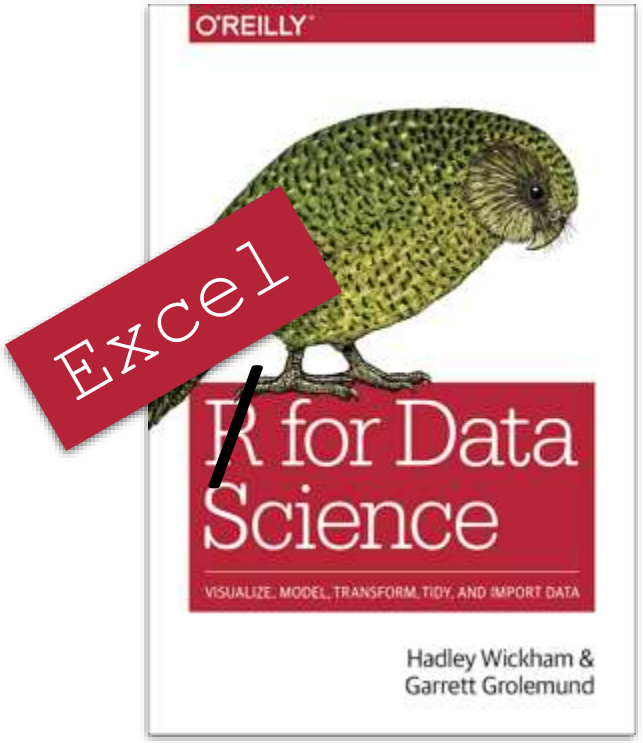
# When to stop using Excel and start using R

