Visualization III

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

Business Intelligence
Spring 2021
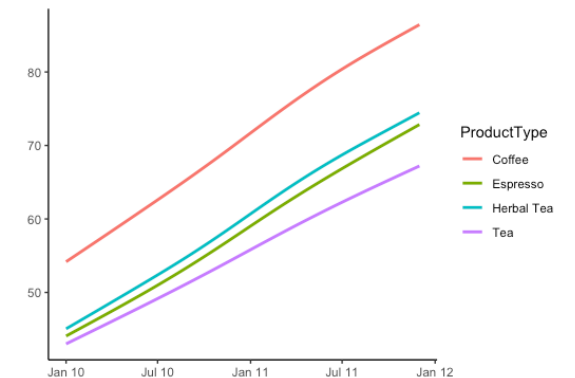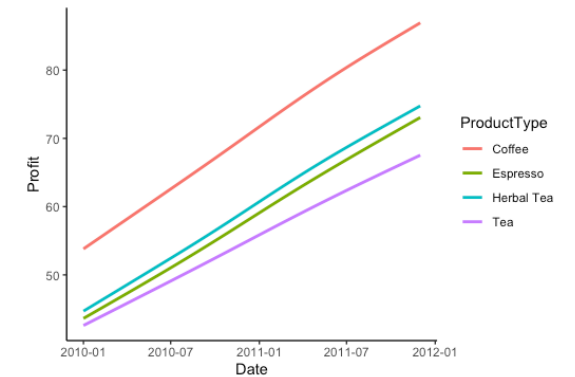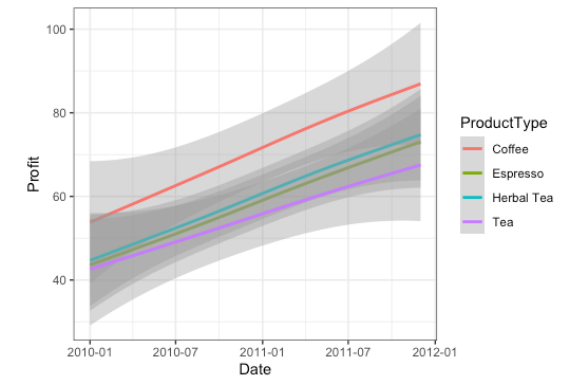
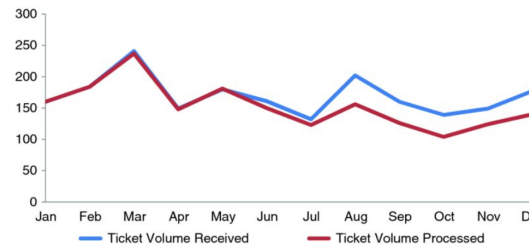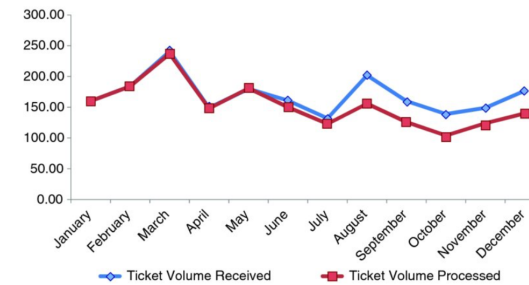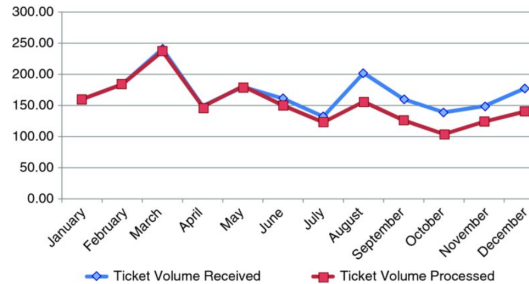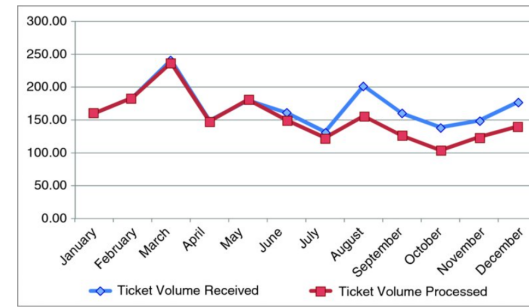**Terry College of Business**
UNIVERSITY OF GEORGIA

"The larger the share of a graphic's ink devoted to data, the better (other relevant matters being equal)"

Edward Tufte

Clutter is your enemy!

Maximize the data-ink ratio (a.k.a., maximize signal-to-noise ratio)

A Few Lessons from Knaflic (2015)

