



Society of Actuaries in Ireland

R For Actuaries: What, Why and Where?

26th October 2017

Disclaimer

**The views expressed in this presentation are
those of the presenters and not necessarily
of the Society of Actuaries in Ireland**



What is this and who are these people?

- What is this?
 - Recent Qualifiers suggestion.
 - Wider Fields organised a working group.
 - The first of many talks!
 - Visualisations, data handling, GLMs, Chain Ladder...
- Who are these people?
 - Kieran Walsh is a member of the R working group.
 - Pedro Écija Serrano is a member of the Wider Fields Committee and leads the R working group.



Title

- What is R?
- Why bother?
- Where can I find more?



What is R?

- R is a free, portable, open source language for statistical programming.
 - Free: no fees for the base product and the vast majority of packages.
 - Portable: it runs on any OS (Windows, Linux, Mac OS...)
 - Open Source: anyone can contribute to its development.
 - Programming Language: set of instructions to implement algorithms.
 - Statistical: designed for statistical computing (linear and non-linear modelling, time series analysis, classification, clustering, etc.)



What is R?

R on its own can be ugly.

```
Terminal - gpcarpen@vm64-173: ~ (on vm64-173.iplantcollaborative.org)
File Edit View Terminal Go Help
gpcarpen@vm64-173:~$ R

R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-unknown-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

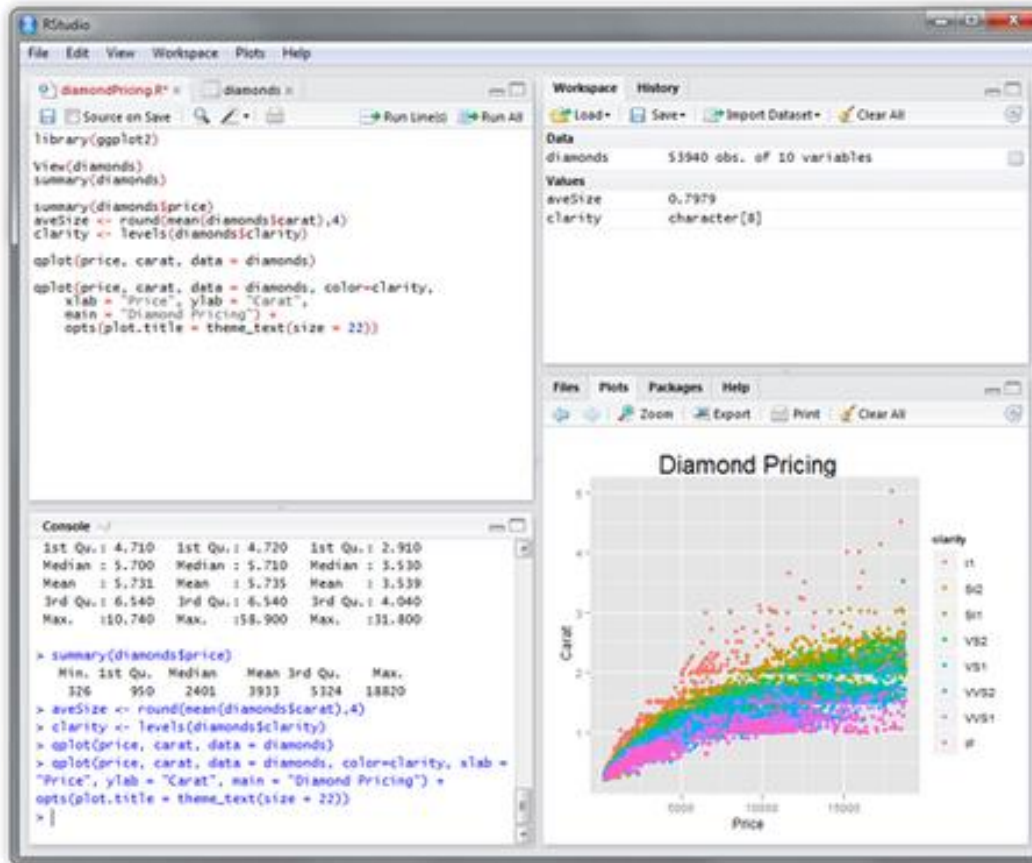
> install.packages("/home/gpcarpen/Desktop/randomForest_4.6-6.tar.gz", type="source")
Warning in install.packages("/home/gpcarpen/Desktop/randomForest_4.6-6.tar.gz",
:
  'lib = "/usr/local/lib/R/library"' is not writable
Would you like to use a personal library instead? (y/n) y
Would you like to create a personal library
~/R/x86_64-unknown-linux-gnu-library/3.0
to install packages into? (y/n) y
```

<https://pods.iplantcollaborative.org/wiki/pages/viewpage.action?pageId=11445211>



What is R?

