

Getting Started with R

Actuaries Club of Hartford and Springfield

May 14, 2019

Joy Chen, FSA, CERA

Matt Heaphy, FSA, MAAA



Agenda

- 1** What is R?
- 2** Why should we learn R?
- 3** Sample use cases and live demonstration
- 4** Conclusion

R Introduction

R is a free, open source programming language built for statistics, graphics, and data analysis

We have Excel...

We have Access...

Why should we use R?

- R is powerful
- R is flexible
- R is reproducible
- R is easy to learn and easy to read
- R is fun!

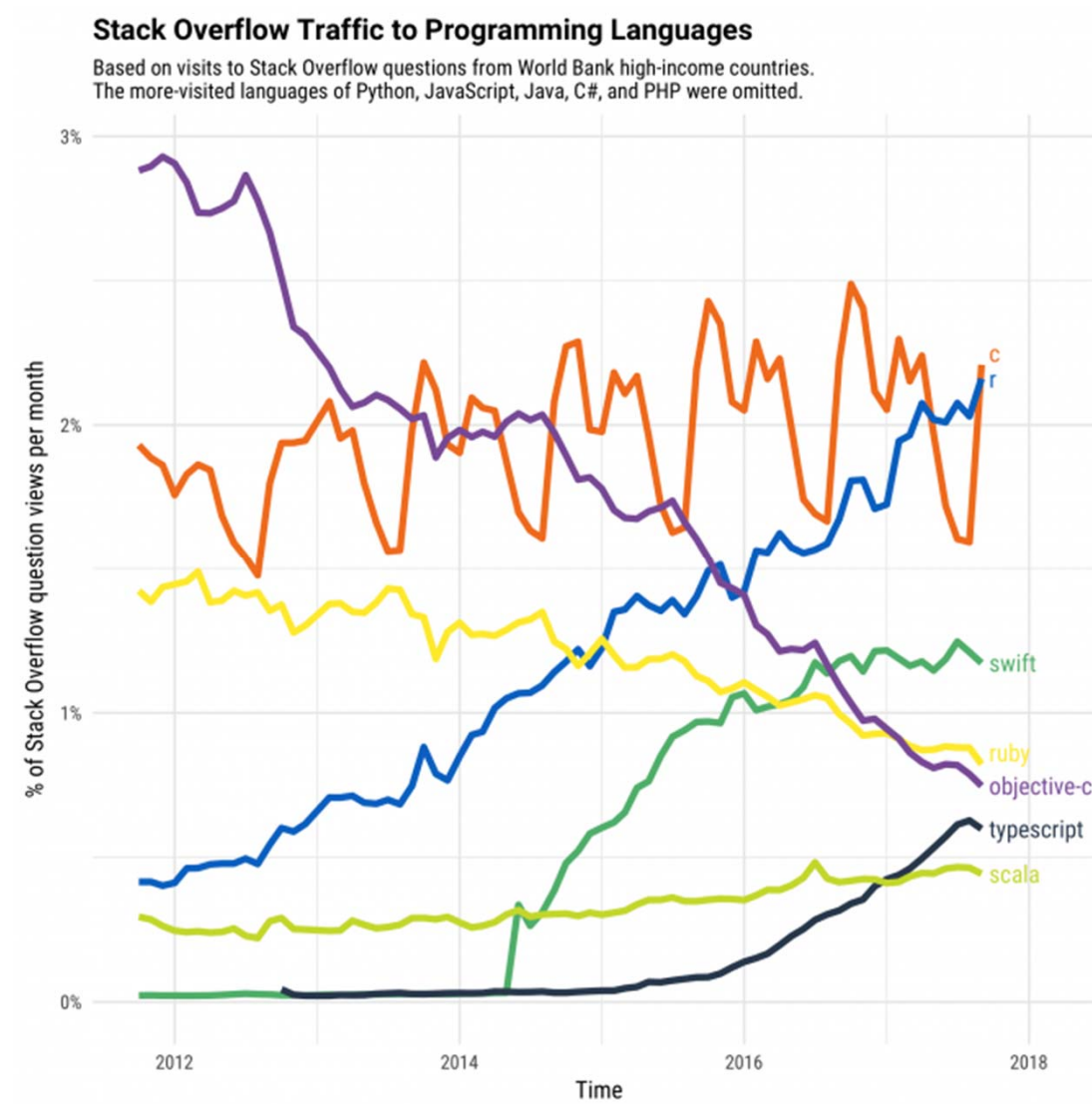


Why the name 'R'?

R follows S - R was built from the S programming language. S was developed by John Chambers at Bell Labs in 1976.