Nathan Stephens

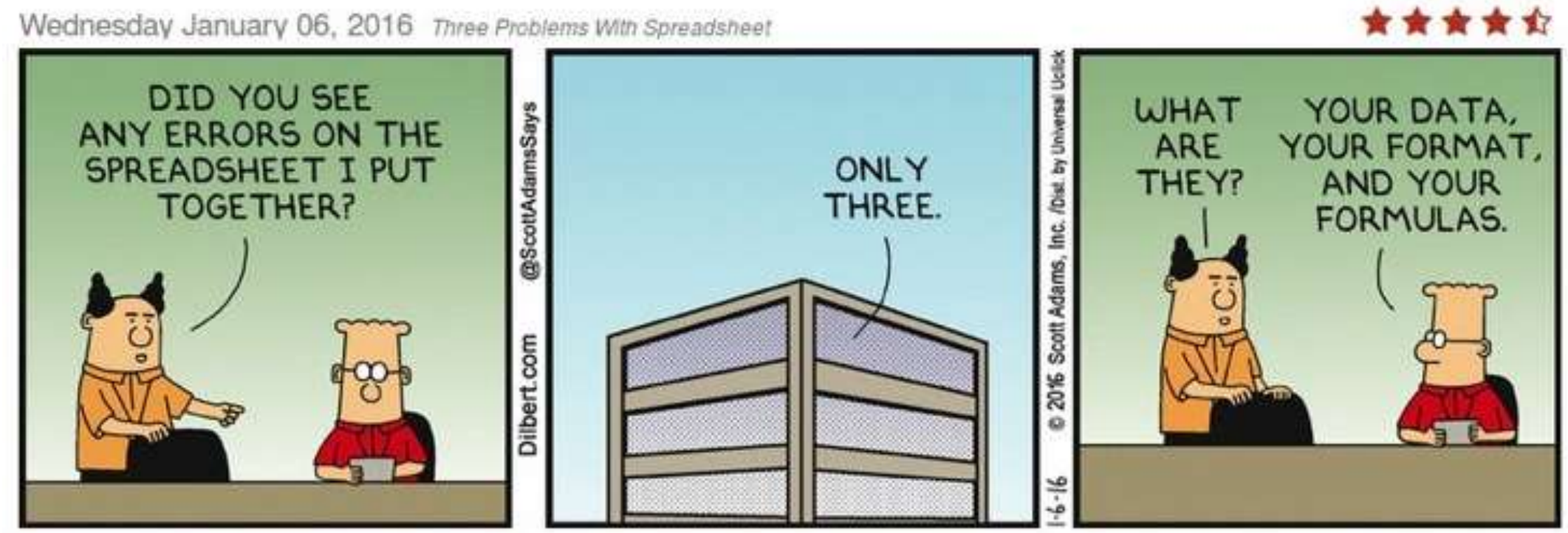
February 2022

# Why do we use Excel?

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

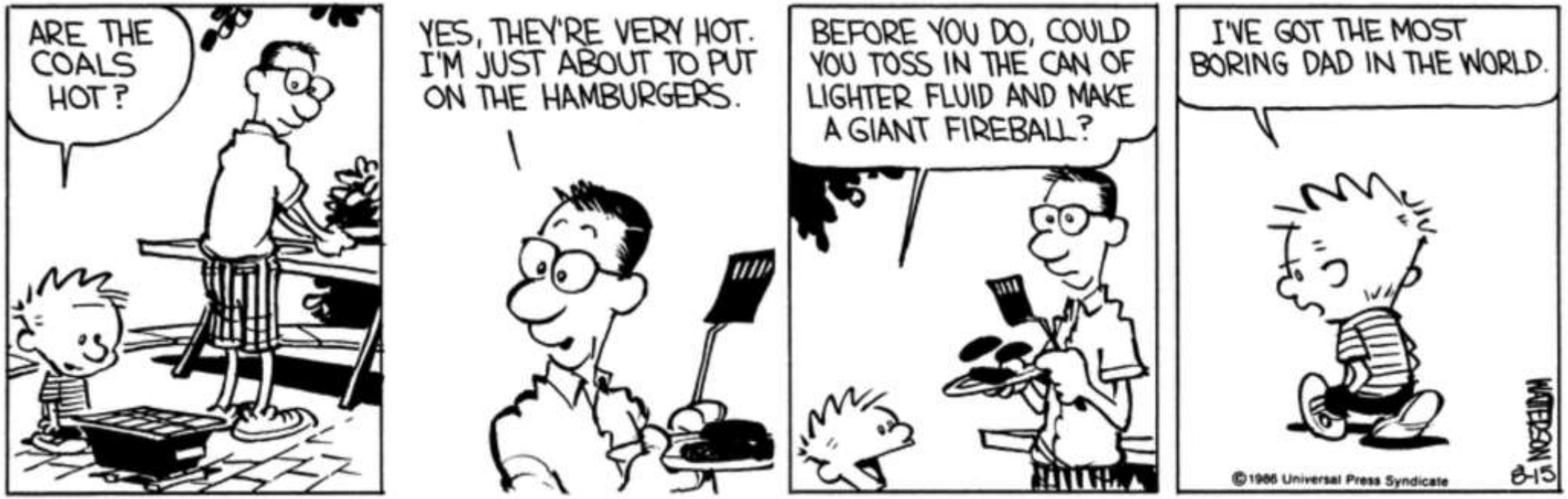


# [A few] Problems using Excel



# **Problem 1:** Scaling your data

Can be like playing with fire



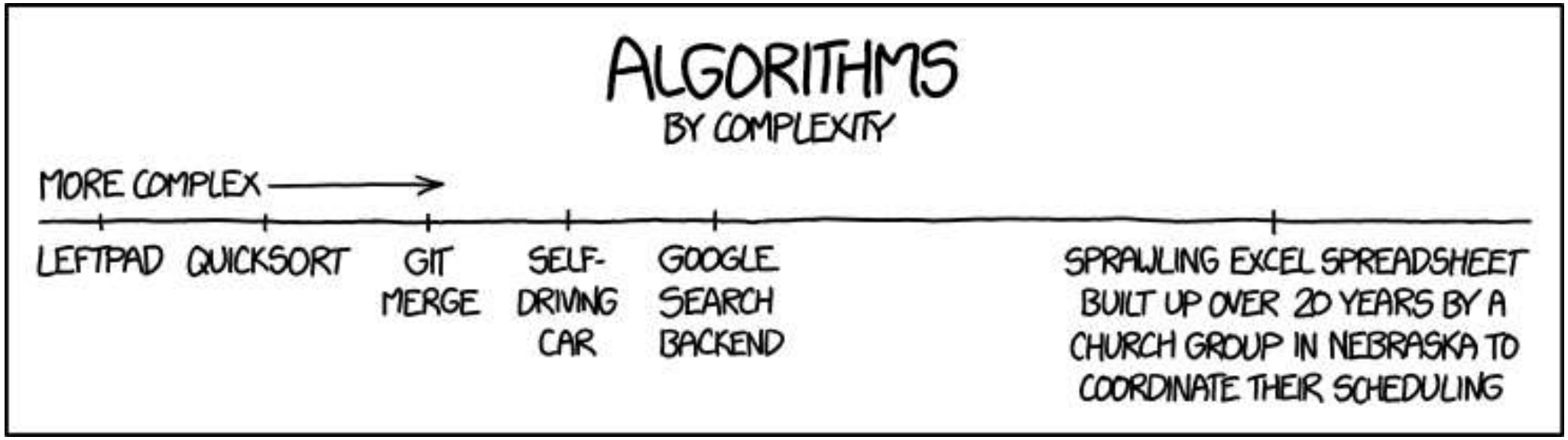
# Hard limits in Excel

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

