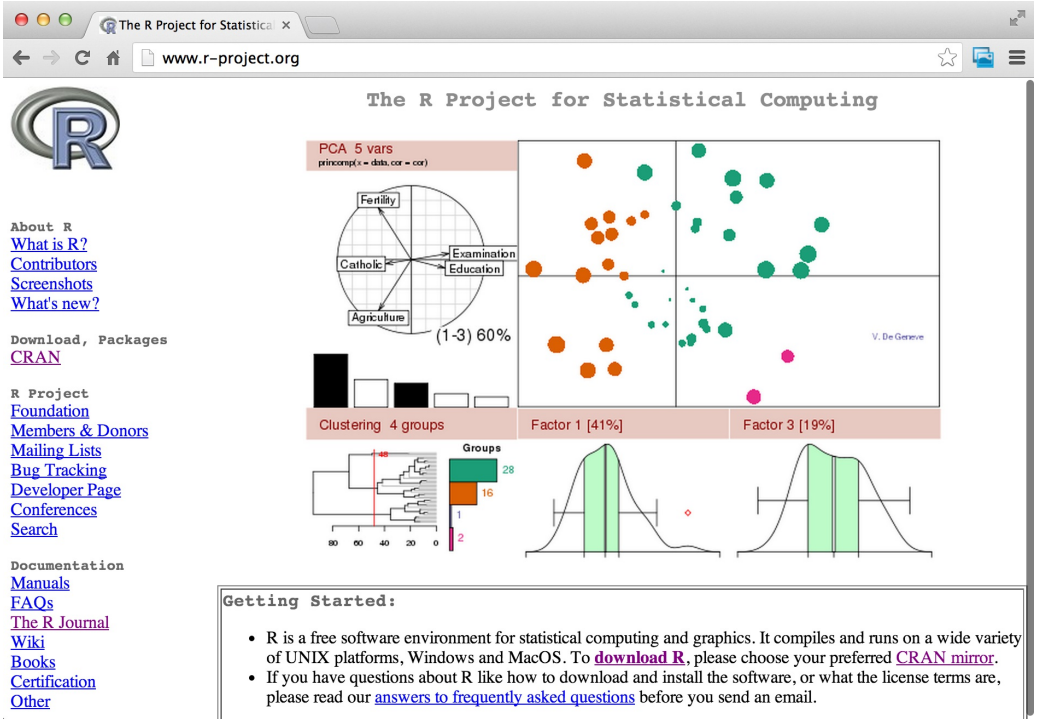
The Data Scientist's Toolbox

Johns Hopkins Bloomberg School of Public Health

What do data scientists do?