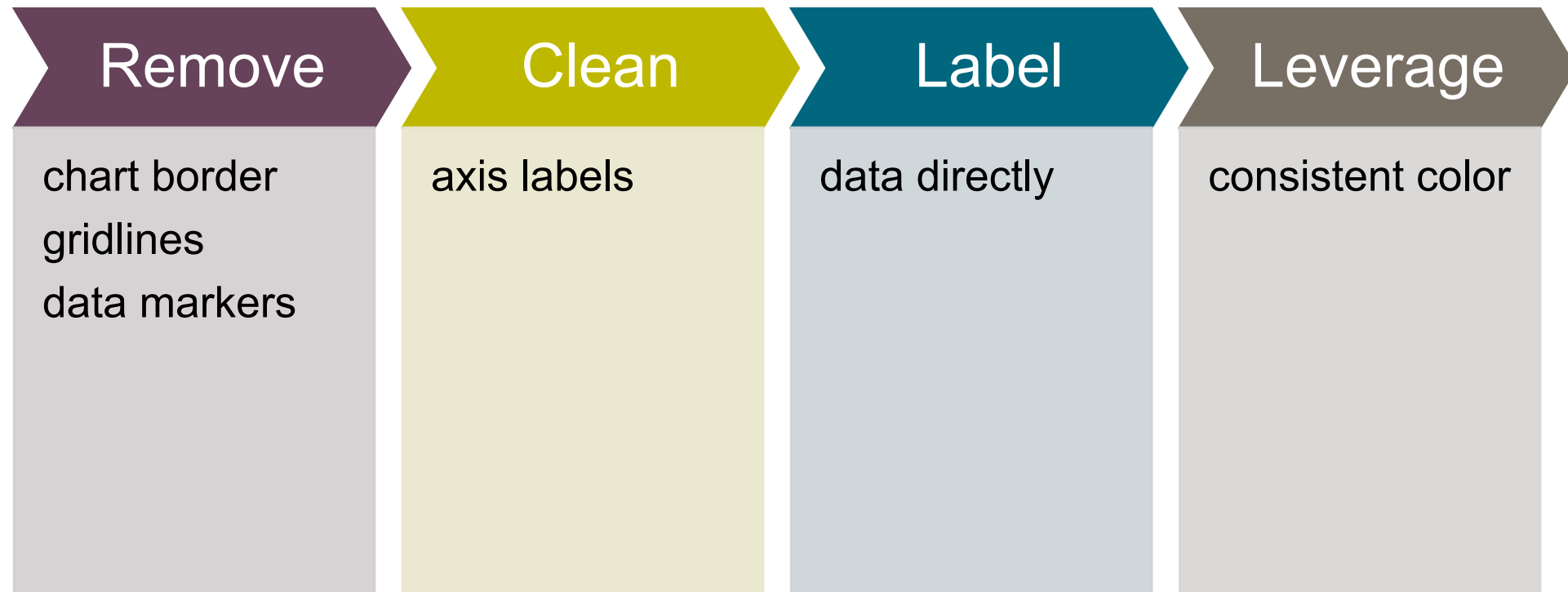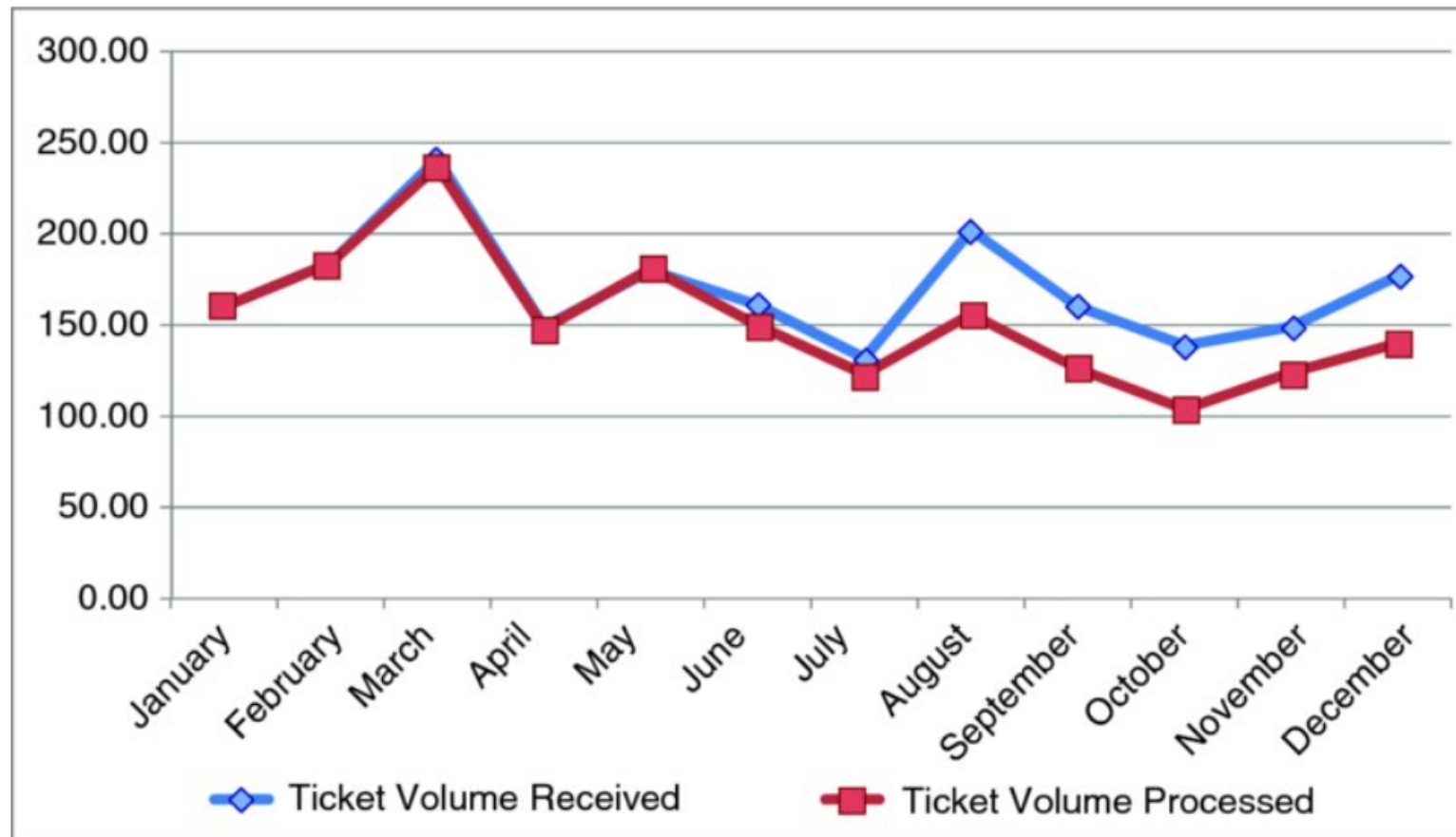| 1	2	3	4	5	6
Understand the context	Choose an appropriate visual display	Eliminate clutter	Focus attention where you want it	Think like a designer	Tell a story

Decluttering, step-by-step (RCLL)

Remove	Clean	Label	Leverage
chart border	axis labels	data directly	consistent color
gridlines			
data markers			

