

Intro to R

Héctor Corrada Bravo
University of Maryland



http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?_r=2&pagewanted=1

<http://www.forbes.com/forbes/2010/05/06/opinions-software-norman-nie-spss-ideas-opinions.html>



http://www.theregister.co.uk/2010/05/06/revolution_commercial_r/

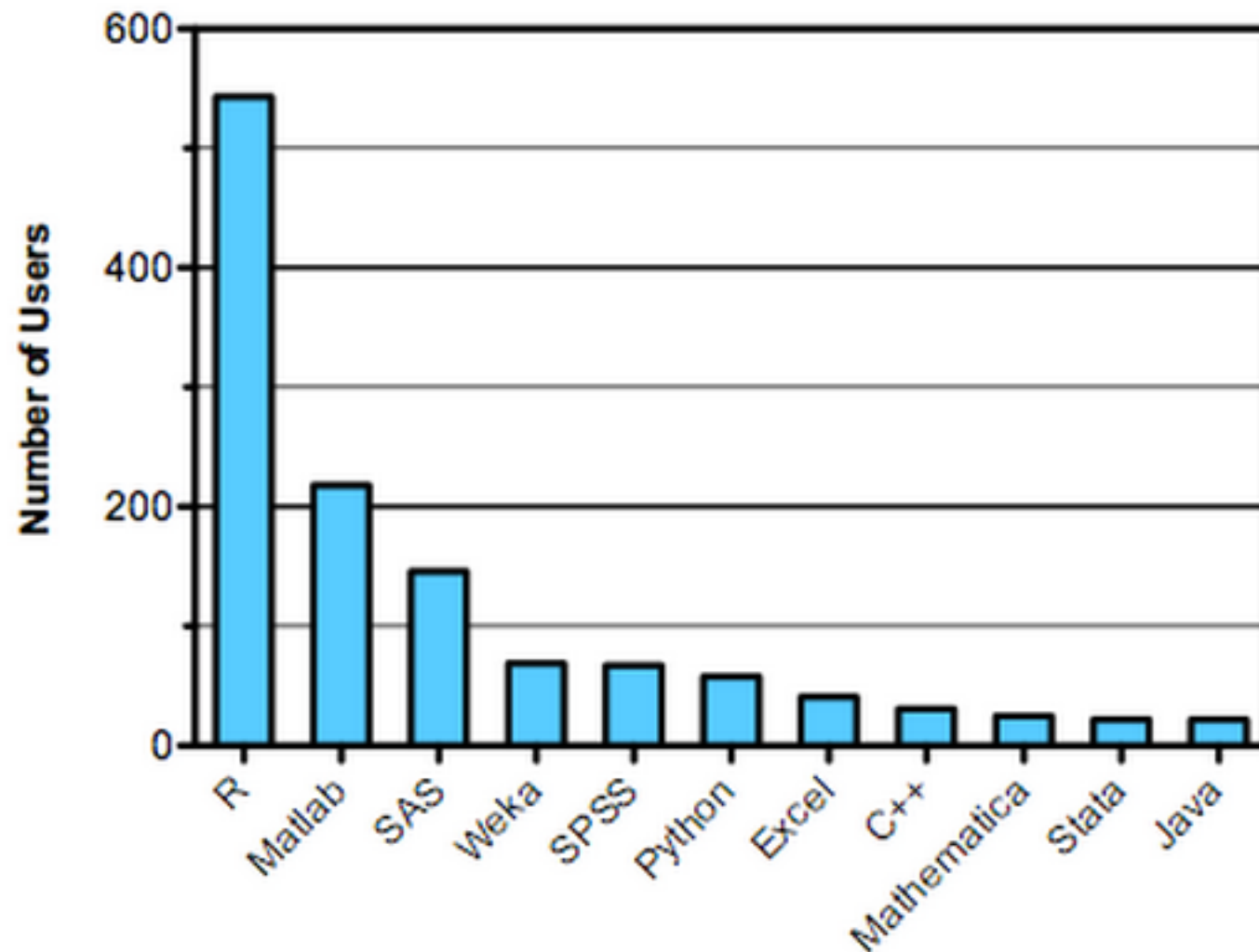
Some history

- John Chambers and others started developing the “S” language in 1976
- Version 4 of the language definition(currently in use) was settled in 1998
- That year, “S” won the ACM Software System Award

Some history

- Ihaka and Gentleman (of NYTimes fame) create R in 1991
- They wanted lexical scoping (see NYTimes pic)
- Released under GNU GPL in 1995
- Maintained by R Core Group since 1997