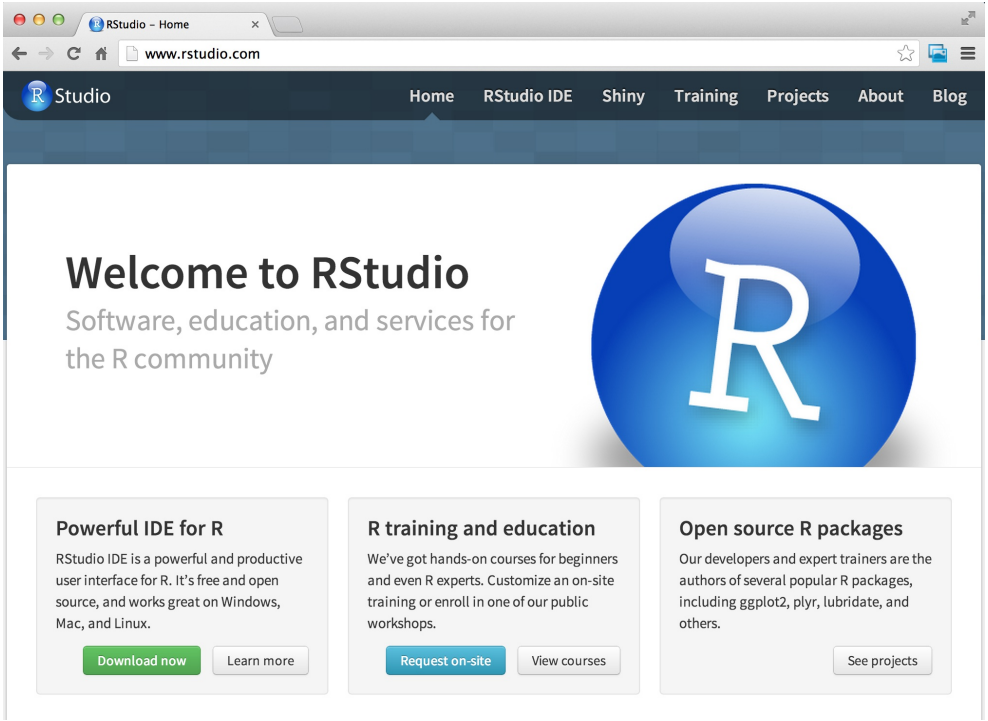
- Define the question
- Define the ideal data set
- Determine what data you can access
- Obtain the data
- Clean the data
- Exploratory data analysis
- Statistical prediction/modeling
- Interpret results
- Challenge results
- Synthesize/write up results
- Create reproducible code
- Distribute results to other people

