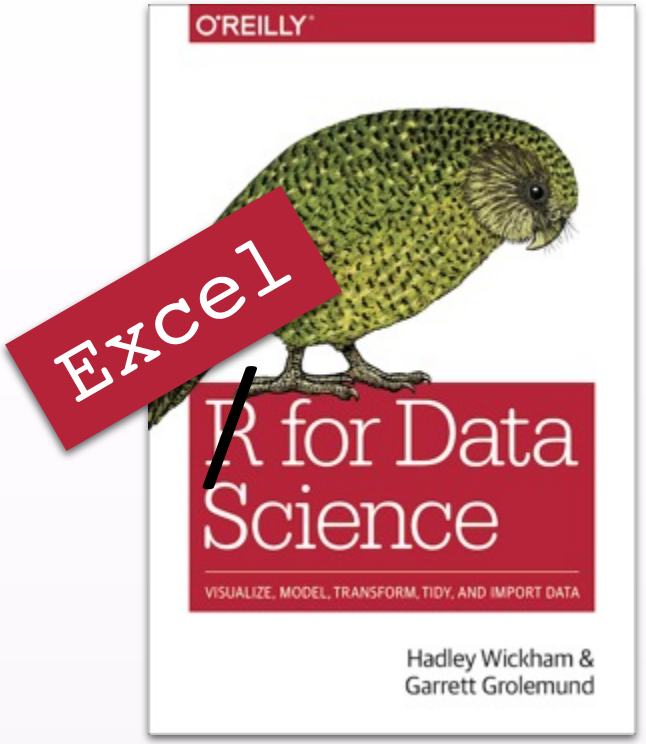
Scaling Spreadsheets with R

Nathan Stephens

October 2021

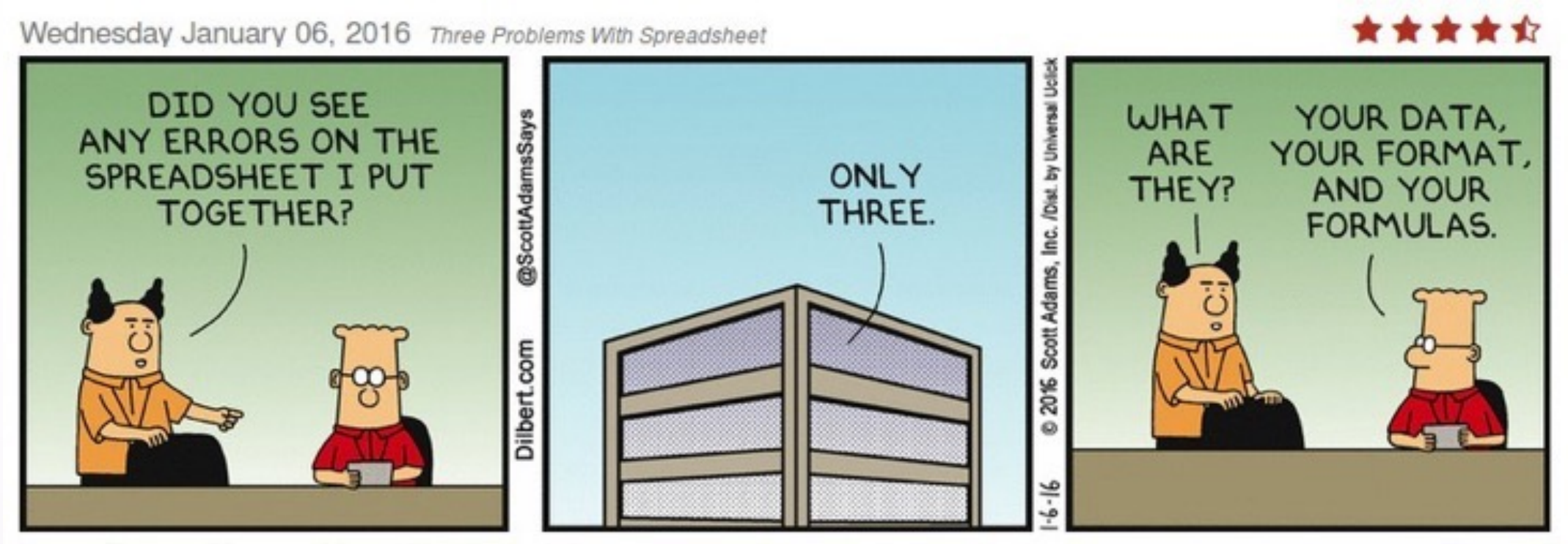
Why do we use Excel?