*The authors of the first R release in 1995 are **R**oss Ihaka and **R**obert Gentleman*

The Impressive Growth of R



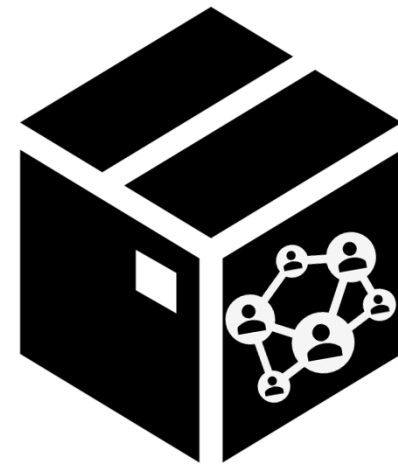
R and Friends



The R language



Integrated Development Environment



Over 10,000 packages
developed by users

Third Party Packages

- R's greatest strength is its extensibility
- Users of R can create and contribute packages to add new functionality for data manipulation, plotting, statistics, modeling, etc.
- Tidyverse – a family of packages that focuses on easy-to-read, yet powerful code

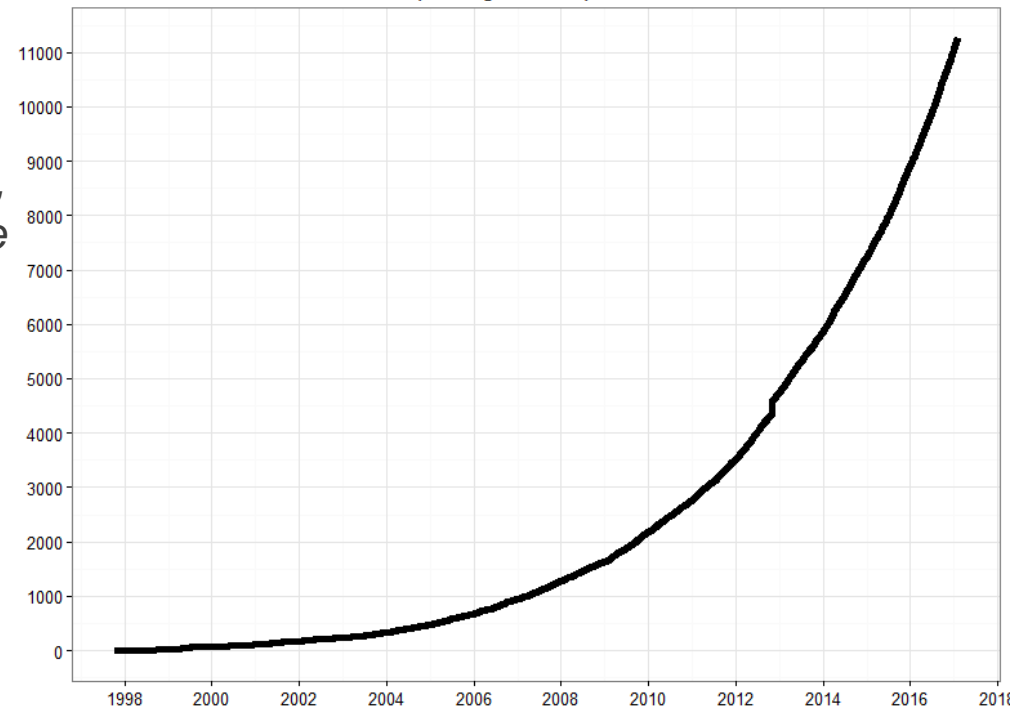


So Many Packages!

Whatever you want to do, chances are someone already built a package for that

data manipulation, data visualization (plots, maps, images, interactivity), statistical models, predictive modeling, machine learning, deep learning, financial analysis, programming (iteration, C++ integration, parallelization), reporting (HTML, pdf, Word, PowerPoint), text mining, natural language processing, forecasting, network analysis, importing data (databases, API's, websites, text files, Excel, pdf, archives, HDF5, JSON, etc...)

