



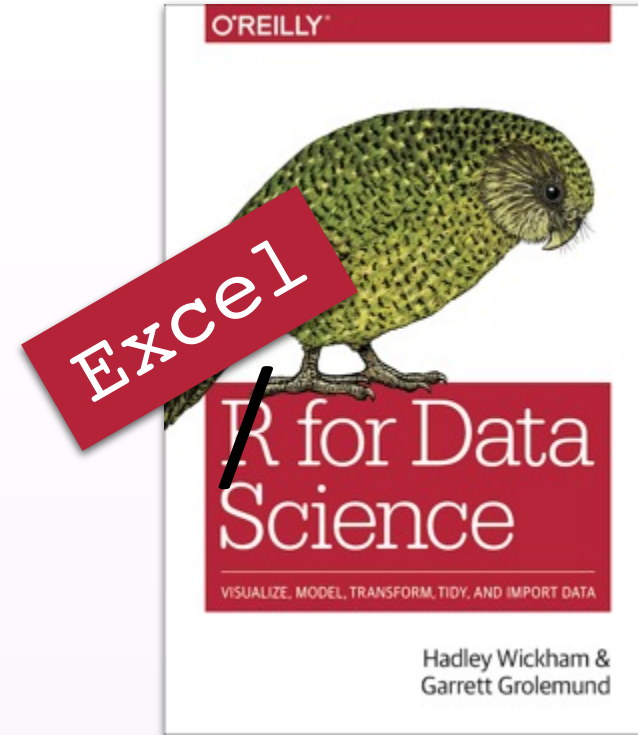
Scaling Spreadsheets with R

Nathan Stephens

October 2021

Why do we use Excel?

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

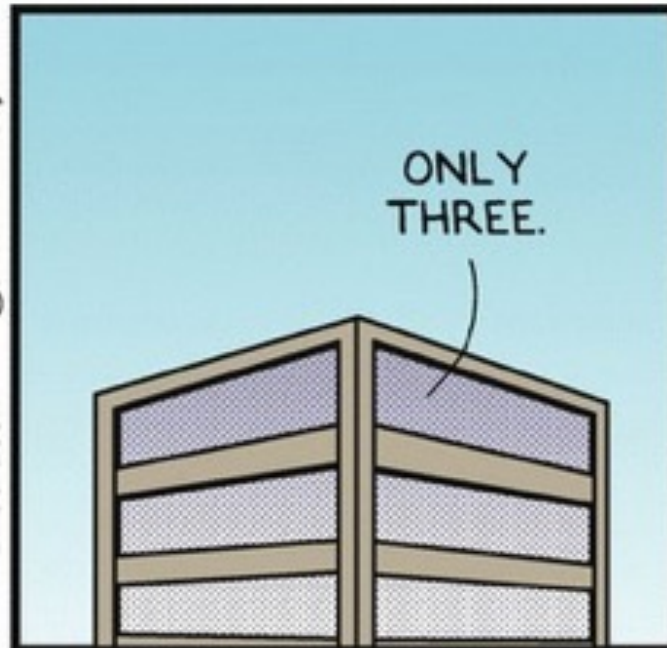


[A few] Problems using Excel

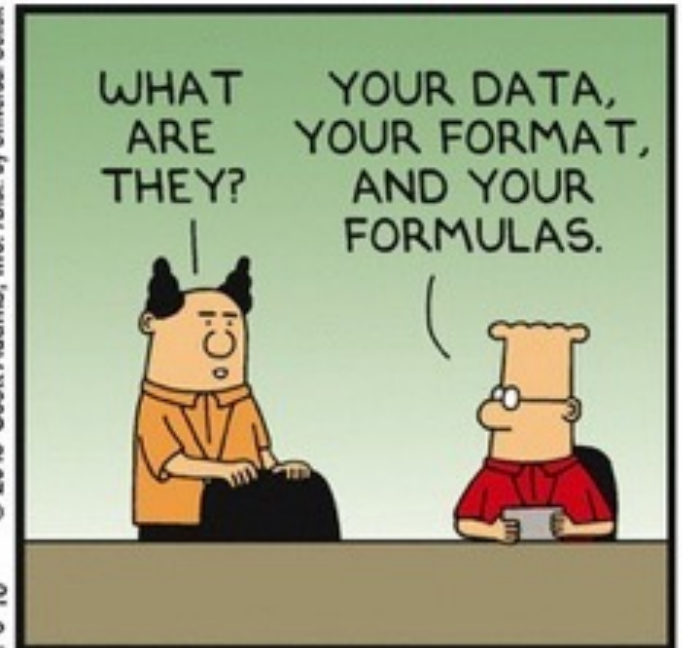
Wednesday January 06, 2016 *Three Problems With Spreadsheet*