# Memory limitations

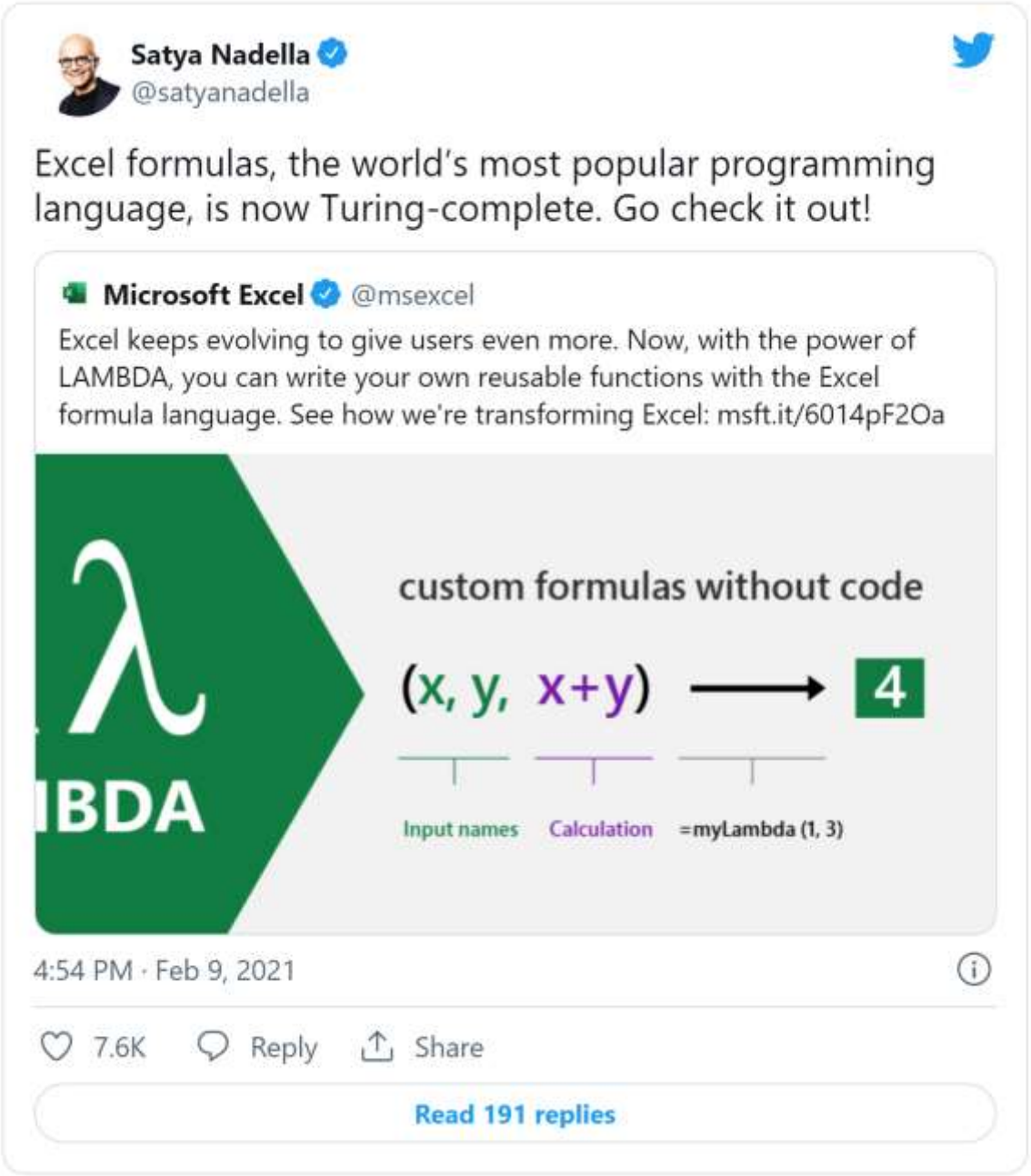
| 1’s Mb | 10’s Mb | 100’s Mb | 1’s Gb | 10’s Gb | 100’s Gb |
|---|---|---|---|---|---|

# Problem 2: Scaling your code



# Excel is Turing Complete

- A system that is able to recognize or decide other data-manipulation rule sets.
- In principle -- **although often not in practice** -- it can be used to solve any computation problem.



