WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

A very short introduction to R

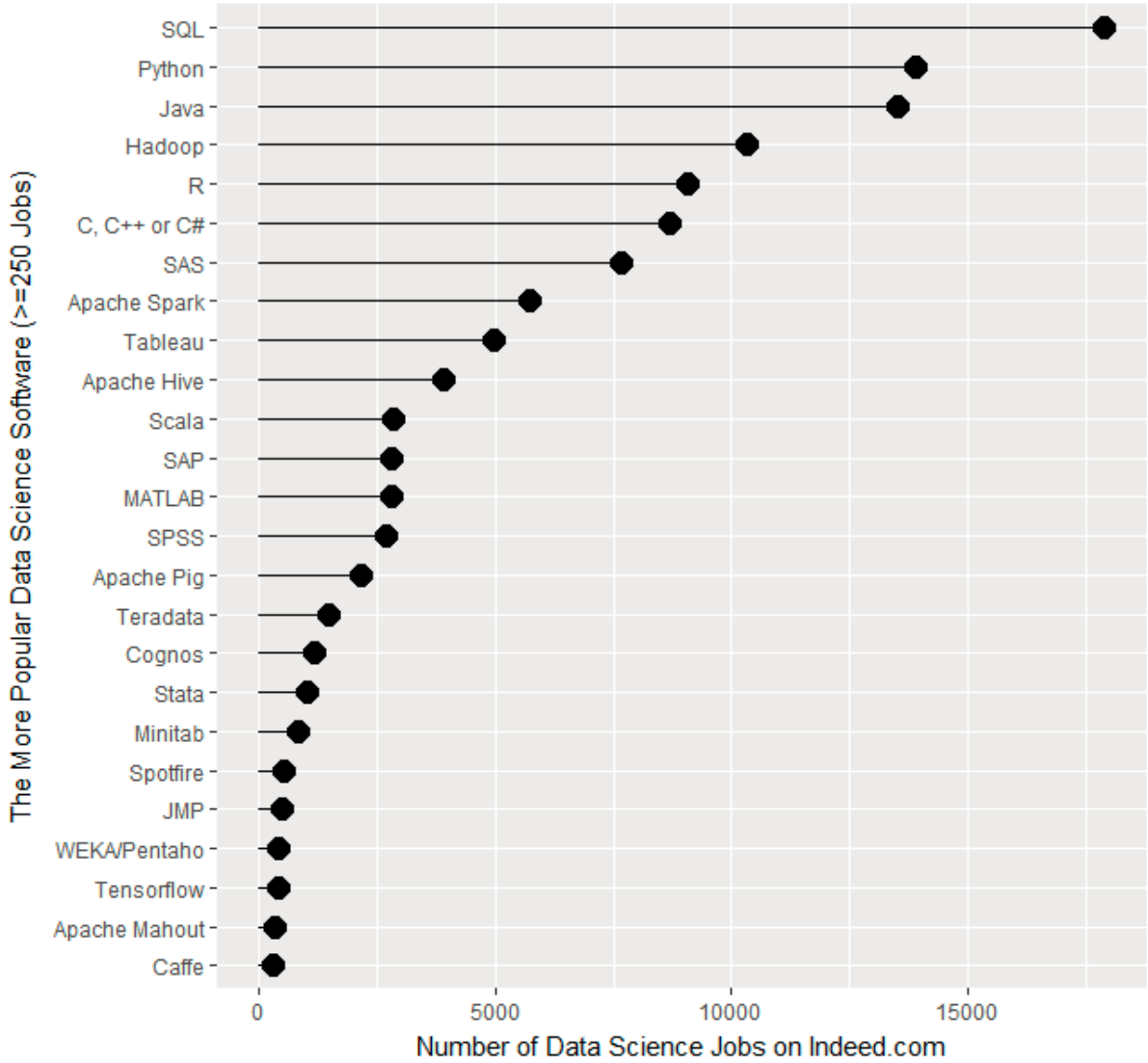
Class 3: Marketing Research

Service and Digital Marketing
March 6, 2017

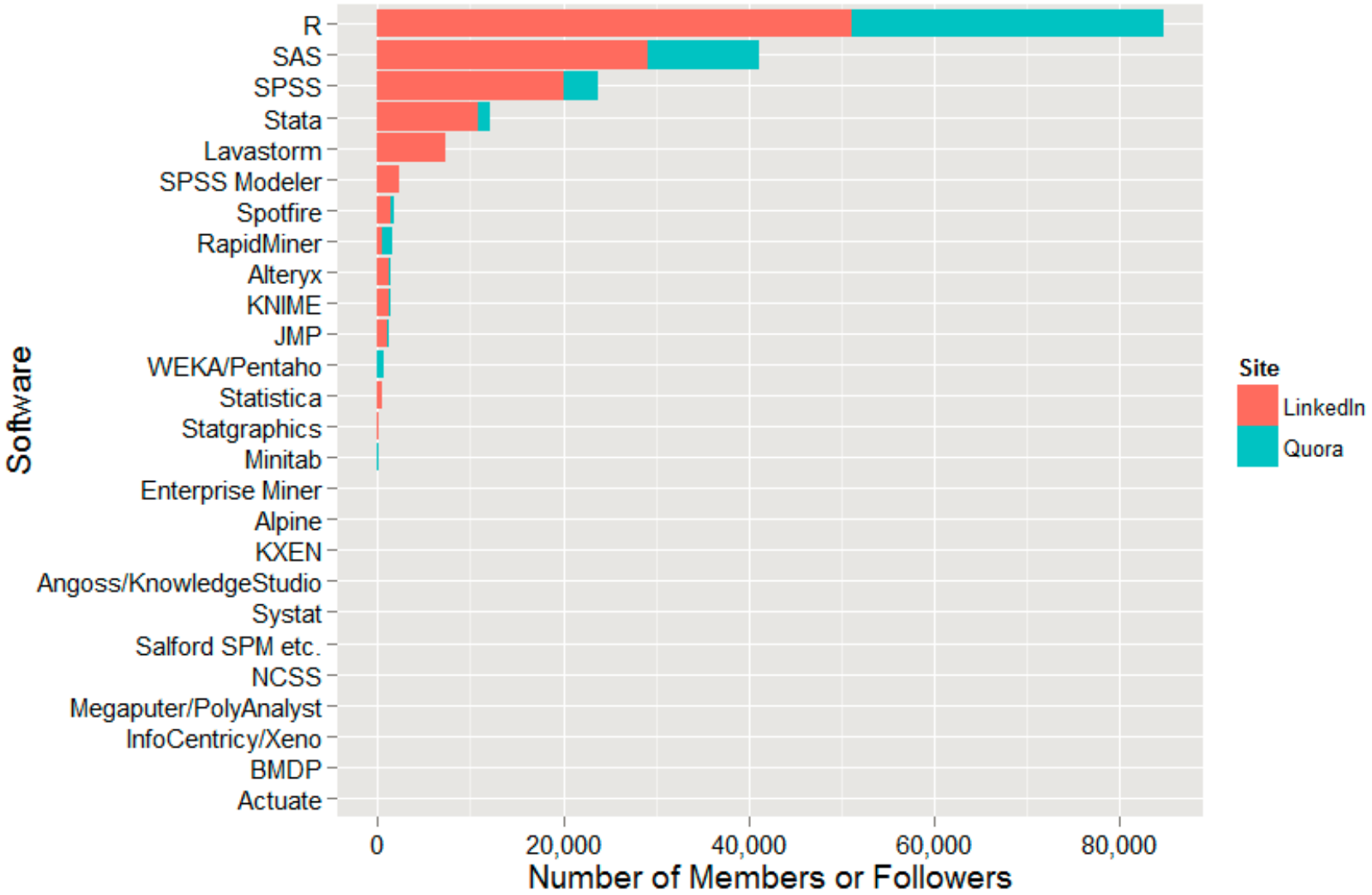
What is R?

- R is a free implementation of a dialect of the S language, the interactive statistics and graphics environment developed at Bell Labs.
- R/S are probably the most widely used software for research in statistical methodology and in genomics, and is popular in financial modeling and medical statistics.
- John Chambers won the 1999 ACM Software Systems award for S, which *will forever alter the way people analyze, visualize and manipulate data.*