But it is easy to make it look better.



<https://www.r-bloggers.com/rstudio-new-open-source-ide-for-r/>



Why BotheR? – Summary

- Fit for Purpose
- Project Organisation
- Community
- Data Management
- Reproducibility and Expandability
- Graphics
- Missing Values and Cleaning
- Packages Relevant to Actuaries



Fit for Purpose

Excel is designed for

Limited data storage in tabular form

Basic arithmetic and mathematical analysis

Basic analysis and visualisations

R is designed for

Data manipulation of data files in different formats

Advanced statistical modelling

Comprehensive visualisation of results

Actuaries have to

Manipulate relatively large datasets


Apply statistical models

Communicate effectively - visualisations are useful



Project Organisation – Excel

- If you've ever come across the following:

 Pricing report v3.5 - final (2) - TM updates - correct 20170812 - to be pdf v7 13/09/2017 11:38 Microsoft Excel W... 1,773 KB

- You'll know it's sometimes hard to maintain consistent file naming conventions in Excel

Organisational Tools Lacking in Excel

- Version control
- Debugging software
- Compact data objects





Project Organisation – R



Version
control tool
with R
integration



IDE
debugging
software



Standardised
project
development



IDE data
visibility
from high to
low level





Community

Learning

Vast range
of free,
good
quality
books and
courses

Packages

Huge
quantity of
community
packages

Support

Widespread
online
support

- If you're ever stuck, someone has already solved the problem online (much like Excel/VBA)



Data Management – Excel Issues

- Virtually anything beyond simple operations on a small dataset
- Let's take an example...

Inconsistent Dates

- Multiple text to columns
- Multiple “IF(LEFT...)” etc.

Join Tables

- Vlookup for each column you want to add it

Run Pareto Distribution

- Available add-in
- Define manually

Varying Assumptions

- Link to new parameters
- VBA loop

- In summary, a messy nightmare!



Data Management – R Solutions

- Handle vast complications on vast data in multiple dimensions

Inconsistent Dates

- Single line regex
- Lubridate package

Join Tables

- SQL join functionality

Run Pareto Distribution

- Built-in functions

Varying Assumptions

- Vector operations
- Grouped in single object

- Ahh... that's much better



Reproducibility and Expandability

- Take initial example of applying Pareto distribution to messy data
- What if we were to receive another set of files with 10x as many rows?

Excel

Copy new data in

Ensure formulae look over correct range

Extend each assumption run

Ad-hoc documenting in separate location

Occasionally difficult to peer review

R

Replace input files

Documentation and code in single location (RMarkdown)

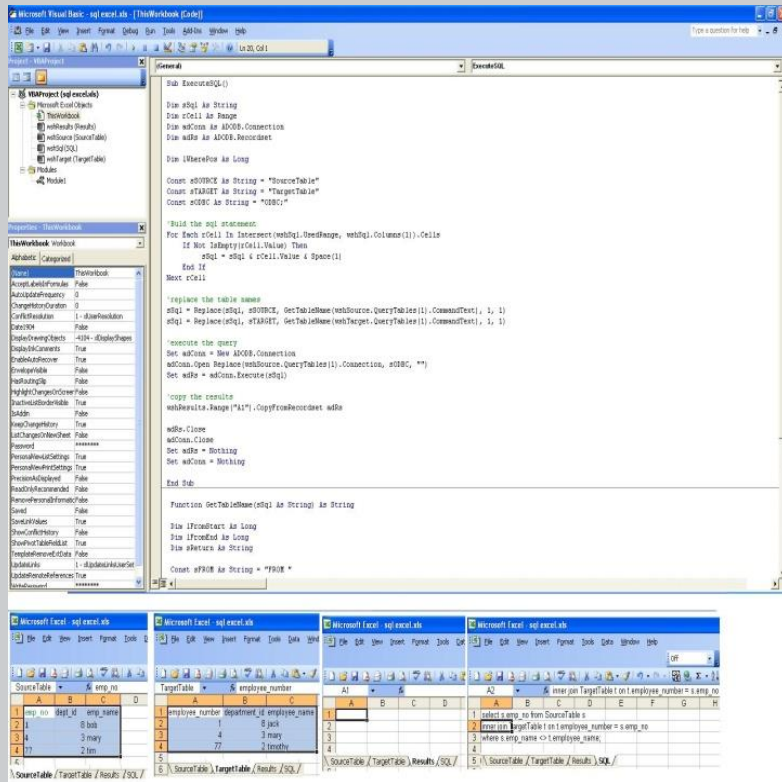
It has been used for statistical analysis under regulatory peer review for many years



