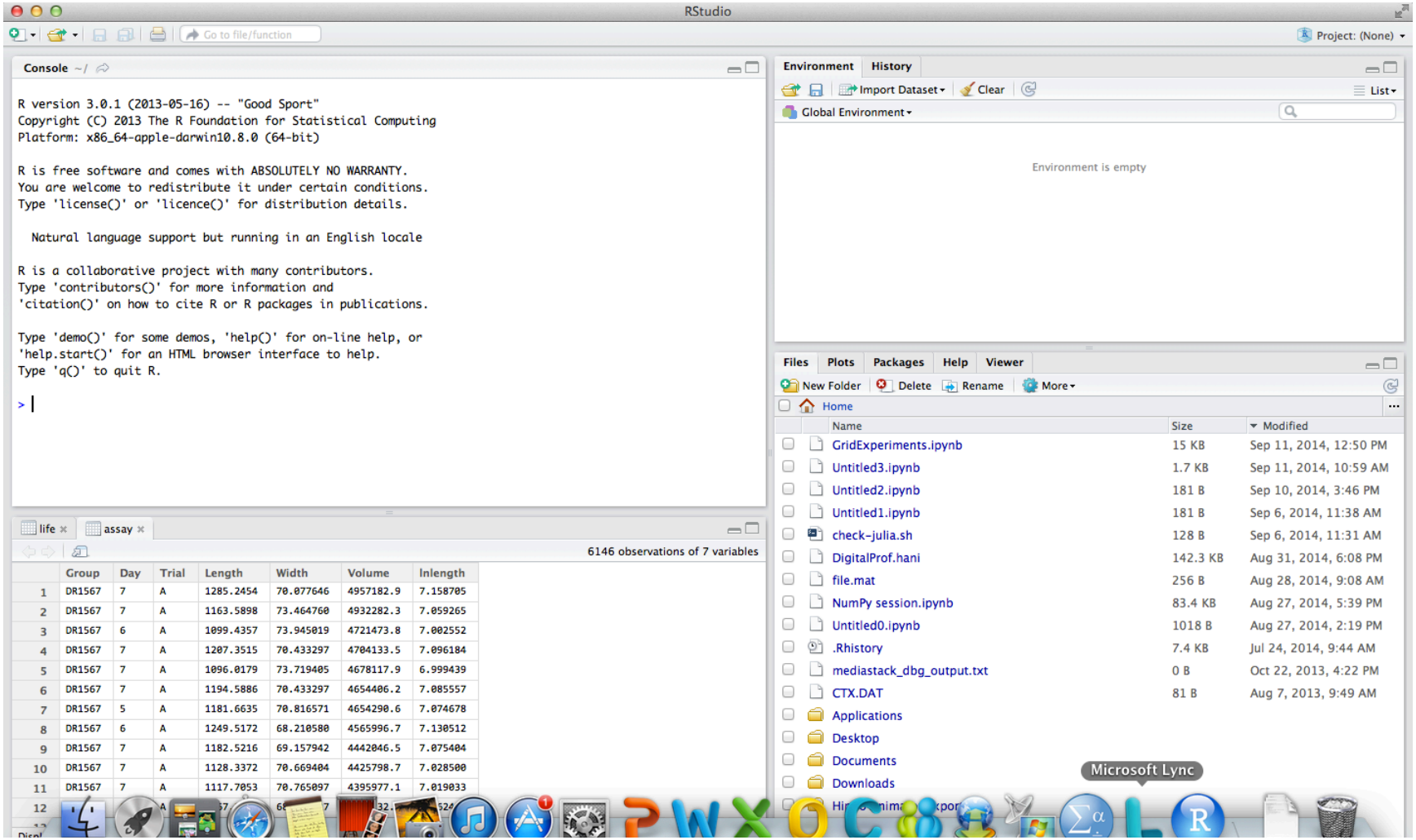
- R is an open source, free to use (but should be cited in publications) programming language for statistical analysis and mathematical modelling.

- R is interactive like other modern languages and is modular, there are thousands of packages that can be added to the base install for extra functionality such as Bayesian statistics and MCMC modelling.

- Although it is a programming language, R can be utilised through a menu system as opposed to commands using the RCmndr package developed initially by John Fox of the R Foundation.