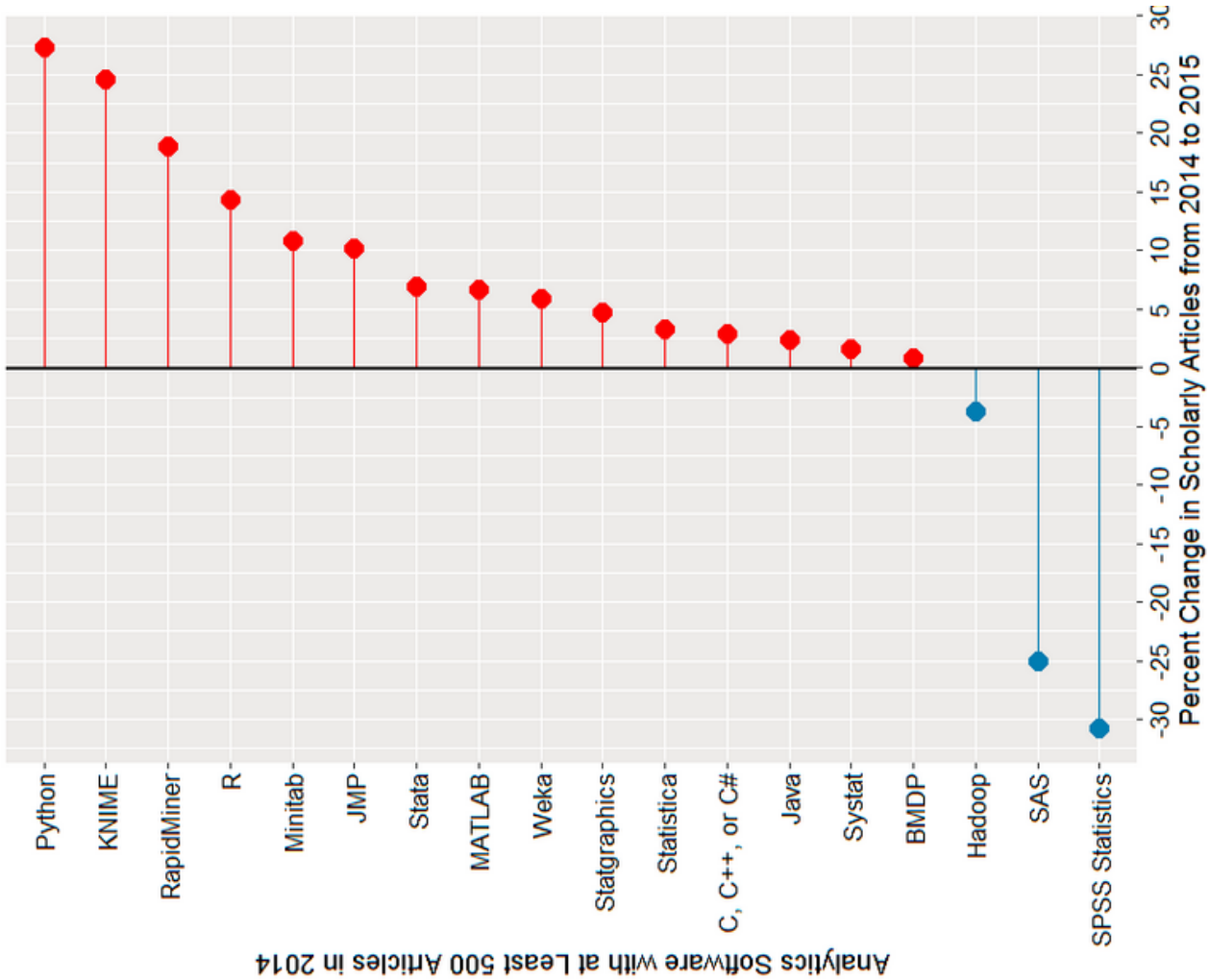
What is R?


SQL
Python
Java
Hadoop
R
C, C++ or C#
SAS
Apache Spark
Tableau
Apache Hive
Scala
SAP
MATLAB
SPSS
Apache Pig
Teradata
Cognos
Stata
Minitab
Spotfire
JMP
WEKA/Pentaho
Tensorflow
Apache Mahout
Caffe
The More Popular Data Science Software (>=250 Jobs)
0
5000
10000
15000
Number of Data Science Jobs on Indeed.com

What is R?