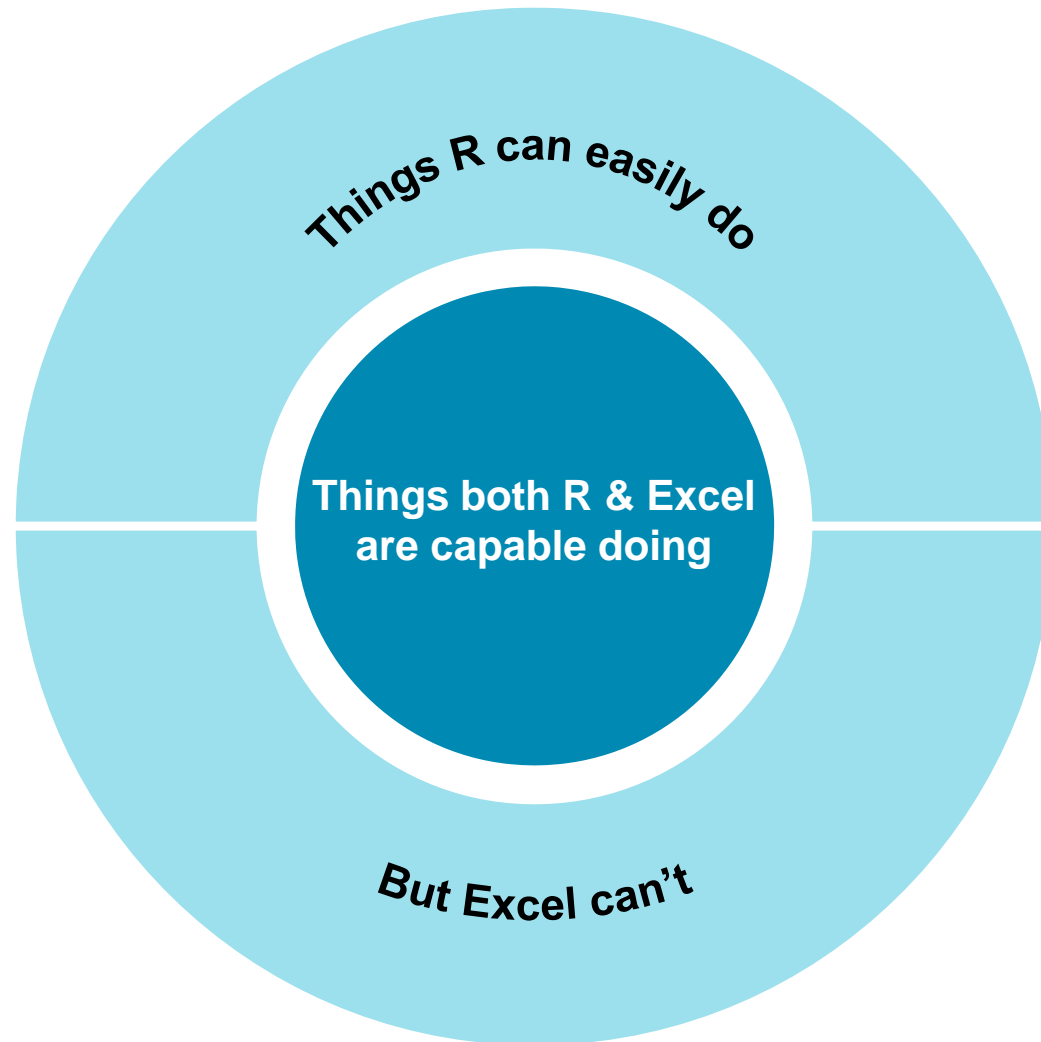
Number of R packages ever published on CRAN



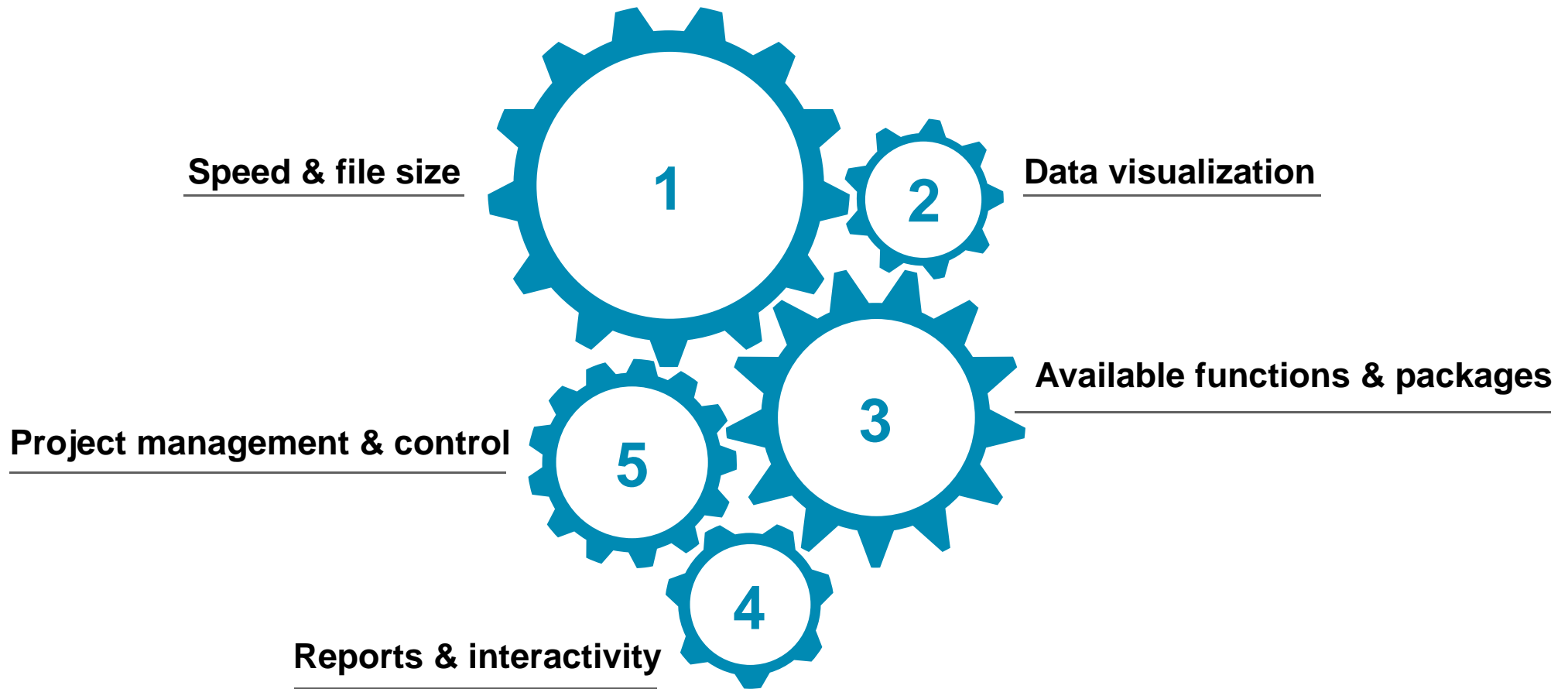
Why should we learn **R**?