Dilbert.com @ScottAdamsSays

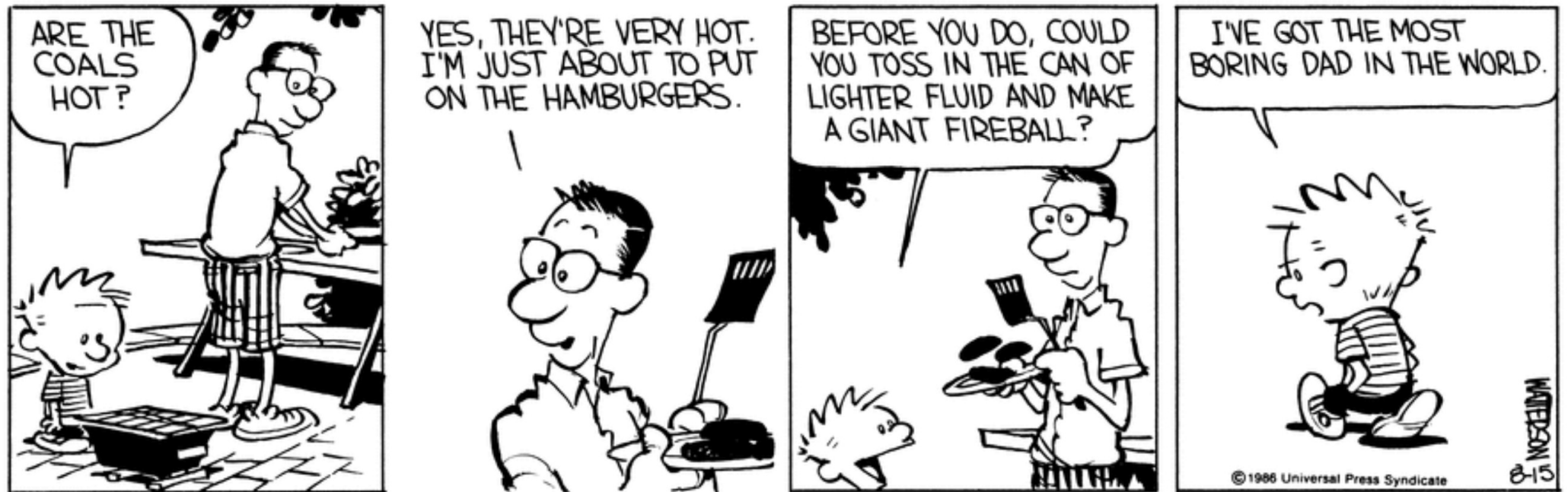


I-6-16 © 2016 Scott Adams, Inc. /Dist. by Universal Uclick



Problem 1: File size

Excel handles kilobytes and megabytes comfortably



Hard limits in Excel

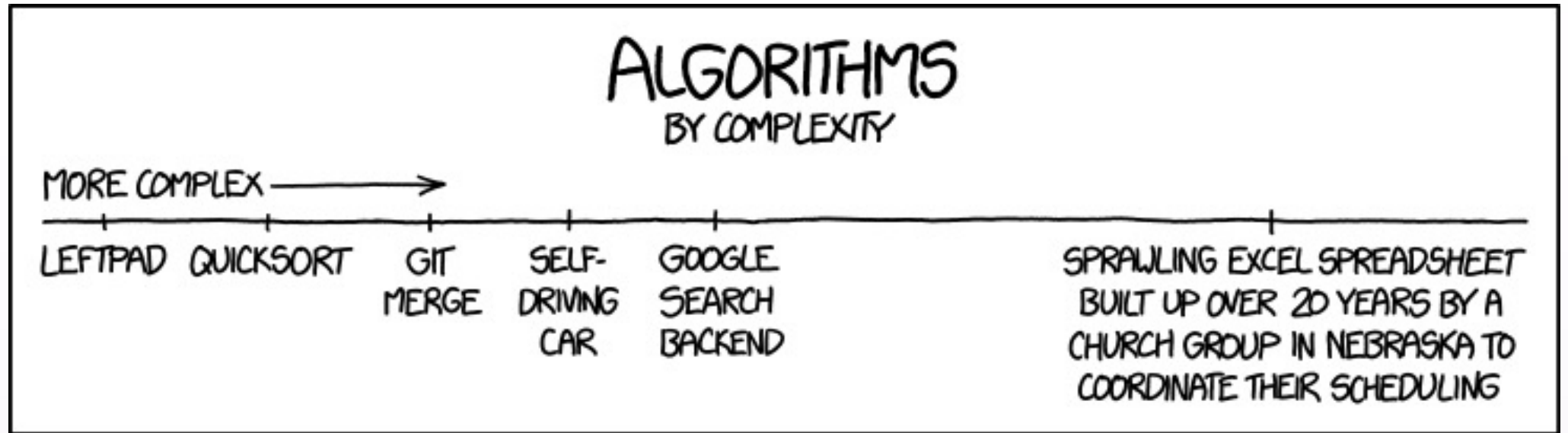
- Maximum rows: 1,048,576
- Maximum columns: 16,384



Memory limitations

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

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)

Why scale spreadsheets with R?

