

Software for Analytics

Is it time for R?

Bob Stine

Department of Statistics
The Wharton School
University of Pennsylvania

What do you want?

- What features do you want in software for business analytics?
 - Toolset, functionality
 - Easy to do the “right things”, has functionality you need
 - Mix of graphics, models, and data manipulation
 - Active community, current
 - Ease of instruction Excel
 - Learning curve should not be too steep: class time teaching software?
 - What skills with software do students bring to the course?
 - Incorporated into textbook and other course material
 - Reports Excel
 - Workflow for “reproducible research”
 - Importance of communication in analytics
 - Linkage to other courses
 - Visibility in other course

What do students want?

- What features do your students want in software for business analytics?

- Ease of use

Many don't come to stat course with programming skills for a reason

Excel

- Support

An active community with prominent on-line help
Available help when needed from course TAs

Excel

- Cost

“Cheap is good. Free is better”

- Enhance resume

Software training perceived as valuable skill
Adds monetary value to resume

Excel

Other Relevant Issues

- Other issues influence software choice
- Level of the course
 - One size fits all?
 - Different software for required intro level course, but others for
E.g. Excel for basic course, SAS for more advanced
- Colleagues' preferences
 - You probably have to share the decision of what software to use with others
 - Expectations, needs of subsequent courses

Does R make the cut?

Feature	R
Toolset	All statistics, data manipulation, presentation graphics
Teachable	Depends on what they know, what you expect
Reports	RStudio notebooks change playing field
Linkages	Depends on other analytics courses
Ease of use	"I wanted to learn statistics, not programming"
Support	Very active user community
Cost	Free
Resume	Students see it as value added

R Studio Notebook

- What is it?
 - Document your analysis
 - Mix text, R commands, output
 - Markdown format
 - Can embed Latex if desired, with “instant” rendering
 - Saved as text file
 - Rendered in browser (html)
- Similar to
 - Jupyter notebook for Python
 - Mathematica notebook front end
- Free part of R Studio environment

Notebook Illustration

- Two examples
 - Simple regression
 - Lattice graphics
- Illustrate
 - HTML rendering
 - Analysis
 - Output, cmds, graphs
 - Story
 - Comments, directions
 - Ease of update
 - One button recompute

DSI Example

Bob Stine
November 19, 2017

This example fits simple regression models to small data sets. The second shows lattice/trellis plots.

```
library(ggplot2)
library(car)
library(lattice)
```

Example 1: Locating a Franchise Outlet

Read the data into a data frame. The data frame has 80 observations of two variables.

```
Franchise <- read.csv("Data/21_4m_franchise.csv")
```

Each row of the data frame describes the amount of gasoline sales (in thousands of gallons) and the traffic volume. For example, sales in the first week were 6640 and 7760 in the second week. Traffic volume was 3.3610^4 in the

Franchise	Sales
	<dbl>
	6.64
	7.76
	9.56

Greater Integration

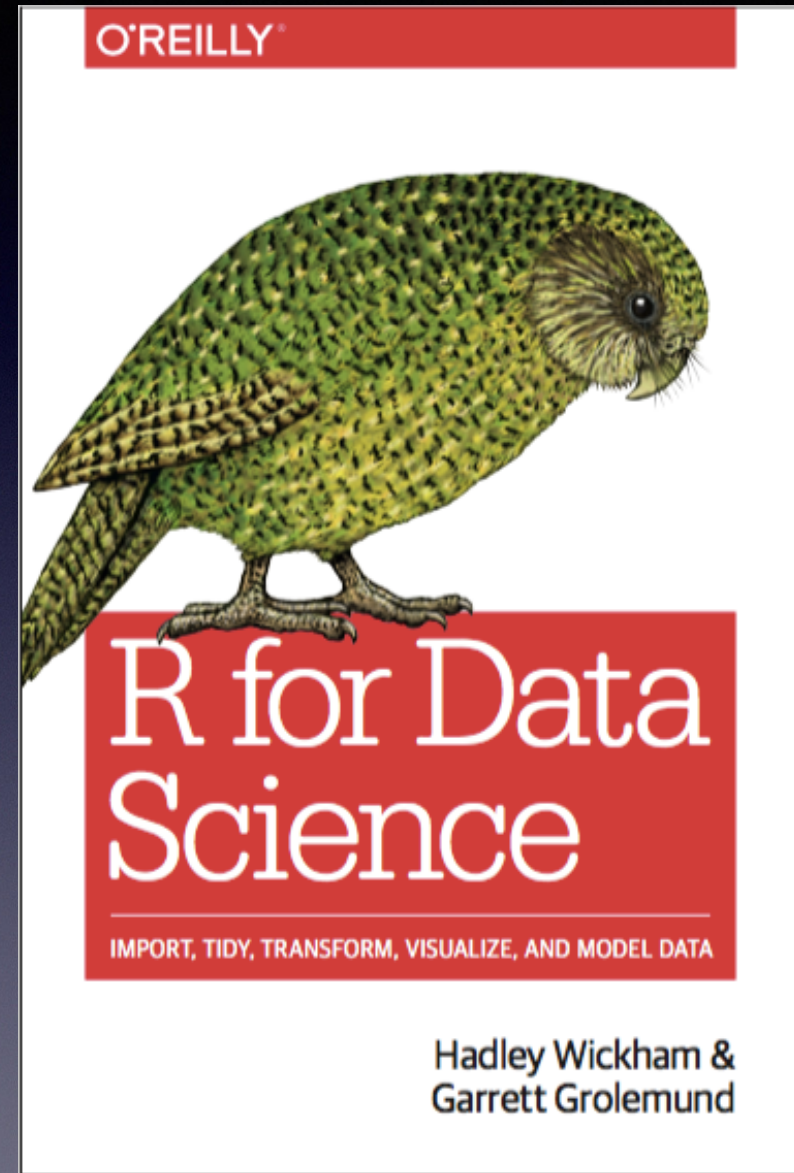
- knitr
 - Combine R with Latex
 - Reproducible research documents
 - S-weave
- bookdown
 - Build a book from Rmd notebooks
 - Constructs web site of linked documents
- Example
 - Textbook supplement... “R-Companion”

Other Packages

- R is the basic language
- Packages are the supplements available for specific tasks
- Shiny from R-Studio
 - Interactive graphics, web pages
- R-commander by John Fox
 - Point and click interface to R
- “Tidyverse” of related packages

An R Shout Out

- Interested in how to use R for analytics?
- Check out this book...
 - different style of R
 - ggplot
 - rplyr
 - and many others



Conclusion

- What did you decide to do?
- Currently...
 - JMP used in basic required courses:
 - “Point-and-click”
 - Plot linking-brushing
 - Emphasis on graphics
 - R used in more advanced courses, grad courses
 - Offer half-semester courses on R for undergrads and MBA students
 - Very popular course now
 - Python a strong competitor
 - Popular among CS students, “machine learning” community