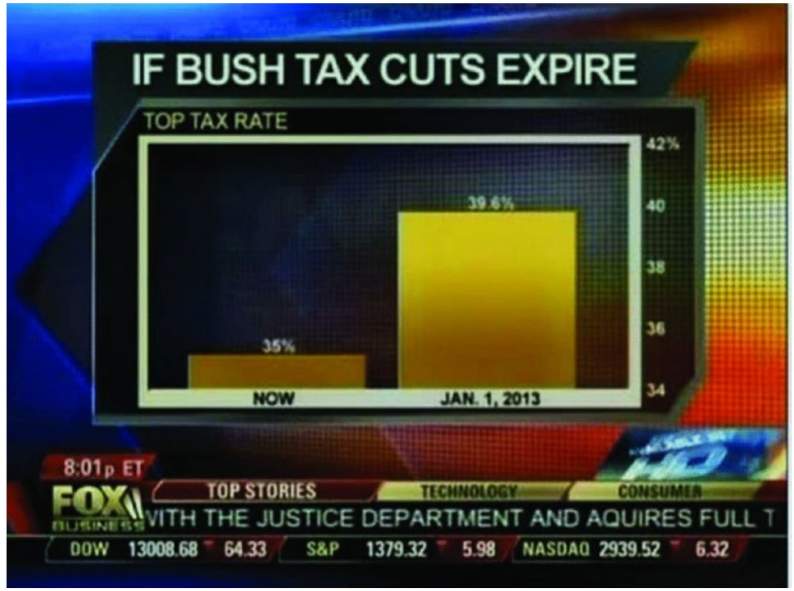
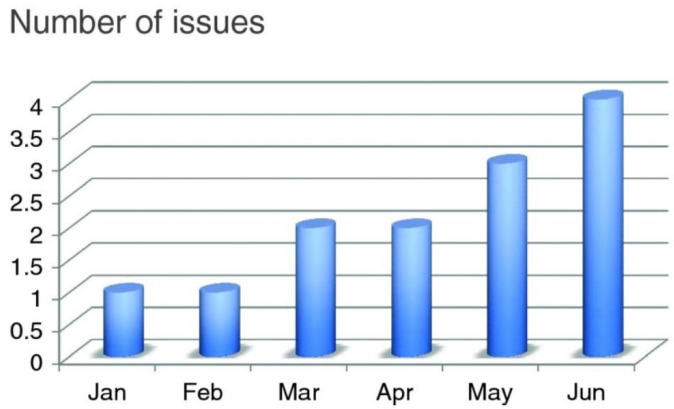
Data Visualization I

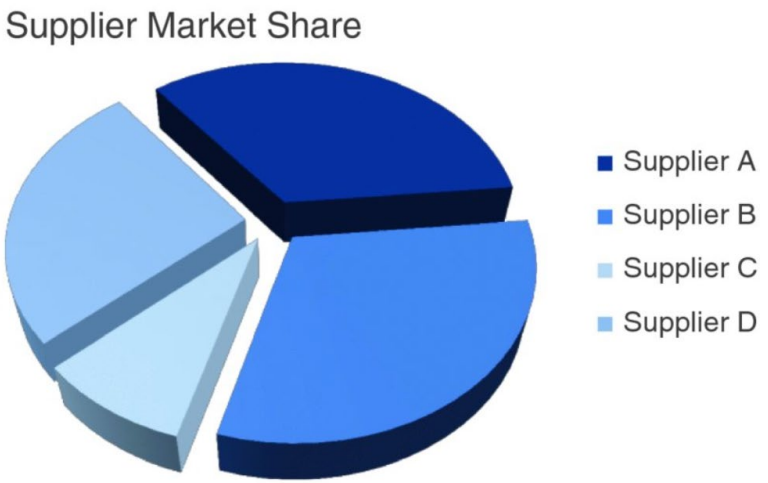
John Rios

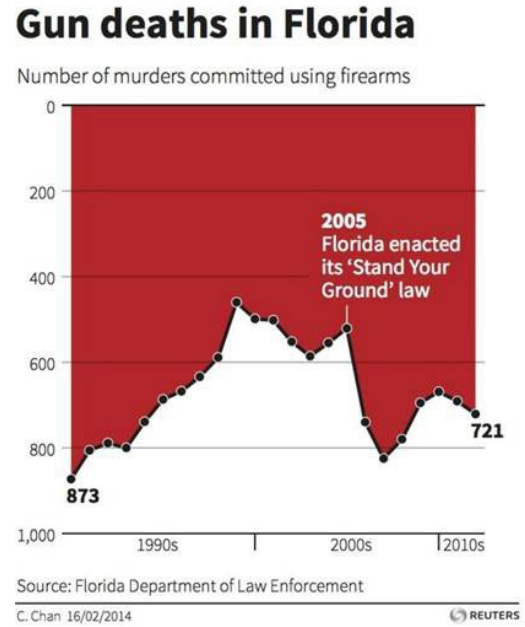
Business Intelligence and Analytics

Terry College of Business
UNIVERSITY OF GEORGIA









Exploratory versus explanatory

- Easily-generated
- Data-heavy
- Small specialist audience (yourself/colleagues)
- Graphical data analysis.

- Labor-intensive
- Data-specific
- Broader audience (publications or presentations)
- Part of the communications process.

A Few Lessons from Knaflic (2015)

1. Understand the context
2. Choose an appropriate visual display
3. Eliminate clutter
4. Focus attention where you want it
5. Think like a designer
6. Tell a story

Knaflic, C. G. (2015). Storytelling with Data.

The importance of context

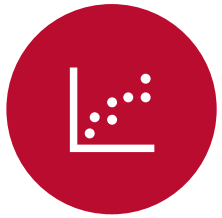
- Who, what, and how
- **Who** is your audience?
- **What** do you need them to know or do?
- **How** can you use data to help make your point?
- Understand the situational context:
  - Audience
  - Communication mechanism
  - Desired tone

A Few Lessons from Knaflic (2015)

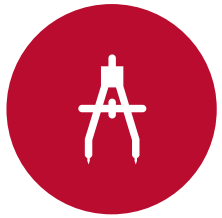
1. Understand the context
2. Choose an appropriate visual display
3. Eliminate clutter
4. Focus attention where you want it
5. Think like a designer
6. Tell a story

Knaflic, C. G. (2015). Storytelling with Data.