Ticket volume example

Remove chart border
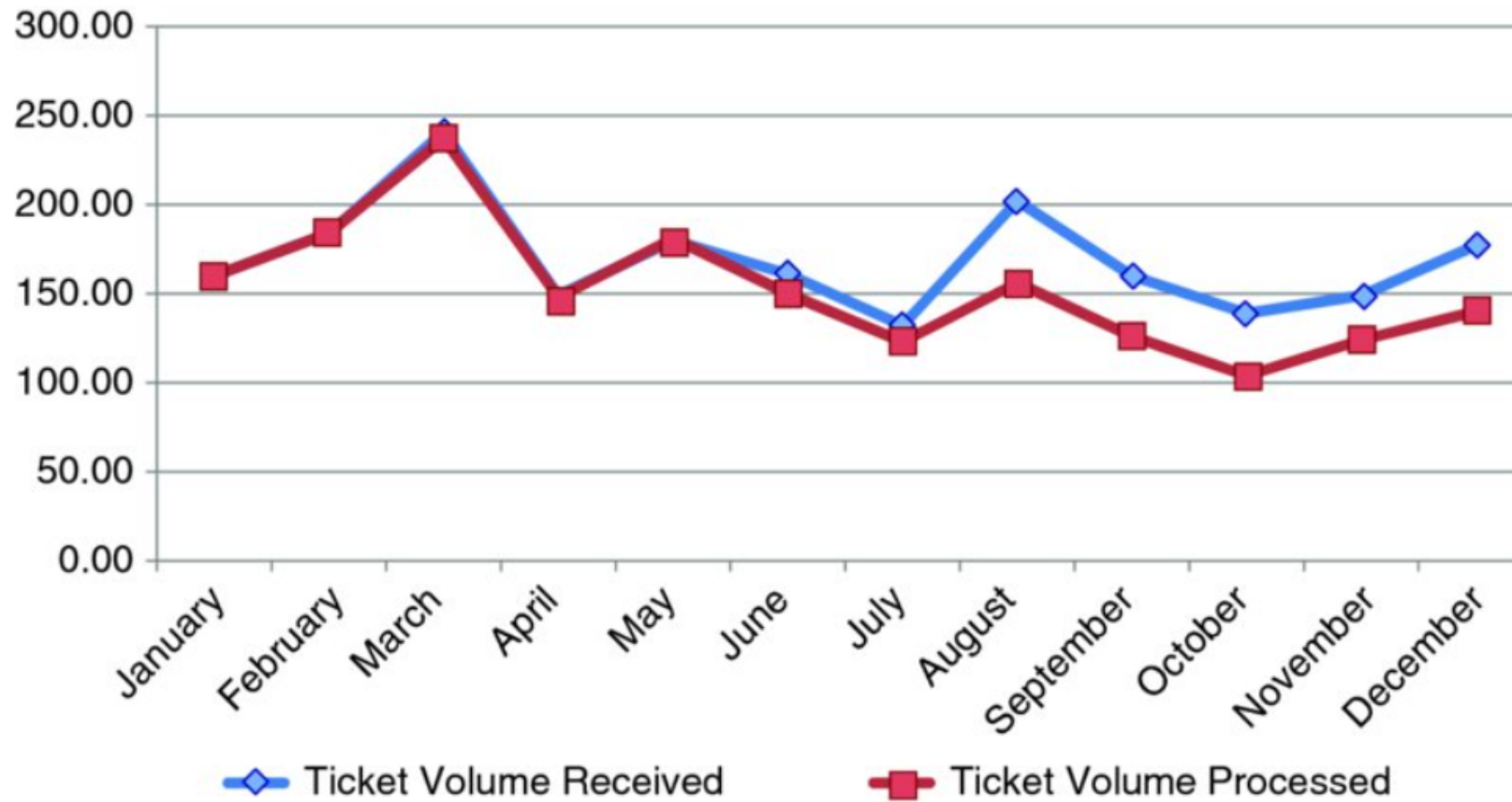
Then gridlines

Next data markers (and clean axis too!)