The main workhorse of data science



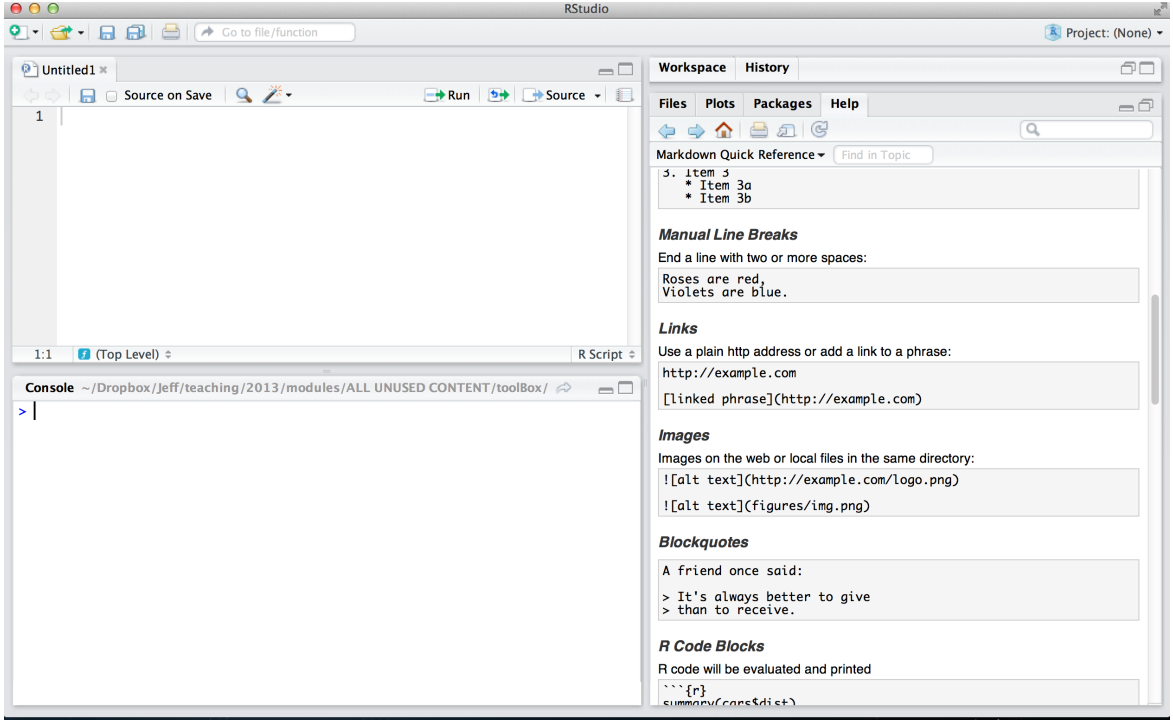
http://www.r-project.org/

Where we will work on coding



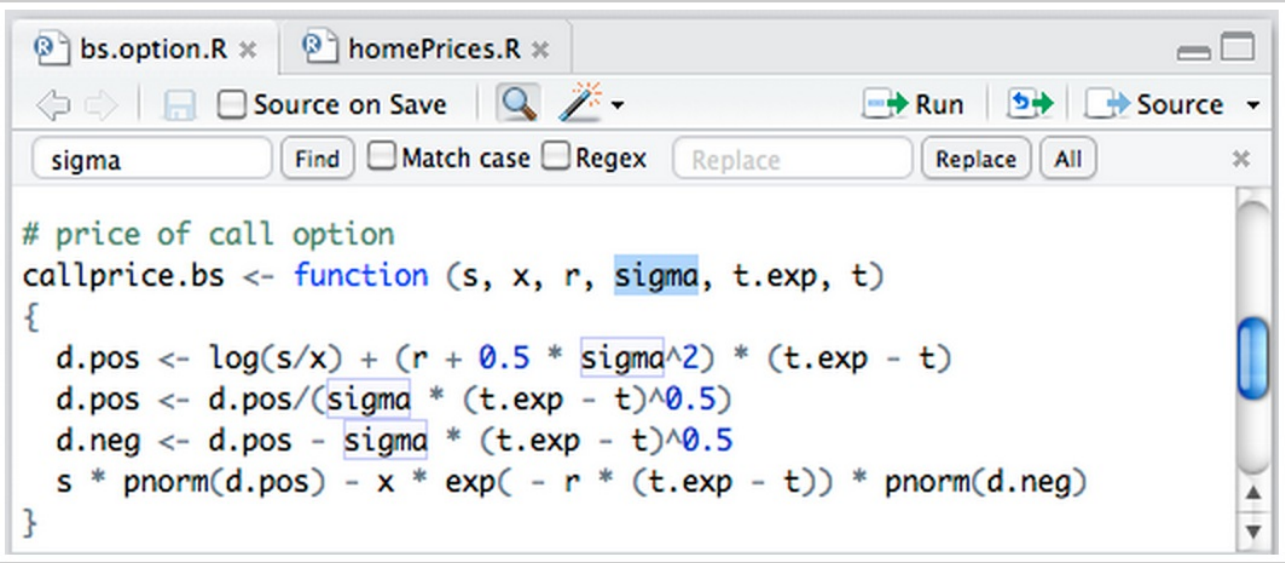
http://www.rstudio.com/

Rstudio's interface



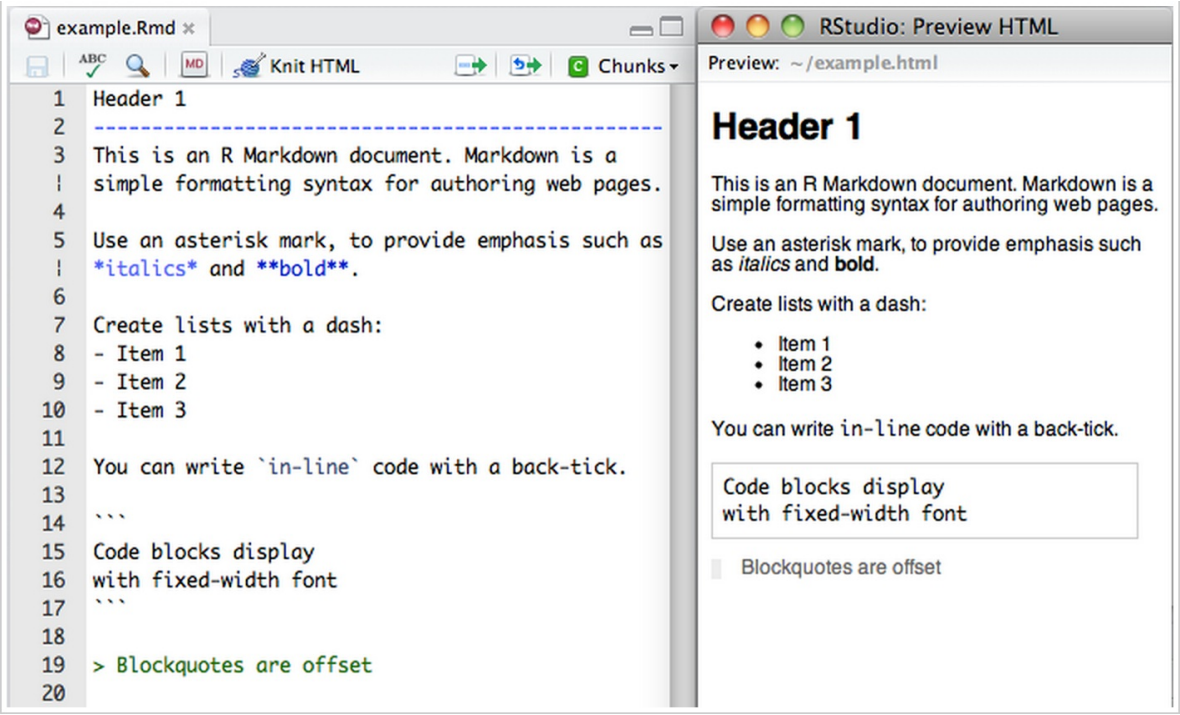
http://www.rstudio.com/

Primary file types - R script