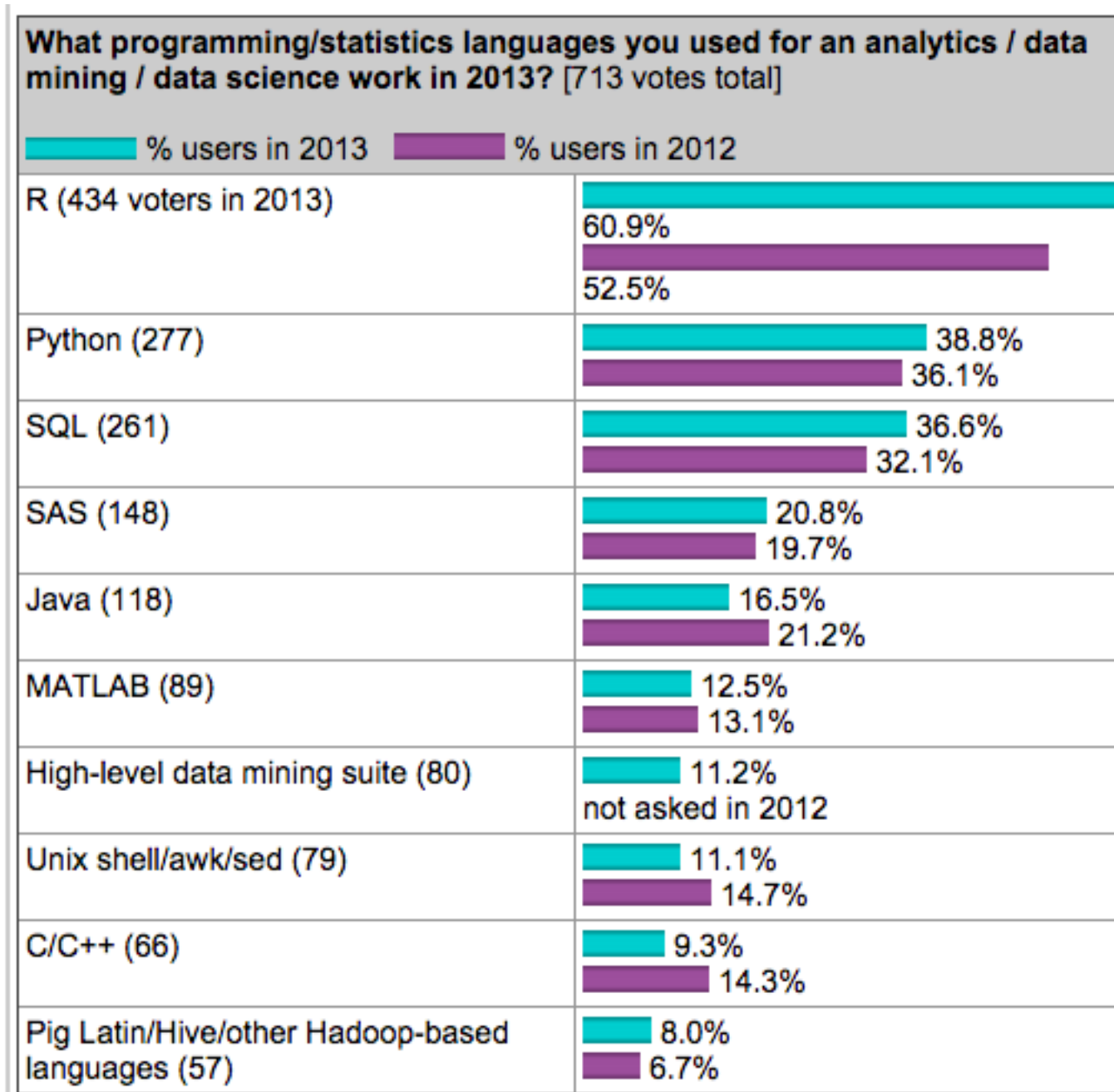
2011



Languages used in Kaggle (prediction competition site)