Choosing an effective visual – graphs



POINTS



LINES

BARS

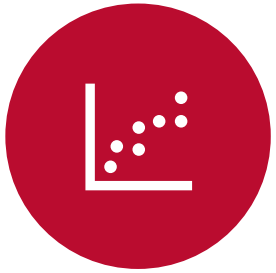
AREA

PIES

DONUT

3D

Choosing an effective visual – graphs



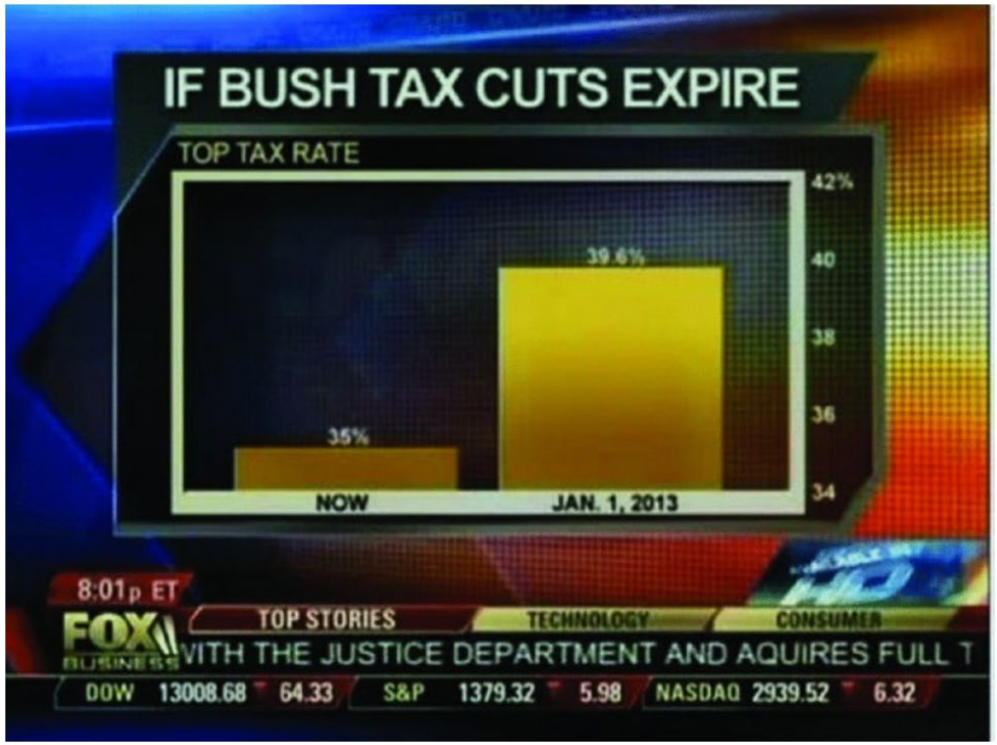
POINTS

LINES

BARS

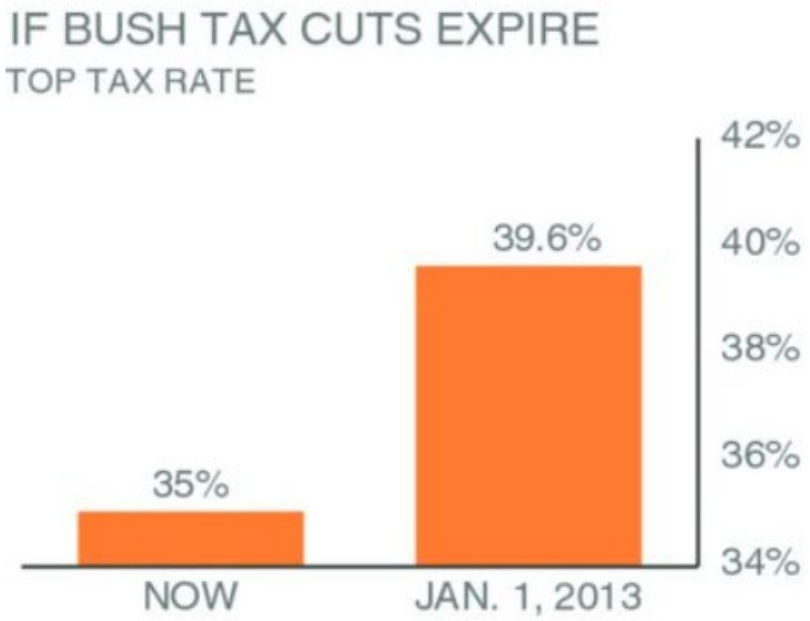
AREA

Easy for our eyes to read

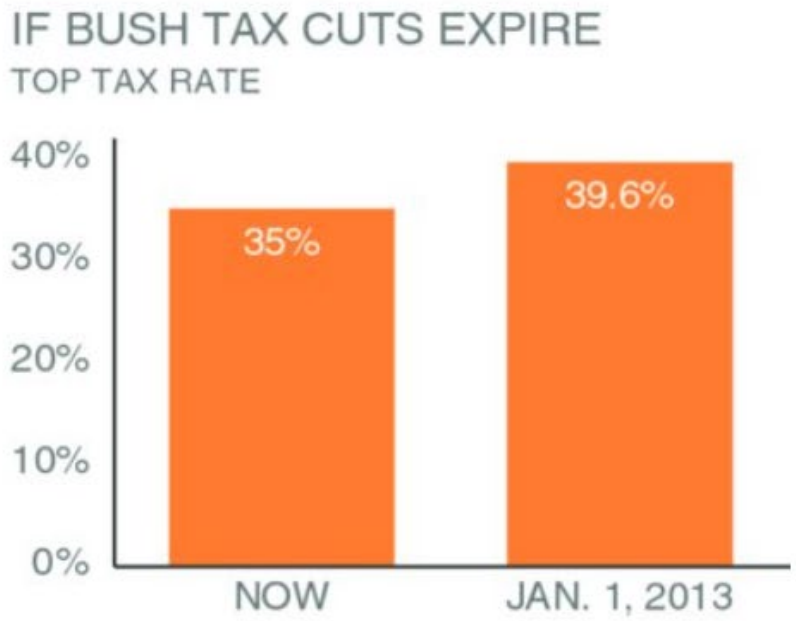


Always have a zero baseline

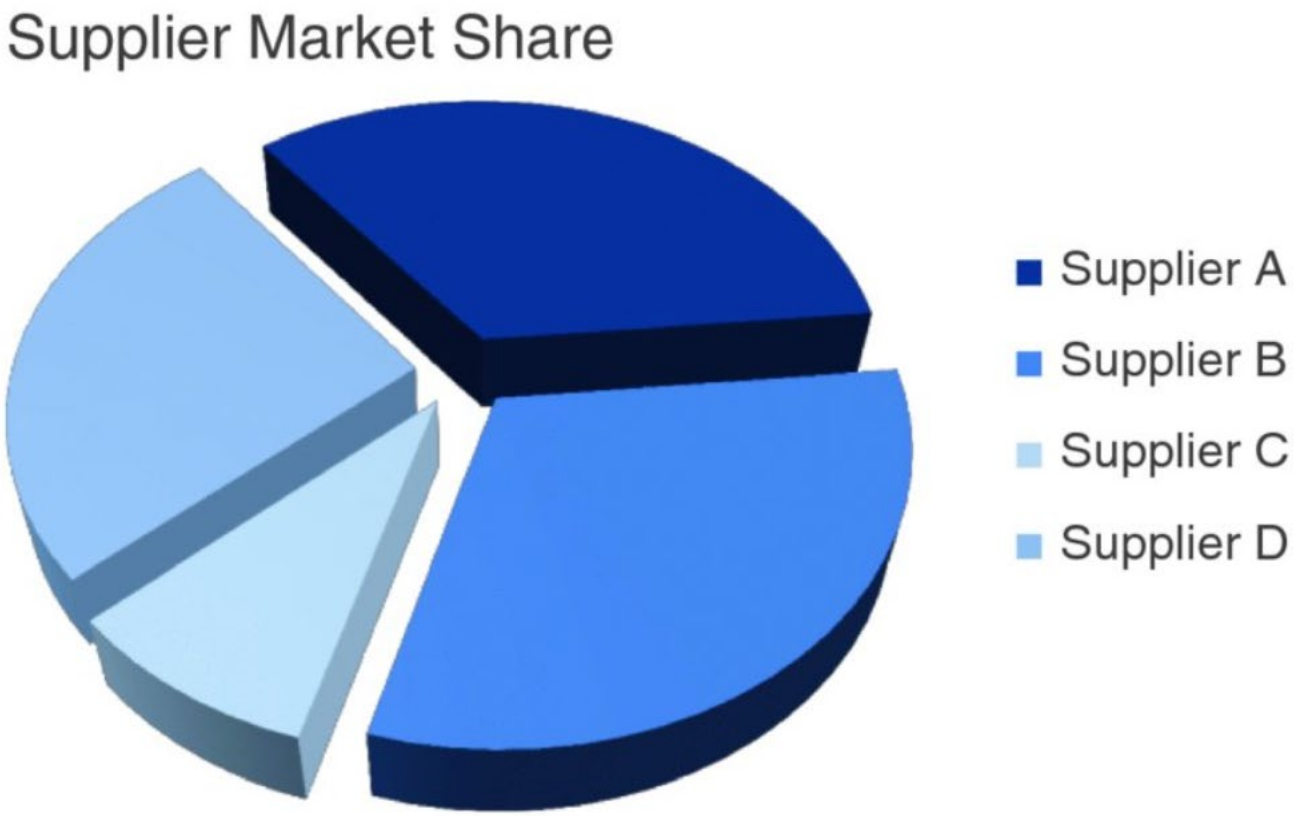
Non-zero baseline: as originally graphed



Zero baseline: as it should be graphed



Pie charts are evil

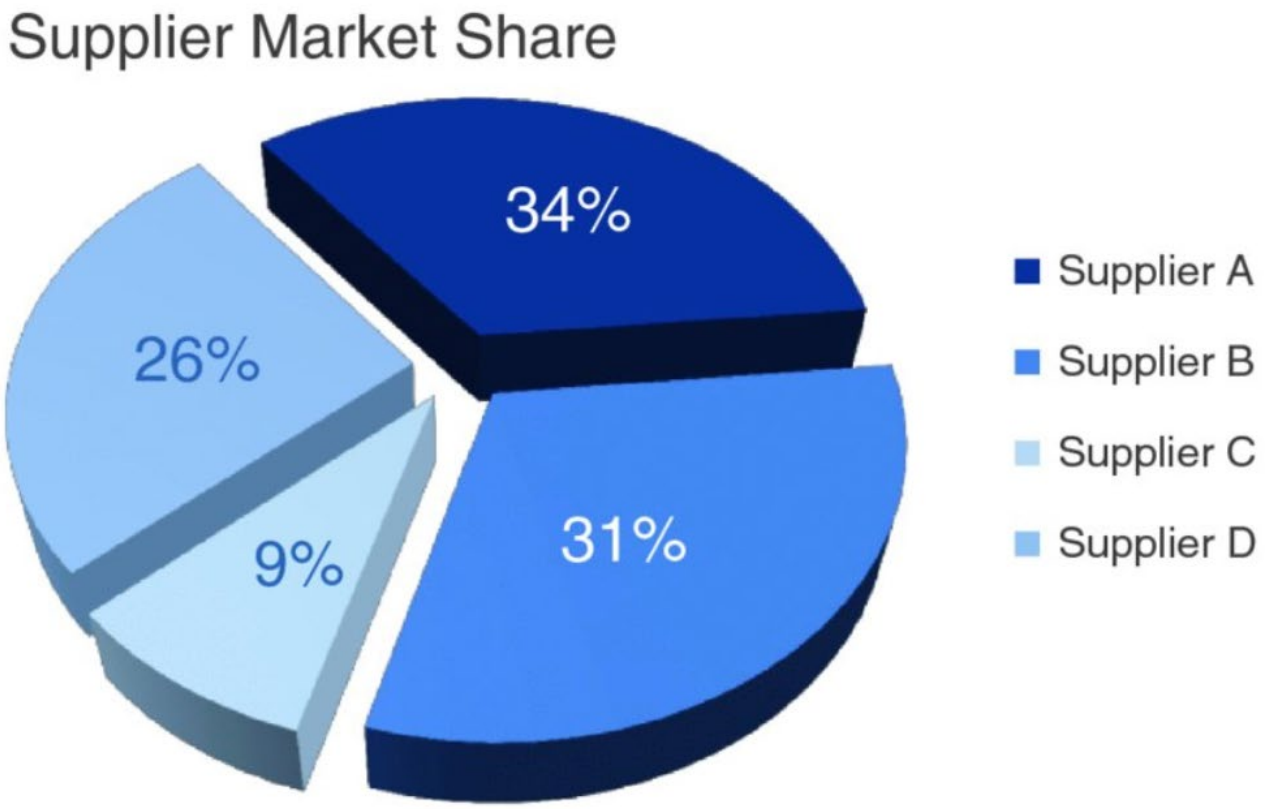


Which is largest?

If you had to estimate by what proportion, what percent would that be?

Knaflic, C. G. (2015). Storytelling with Data.