R
SAS
SPSS
Stata
Lavastorm
SPSS Modeler
Spotfire
RapidMiner
Alteryx
KNIME
JMP
WEKA/Pentaho
Statistica
Statgraphics
Minitab
Enterprise Miner
Alpine
KXEN
Angoss/KnowledgeStudio
Systat
Salford SPM etc.
NCSS
Megaputer/PolyAnalyst
InfoCentricy/Xeno
BMDP
Actuate
Software
0
20,000
40,000
60,000
80,000
Number of Members or Followers
Site
LinkedIn
Quora

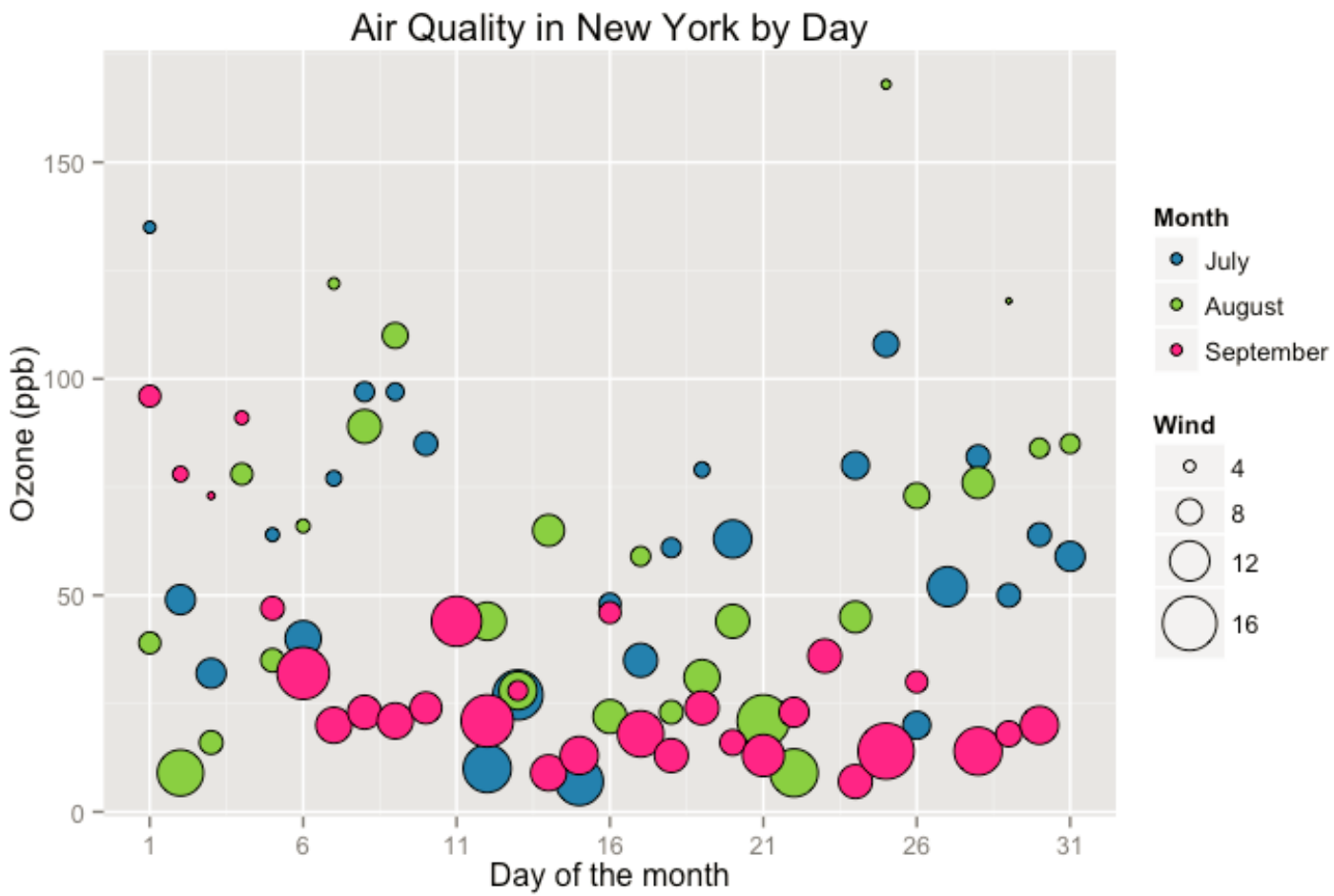
What is R?