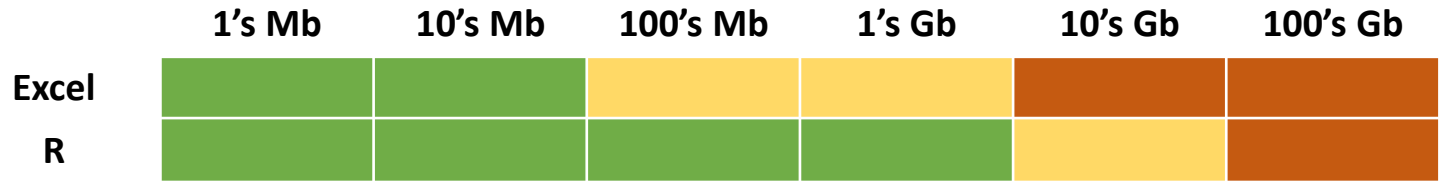
# Code complexity

**Complexity**

| Example | Simple Workbook | Advanced 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) |

# Use R to scale your spreadsheets

- R handles more data

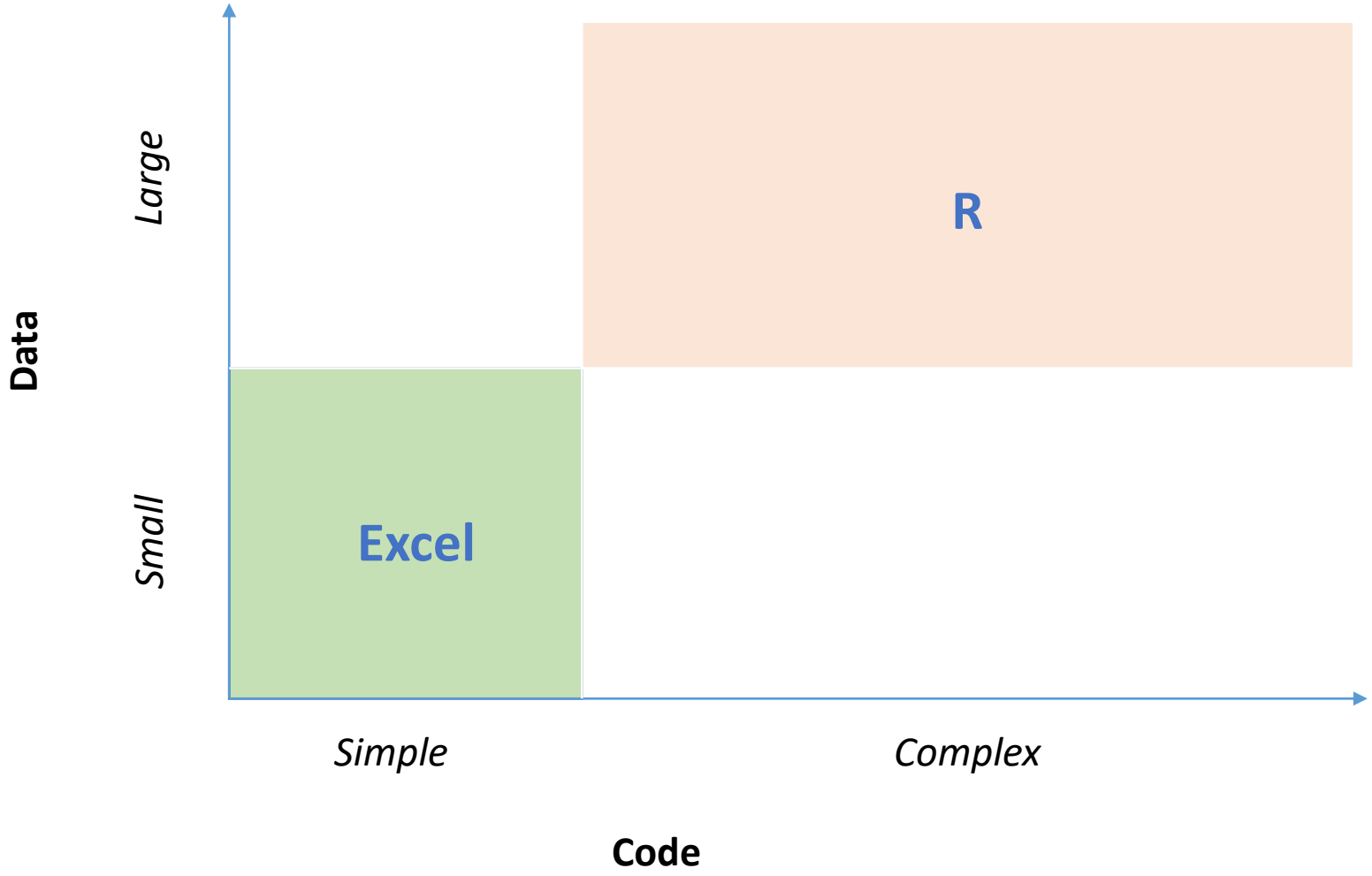



- R code simplifies complex tasks

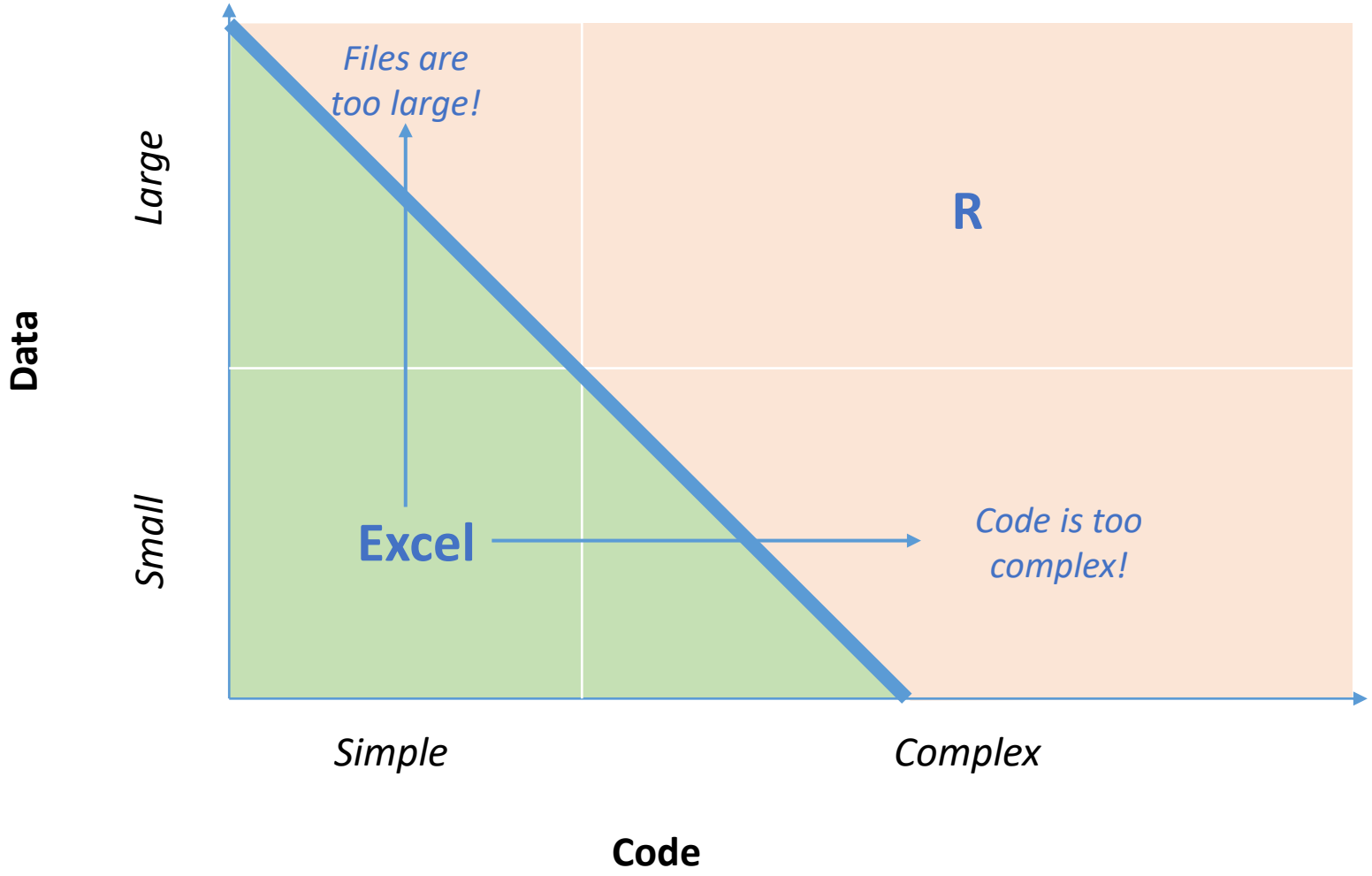
**Simple in R**

| Example | Simple Workbook | Advanced 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) |