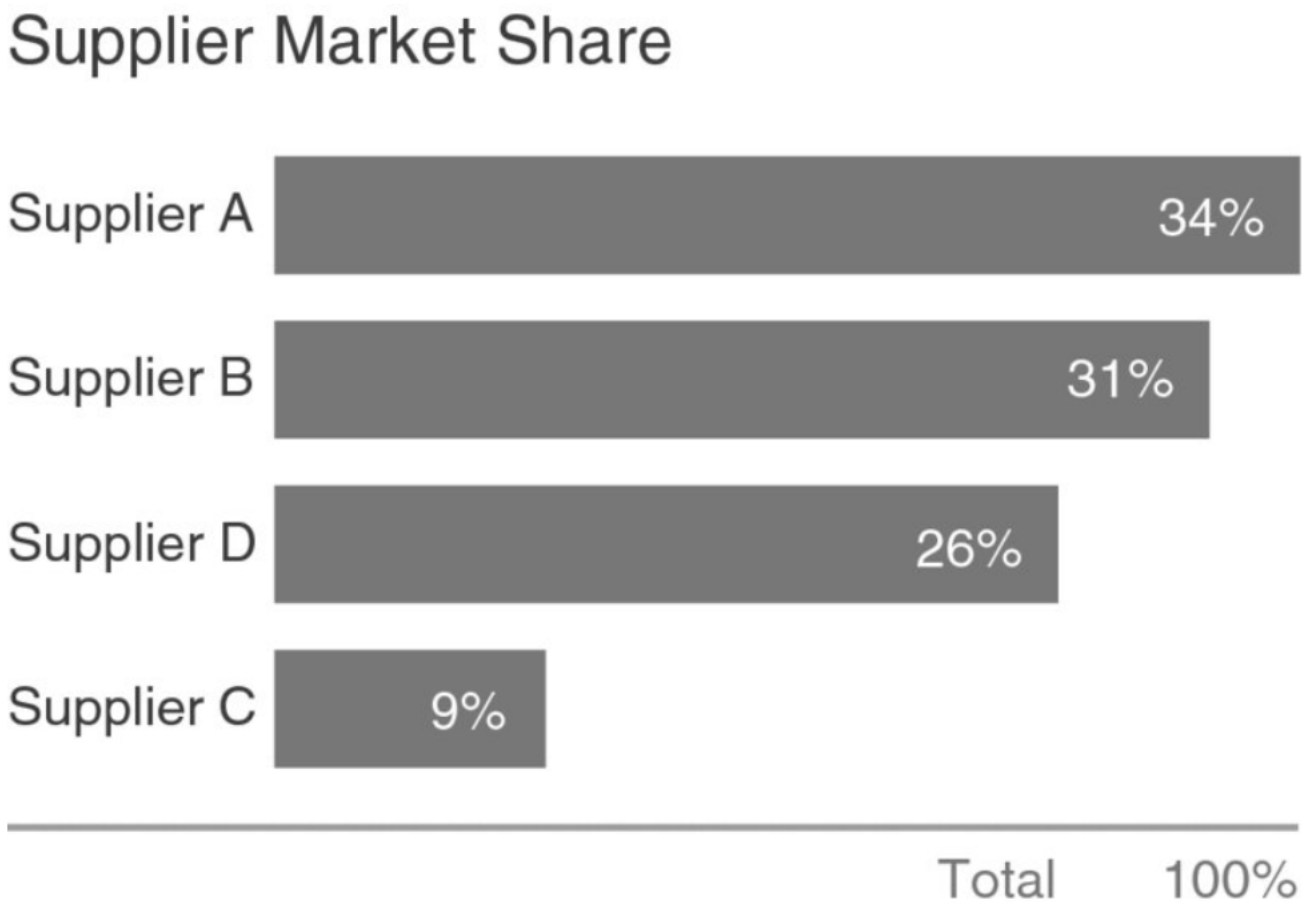
Pie charts are evil



Hard to read

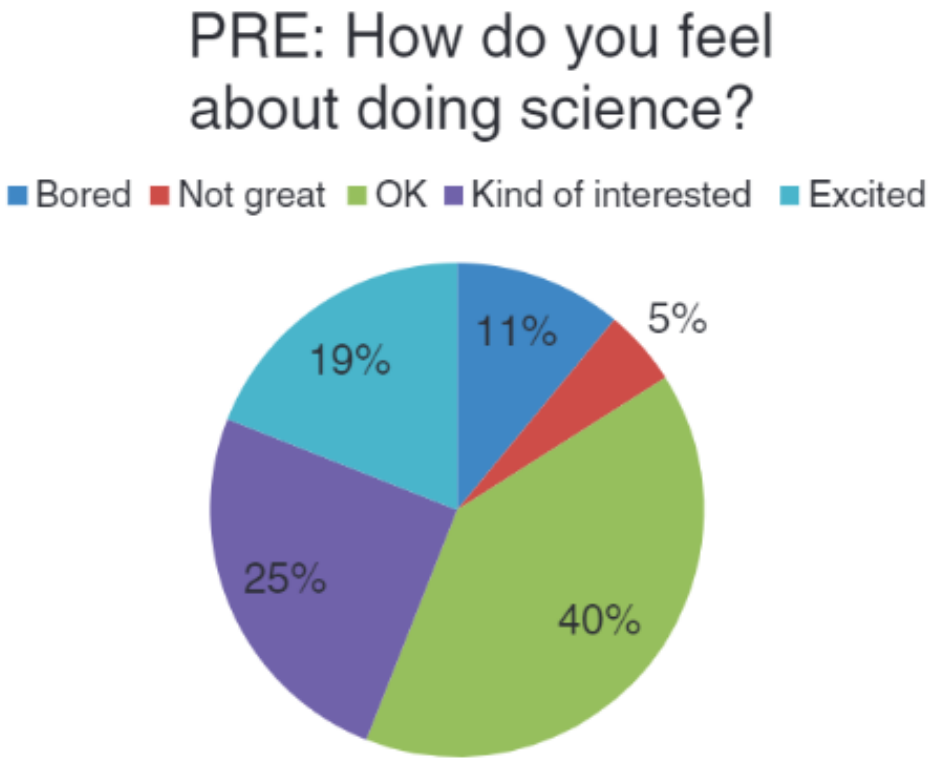
If you *really* want to use pie charts, add data labels

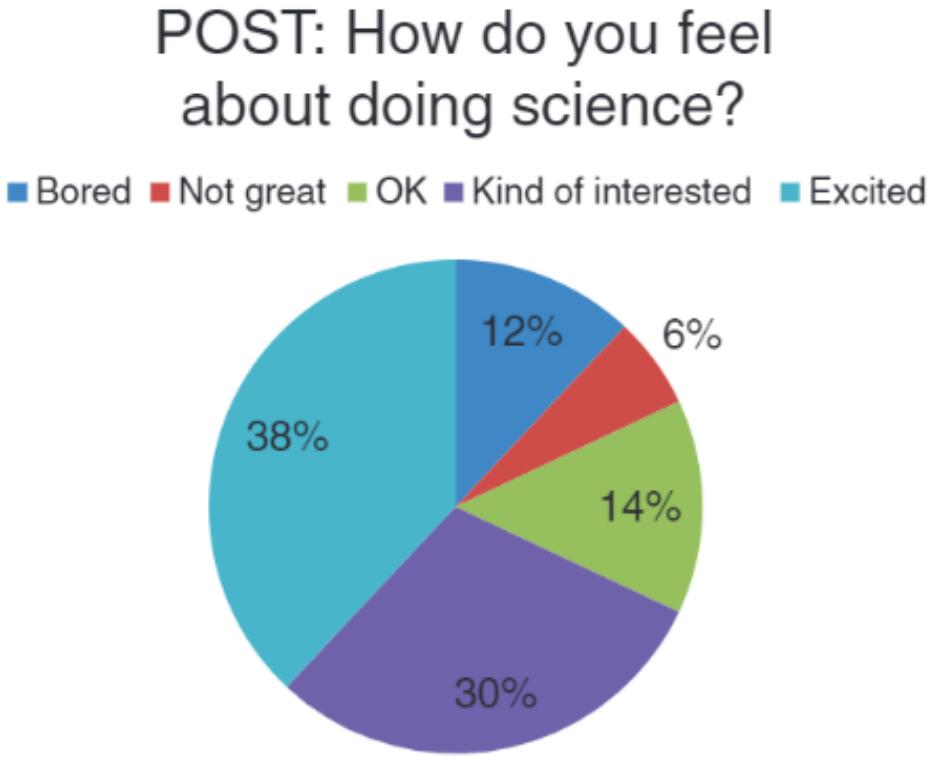
Use bar chart instead



Knaflic, C. G. (2015). Storytelling with Data.