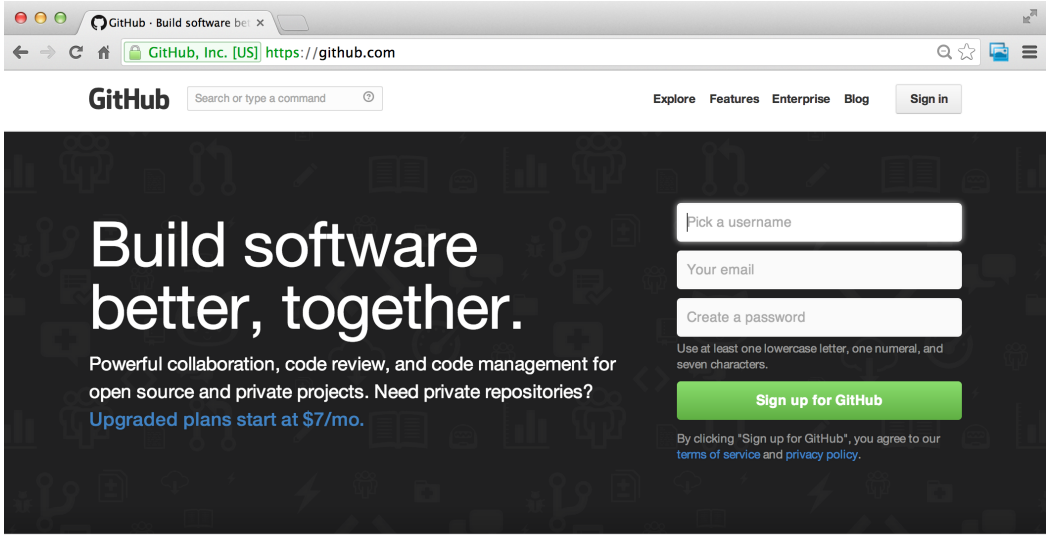
http://www.rstudio.com/ide/docs/using/source

Primary file types - R markdown document



http://www.rstudio.com/ide/docs/authoring/using_markdown

Sharing your results - Github & Git



Why you'll love GitHub.

Powerful features to make software development more collaborative.

Where to run Github commands - the shell