

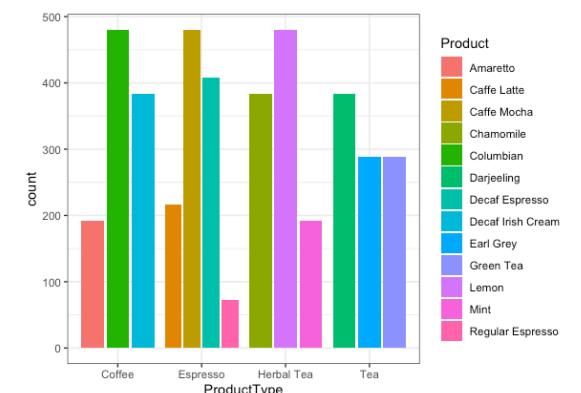
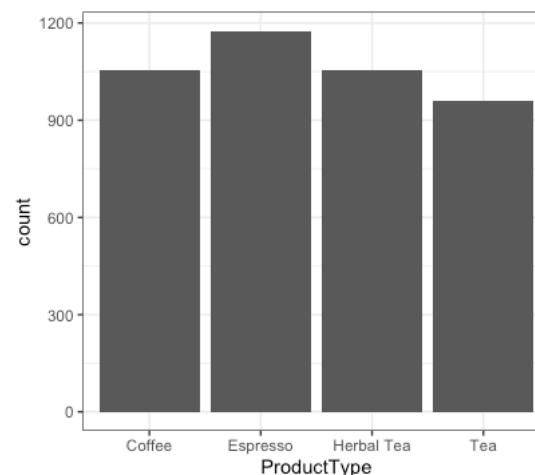
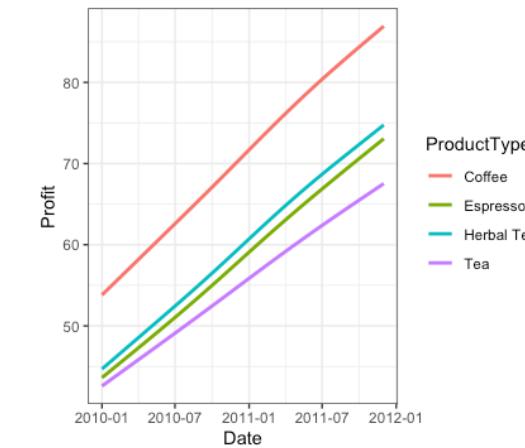
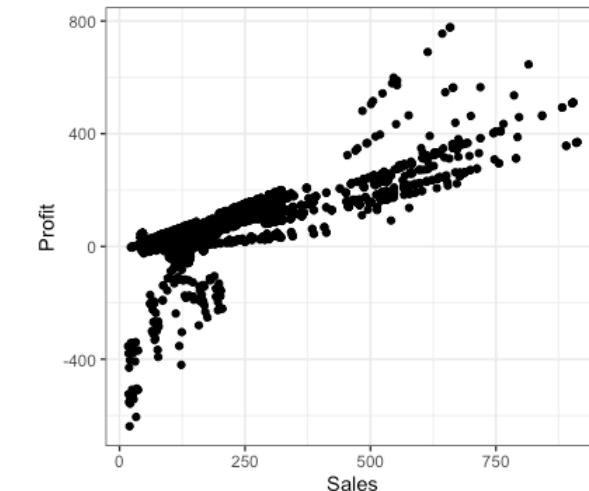
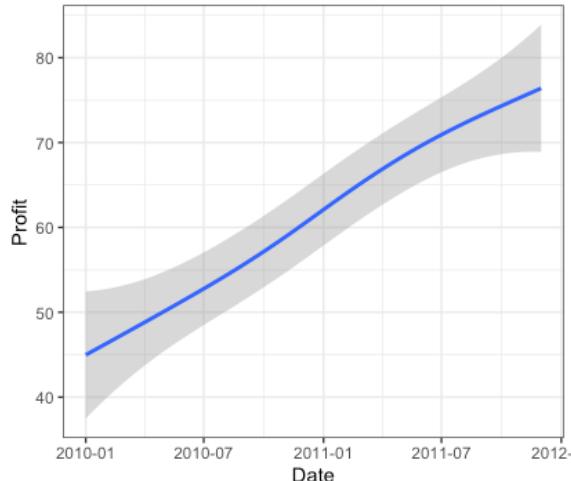
Visualization I

Carolina A. de Lima Salge
Assistant Professor
Terry College of Business
University of Georgia

Business Intelligence
Spring 2021



Terry College of Business
UNIVERSITY OF GEORGIA



Two Notes on Transformation

Sometimes you want to filter observations on a list of values, and not only on a single specific value

```
library(tidyverse)
orderDetails <- read_csv("Order Details.csv")

orderDetails %>%
  filter(., Category %in% c("Furniture", "Clothing"))
# A tibble: 1,192 x 6
`Order ID` Amount Profit Quantity Category `Sub-
Category`<chr>    <dbl>  <dbl>    <dbl> <chr>    <chr>
1 B-25601      1275 -1148     7 Furniture Bookcases
2 B-25601        66   -12      5 Clothing  Stole
3 B-25601        8    -2      3 Clothing  Hankerchief
4 B-25602      561    212      3 Clothing  Saree
5 B-25602      119     -5      8 Clothing  Saree
6 B-25603     1355    -60      5 Clothing  Trousers
7 B-25603       24    -30      1 Furniture Chairs
8 B-25603      193   -166      3 Clothing  Saree
9 B-25603      180      5      3 Clothing  Trousers
10 B-25603     116     16      4 Clothing  Stole
# ... with 1,182 more rows
```