Label data directly
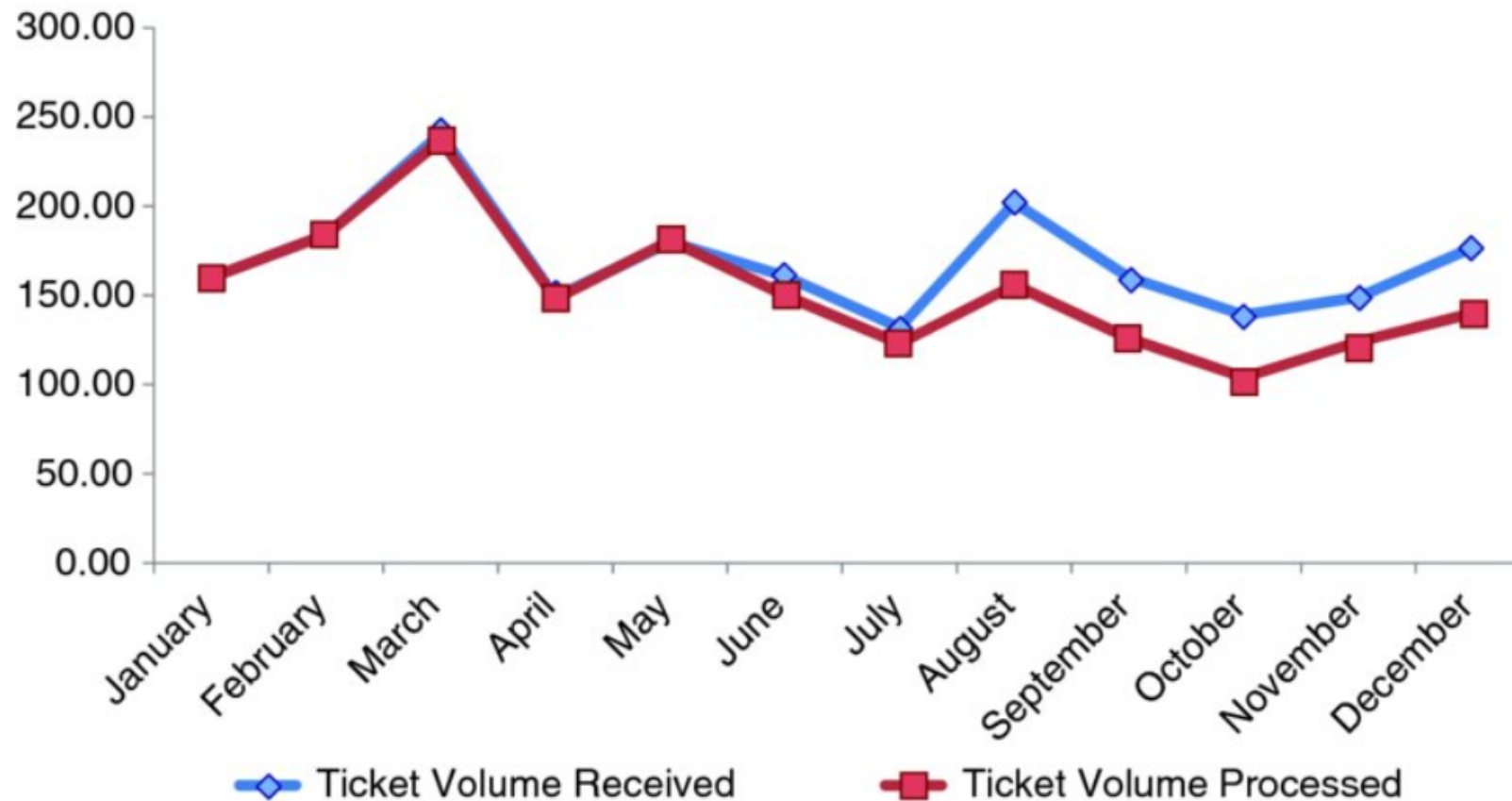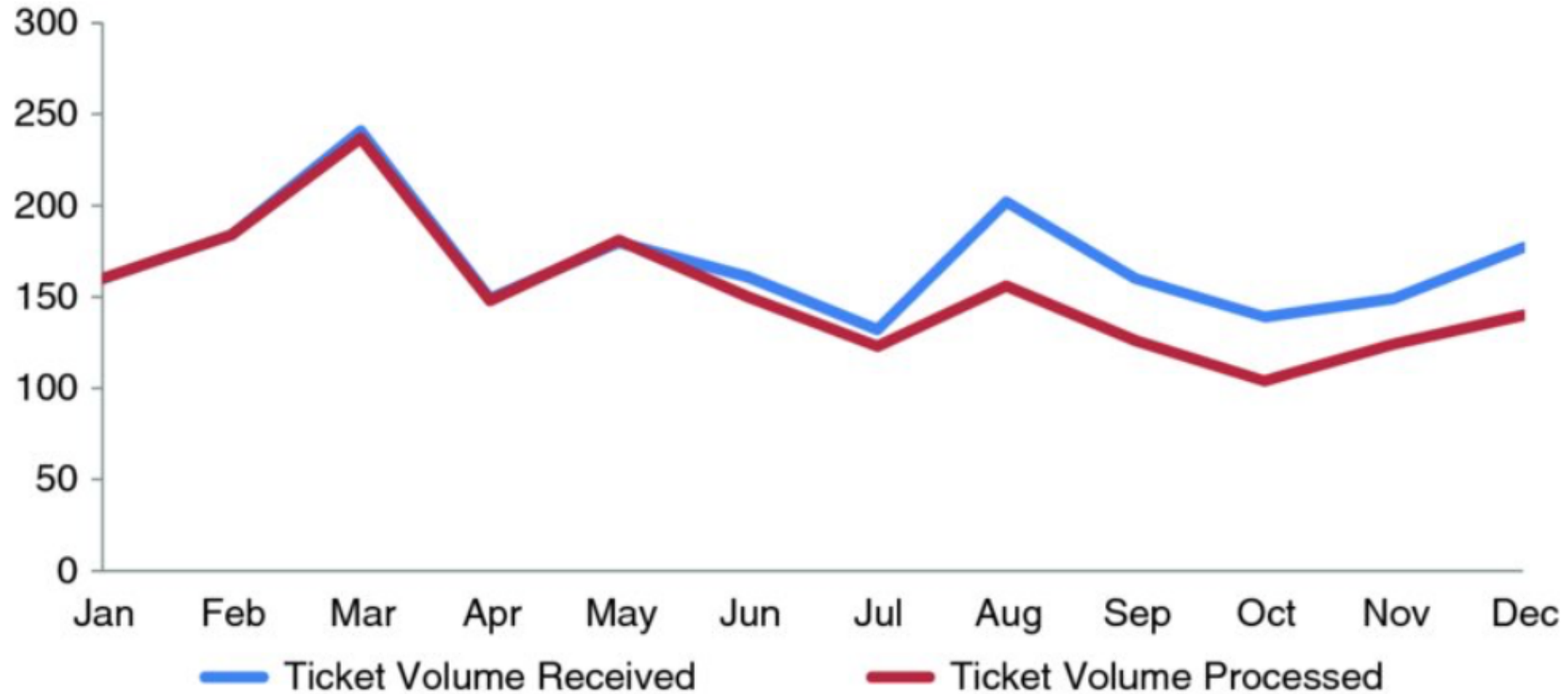