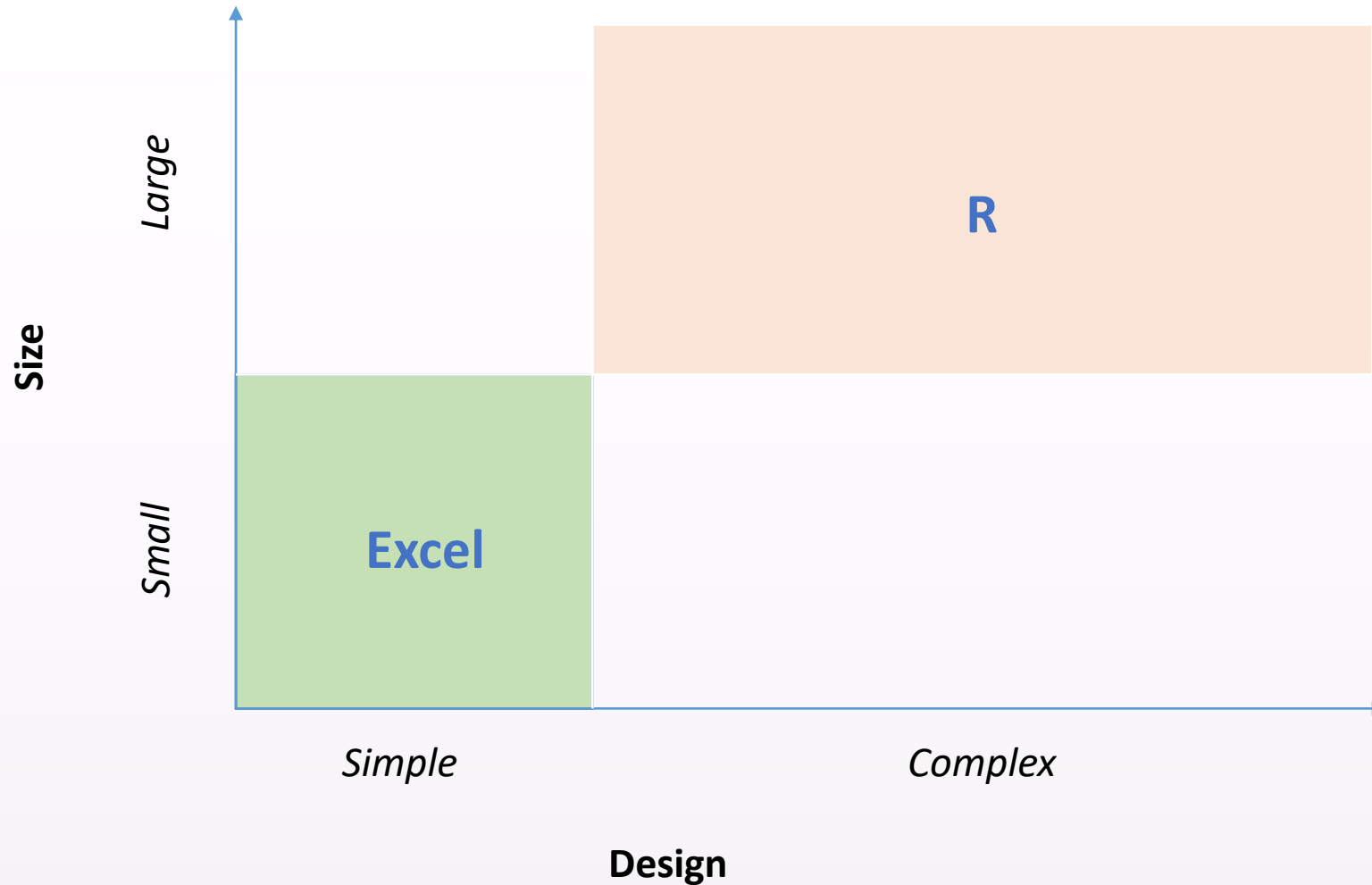
- R can handle far more data than Excel

	1's Mb	10's Mb	100's Mb	1's Gb	10's Gb	100's Gb
Excel						
R						