Two Notes on Transformation

Sometimes you want to create a new variable based on some conditional criteria of another variable

```
orderDetails %>%
  mutate(., ProfitClass = case_when(
    Profit > 0 ~ "Positive",
    Profit < 0 ~ "Negative",
    Profit == 0 ~ "Zero"))
# A tibble: 1,500 x 7
`Order ID`  Amount Profit Quantity Category `Sub-Category` ProfitClass
<chr>       <dbl>   <dbl>   <dbl> <chr>   <chr>          <chr>
1 B-25601     1275   -1148      7 Furniture Bookcases Negative
2 B-25601       66     -12       5 Clothing  Stole      Negative
3 B-25601        8      -2       3 Clothing  Hankerchief Negative
4 B-25601       80     -56       4 Electronics Electronic Games Negative
5 B-25602     168     -111      2 Electronics Phones      Negative
6 B-25602     424     -272      5 Electronics Phones      Negative
7 B-25602    2617     1151      4 Electronics Phones      Positive
8 B-25602     561      212      3 Clothing   Saree      Positive
9 B-25602     119      -5       8 Clothing   Saree      Negative
10 B-25603    1355     -60       5 Clothing  Trousers      Negative
# ... with 1,490 more rows
```



“The simple graph has brought more information to the data analyst’s mind than any other device”

John Tukey



What Makes a Good Visual?

A good visual easily conveys information – low cognitive effort is required to see the data

- Oxygen at Google – What makes a good manager?
- Need to communicate findings to various audiences
 - From engineers who are often skeptical on methodology and want to dig into details, to managers wanting to understand the big-picture findings and how to put them to use
- **Communication, communication, communication!**
- How to best show very complicated stuff in a way that appeases everyone?



A Few Lessons from Knafllic (2015)

1

Understand the context

2

Choose an appropriate visual display

3

Eliminate clutter

4

Focus attention where you want it

5

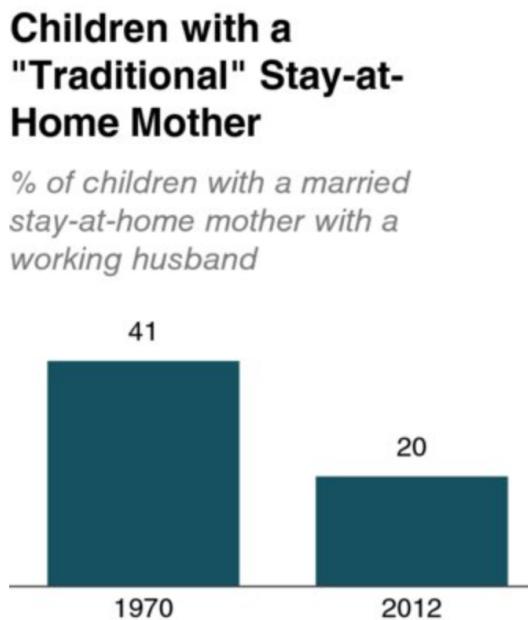
Think like a designer

6

Tell a story



Choosing an effective visual – simple text



Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013

Adapted from PEW RESEARCH CENTER

20%

of children had a
traditional stay-at-home mom
in 2012, compared to 41% in 1970



Choosing an effective visual – tables & heatmaps

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Minimal borders

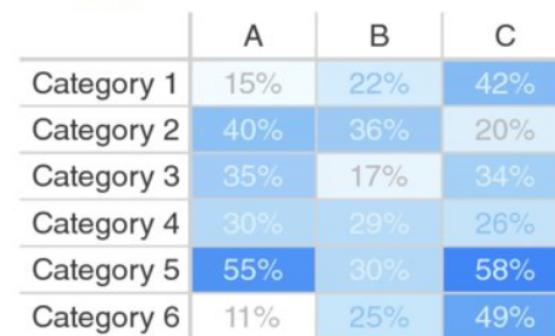
Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Table

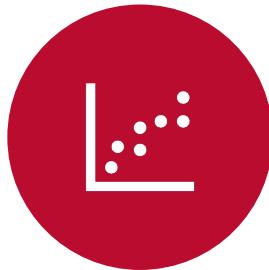
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