- Wrangle
- Visualize
- Transform
- Analyze
- Communicate insights about your data



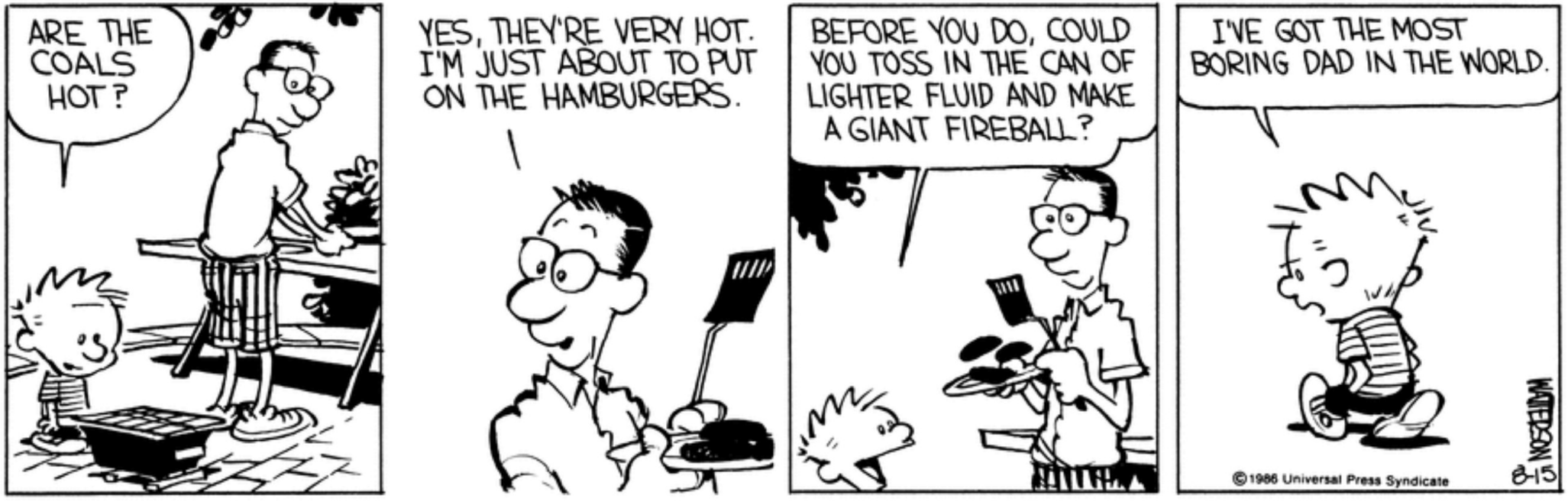
RStudio

[A few] Problems using Excel



Problem 1: File size

Excel handles kilobytes and megabytes comfortably



Hard limits in Excel

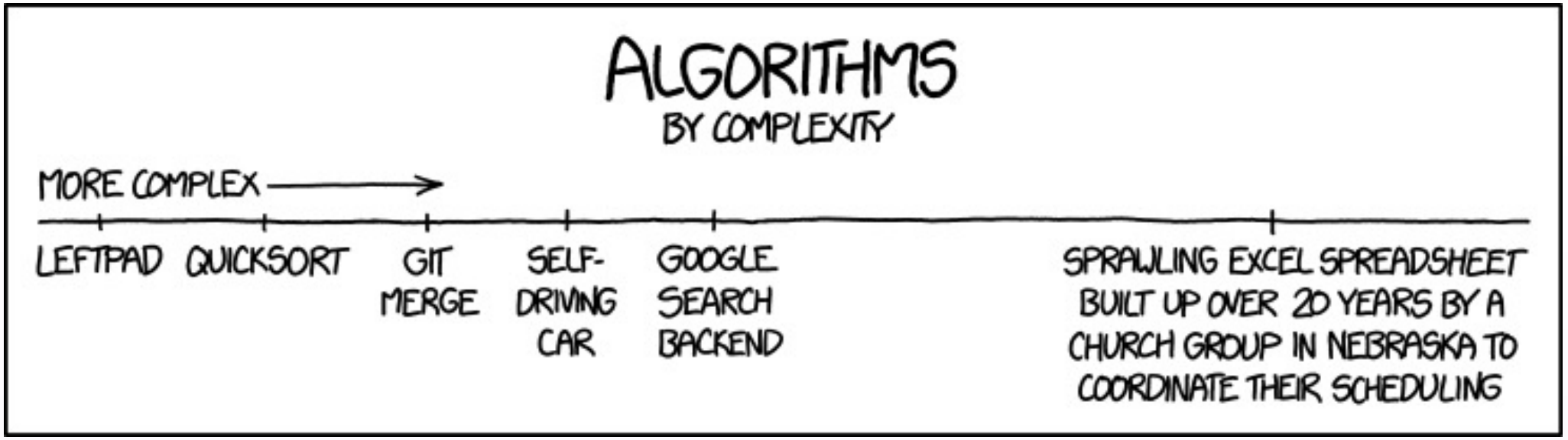
- Maximum rows: 1,048,576
- Maximum columns: 16,384

Memory limitations

1's Mb	10's Mb	100's Mb	1's Gb	10's Gb	100's Gb

RStudio

Problem 2: Complexity



Workbook Design