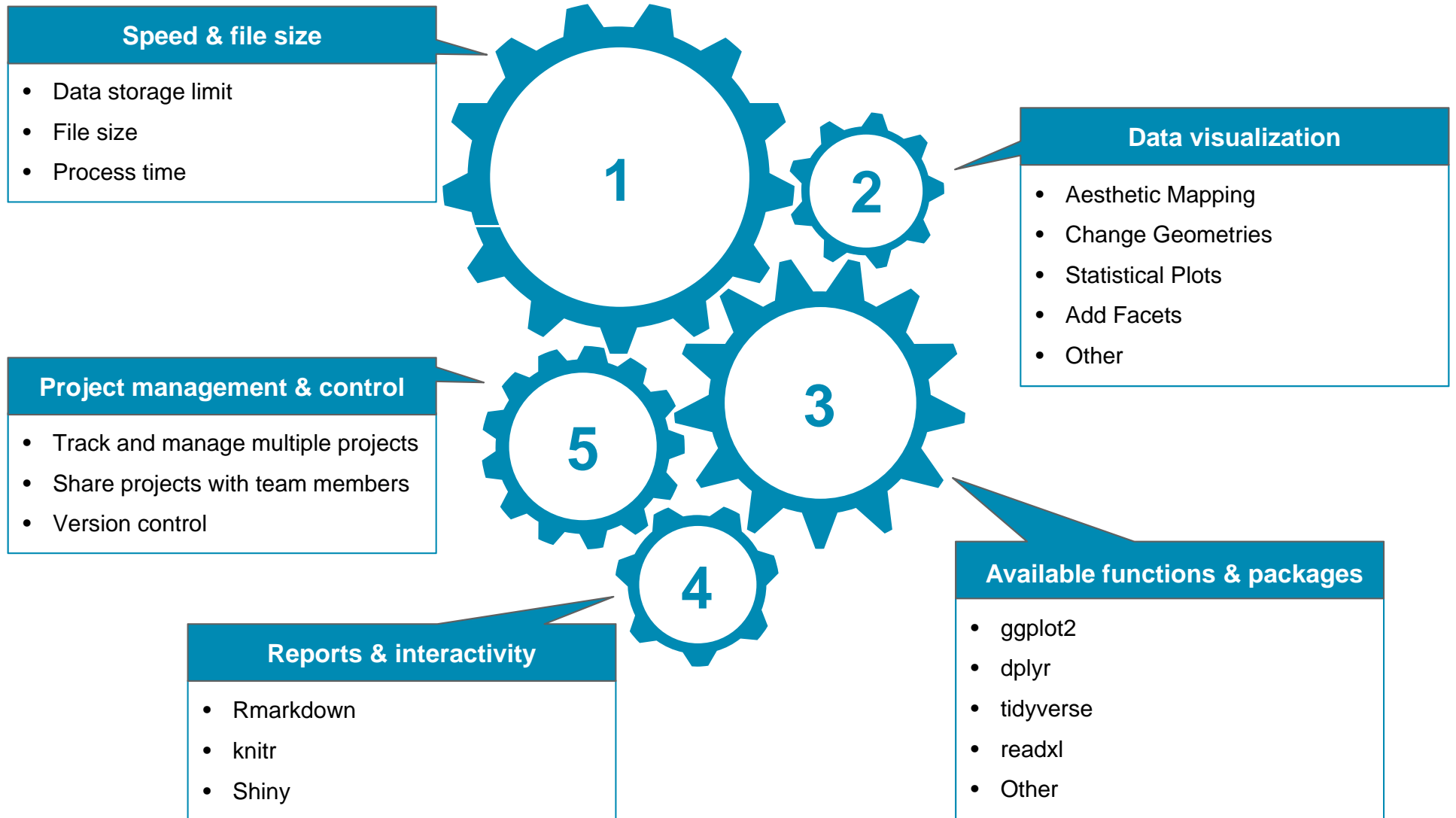
Why should we learn R?