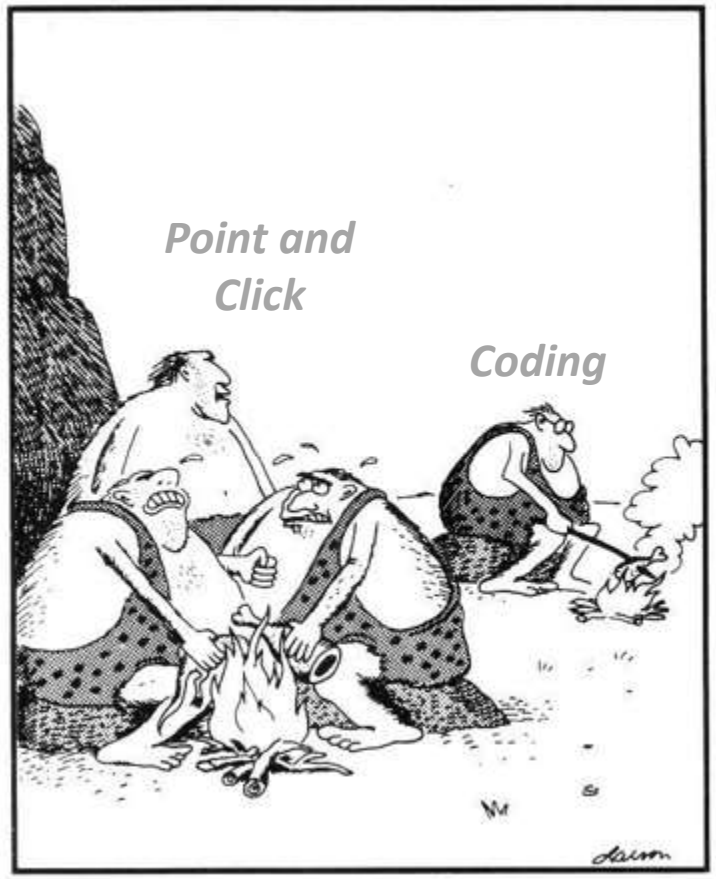
# When to stop using Excel and start using R



# When to stop using Excel and start using R



# For complex code there is a better way



"Hey! Look what Zog do!"

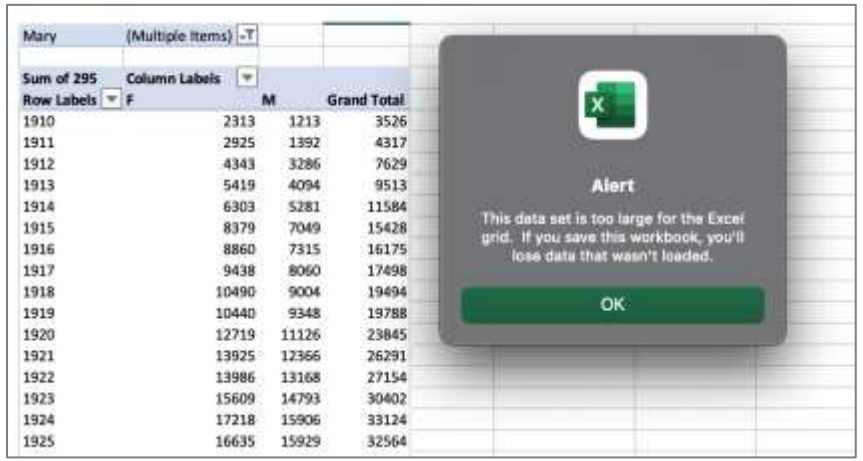
## Benefits of using reproducible code

- Wider variety of tools
- More powerful tools
- Reproducible by others
- Easier to inspect for errors

# Demos

## 1) File is too large

- [US State Specific Baby Names]




## 2) Code is too complex

- [Customer Tracker]