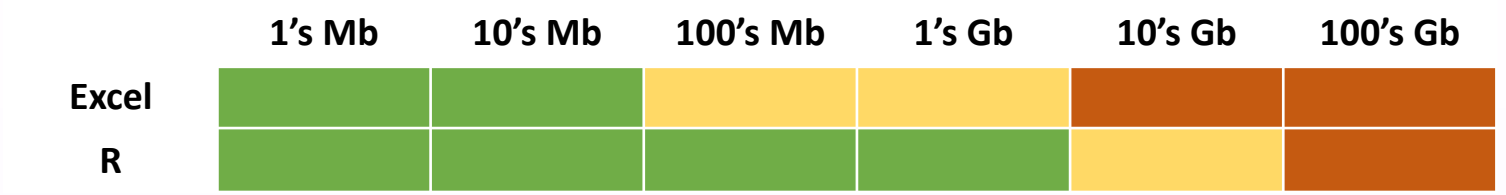
Complexity →

Example	Simple Workbook	Complex Workbook
Data sources	Single	Multiple
Analytics	Summary statistics	Advanced algorithms
Data updates	Static or none	Dynamic
Formulas and Functions	One or two arguments	Multi-line and nested
Advanced tools	Single drop-down list or none	Scripts
Pivot Tables and Charts	One or two	Several
Spreadsheets	Few (less than a dozen)	Many (more than a dozen)

RStudio

Why scale spreadsheets with R?