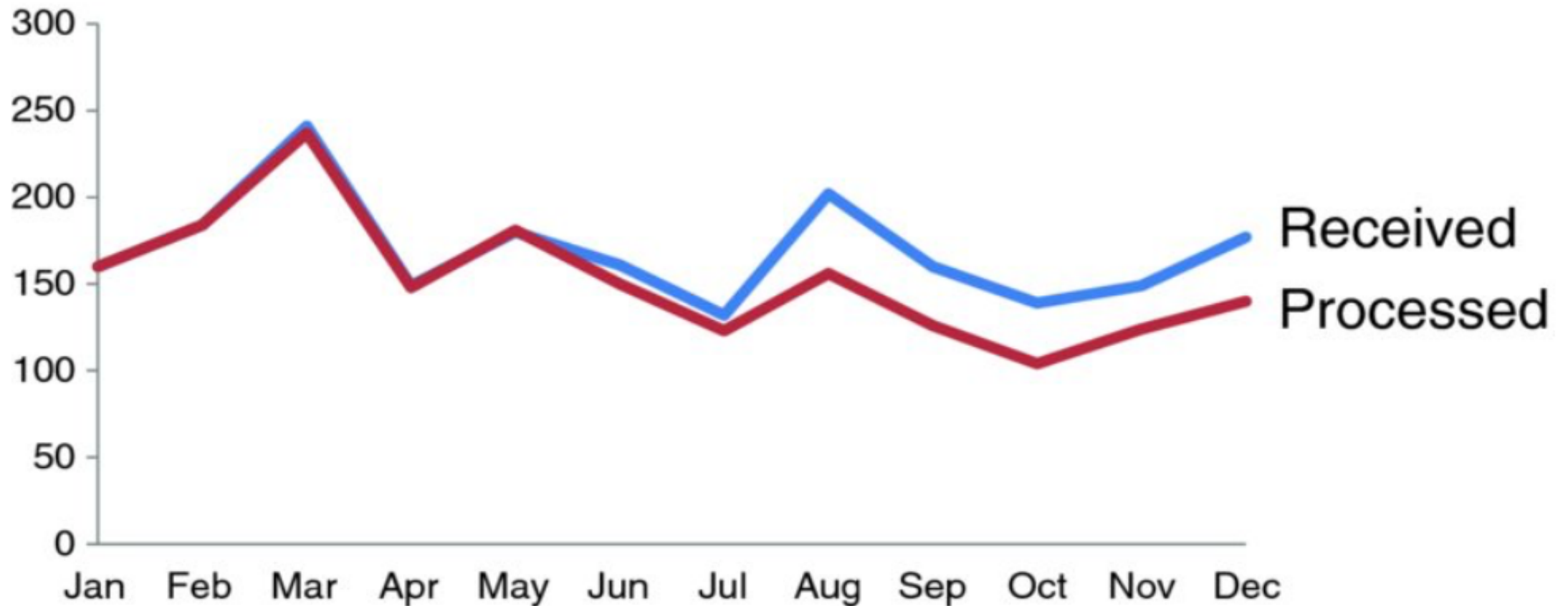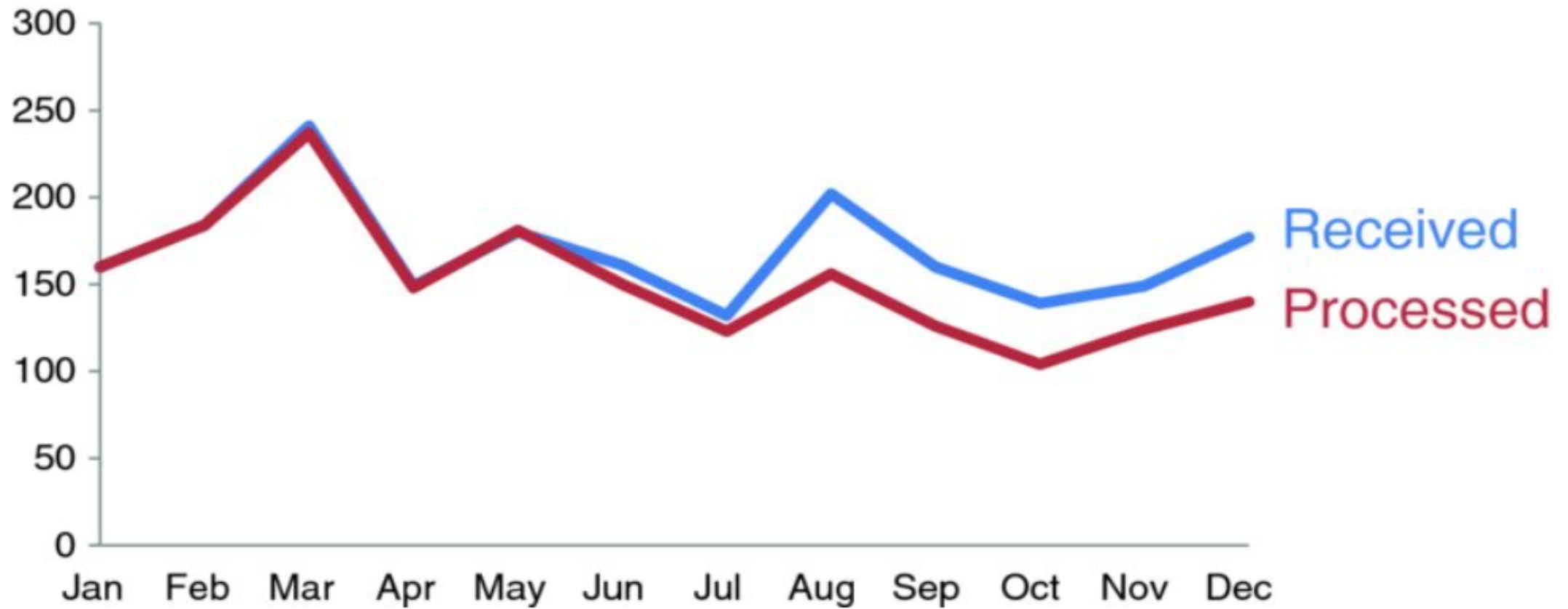
Leverage consistent color