Advantages of R over Excel

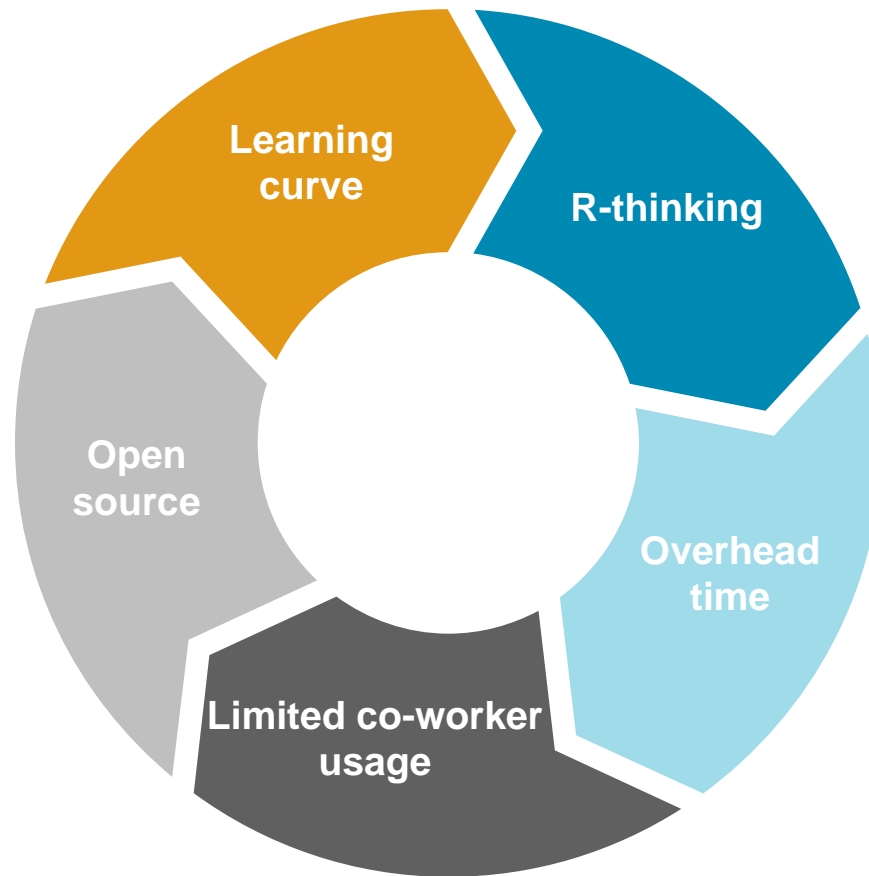


Advantages of R over Excel



Things to consider when transition to R

The transition effort should not be underestimated



Start with things you often found difficult doing in Excel, and engage your colleagues