<http://blog.kaggle.com/2011/11/27/kagglers-favorite-tools/>

2013

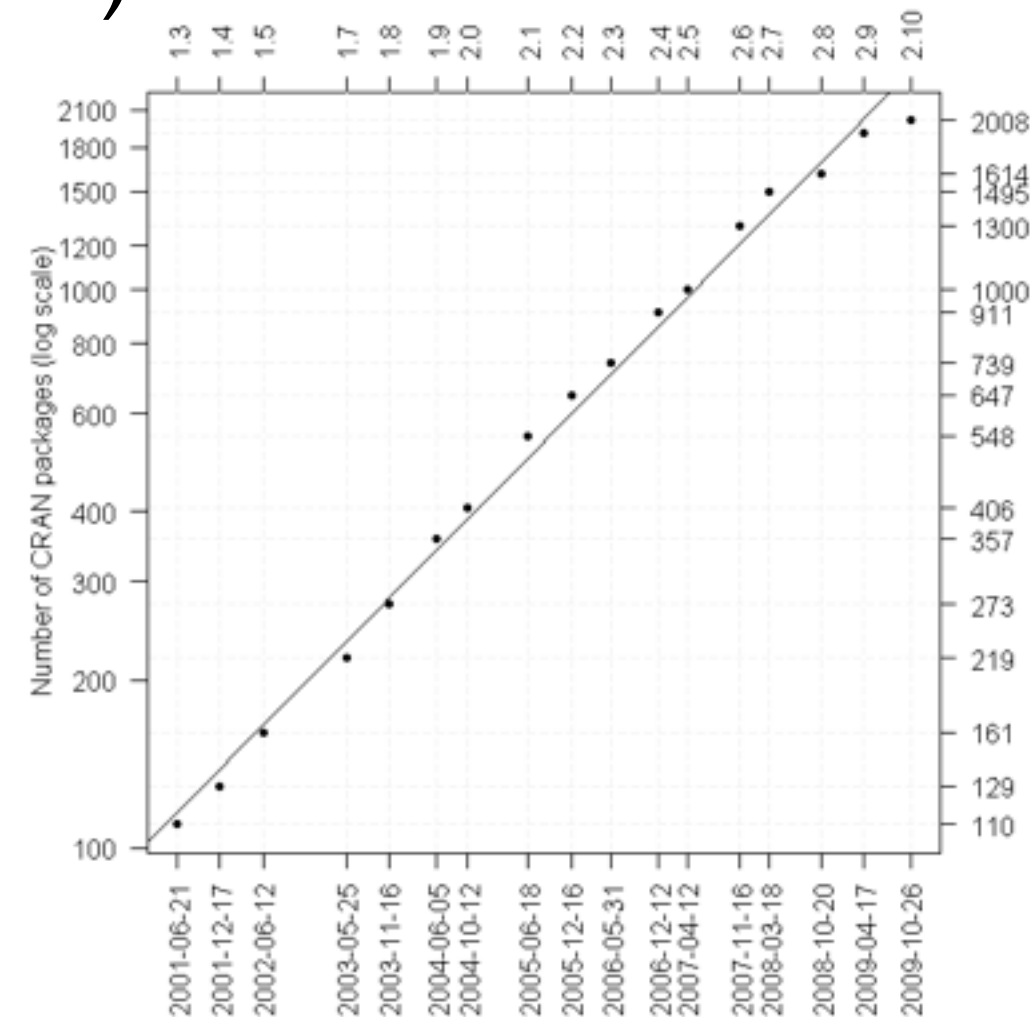


<http://www.kdnuggets.com/polls/2013/languages-analytics-data-mining-data-science.html>

- Freely available: <http://www.r-project.org/>
- IDEs:
 - [cross-platform] <http://rstudio.org/>
 - [Windows and Linux] <http://www.revolutionanalytics.com/>
 - Also bindings for emacs [<http://ess.r-project.org/>] and plugin for eclipse [<http://www.walware.de/goto/statet>]

- Resources:
 - The swirl tutorials: <http://swirlstats.com/>
 - Manuals from r-project <http://cran.r-project.org/manuals.html>
 - Chambers (2008) *Software for Data Analysis*, Springer.
 - Venables & Ripley (2002) *Modern Applied Statistics with S*, Springer.

- Uses a package framework (similar to Python)
- Divided into two parts
 - base: what you get when you download R (base package, and other packages like stats, graphics, utils, Matrix, boot, codetools)
 - everything else:
- [<http://cran.r-project.org/>]



- **Documentation system:**

> `help("sapply")` # bring up help page

> `?sapply` # shortcut

> `??sapply` # search for string in docs

> `help.start()` # open doc index

- Three ways of thinking required
 - Numerical computing (e.g., like Matlab)
 - Functions and lists (e.g., like Lisp and Scheme)
 - **Data tables (e.g., like Excel)**

- Support for literate programming: http://en.wikipedia.org/wiki/Literate_programming
- knitr and rmarkdown: integrates Markdown and R code
- Sweave: integrates Latex and R code

- Alternatives:
 - Python (with Pandas library, <http://pandas.pydata.org/>)
 - Julia (<http://julialang.org/>)