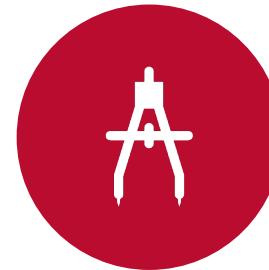
LOW-HIGH



Choosing an effective visual – graphs



POINTS



LINES



BARS

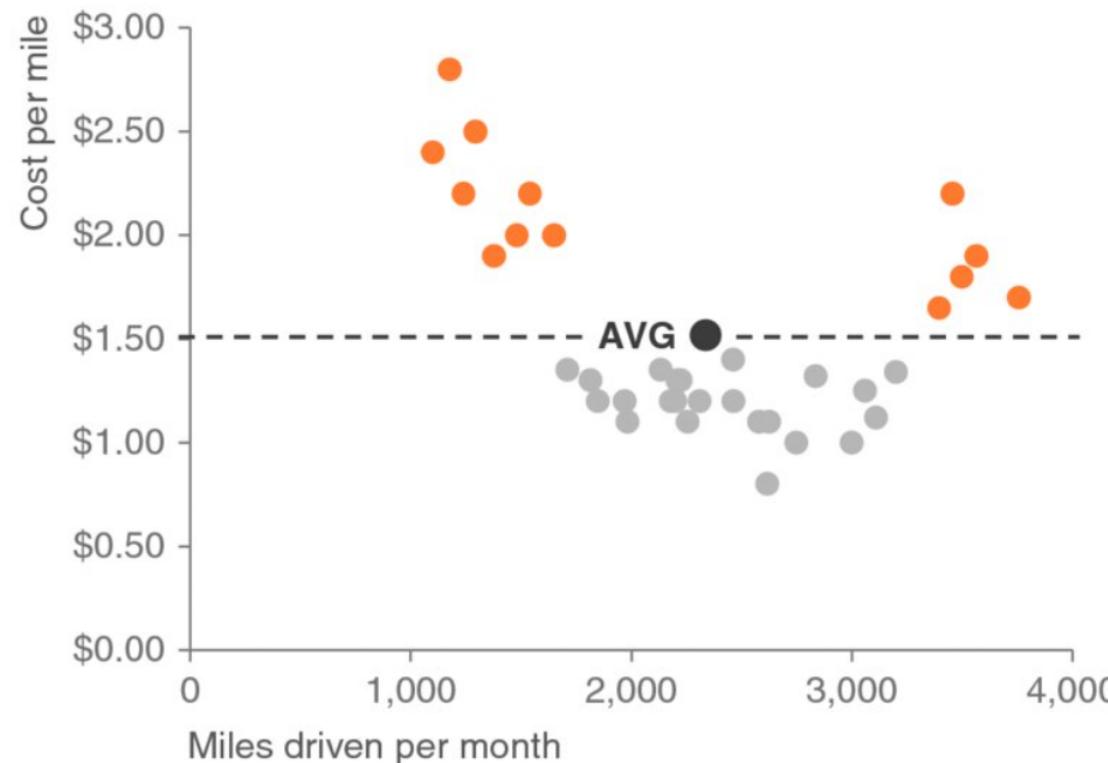


AREA

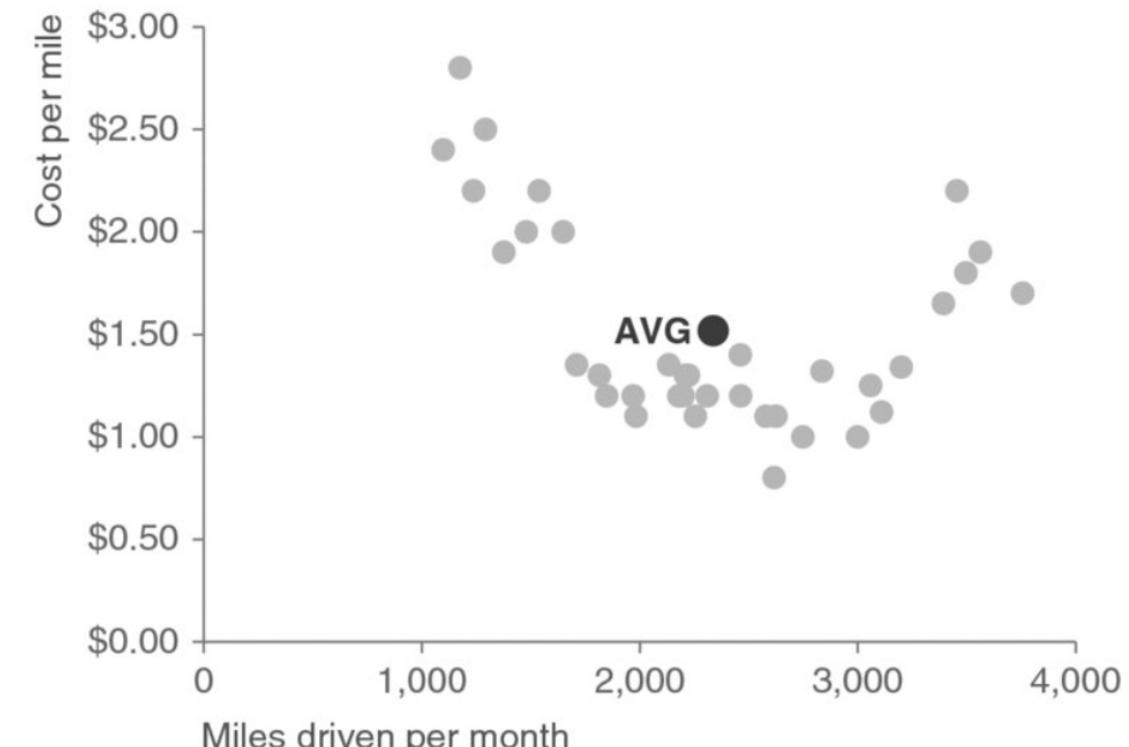
Points - scatterplot

Useful for showing relationship between two things

Cost per mile by miles driven



Cost per mile by miles driven

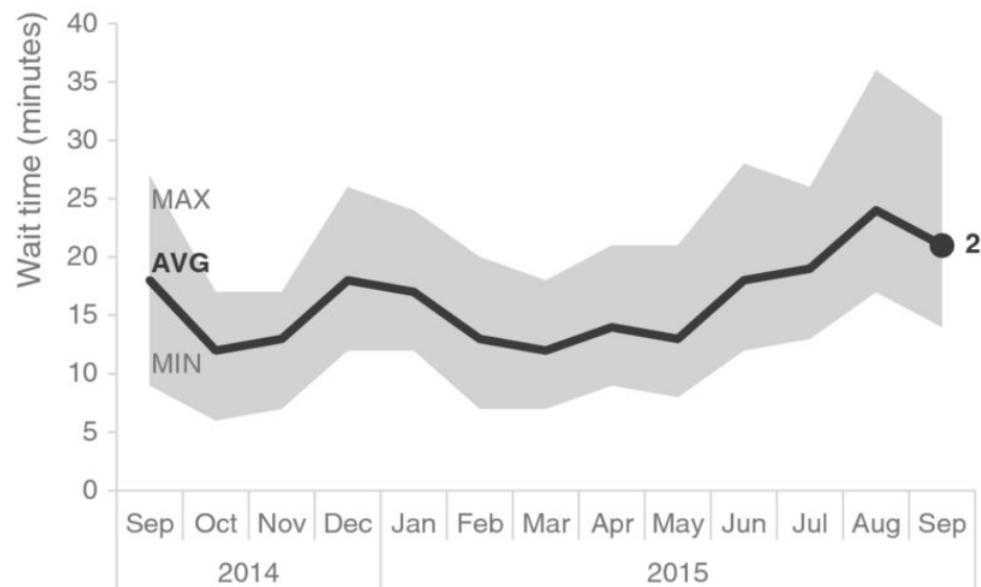


Lines