Terry College of Business
UNIVERSITY OF GEORGIA

Before and after
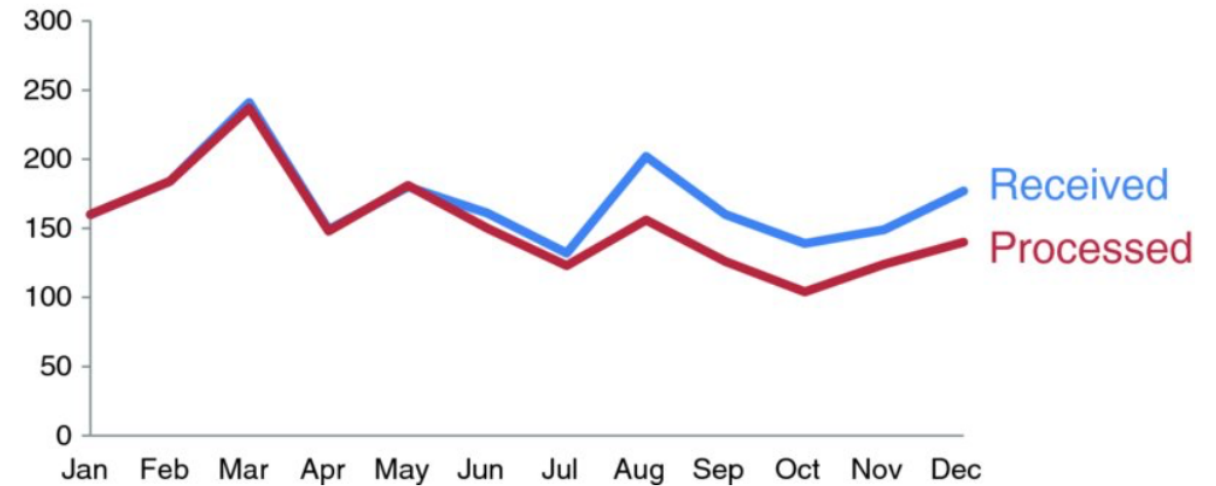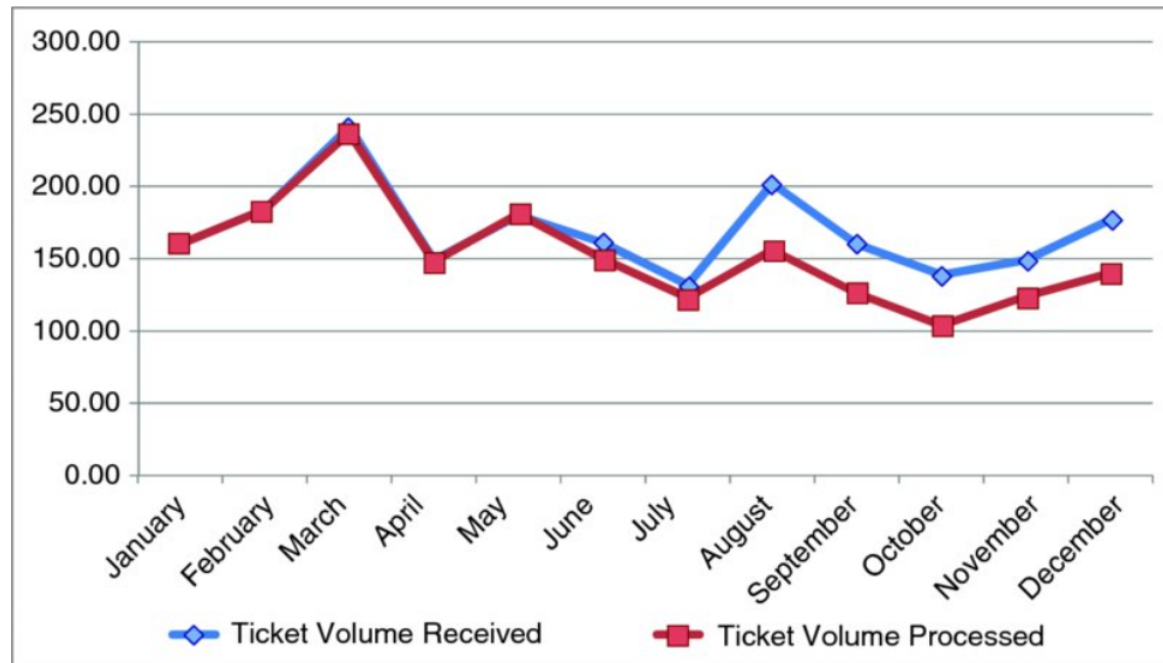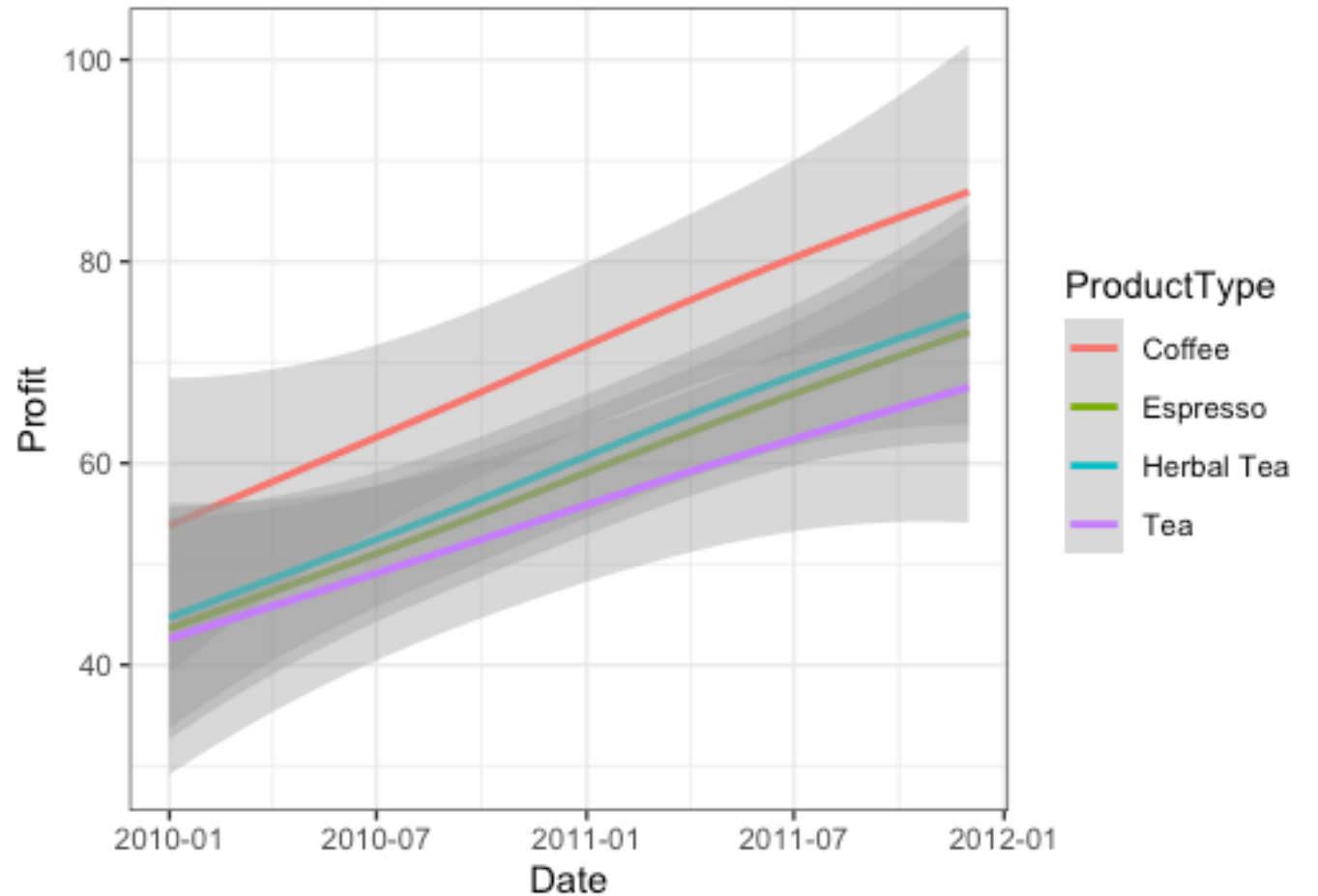
Decluttering (Lines) in R