Pie charts are evil

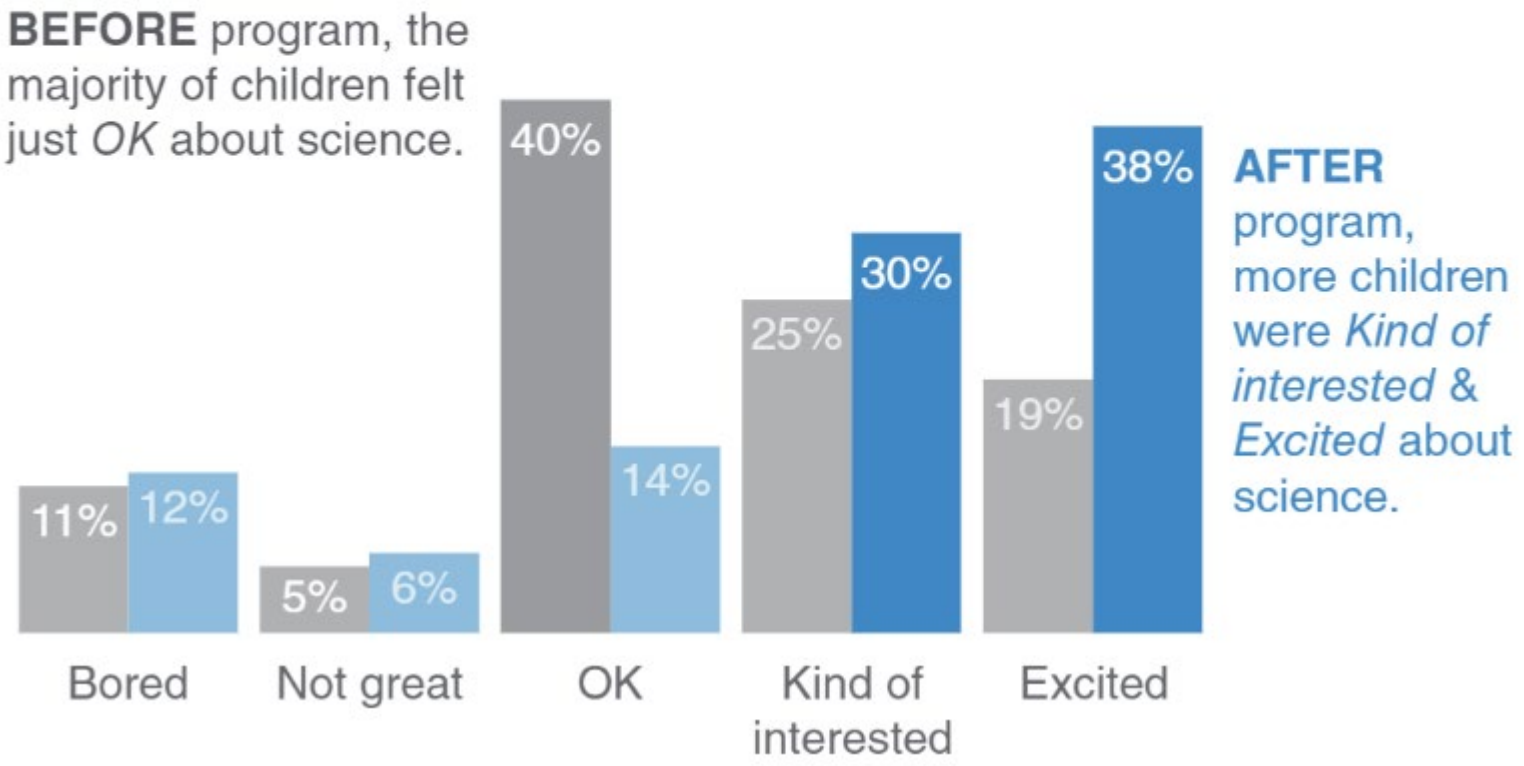




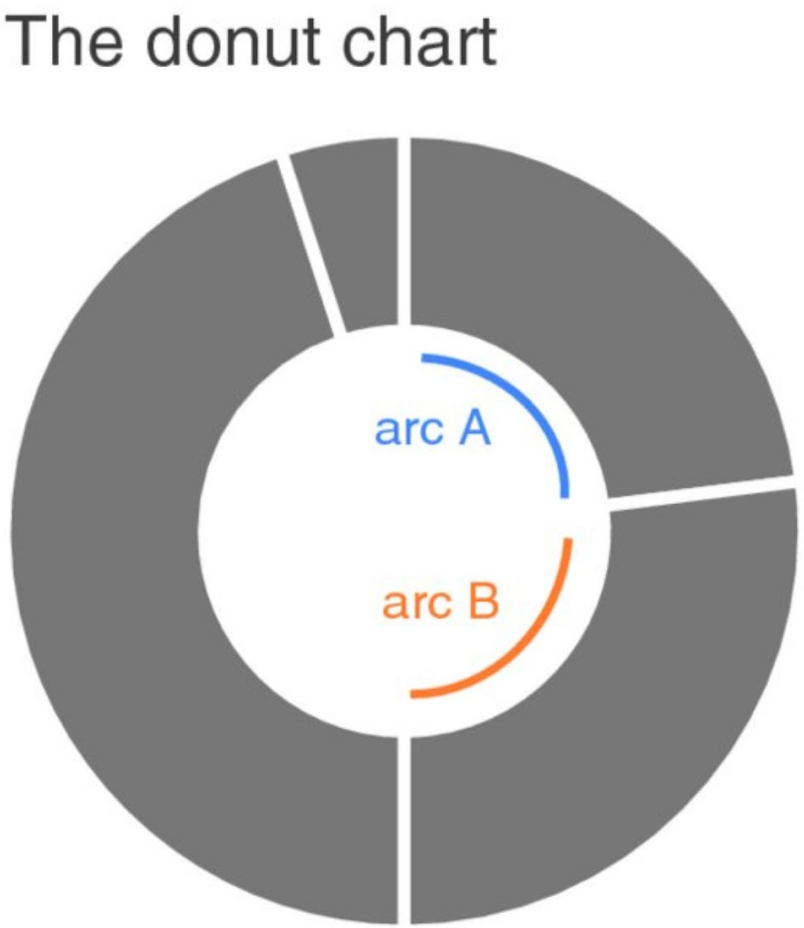
Knaflic, C. G. (2015). Storytelling with Data.

Pilot program was a success

How do you feel about science?



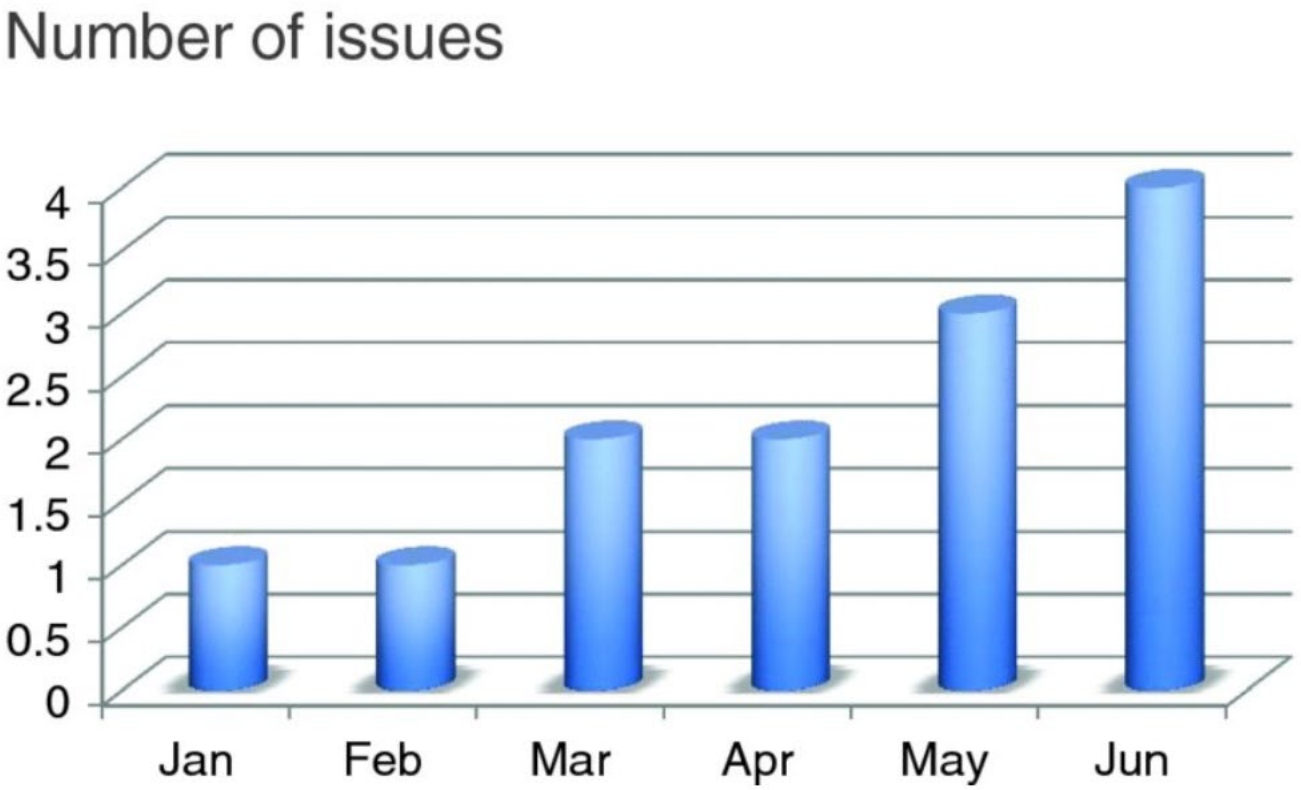
Don’t use the donut chart



Hard to estimate quantities

Knaflic, C. G. (2015). Storytelling with Data.

Never use 3D

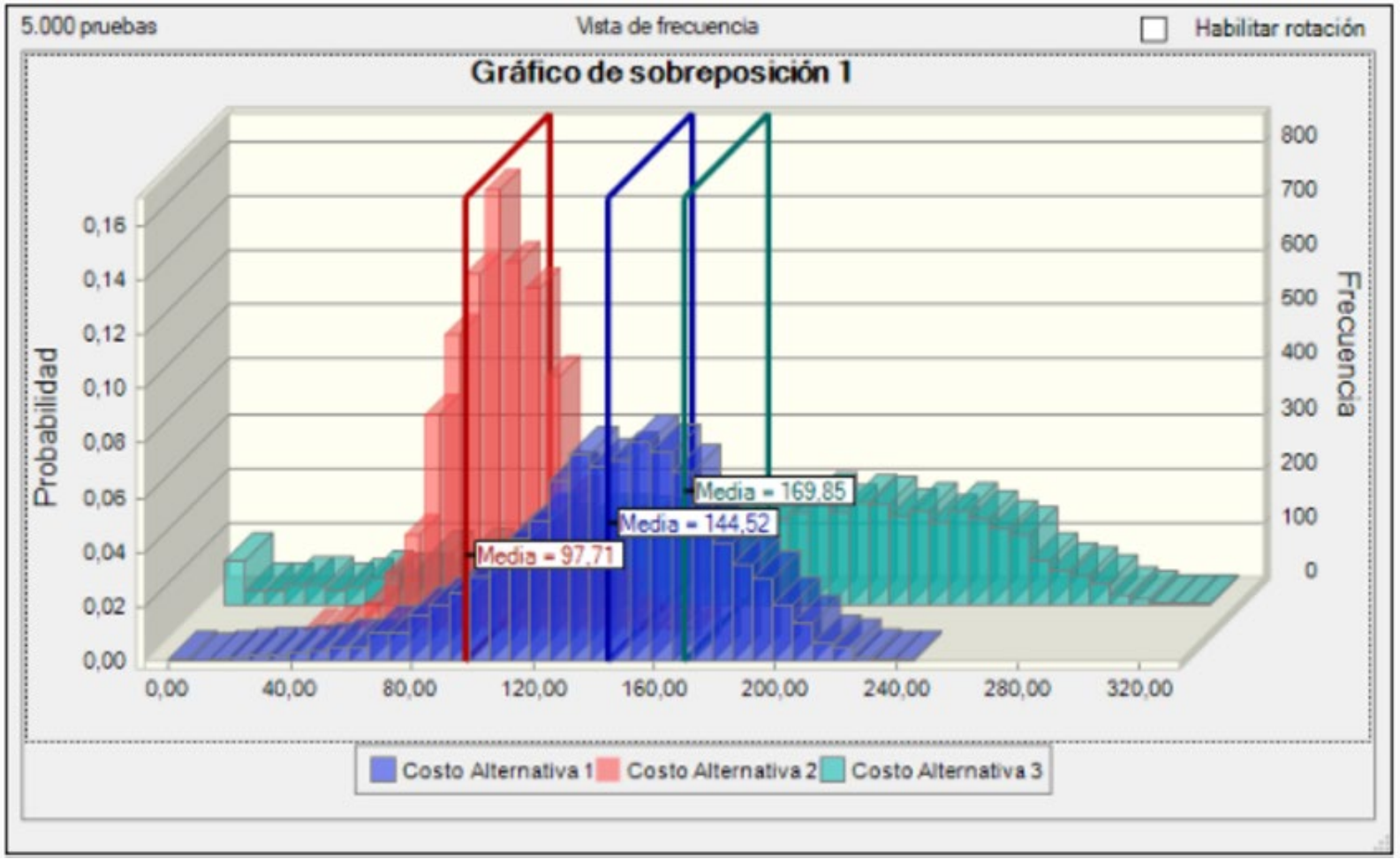


Plotting a third dimension?