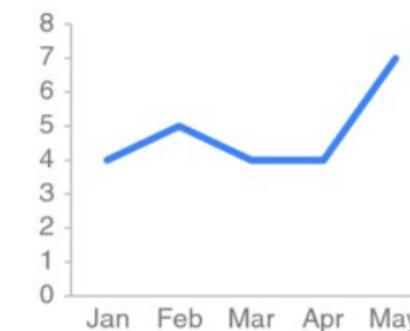
Most used to plot continuous data

Passport control wait time

Past 13 months



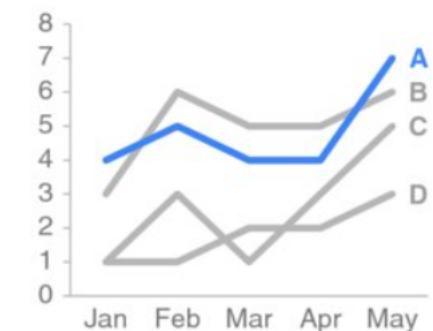
Single series



Two series



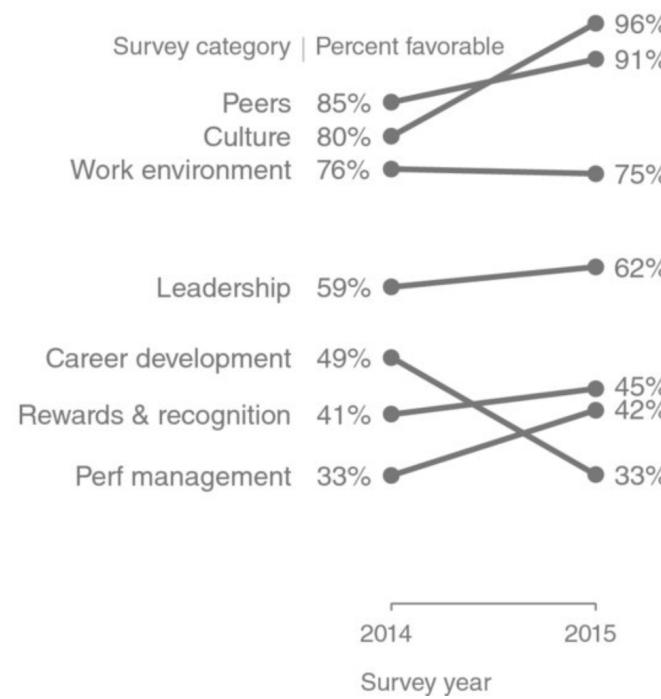
Multiple series



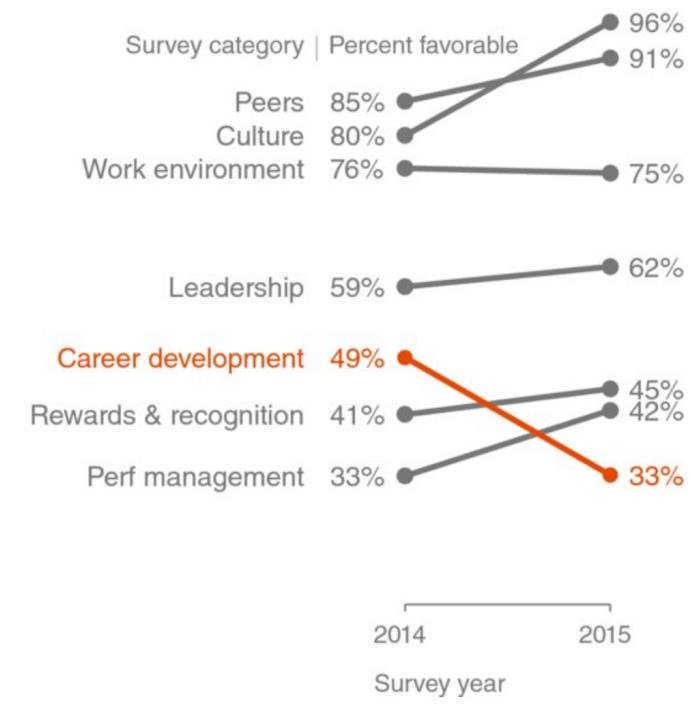
Slopegraph

Useful when you have two time periods or points of comparison and want to quickly show relative increases or decreases