- R can handle far more data than Excel



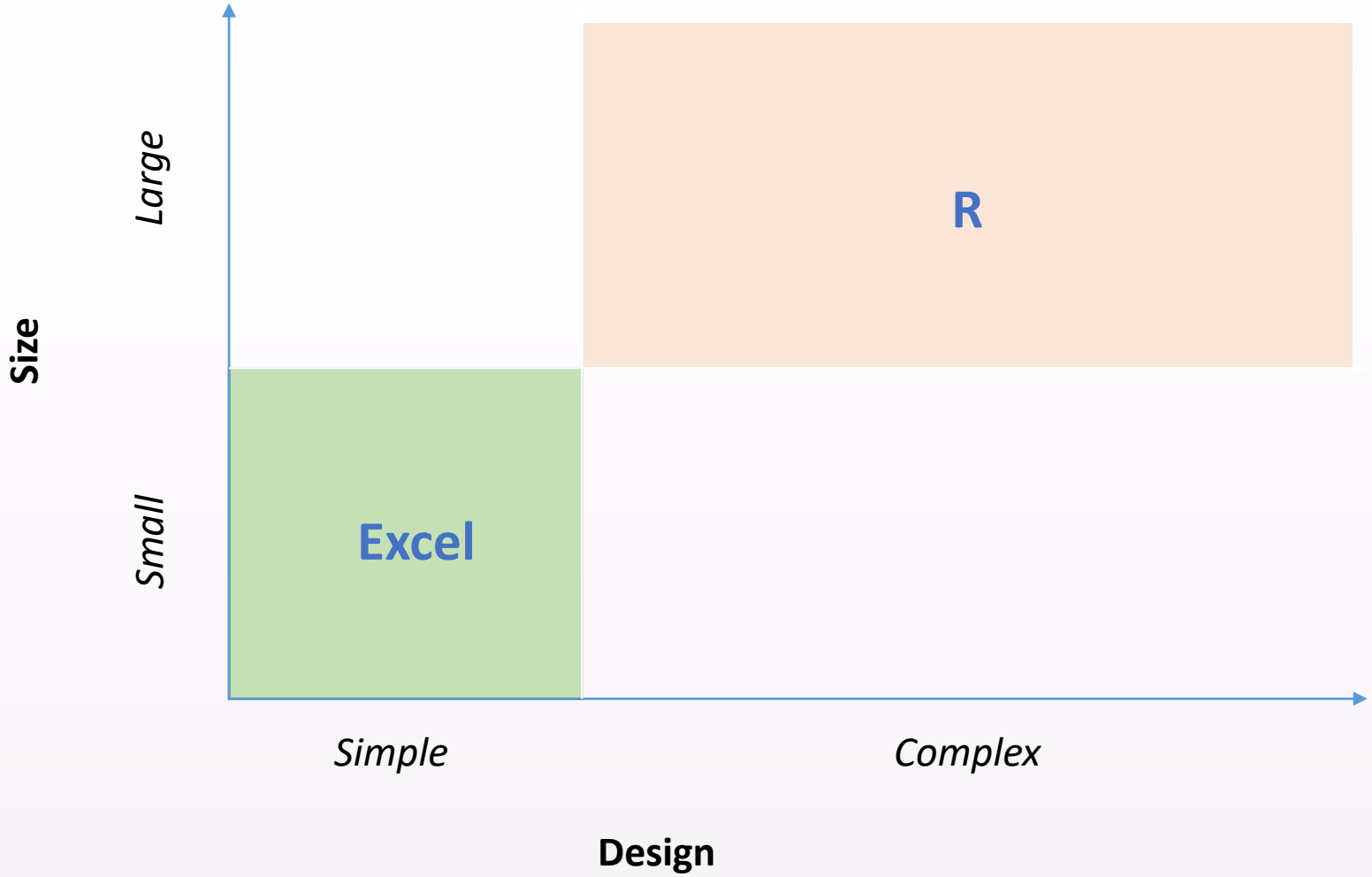
- Complex Workbook in Excel == Simple script in R

