Introduction to R

VP Nagraj

Why R?

- R is free
- R is flexible
- R is everywhere
- R is awesome

R is free

Software	Cost
STATA	$1,140 - $4,370 + maintenance
SAS	$8,700 - $140,000 / year
SPSS	$2,390 - $40,600 / year
Matlab	$2,150 + $1,000s for modules
R	$0

estimates ca 2014[1]

R is free

free as in speech, not free as in beer

- Available under GNU General Public License[2]
- Can be modifed, distributed and used for commercial purposes

R is flexible

R is made to be modified

- > 10000 add-on packages currently available on the Comprehensive R Archive Network (CRAN)[3]
- > 1000 add-on packages for genomic analyses currently available on Bioconductor[4]
- New features are being added (almost) daily

R is flexible

R is a programming language

- Dialect of the S language
- Source code written in C and Fortran
- https://cran.r-project.org/doc/manuals/r-release/R-lang.html
- **TAKEAWAY** R is not "point and click" software, R is a programming language

R is flexible

> Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying and information privacy. The term often refers simply to the use of predictive analytics or certain other advanced methods to extract value from data, and seldom to a particular size of data set. (Wikipedia)
>
> When computing the answer takes longer than the cognitive process of designing the model … (Hadley Wickham)

R is flexible

R code can be optimized for handling "big data"

- *sparklyr* (R interface for connecting to Spark instance)
- *bigmemory* (store large objects in memory and files with external pointer, enabling transparent access from R to large objects)
- *ff* (access datasets too large to fit into memory)
- *foreach* (iterate over a collection without loop counter)
- *multicore* (run parallel computation on computers with multiple cores without

R is everywhere

the R community is huge

- "becoming ... lingua franca" for data analysts[5]
- useRs in just about any field you can imagine
- http://www.r-bloggers.com/ features > 500 contributors from a variety of research disciplines

R is everywhere

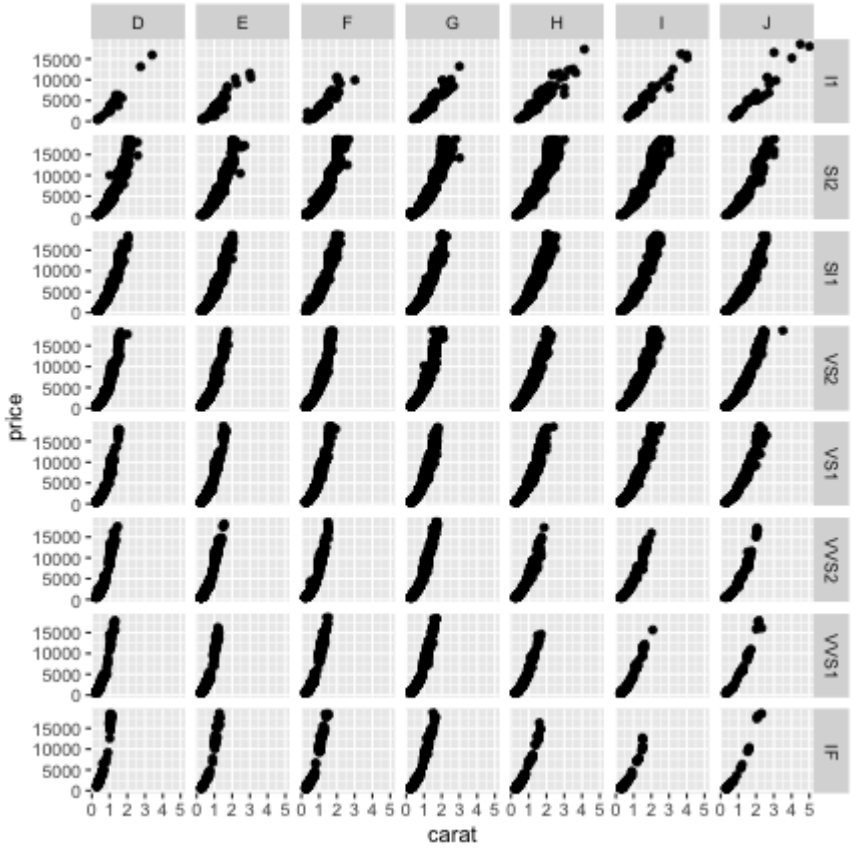
R at UVa

- StatLab [http://data.library.virginia.edu/statlab/]
- Data @ HSL [https://guides.hsl.virginia.edu/data/services]
- UVa R Users Group [http://www.meetup.com/UVa-R-Users-Group/]

R is awesome

R is capable of high-dimensional visualizations

- *ggplot2* and the "Grammar of Graphics"[6]



References

1. https://speakerdeck.com/stephenturner/introduction-to-r-for-life-scientists
2. https://www.tldrlegal.com/l/gpl2
3. https://cran.r-project.org/
4. https://www.bioconductor.org/
5. http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html
6. http://byrneslab.net/classes/biol607/readings/wickham_layered-grammar.pdf