Analytics Software with at Least 500 Articles in 2014
Python
KNIME
RapidMiner
R
Minitab
JMP
Stata
MATLAB
Weka
Statgraphics
Statistica
C, C++, or C#
Java
Systat
BMDP
Hadoop
SAS
SPSS Statistics
Percent Change in Scholarly Articles from 2014 to 2015
30
25
20
15
10
5
0
-5
-10
-15
-20
-25
-30

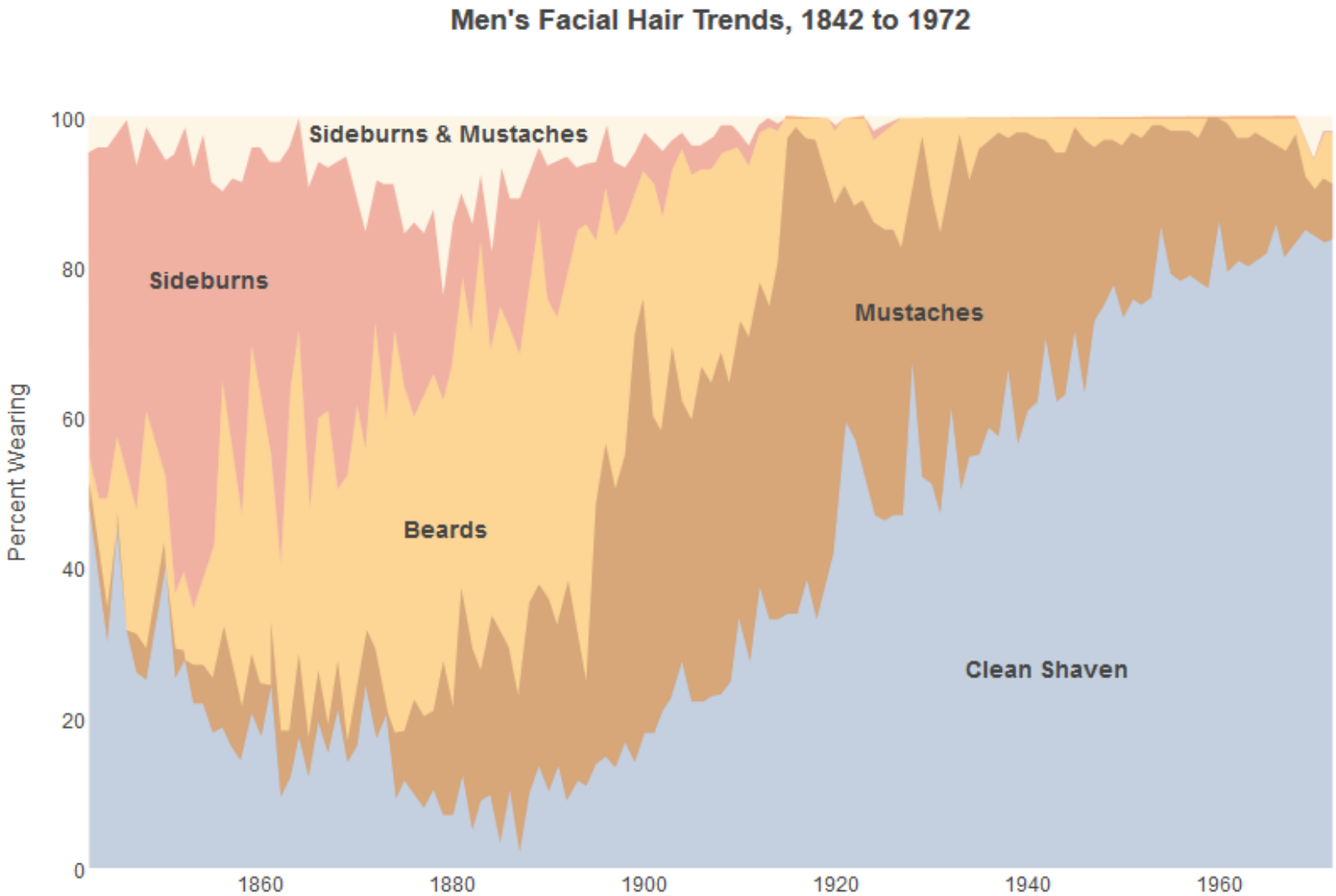
What is R good at?

- **Graphics:** publication-quality 2-d graphics, designed based on visual perception research at Bell Labs and elsewhere.
- **Range of methods:** In addition to many built-in features, over 10.000 add-on packages are available for more specialized analyses.
- **Flexibility:** Data analysis uses the same programming language that R is written in. There is a smooth transition from simple data analyses to customization of analyses to programming.