Simple in R

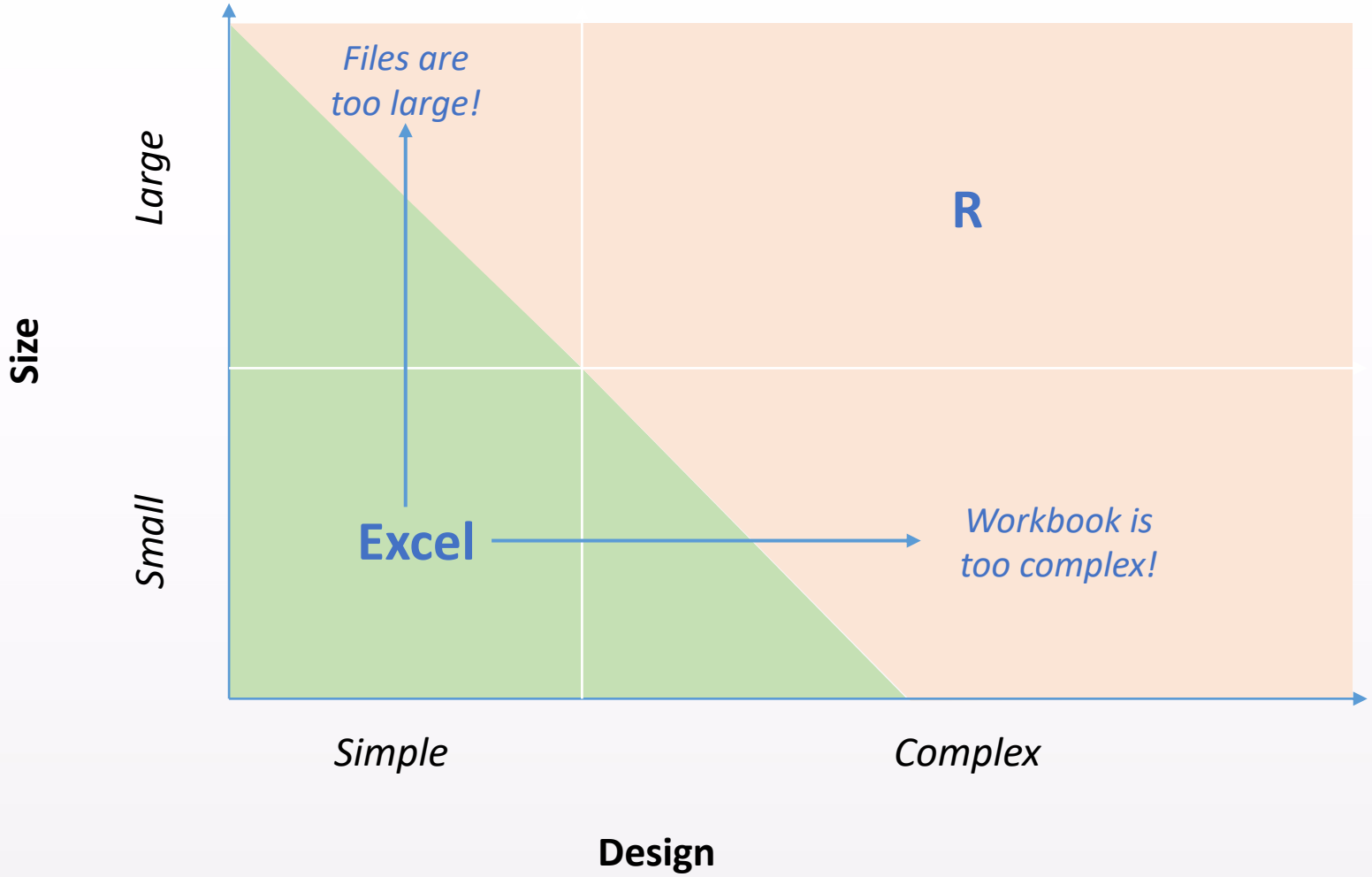
Example	Simple Workbook	~~Complex Workbook~~
Data sources	Single	Multiple
Analytics	Summary statistics	Advanced algorithms
Data updates	Static or none	Dynamic
Formulas and Functions	One or two arguments	Multi-line and nested
Advanced tools	Single drop-down list or none	Scripts
Pivot Tables and Charts	One or two	Several
Spreadsheets	Few (less than a dozen)	Many (more than a dozen)