Employee feedback over time

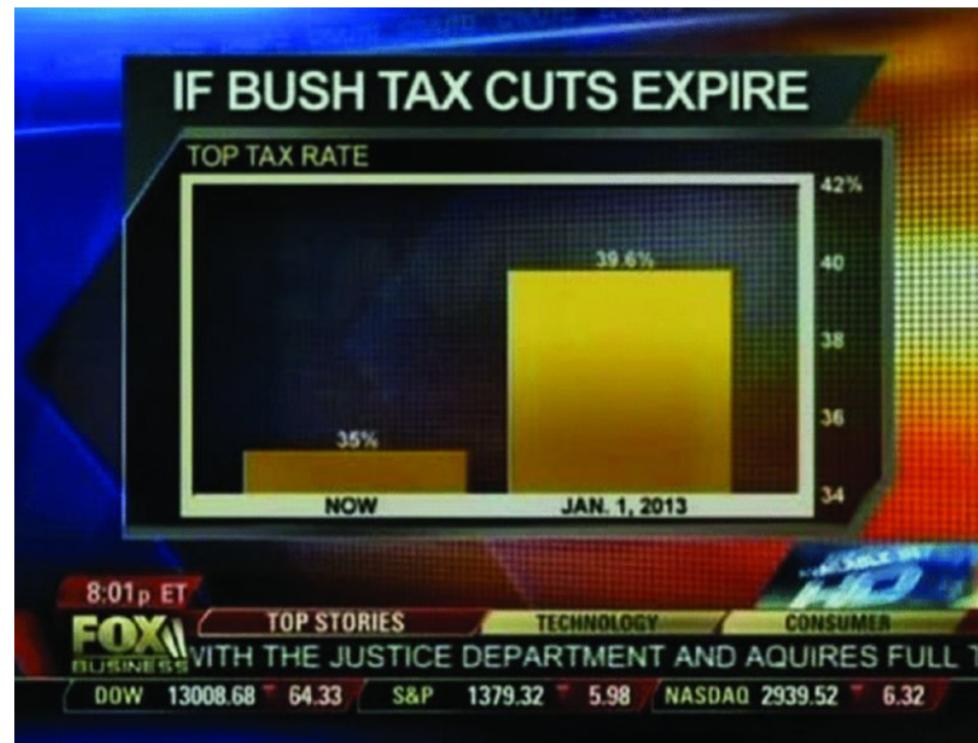


Employee feedback over time



Bars

Easy for our eyes to read

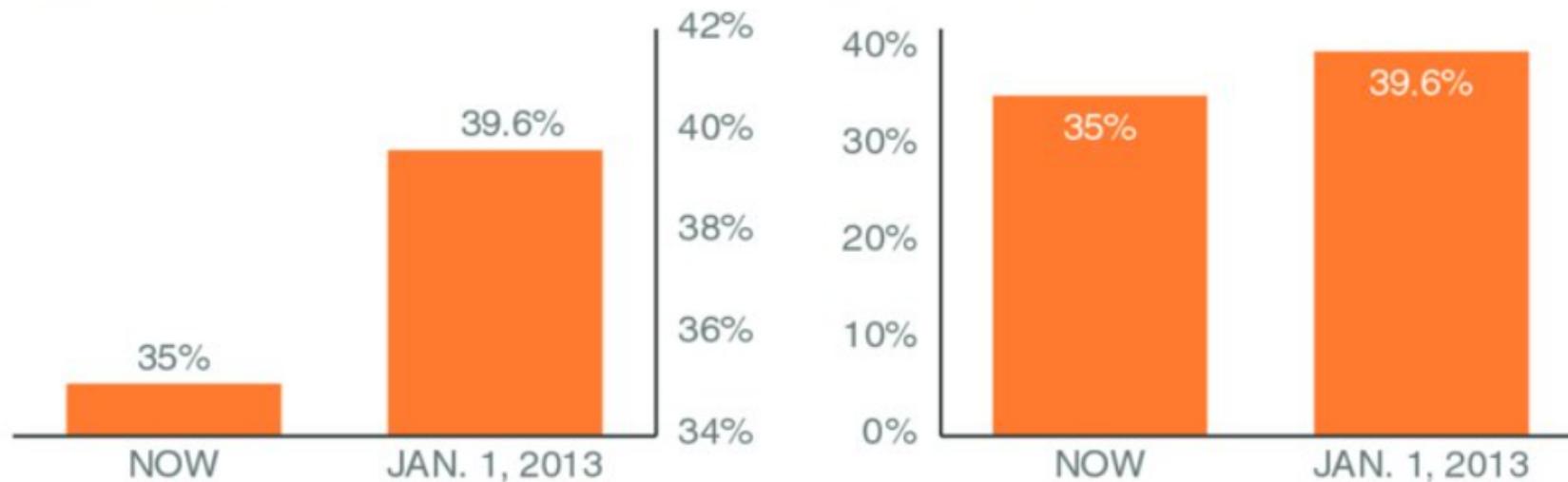


Bars

Always have a zero baseline

Non-zero baseline: as originally graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE

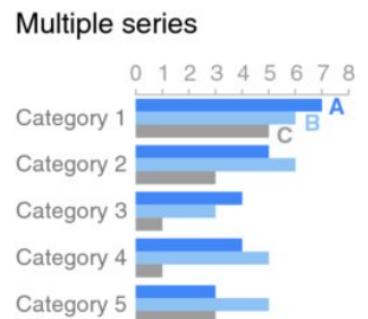
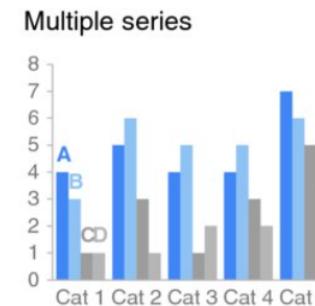
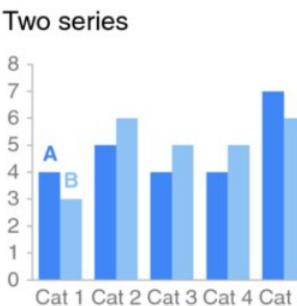
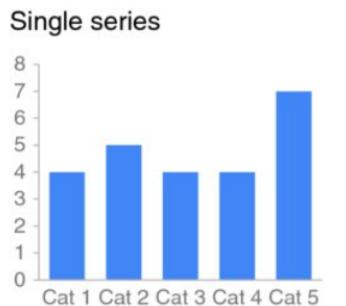
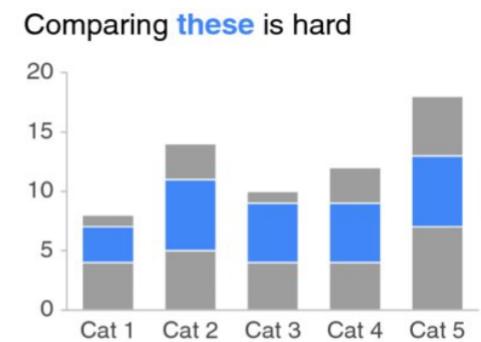
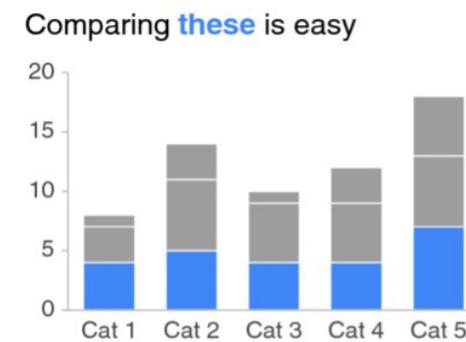
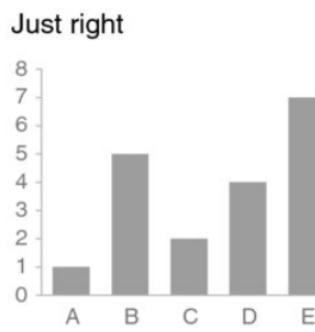
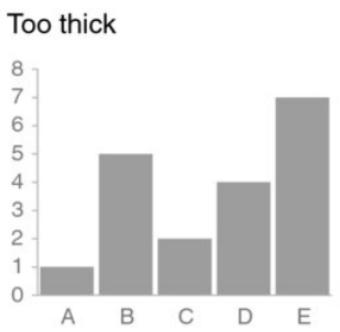
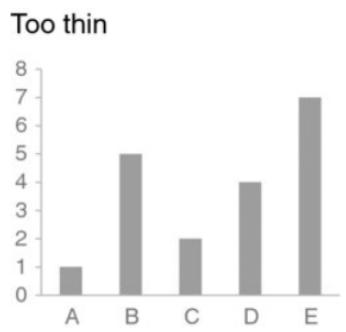