Most used to plot continuous data

Has profit changed over time?

```
ggplot(data = CoffeeChain,
       mapping = aes(x = Date,
                     y = Profit,
                     color = ProductType)) +
  geom_smooth() +
  theme_bw()
```
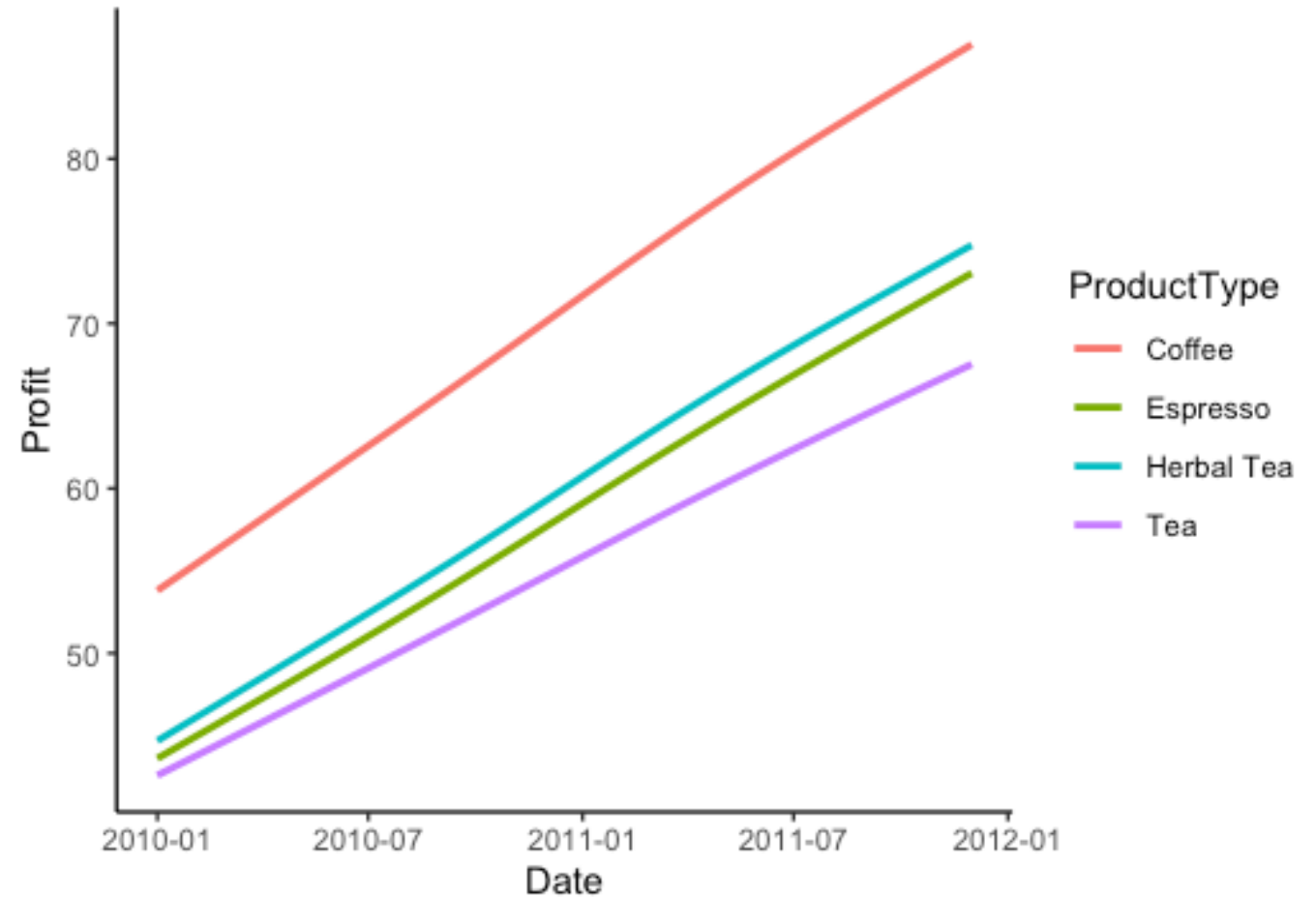
Remove border, gridlines, and data markers

Most used to plot continuous data

Has profit changed over time?