Sample use cases and live demonstration



Data Visualization and Reporting Case Study

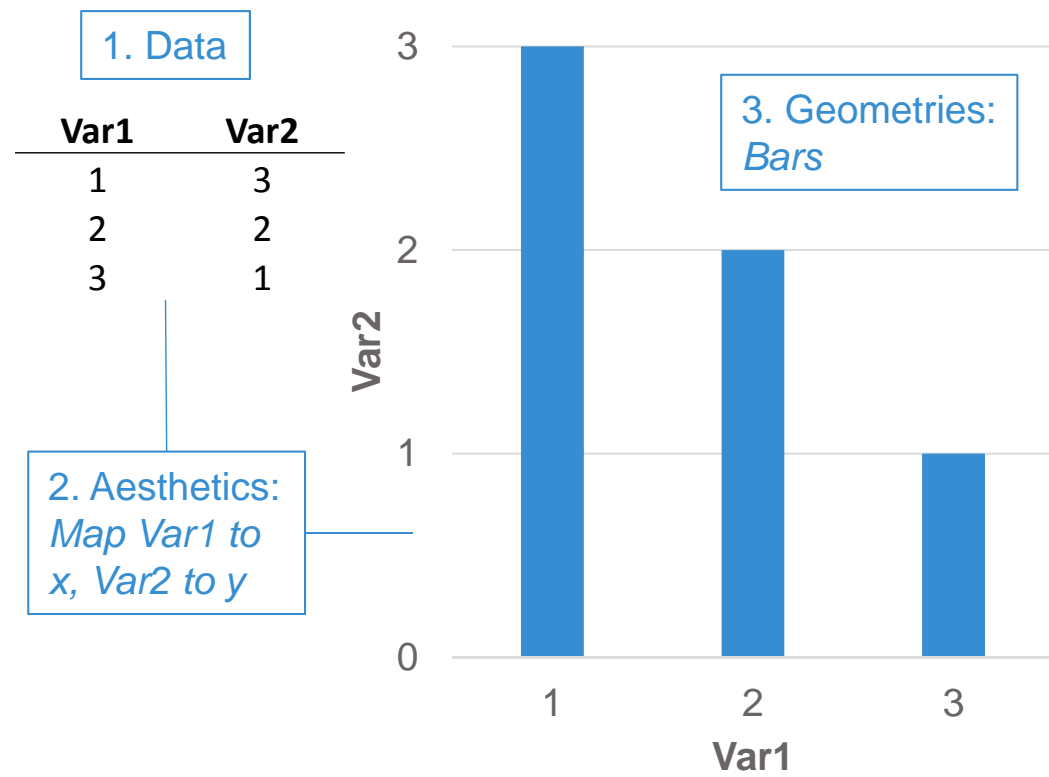
The SOA's 2009-2015 Individual Life Insurance Mortality Experience Report was released in Q4 2018. Your manager has asked you to identify any interesting trends in mortality rates and A:E* ratios by age, gender, underwriting class, etc. that should be considered in future assumption development. You've been asked to document your findings in a formalized report.

- The following examples will showcase:
 - The ggplot2 package for data visualization, and how it differs from creating plots in Excel
 - How R Markdown and Shiny can be used to create professional reports
- The experience report includes a 10+ GB text file of results with 30+ million records. For illustrative purposes, some data preprocessing was applied. We're going to focus on term insurance only and apply some grouping to scale the dataset down to 2.3 million records

Data Visualization with ggplot2

- ggplot2 is one of the most popular plotting packages in R
- It is based on the [grammar of graphics*](#), a consistent approach to describing any plot

1. Data
2. Aesthetics
3. Geometries
4. Facets
5. Statistics
6. Coordinates
7. Themes



Reporting with R Markdown

- R Markdown can be used to combine technical analysis performed in R with written commentary to create professional reports
- Highly transparent, reproducible, and auditable end-to-end analysis
- Output to HTML, pdf, Word, or PowerPoint
- Easy to learn – no requirement to learn or write HTML



Markdown Document Header

```
1 ---
2 title: "2009-2015 Individual Life
3 Insurance Mortality Experience Report"
4 output:
5   word_document:
6   powerpoint_presentation:
7   html_document:
8     toc: yes
9 editor_options:
10  chunk_output_type: inline
```

All R Markdown documents start with a YAML* header defining the title, output options, and other document features (date, author, table of contents...)

2009-2015 Individual Life Insurance Mortality Experience Report

- The Data Set
- Function to calculate mortality rates and AE
- Basic Plotting

“Knitting” a document will generate a report

* Depending on who you ask, this stands for YAML Ain't Markup Language or Yet Another Markup Language

Markdown and Inline R Code

- Most Markdown is written in plain English
- Formatting can be added using simple code*
 - Headings, font, bullets, hyperlinks, images, etc.
- R code can be executed in the middle of a line to using ``r [code]``

`*italics*` → *italics*
`**bold**` → **bold**
`#` → Header Level 1
`##` → Header Level 2

R Markdown

```

60 ## The Data Set
61
62 The [2009-2015 Individual Life
Insurance Mortality Experience
Report](https://www.soa.org/resources/r
esearch-reports/2019/2009-2015-individu
al-life-mortality/) was published by
the SOA in Q4 2018. The SOA provided a
detailed text file of experience
results for expanded analysis. For this
tutorial, we have filtered the data to
only focus on term insurance. We
applied additional grouping to reduce
the size of the file.
63
64 There are `r round(nrow(dat) / 1000000,
1)` million records, `r
round(sum(dat$number_of_deaths) /
1000)` thousand deaths, and `r
round(sum(dat$policies_exposed) /
1000000, 1)` million exposure years.
  
```

HTML Document Output

The Data Set

The 2009-2015 Individual Life Insurance Mortality Experience Report was published by the SOA in Q4 2018. The SOA provided a detailed text file of experience results for expanded analysis. For this tutorial, we have filtered the data to only focus on term insurance. We applied additional grouping to reduce the size of the file.