Reproducibility and Expandability

- Documentation and peer review: Excel vs. R Markdown

Excel

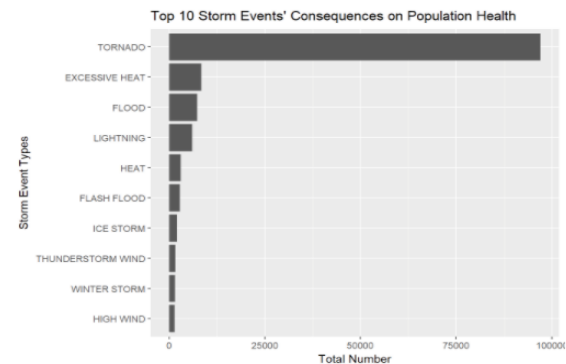


R

3. Results

3.1 Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

```
ggplot(head(storm_health, 10), aes(reorder(EVTYPE, TOTAL_HARM), TOTAL_HARM)) +  
  geom_col() +  
  coord_flip() +  
  labs(title = "Top 10 Storm Events' Consequences on Population Health",  
       y = "Total Number",  
       x = "Storm Event Types")
```



The graph's y-axis contains the 48 different storm events types while its x-axis displays the total count for each type.

According to the graph, tornado, excessive heat, flood and lightning are most harmful with respect to population health, with tornado having the most harmful effects.

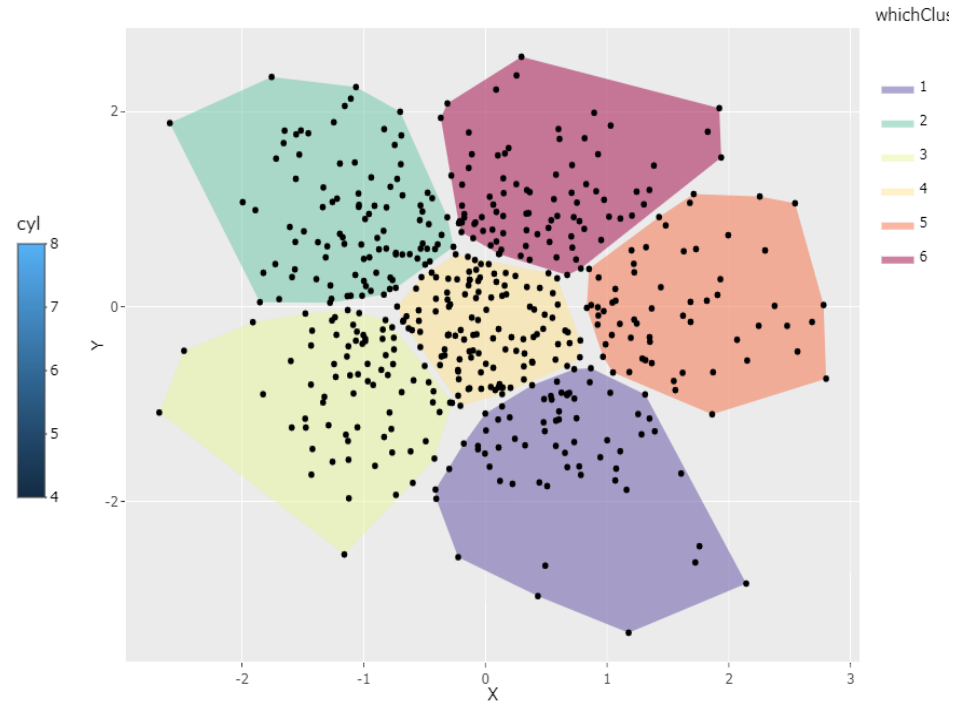
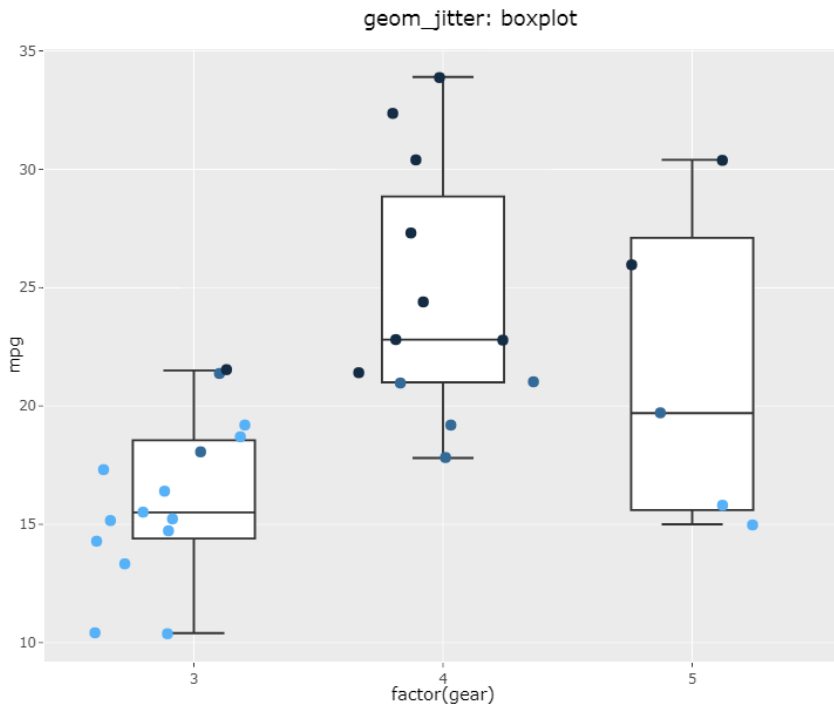
3.2 Across the United States, which types of events have the greatest economic consequences?

```
ggplot(head(storm_econ, 10), aes(reorder(EVTYPE, TOTAL), TOTAL)) +  
  geom_col() +  
  coord_flip() +  
  labs(title = "Top 10 Economic Damage Caused by Storm Events",  
       y = "Damage Amount($)",  
       x = "Storm Event Types")
```