- Unlike stats software packages such as SPSS, R does not generate spurious output. The user only receives the output they request explicitly. Plots produced are of a good quality and are suitable for publishing.

R Studio


RStudio
Go to file/function
Project: (None)
Console ~/
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin10.8.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Import Dataset
Clear
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Name
Size
Modified
GridExperiments.ipynb 15 KB Sep 11, 2014, 12:50 PM
Untitled3.ipynb 1.7 KB Sep 11, 2014, 10:59 AM
Untitled2.ipynb 181 B Sep 10, 2014, 3:46 PM
Untitled1.ipynb 181 B Sep 6, 2014, 11:38 AM
check-julia.sh 128 B Sep 6, 2014, 11:31 AM
DigitalProf.hani 142.3 KB Aug 31, 2014, 6:08 PM
file.mat 256 B Aug 28, 2014, 9:08 AM
NumPy session.ipynb 83.4 KB Aug 27, 2014, 5:39 PM
Untitled0.ipynb 1018 B Aug 27, 2014, 2:19 PM
.Rhistory 7.4 KB Jul 24, 2014, 9:44 AM
mediastack_dbg_output.txt 0 B Oct 22, 2013, 4:22 PM
CTX.DAT 81 B Aug 7, 2013, 9:49 AM
Applications
Desktop
Documents
Downloads
life
assay
6146 observations of 7 variables
Group Day Trial Length Width Volume lnlength
1 DR1567 7 A 1285.2454 70.077646 4957182.9 7.158705
2 DR1567 7 A 1163.5898 73.464760 4932282.3 7.059265
3 DR1567 6 A 1099.4357 73.945019 4721473.8 7.002552
4 DR1567 7 A 1207.3515 70.433297 4704133.5 7.096184
5 DR1567 7 A 1096.0179 73.719405 4678117.9 6.999439
6 DR1567 7 A 1194.5886 70.433297 4654406.2 7.085557
7 DR1567 5 A 1181.6635 70.816571 4654290.6 7.074678
8 DR1567 6 A 1249.5172 68.210580 4565996.7 7.130512
9 DR1567 7 A 1182.5216 69.157942 4442046.5 7.075404
10 DR1567 7 A 1128.3372 70.669404 4425798.7 7.028500
11 DR1567 7 A 1117.7053 70.765097 4395977.1 7.019033
Microsoft Lync

Basics – setting variables

- x <- 10
- y <- 5

```
z <- c(1,3,5,7,9)
a <- c(2,4,6,8,10)
b <- c(“one”,”two”,”three”)
c <- c(1977,1980,1983)
d(c)<- c(“Episode IV”,”Episode V”,”Episode VI”)
print(x)
print(pi)
cat(“Some numbers ”,b,”\n”)
```

Basics - operators

- x == y
- x != y
- x > y
- x <= y
- x > pi
- x + y
- x – y
- x / y
- x * y
- x * z
- sqrt(x)
- z == a
- z != a
- z > a
- z <= a
- a > pi
- z + a
- z – a
- a / z
- z * a
- z^2
- all(a < pi)

Basics – descriptive statistics

- mean(z)
- median(z)
- sd(z)
- var(z)
- cor(z,a)
- cov(z,a)
- summary(a)

Basics – Data frames

- Sample data sets in-built into R

- data()
- help(pressure)
- head(pressure)

- data(Cars93, package="MASS")
- head(Cars93)

Basics – accessing data frames

- Data frames are not matrices
- They are lists in a rectangular structure

- Access individual elements:
  - dataframe[i,j]

- Access a column of data
  - dataframe$columnname

- Access a row of data
  - dataframe[,j]

Basics - plots

- Plot discrete variables
  - x <- 1:10
  - y <- x^2
  - plot (x,y)
- Plot a data set
  - plot(cars)
- A more complete plot
  - plot(cars, main="cars: Speed vs. Stopping Distance (1920)",
  - xlab="Speed(mph)', ylab="Stopping distance(ft)")

Thanks for listening

- Bibliogrphy
  - R Cookbook – Paul Teetor, O'Reilly
- Check out my GitHub repository, R scripts coming soon:
  - quirksahern
- Find me on twitter:
  - @2standandstare