There are 2.3 million records, 209 thousand deaths, and 121.1 million exposure years.

Markdown and R Code “Chunks”

- Chunks of R code can be added to an R Markdown file to produce graphs, tables, and other visuals
- Chunks start with ```{r}` and end with ````. Several options are available such as showing or hiding the code in the output or specifying plot size

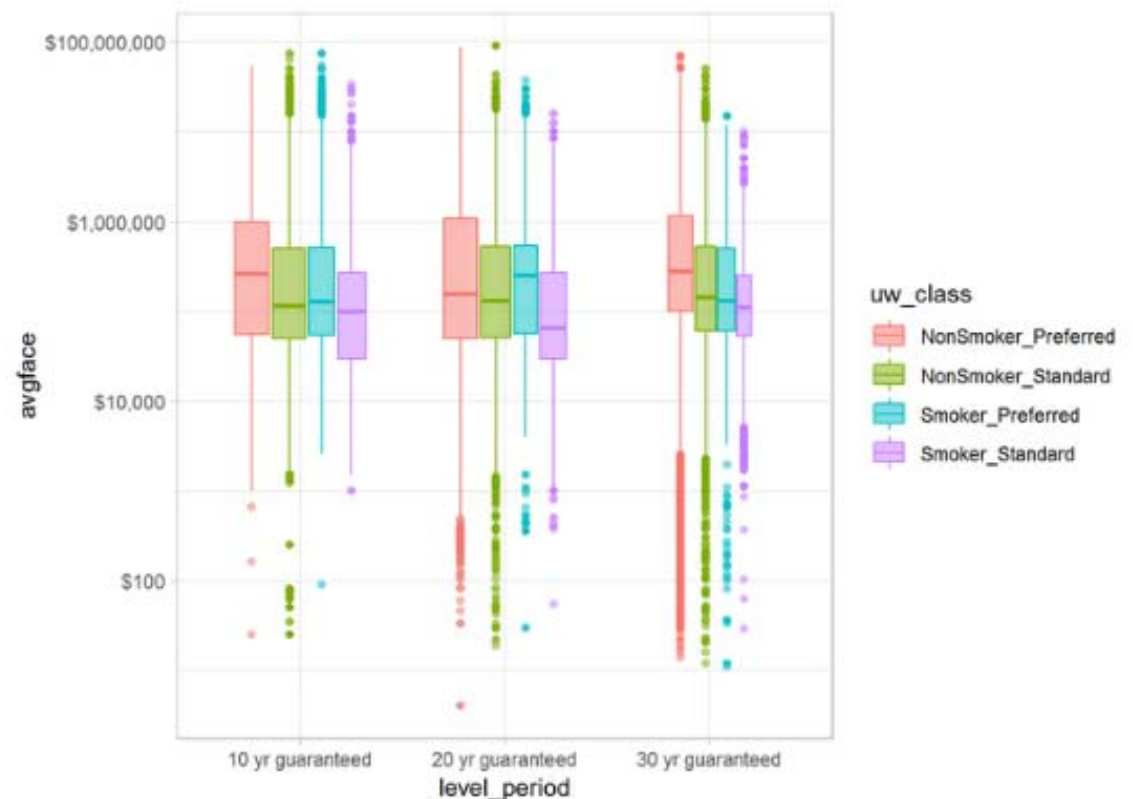
R Markdown

HTML Document Output

```

128 ``{r}
129 dat %>%
130   filter(level_period %in% c("10 yr
131     guaranteed", "20 yr guaranteed",
132     "30 yr
133     guaranteed")) %>%
134   mutate(avgface=amount_exposed /
135     policies_exposed) %>%
136   ggplot(aes(level_period, avgface,
137     color=uw_class, fill=uw_class)) +
138   geom_boxplot(alpha=0.5,
139     varwidth=TRUE) +
140   scale_y_log10(labels=scales::dollar)
141 ``

```



Shiny Apps – Interactive Data Visualization

- Shiny can supercharge reports with interactive data visualizations
- Great for creating dashboards – especially if an app is connected to a live data source
- Shiny can be added to R Markdown documents or used to create standalone web apps
- Large library of fantastic open source HTML widgets
- No need to learn HTML
- Downside: Shiny requires R to be running in the background either on a desktop or on a server



Shiny App Example

2009-2015 Individual Life Mortality Experience



Line Plot

Disclosures

Age Range:



Plot:

- Mortality Rate
- A:E 2015 VBT

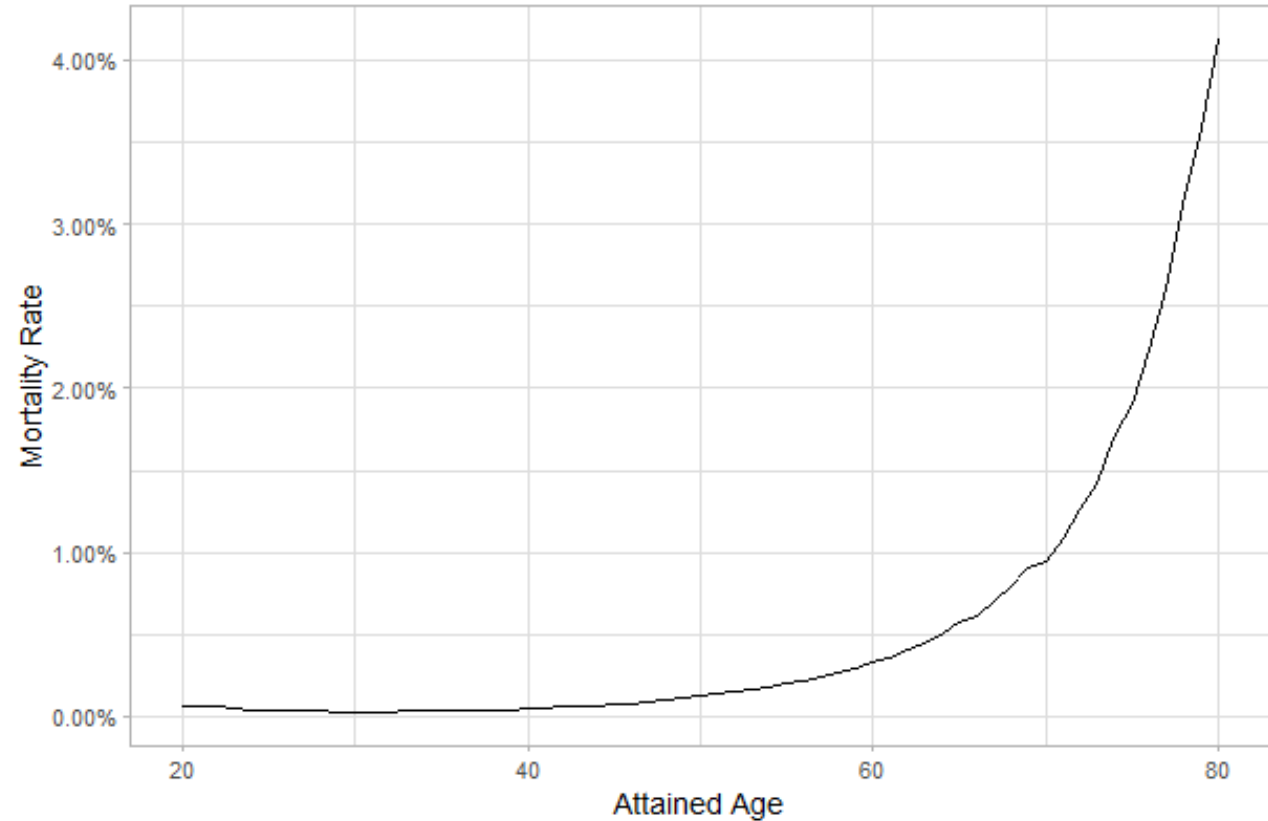
Color by:

None

Facet by:

None

Line Graph



Learning Resources



Predictive Analytics and Futurism Section

Available resources to PAF section members:

- Annual PAF symposium and seminars
- Newsletters
- Podcast on leading edge topics (machine learning, block chain, etc.)
- Data science competition
- LinkedIn discussion group

New ASA requirements effective June 2018

- Statistics for Risk Modeling (SRM) exam
 - 3.5 hour computer-based multiple-choice questions
- Predictive Analytics e-learning modules
- Predictive Analytics exam
 - 5-hour and 15-minute
 - Require analysis of a dataset in the context of a business problem and submission of a report

Predictive Analytics and Futurism Section: <https://www.soa.org/sections/pred-analytics-futurism/pred-analytics-futurism-landing/>
Predictive Analytics Certificate and Exam: <https://www.soa.org/predictive-analytics-certificate/default/>
Actuarial Toolkit's R Console: <https://actuarialtoolkit.soa.org/tool/r-console>

Learning Resources

- CRAN Task Views: <https://cran.r-project.org/web/views/>
- R-bloggers: <https://www.r-bloggers.com/>
- Online Courses from Coursera, Udemy, and others
- Recommended Textbooks
 - R for Data Science (free online): <https://r4ds.had.co.nz/>
 - R for Everyone: <https://www.jaredlander.com/r-for-everyone/>
- R Conferences and Meetups
- Package documentation: <https://www.tidyverse.org/>, <https://rmarkdown.rstudio.com/>, <https://www.rstudio.com/products/shiny/>
- Social Media: #rstats

Thank You!



Questions?

ACHS