Graphics

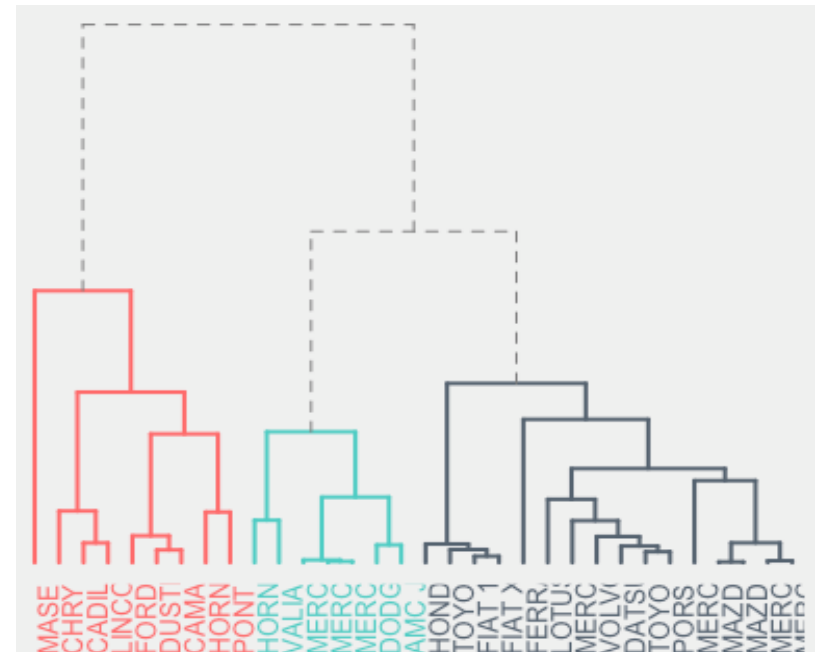
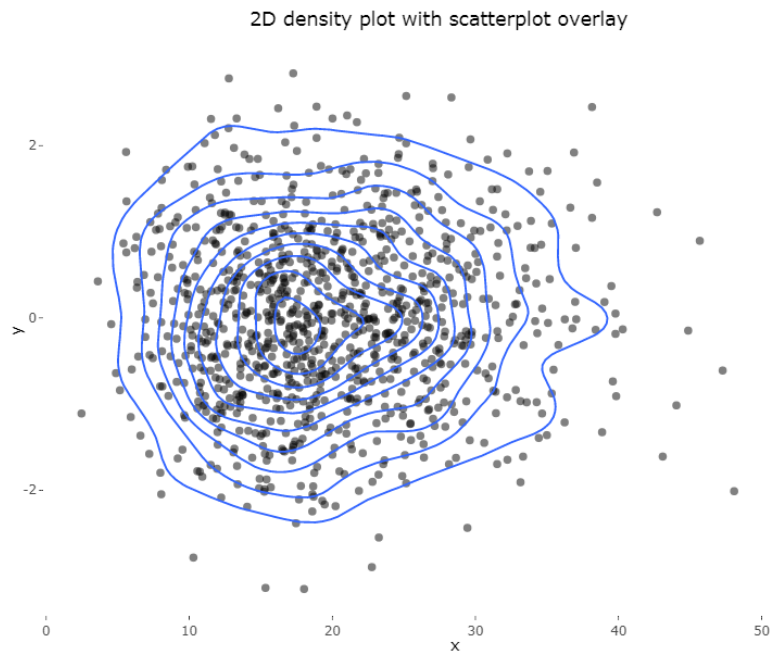
- Let the graphs speak for themselves...