Zero baseline: as it should be graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE

Bars

Width, series, stacked, horizontal



ggplot2

One of the packages in the tidyverse that enables the “static” visualization of data

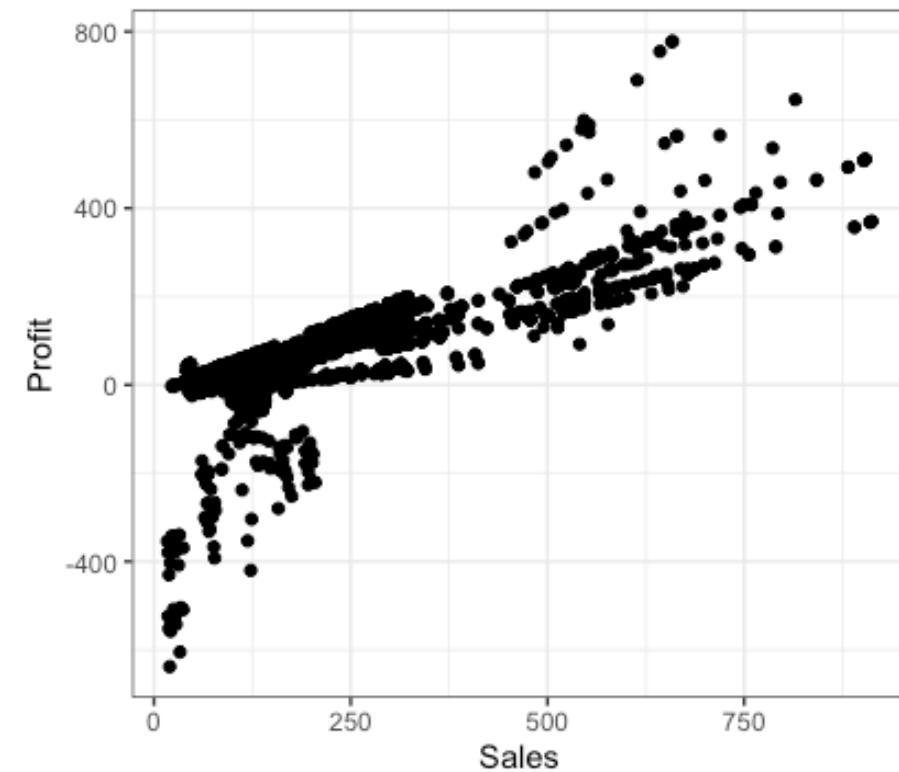
```
library(tidyverse) # load package  
library(readxl)  
  
CoffeeChain <- read_excel("CoffeeChain.xlsx")  
  
ggplot(data = CoffeeChain) +  
  geom_point(mapping = aes(x = Sales, y = Profit)) +  
  theme_bw()
```



Points – scatterplot in R

Useful for showing relationship between two things

Is there a positive
relationship between
sales and profit?

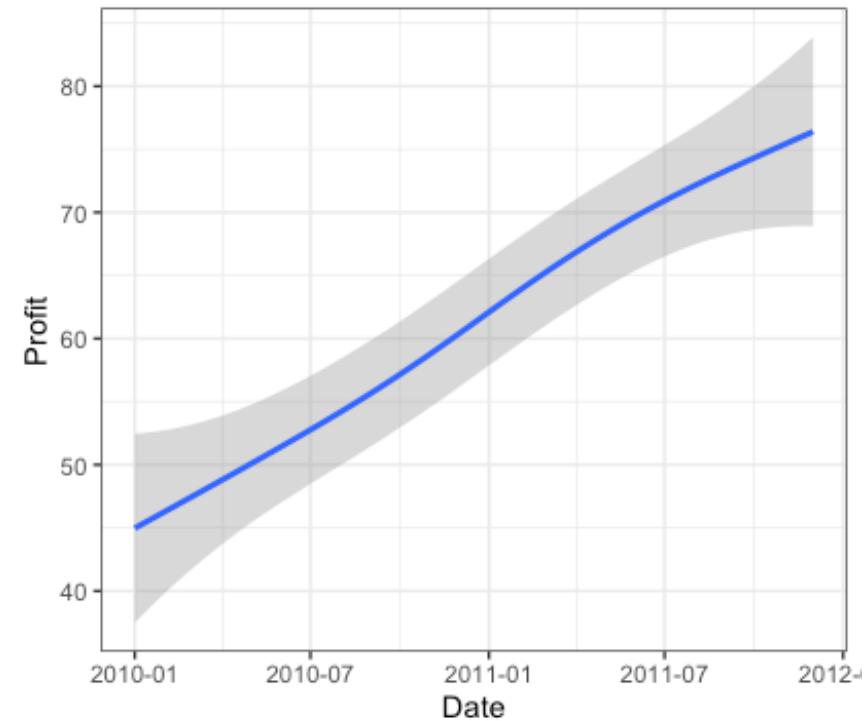


Lines in R

Most used to plot continuous data

Has profit changed over time?