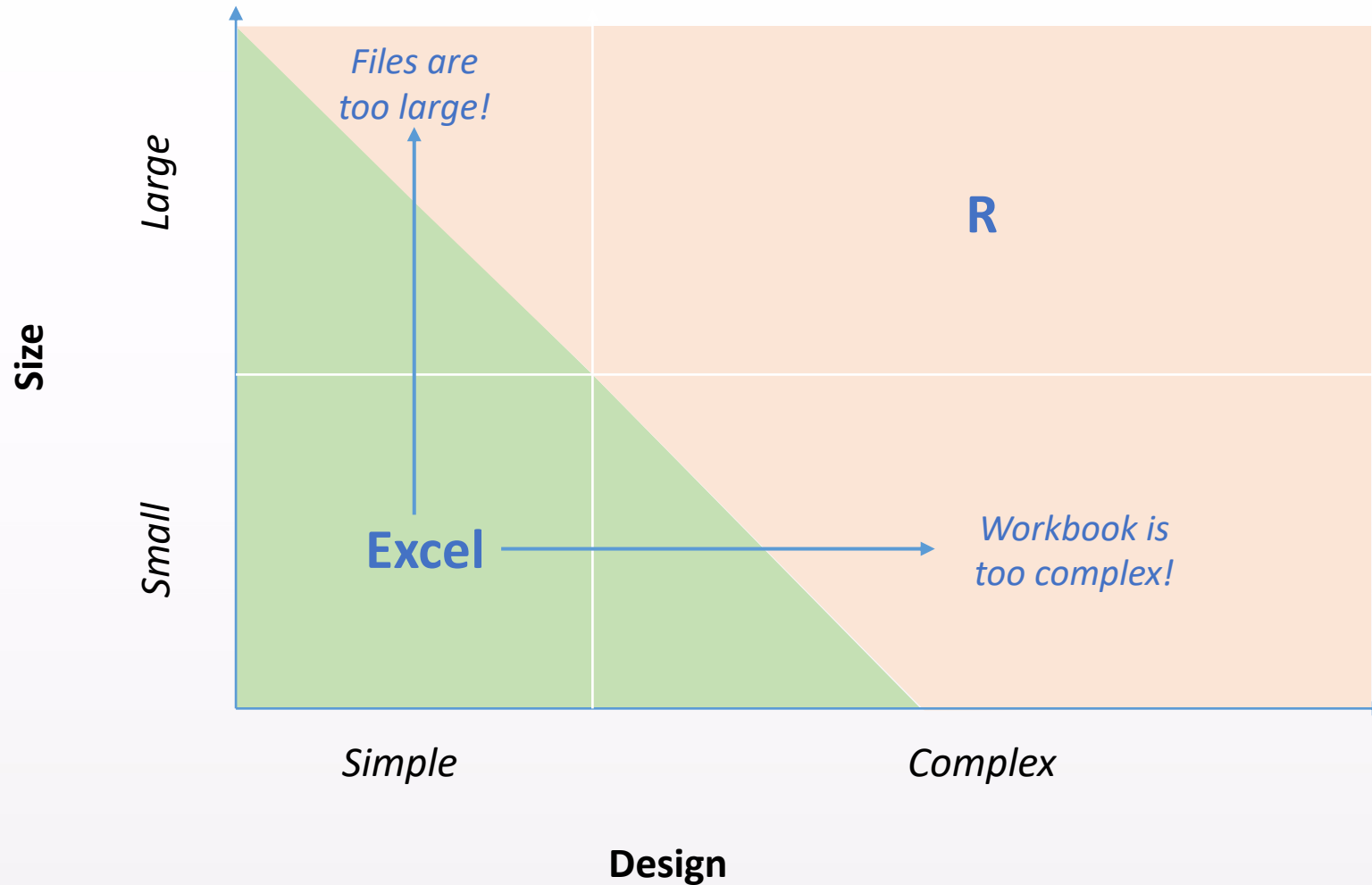
- Complex Workbook in Excel == Simple script in R

Example	Simple Workbook	Simple in R
		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)