RStudio

Scaling spreadsheets with R

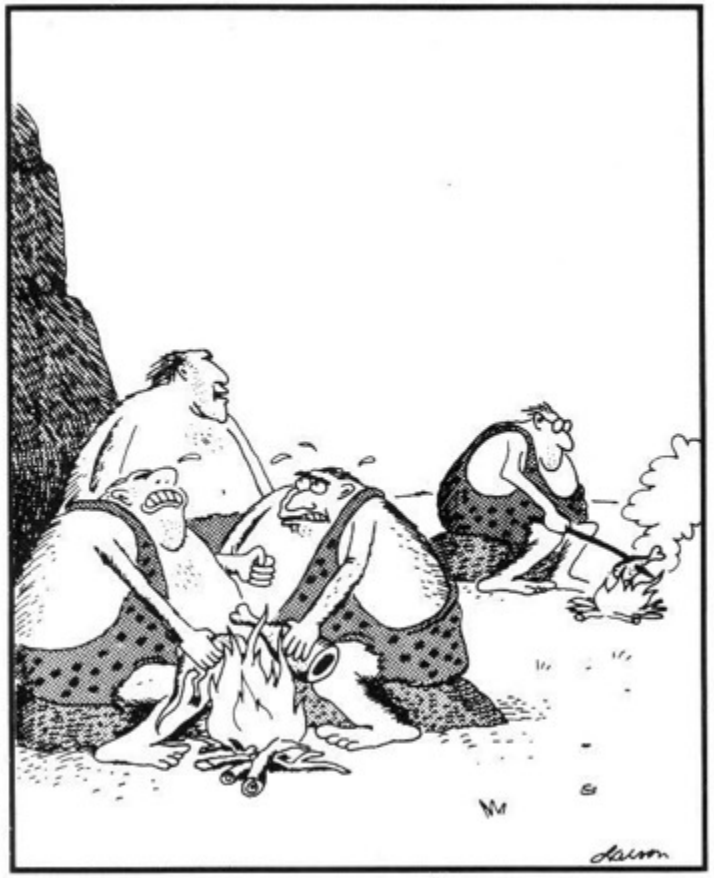


Scaling spreadsheets with R



How do you scale spreadsheets with R?

- By using powerful coding tools that make your work reproducible
- Communicating with Apps and Notebooks



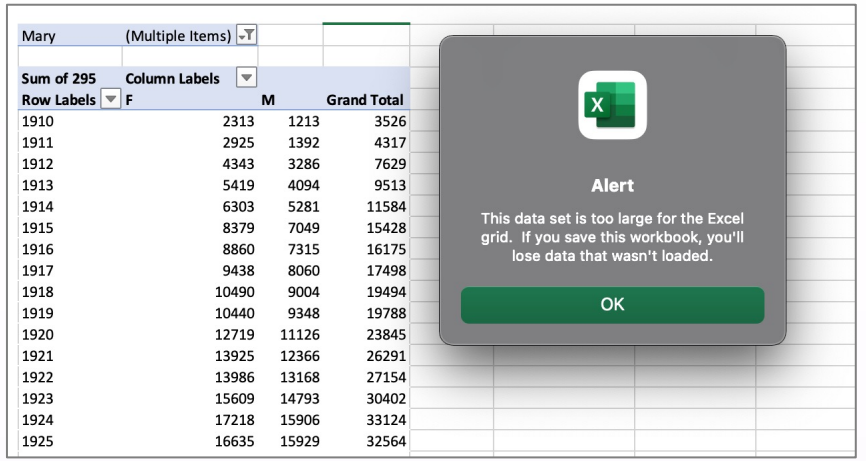
"Hey! Look what Zog do!"

RStudio

Two Demos

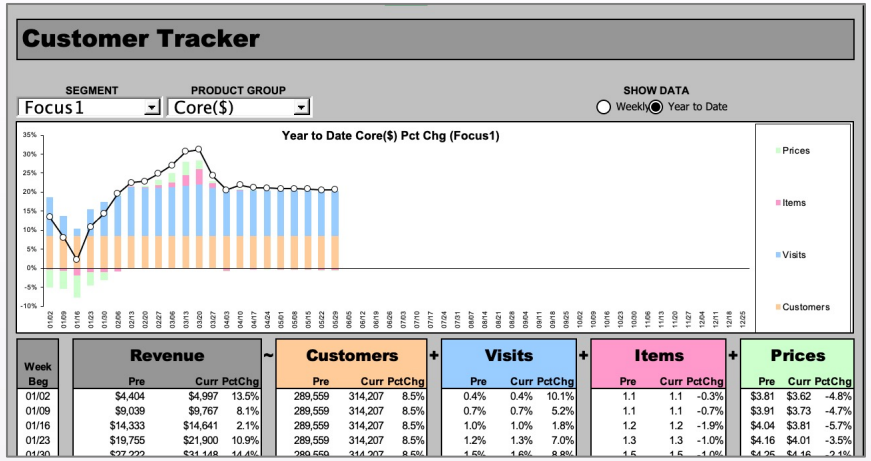
File is too large

- US State Specific Baby Names



Workbook is too complex

- Customer Tracker



Summary

In data science, the source for all your results is

YOUR CODE

Not your report or presentation