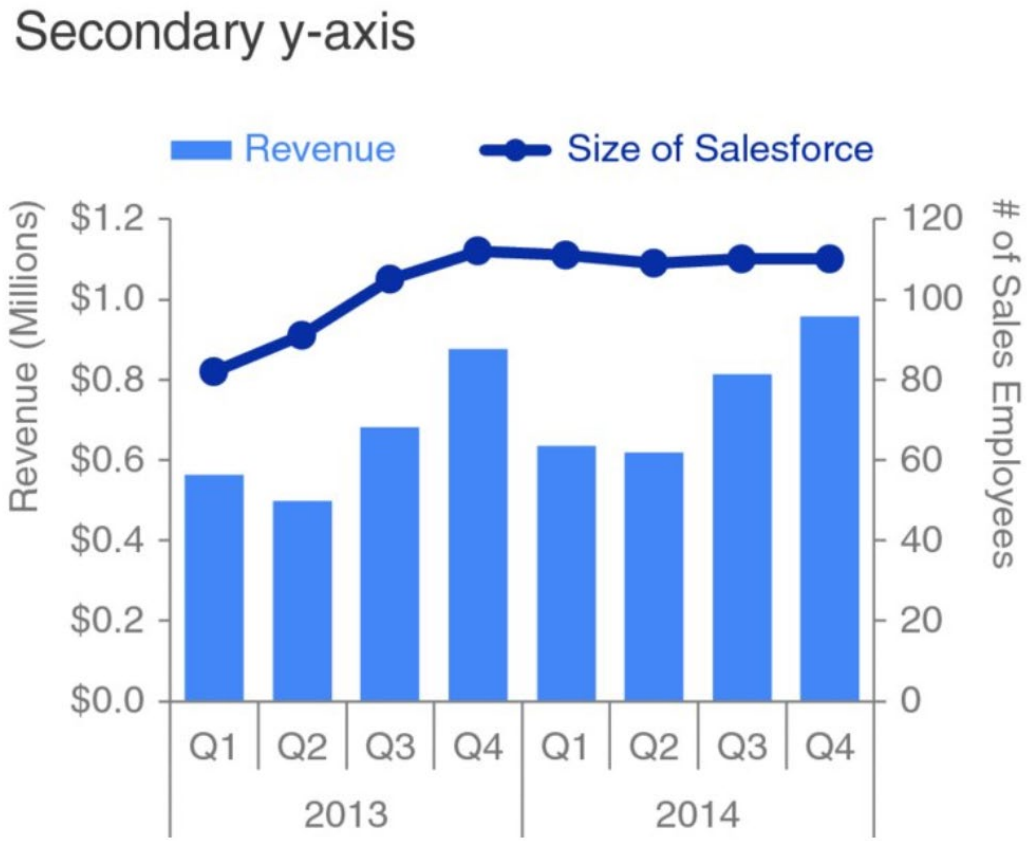
How many issues were in Jan and Feb?

"Never use 3D"

Unless …


5.000 pruebas
Vista de frecuencia
Habilitar rotación
Gráfico de sobreposición 1
Probabilidad
0,16
0,14
0,12
0,10
0,08
0,06
0,04
0,02
0,00
Frecuencia
800
700
600
500
400
300
200
100
0
Media = 169,85
Media = 144,52
Media = 97,71
0,00
40,00
80,00
120,00
160,00
200,00
240,00
280,00
320,00
Costo Alternativa 1
Costo Alternativa 2
Costo Alternativa 3

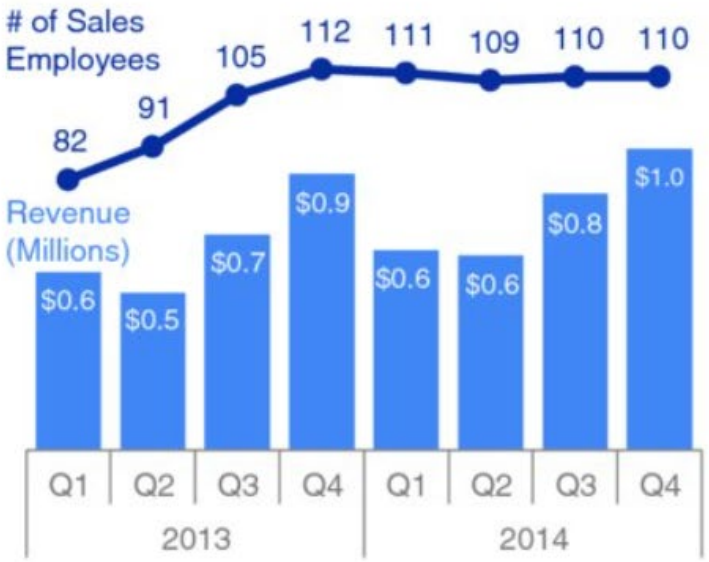
Don’t go for secondary y-axis



Knaflic, C. G. (2015). Storytelling with Data.