Scaling spreadsheets with R



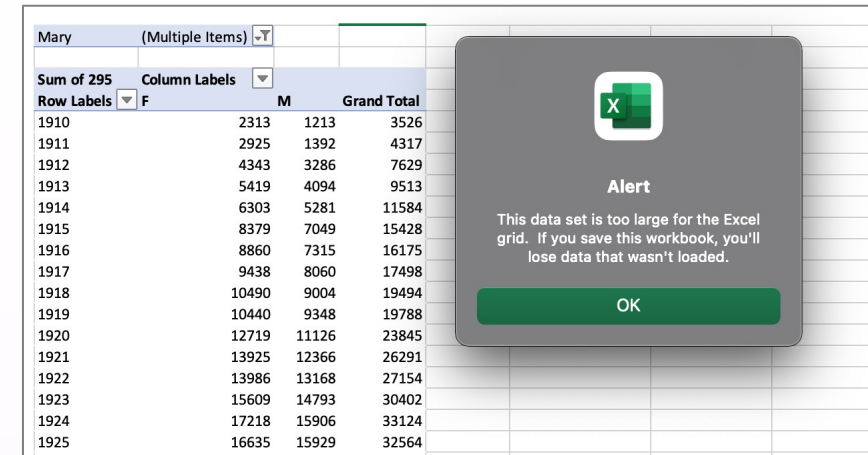
Scaling spreadsheets with R



Two Demos

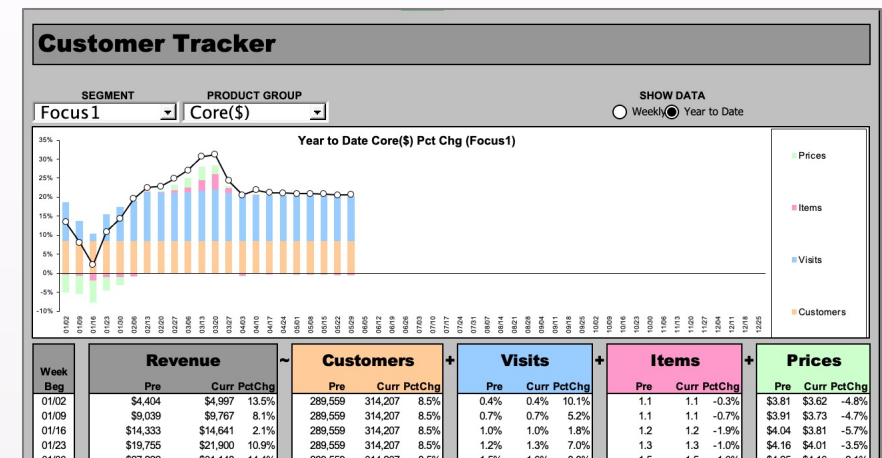
File is too large

- [US State Specific Baby Names](#)



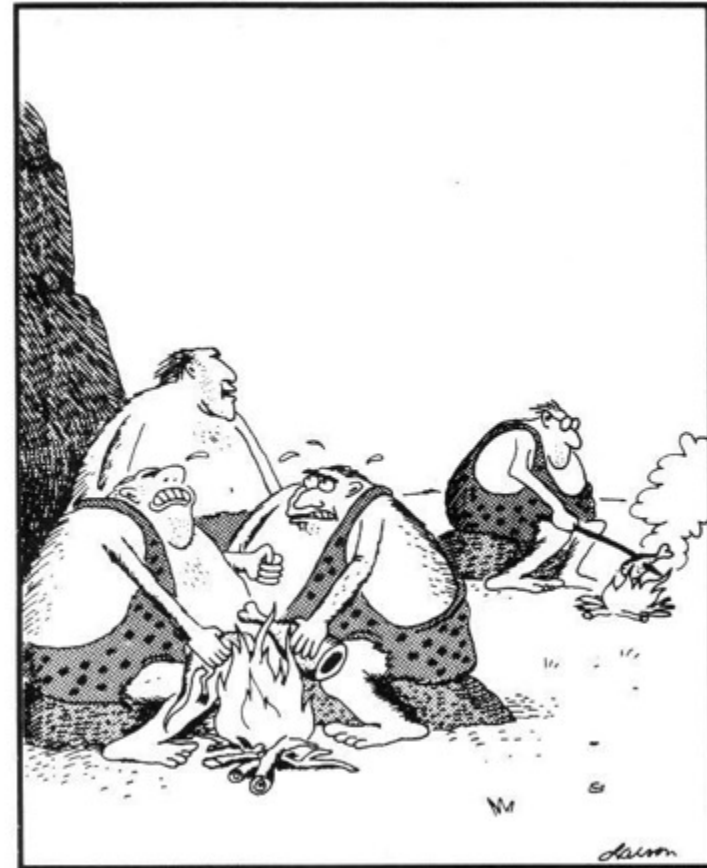
Workbook is too complex

- [Customer Tracker](#)



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!"

Summary

In data science, the source for all your results is

YOUR CODE

Not your report or presentation