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

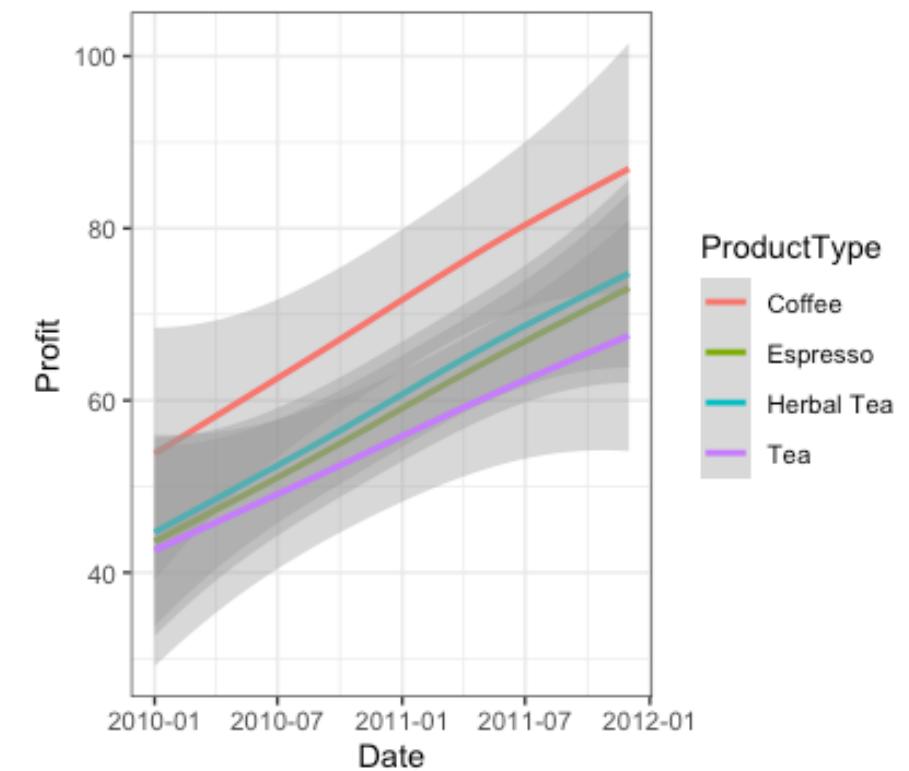


Lines in R

Most used to plot continuous data

Has profit changed over time?

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

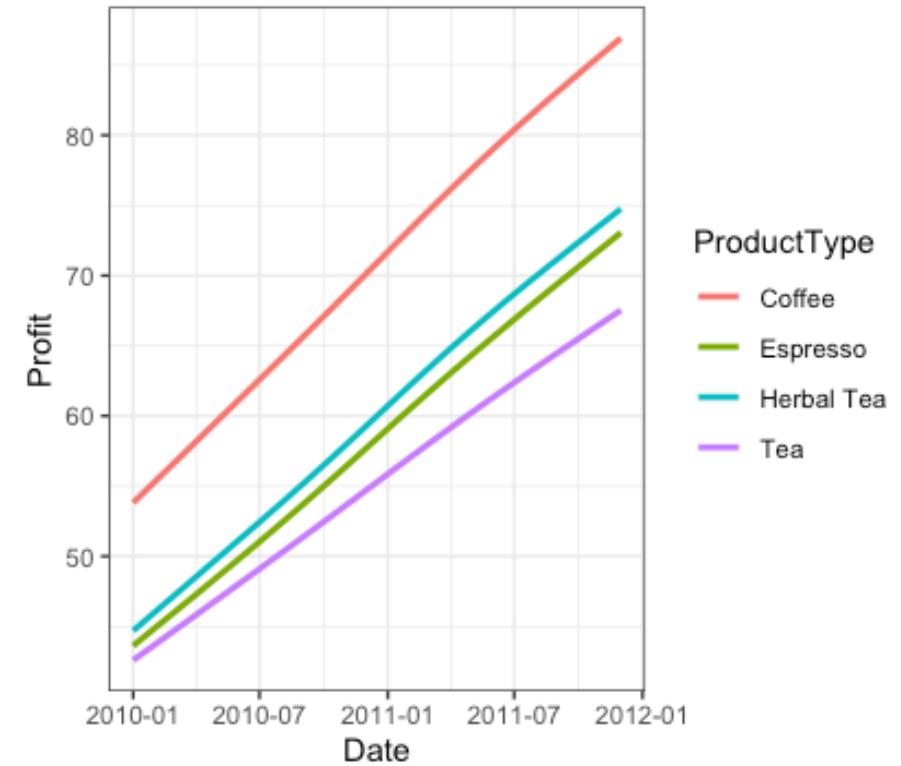


Lines in R

Most used to plot continuous data

Has profit changed over time?