Alternatives

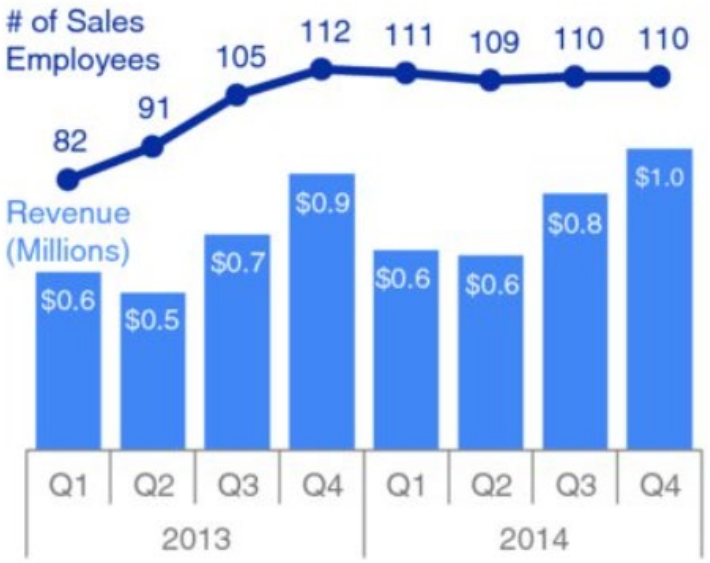
Alternative 1: **label directly**



Alternative 2: **pull apart vertically**

Compare and contrast

Alternative 1: **label directly**



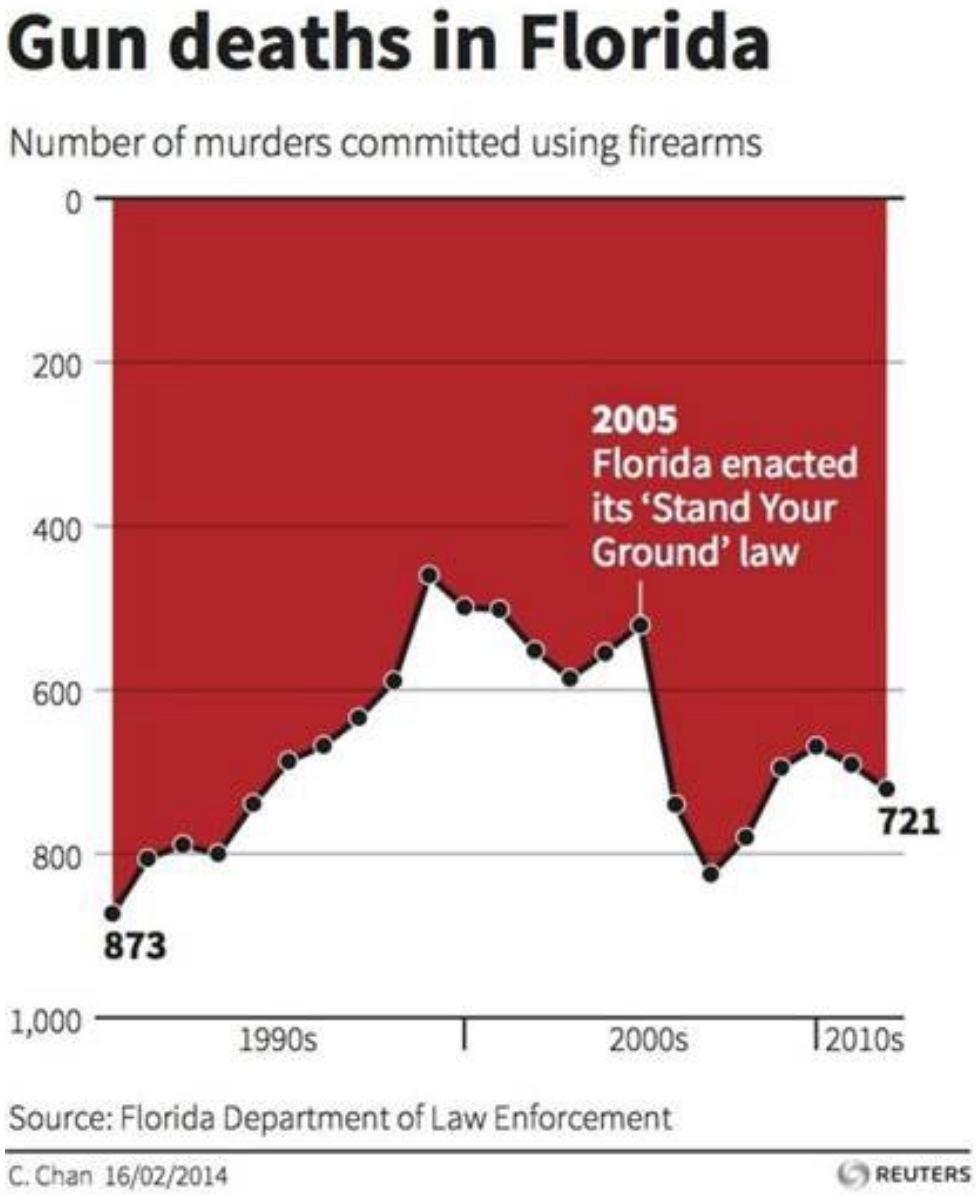
Secondary y-axis

Compare and contrast

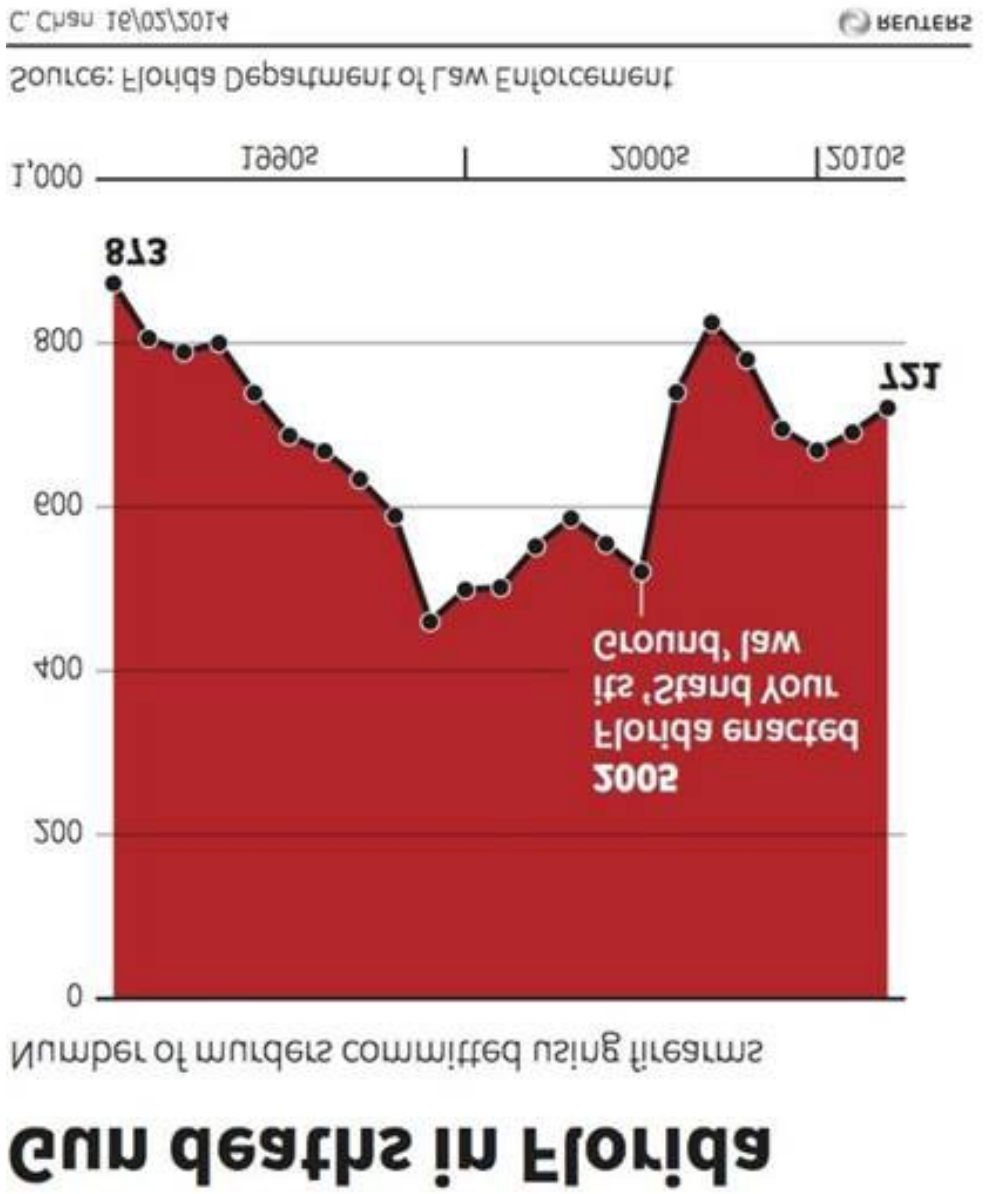
Alternative 2: **pull apart vertically**

Secondary y-axis

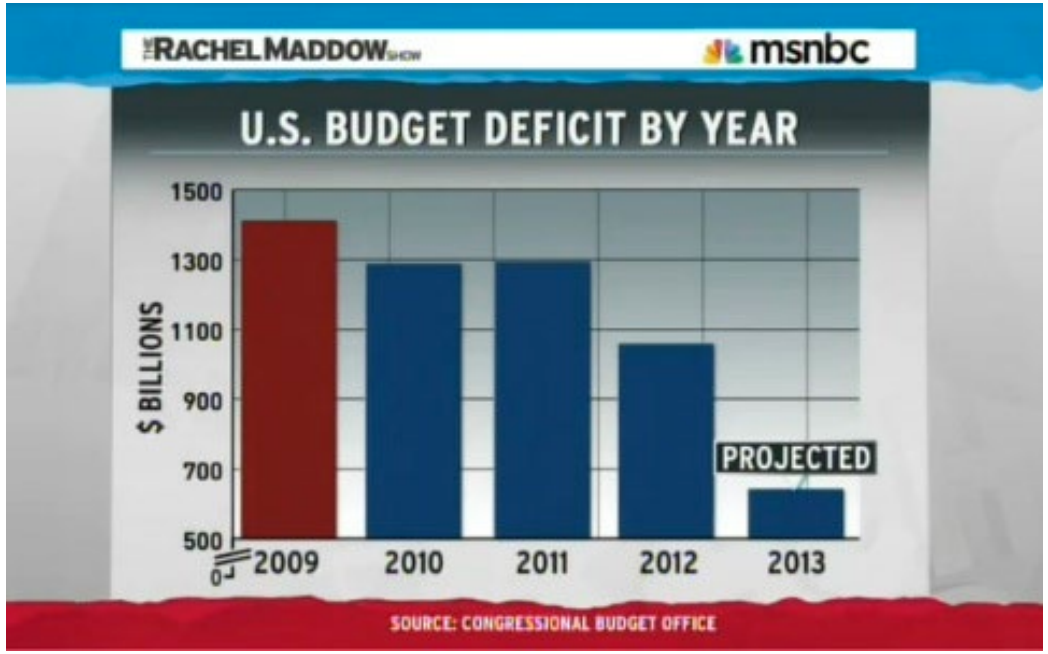
How do you feel?

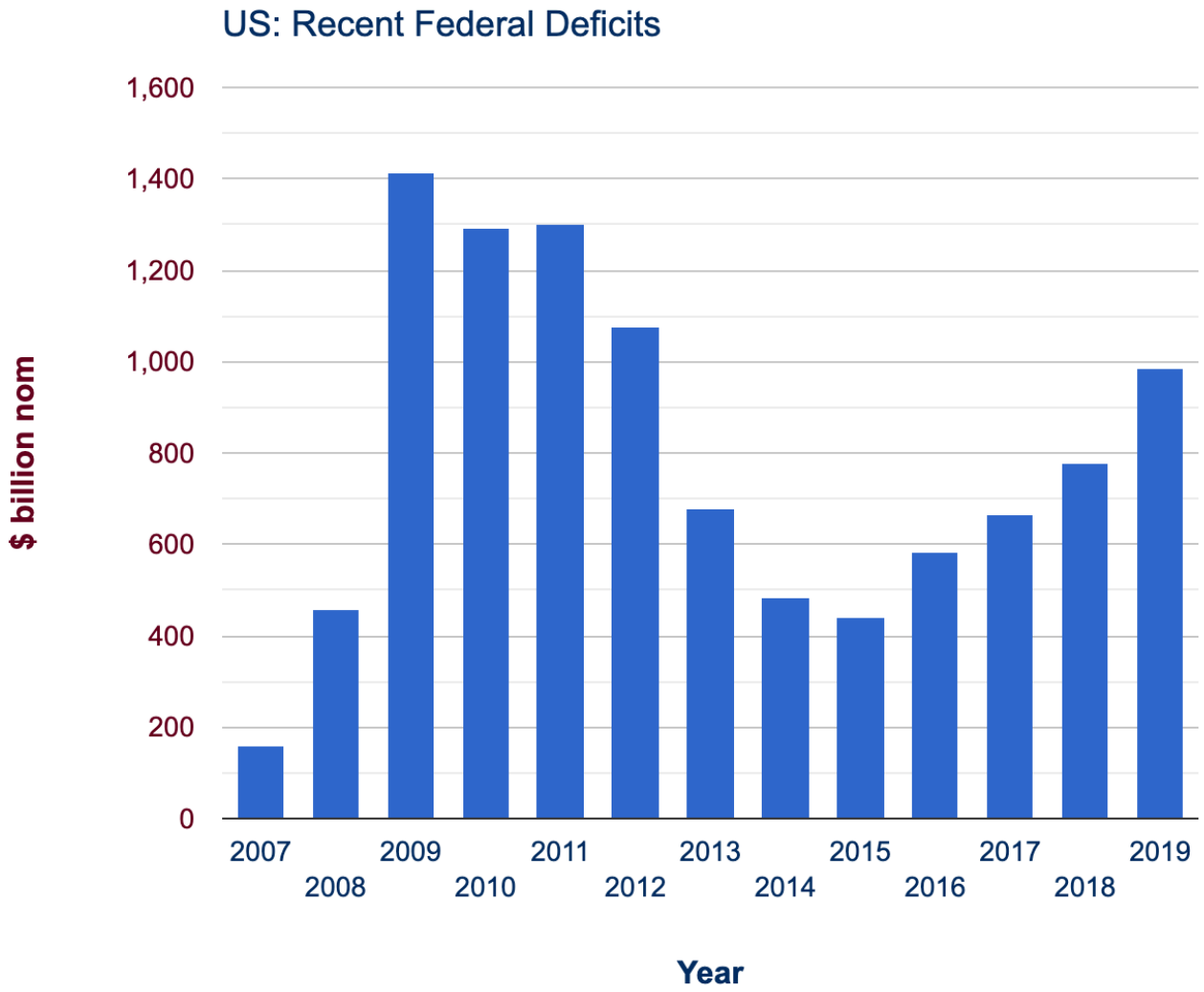


What about now?