Graphics

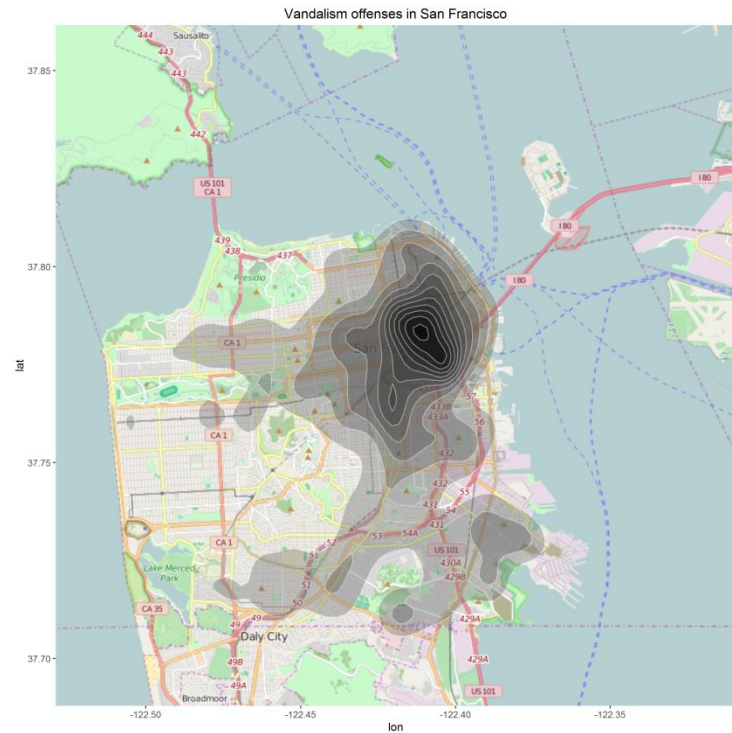
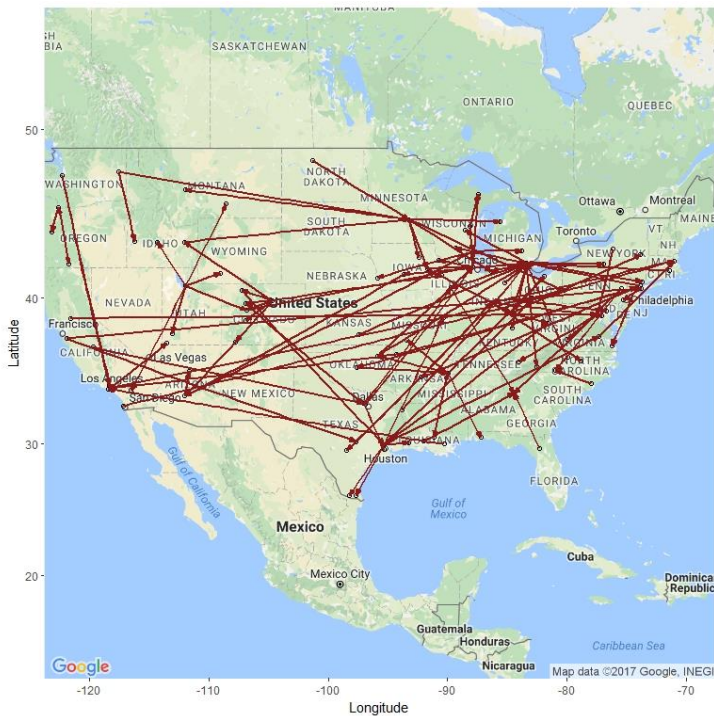
- Let the graphs speak for themselves...