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

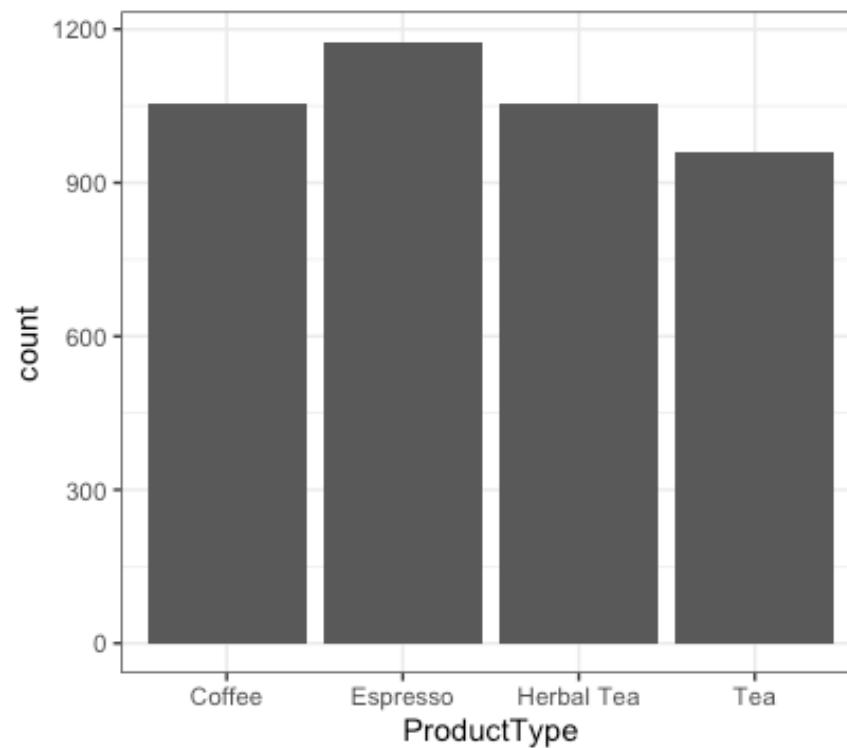


Bars in R

Easy for our eyes to read

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

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

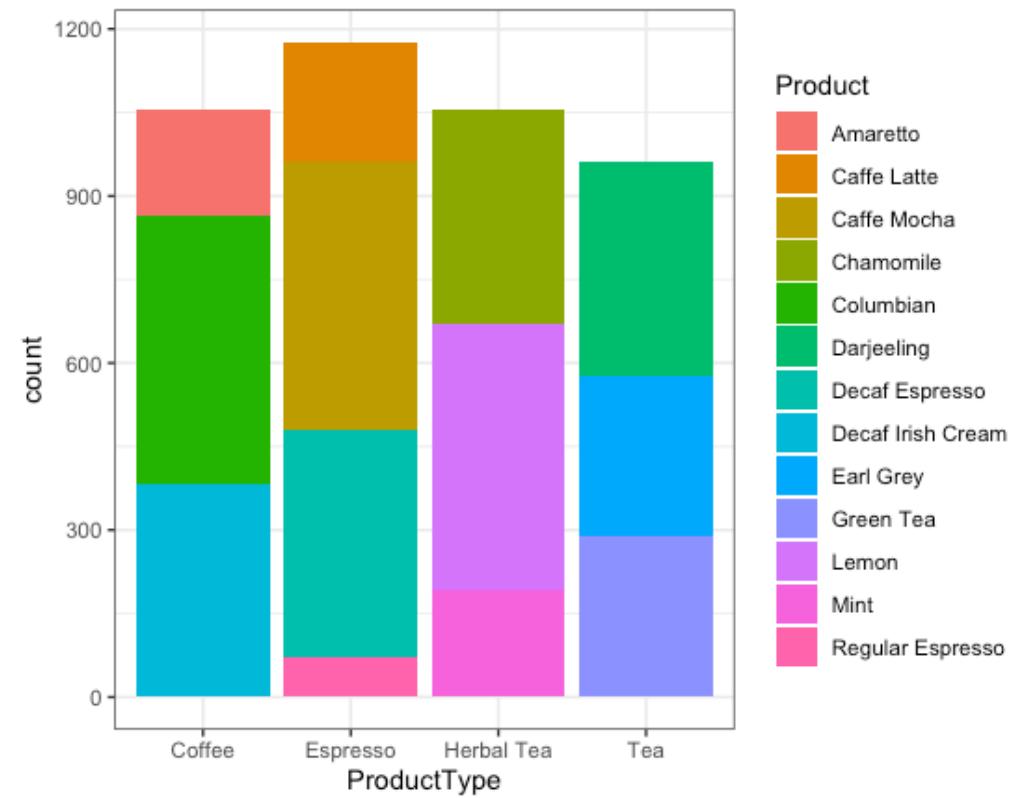


Bars in R

Easy for our eyes to read

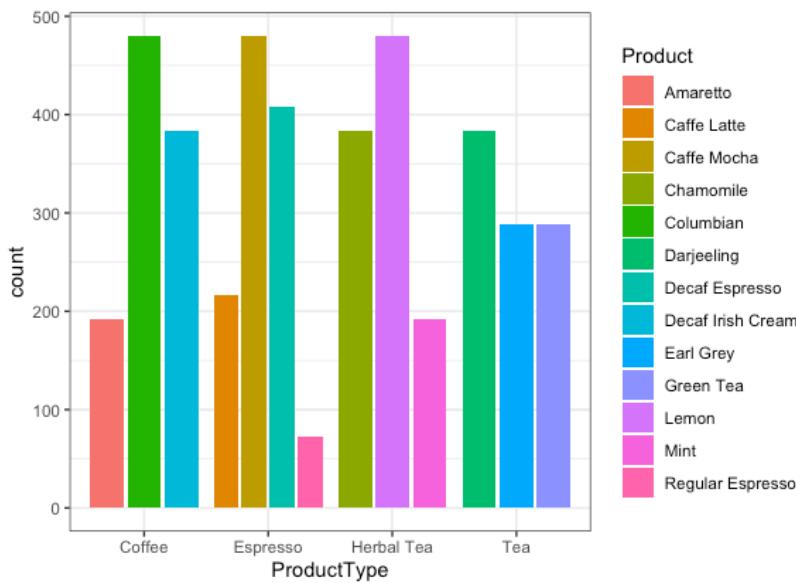
What products and how many of them do we have for every type of product?

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



At-Home Exercises

Copy and paste the code from the slides (starting on slide 16) into an R Markdown document. Execute the code in R, line by line. Are your results like the ones in the slides? If yes, write code to generate the following visual:



Once you are done, try to knit the code to PDF



At-Home Exercises

Read chapter 3 and do the exercises of the R for Data Science book

Read chapter 2 of story telling with data

Check out the ggplot2 cheat sheet: <https://github.com/rstudio/cheatsheets/raw/master/data-visualization-2.1.pdf>

Open the CoffeeChain dataset in Tableau. Next, try to use the software to create visualizations (very similar to how we used R to do it). How long did it take you? Which tool provides better visuals?



Thank You!



Terry College of Business
UNIVERSITY OF GEORGIA