How do you feel?





https://www.usgovernmentspending.com/federal_deficit_chart.html

“The greatest value of a picture is when it forces us to notice what we never expected to see”

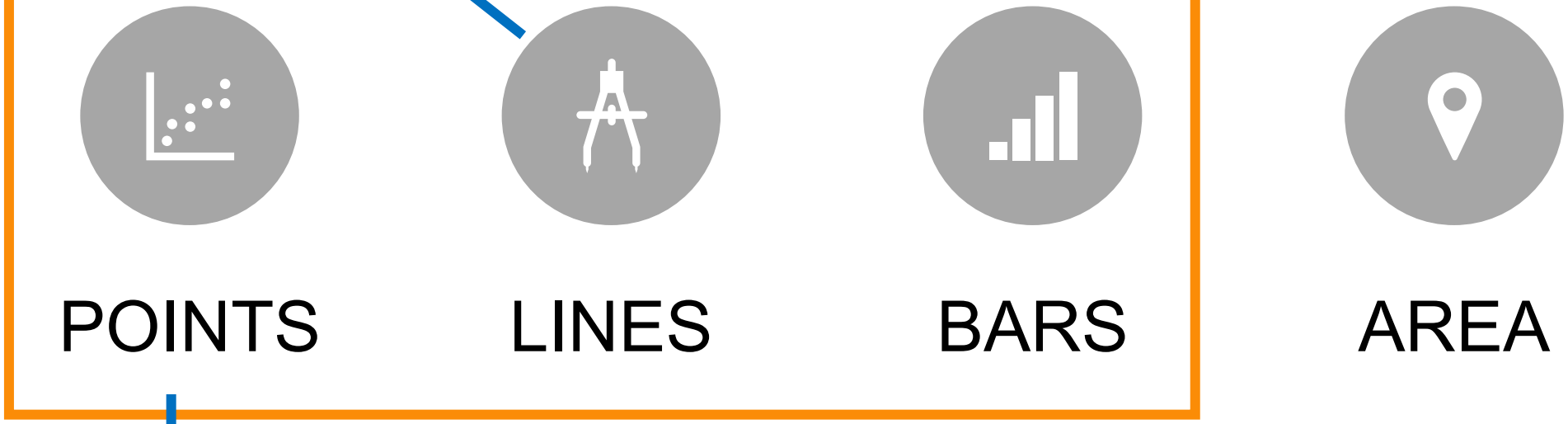
John Tukey

Choosing an effective visual – graphs

Most used to plot continuous data

Most popular

Easy for our eyes to read



Useful for showing relationship between two things

Knaflic, C. G. (2015). Storytelling with Data.

Wilkinson's grammar of graphics

- *Data*
  - A set of data operations that create variables from datasets
- *Trans*
  - Variable transformations
- *Scale*
  - Scale transformations
- *Coord*
  - *A* coordinate system
- *Element*
  - Graph and its aesthetic attributes
- *Guide*
  - One or more guides

ggplot2

- An implementation of the grammar of graphics in R
- The grammar describes the structure of a graphic
- A graphic is a mapping of data to a visual representation
- ggplot2

ggplot2

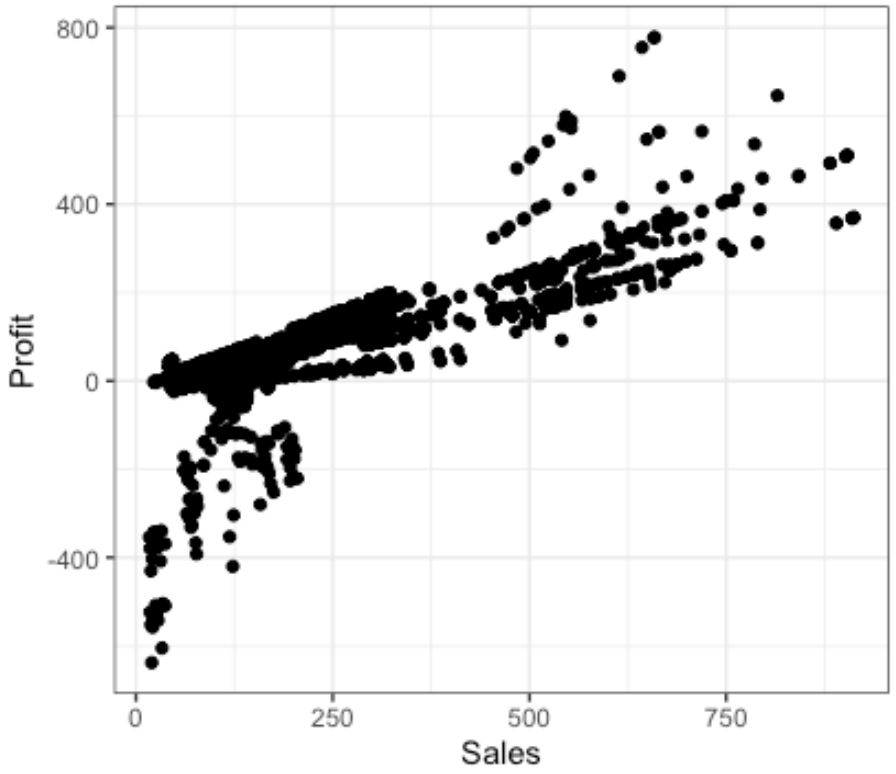
- One of the packages in the tidyverse.
- Enables the “static” visualization of data.

```
library(tidyverse) # load package
library(readxl)

CoffeeChain <- read_excel("CoffeeChain.xlsx")

ggplot(data = CoffeeChain, aes(x = Sales, y = Profit)) +
  geom_point() +
  theme_bw()
```

Points – scatterplot in R

Useful for showing relationship between two things

Is there a positive relationship between sales and profit?

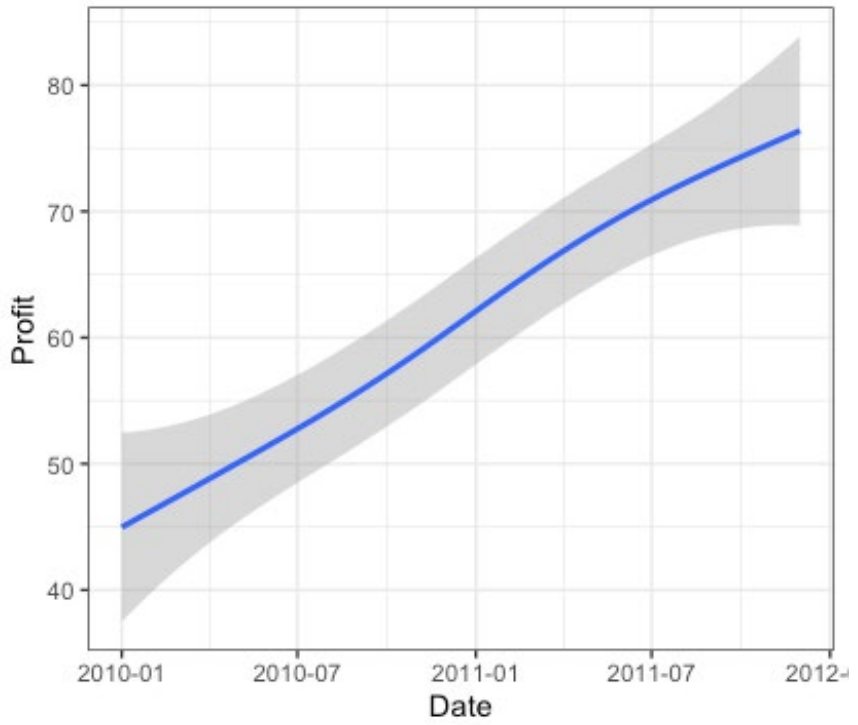


Lines in R

Most used to plot continuous data

Has profit changed over time?

```
ggplot(CoffeeChain, aes(x = Date, y = Profit)) +
  geom_smooth() + theme_bw()
```

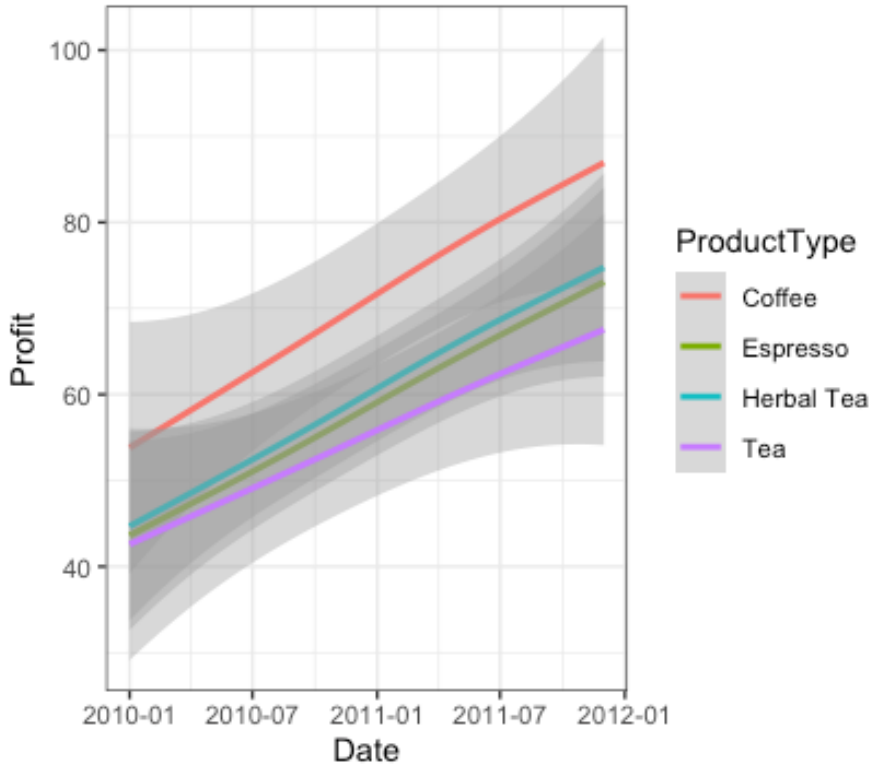


Lines in R

Most used to plot continuous data

Has profit changed over time?

```
ggplot(CoffeeChain, aes(x = Date, y = Profit, color = ProductType)) +
  geom_smooth() + theme_bw()
```

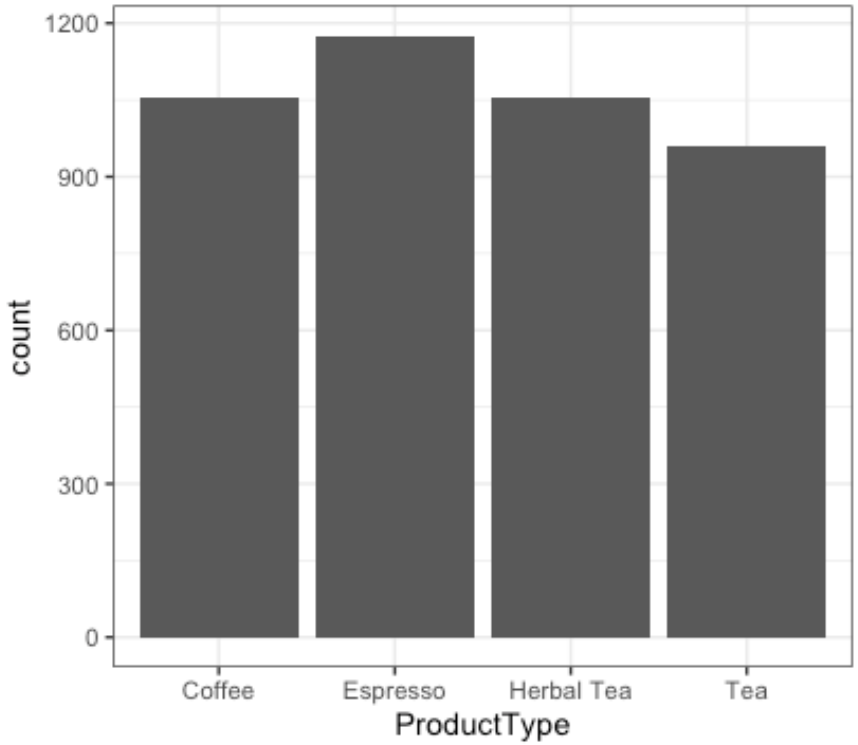


Bars in R

Easy for our eyes to read

How many products do we have for each type of product?

```
ggplot(CoffeeChain, aes(x =
ProductType)) +
  geom_bar() + theme_bw()
```