Graphics

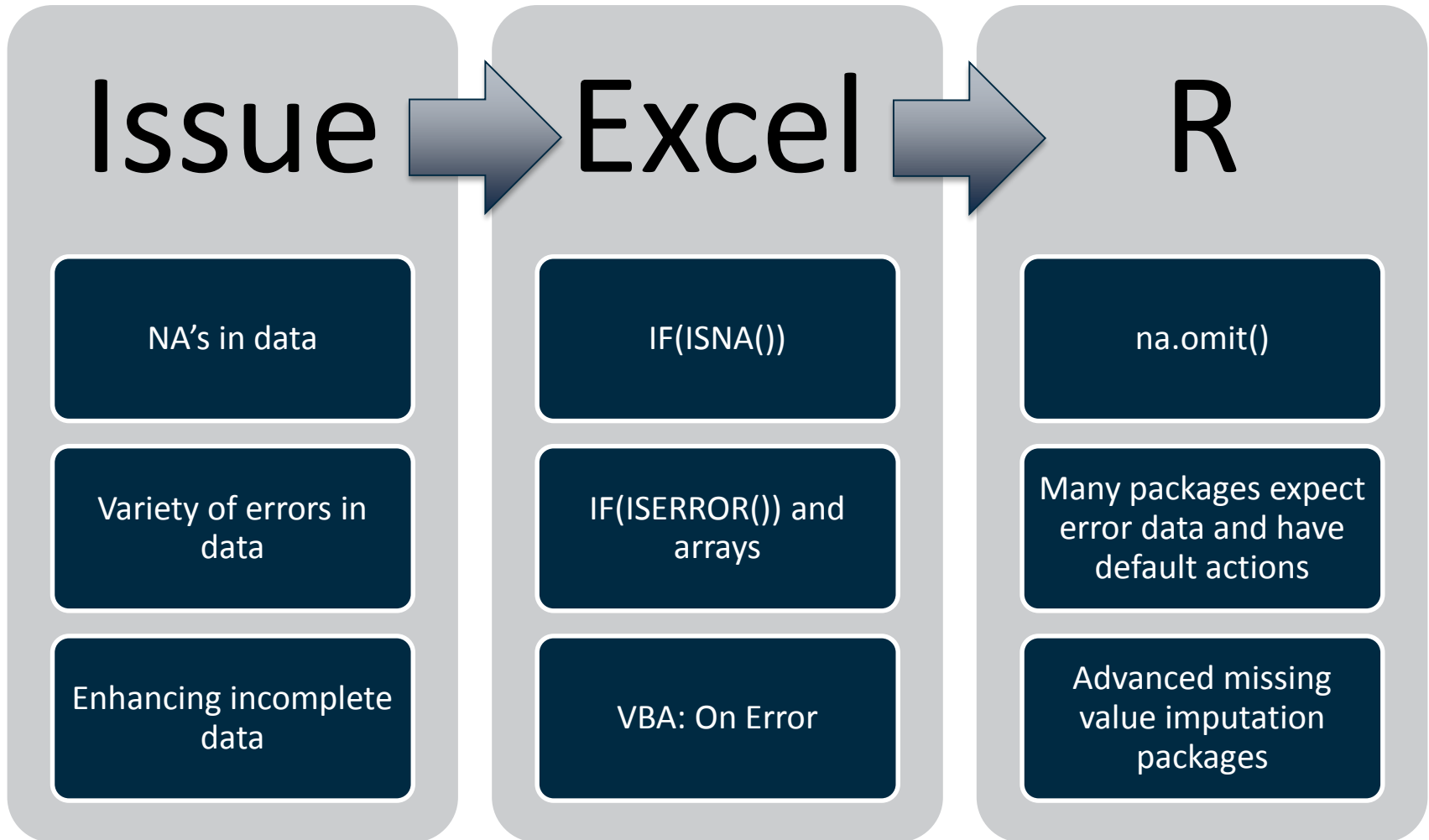
- Let the graphs speak for themselves...