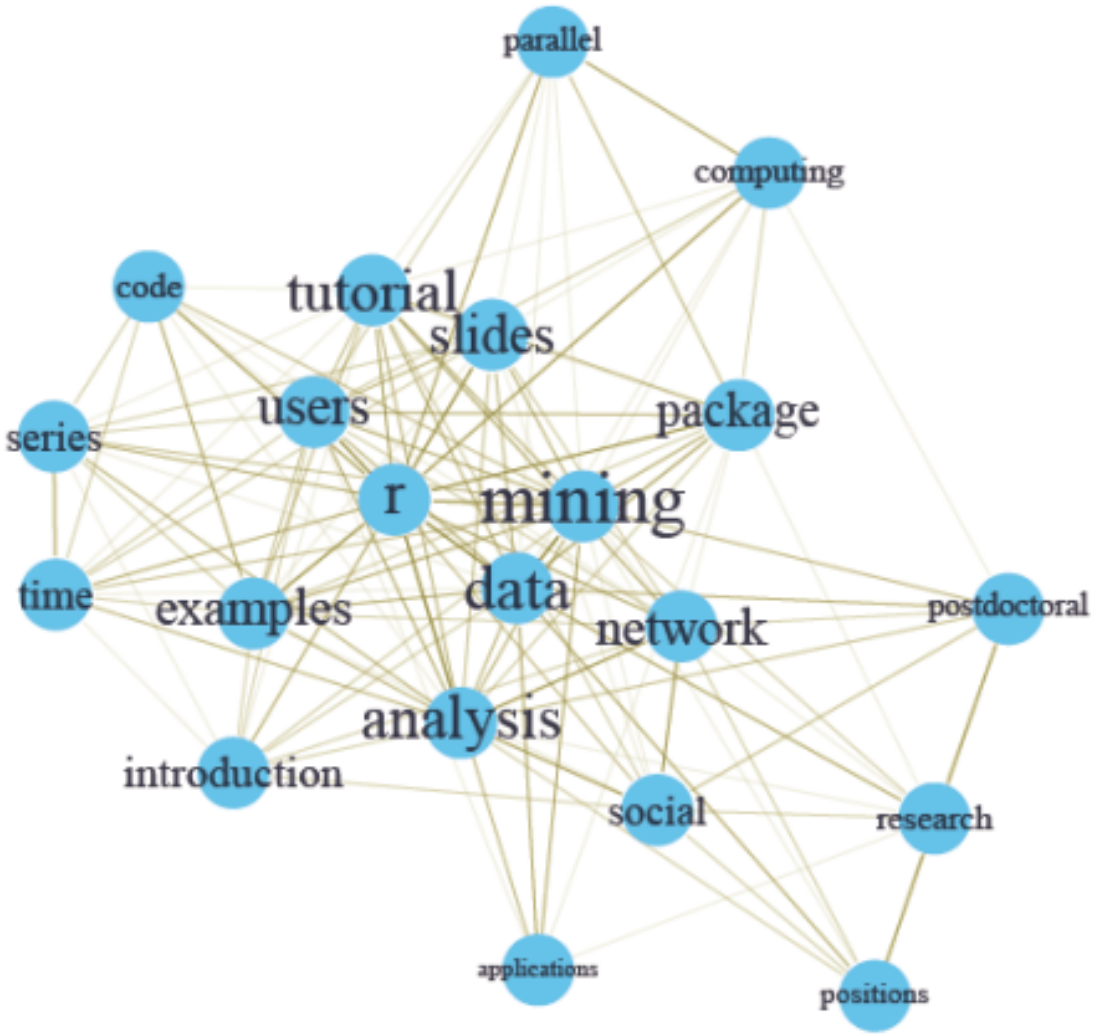
What is R good at? Data visualization



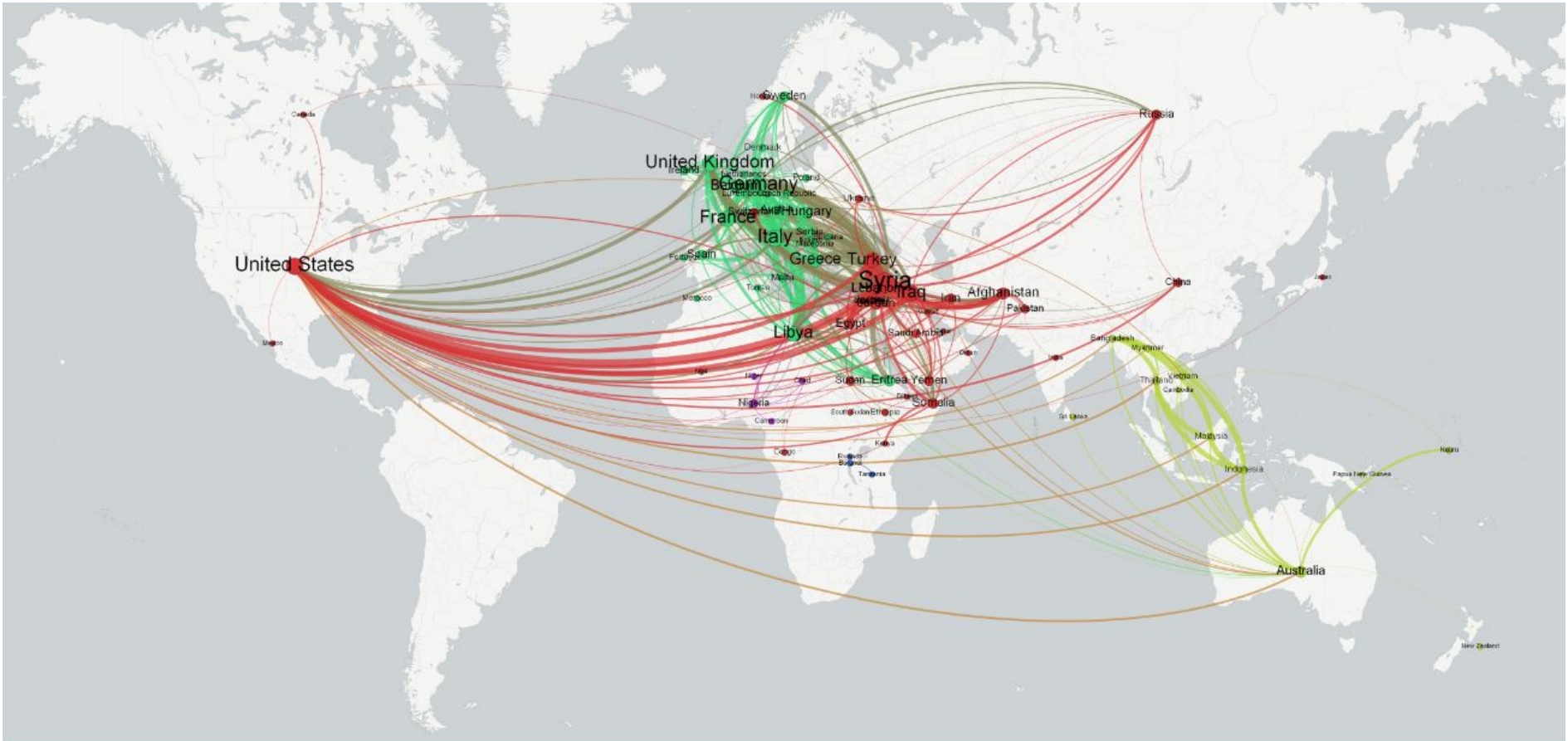
What is R good at? Data visualization



What is R good at? Data mining



What is R good at? Data mining & visualization



Why is R challenging?

- No built-in pointy-clicky analyses, although there are tools to programm them.
- The system is large and parts of it may use terminology from different statistical areas.
- Spaces in commands don't matter (except for readability) but Capitalisation Does matter!!!
- R does not distinguish between commands that do something and commands that compute a value. Everything is a function (returns a value)!

Resources

- Obtaining R: `http://cran.at.r-project.org/`
- Obtaining R-Studio (one recommended editor): `http://www.rstudio.com/`
- a collection of some interesting tutorials
  - `http://www.r-bloggers.com/the-guerilla-guide-to-r/`
  - `http://www.r-bloggers.com/top-3-r-resources-for-beginners/`
- Contributed documentation on CRAN (can be found on the R-project page under Documentation: Manuals)

Recommended books

- Spector, Phil (2008). Data Manipulation with R. Springer.
- Braun, W. John & Murdoch, Duncan J. (2011). A first course in statistical programming with R. Cambridge.
- Albert, Jim (2011). R by Example. Springer.
- Murrell, Paul (2011). R Graphics. Chapmann & Hall.
- Faraway, Julian J. (2005). Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapmann & Hall.
- Faraway, Julian J. (2004). Linear Models with R. Chapmann & Hall.