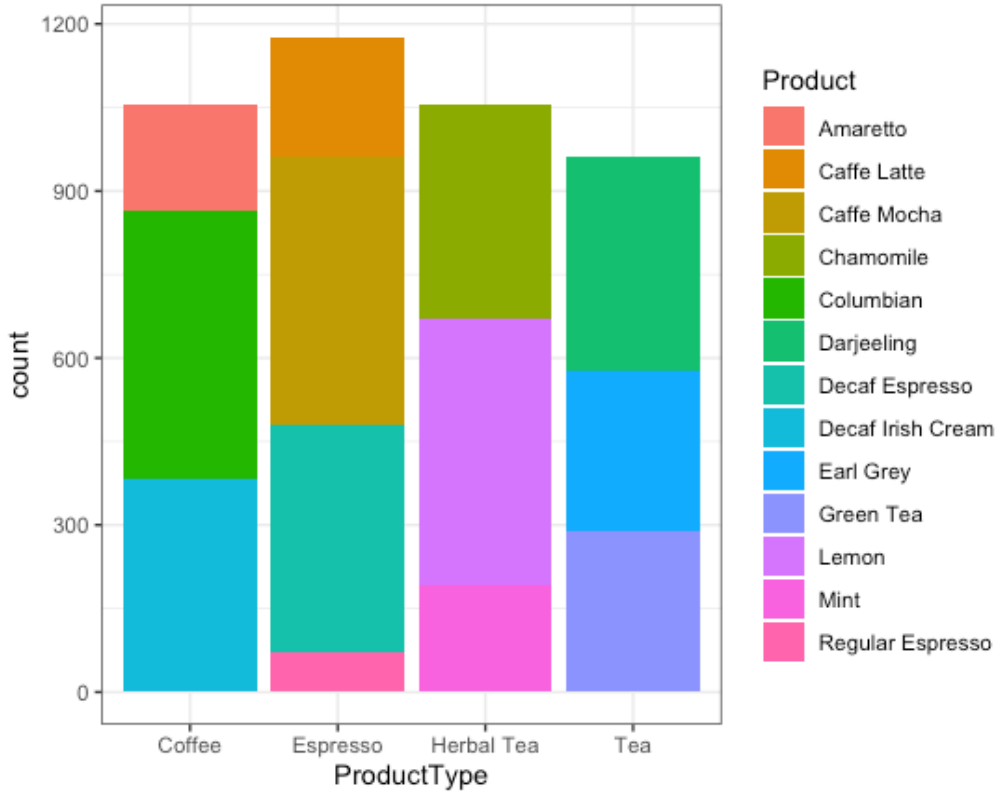


Bars in R

Easy for our eyes to read

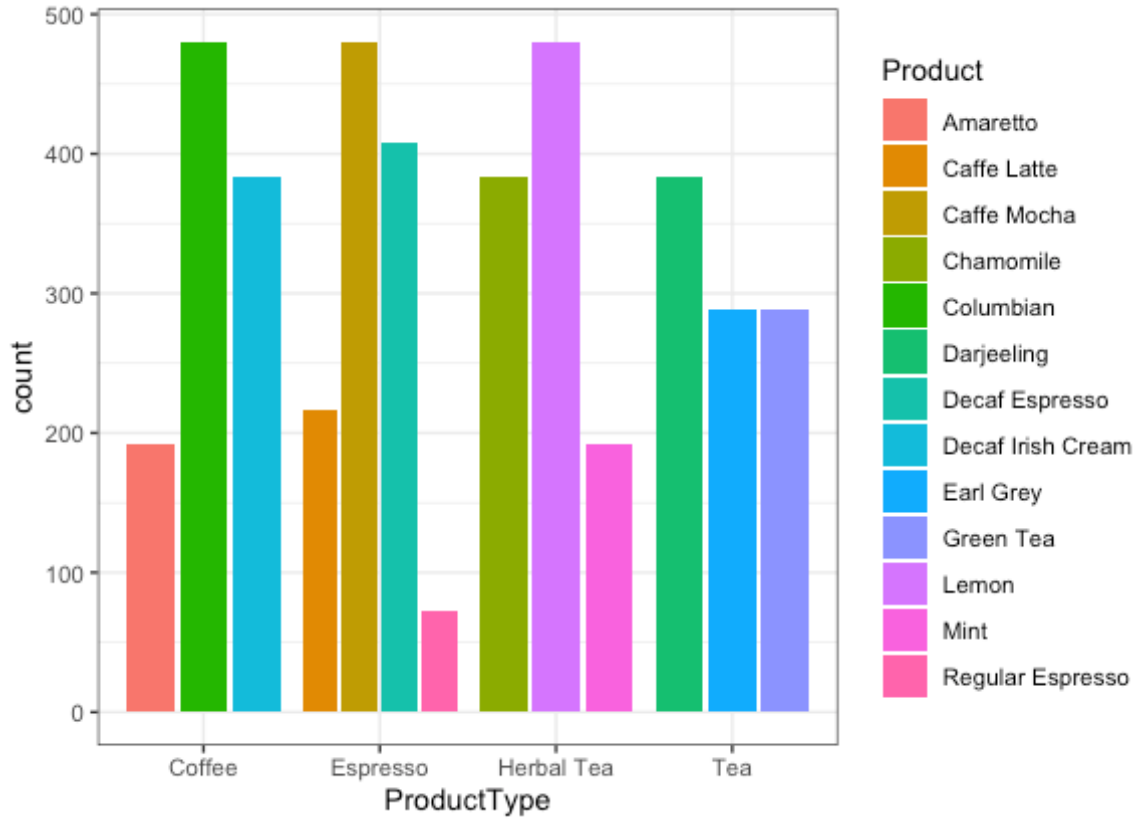
What products and how many of them do we have for every type of product?

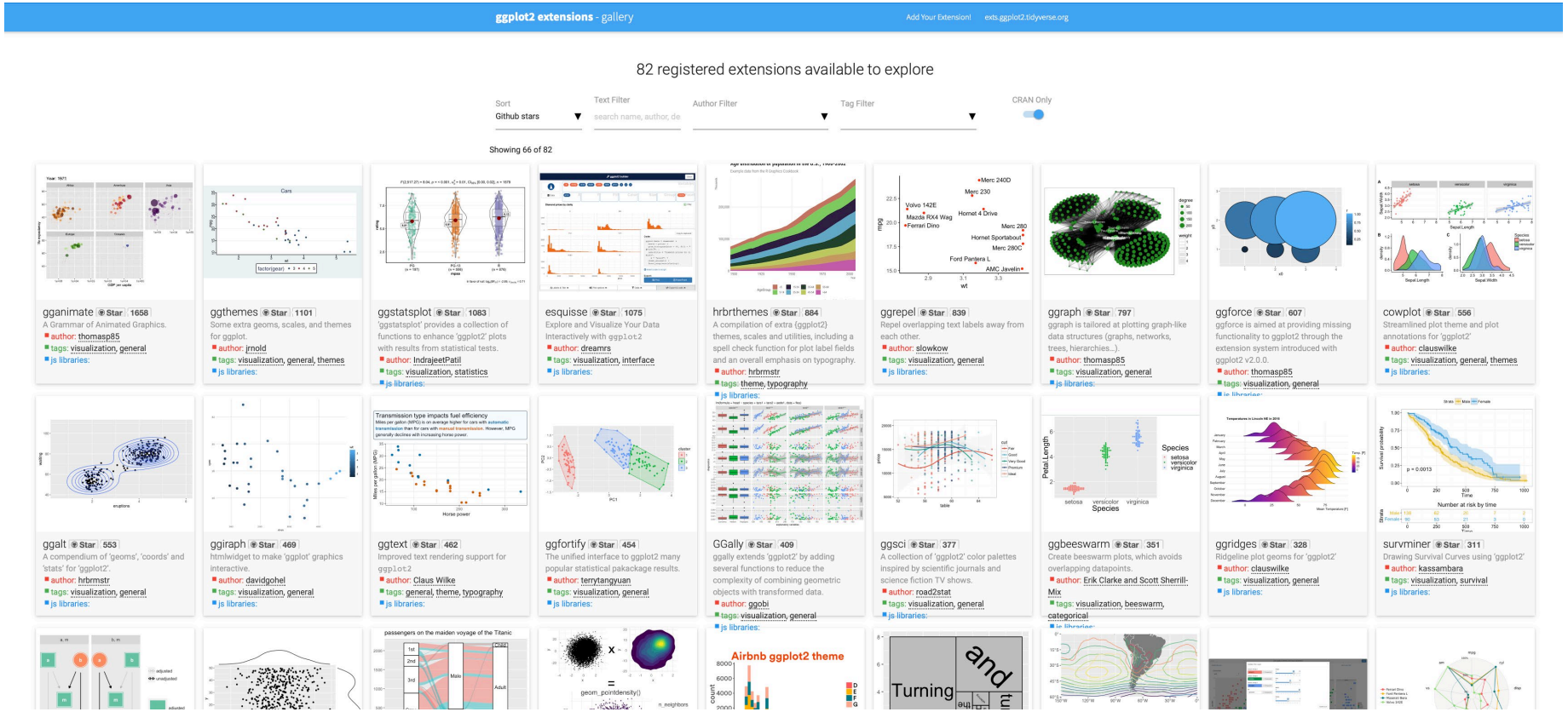
```
ggplot(CoffeeChain, aes(x =
ProductType, fill = Product)) +
  geom_bar() + theme_bw()
```



Challenge

- Stack bars next to each other
- See: https://ggplot2.tidyverse.org/reference/position_dodge.html



A couple of notes in ggplot2

Many extensions available – the sky is the limit!