Missing Values and Cleaning

- Missing or error values can sometimes be difficult to handle in Excel





Packages Relevant to Actuaries

No package required

- GLMs

Actuar

- Loss modelling, risk and ruin theory, credibility theory and hierarchical models

Lifecontingencies

- Life tables, functions for demographic, financial and actuarial mathematics of life insurance

ChainLadder

- Claims reserving in general insurance

ELT

- To build experience life tables

DCL

- Claims reserving in general insurance using the double chain ladder framework

MRMR

- Multiplicative chain ladder and additive model for loss reserving in general insurance

RQuantLib

- Quantitative finance, modelling, trading and risk management of financial assets

Tweedie

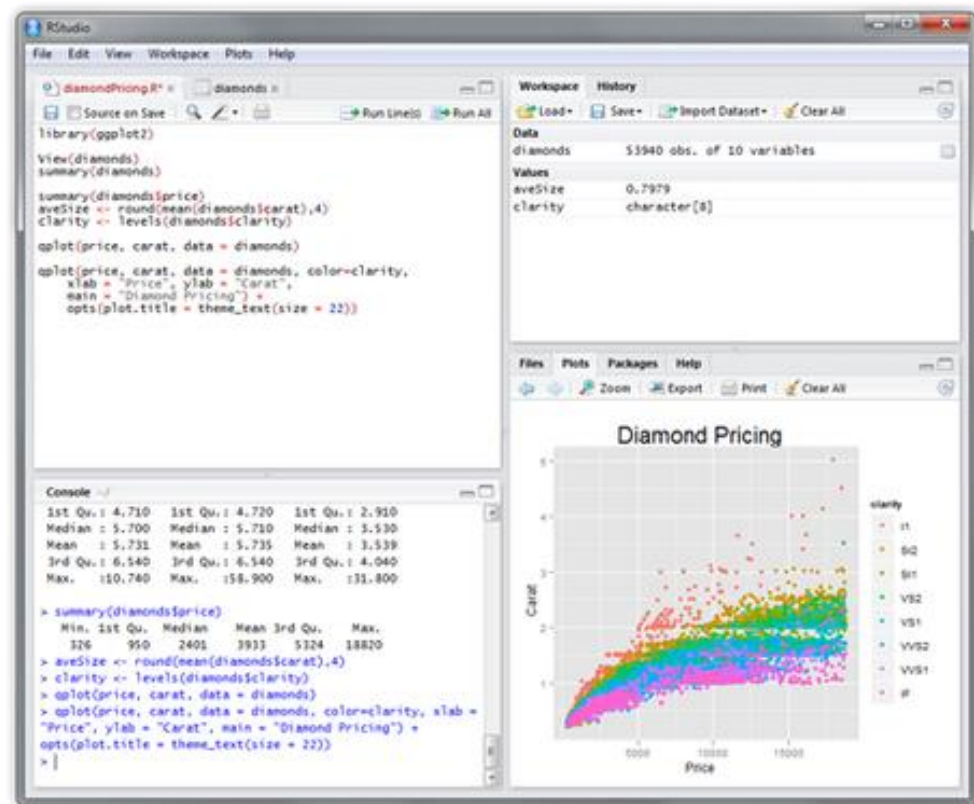
- Tweedie distribution



Where can I find more?



- You need R to start: <https://www.r-project.org/>
- We recommend using an IDE
 - RStudio
 - Jupyter Notebook
 - Emacs
 - Atom
 - Others...



<https://www.r-bloggers.com/rstudio-new-open-source-ide-for-r/>



Where can I find more?

- Learn R:
 - Google “introduction to R” and choose among the many free online courses available.
 - Find a book: <https://www.r-project.org/doc/bib/R-books.html>
 - The power of R comes mostly from its packages: Explore freely and replace base R with packages that provide an easier, more flexible way to do things.
- Some personal suggestions:
 - “Computational Actuarial Science with R”, Arthur Charpentier
 - Anything by Hadley Wickham, even if you are not into data science.
 - Hadley Wickham’s “tidyverse” set of packages.
 - Dplyr for easy and powerful data manipulation
 - Ggplot2 for great and flexible visualisations
 - Subscribe to R-Bloggers: <https://www.r-bloggers.com/>
 - Consider joining a relevant MeetUp group.
 - “R in Insurance” annual conference.



Questions?