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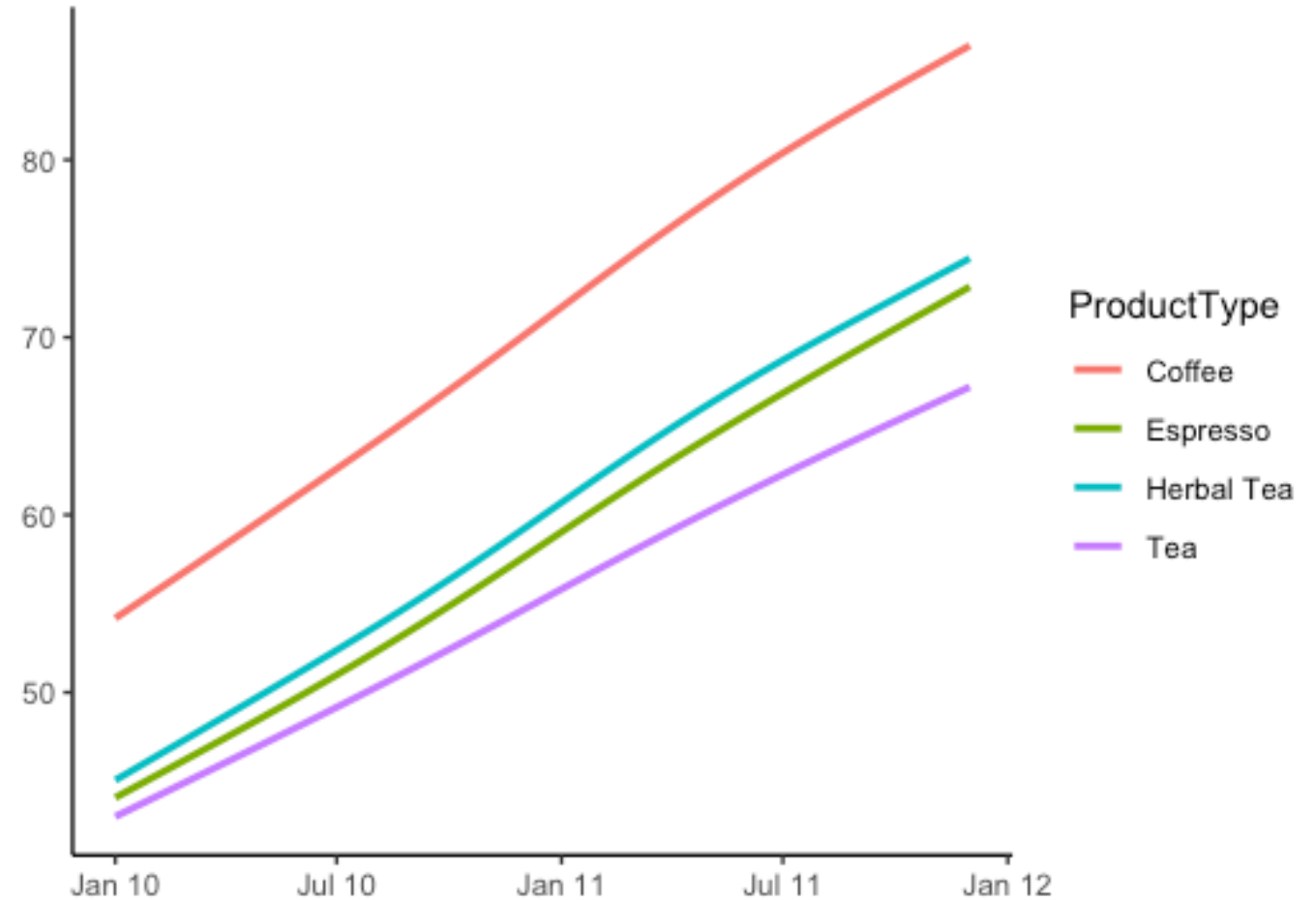
Clean up axis

Most used to plot continuous data

Has profit changed over time?

```r
library(lubridate)
CoffeeChain$Date <- ymd(CoffeeChain$Date)

ggplot(data = CoffeeChain,
       mapping = aes(x = Date,
                     y = Profit,
                     color = ProductType)) +
  geom_smooth(se = FALSE) +
  theme_classic() +
  scale_x_date(date_labels = "%b %y") +
  labs(x = NULL, y = NULL)
```
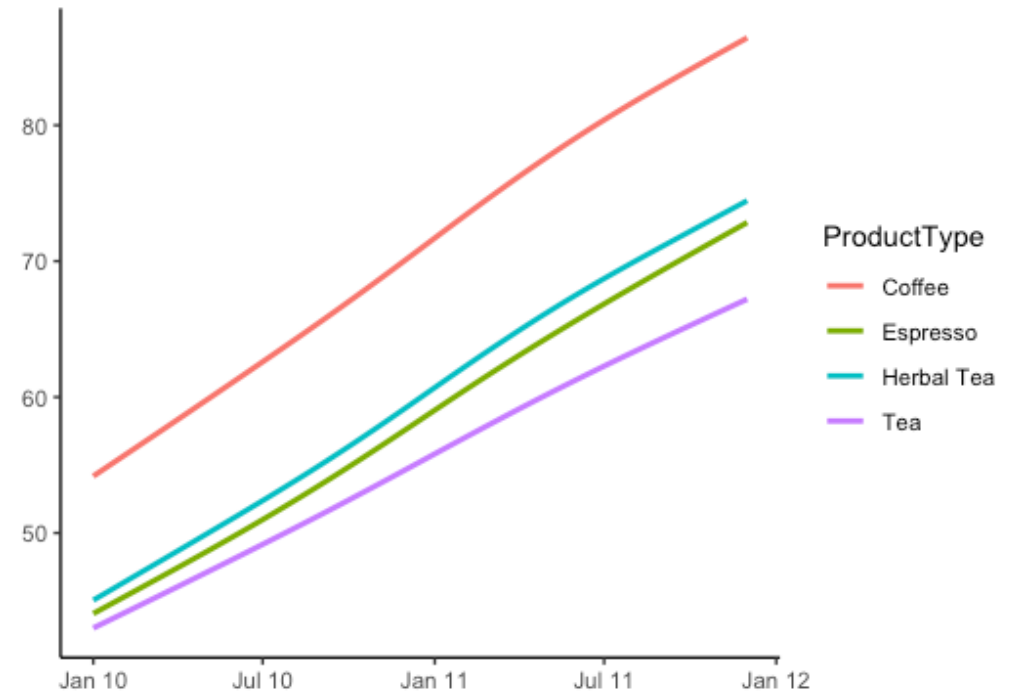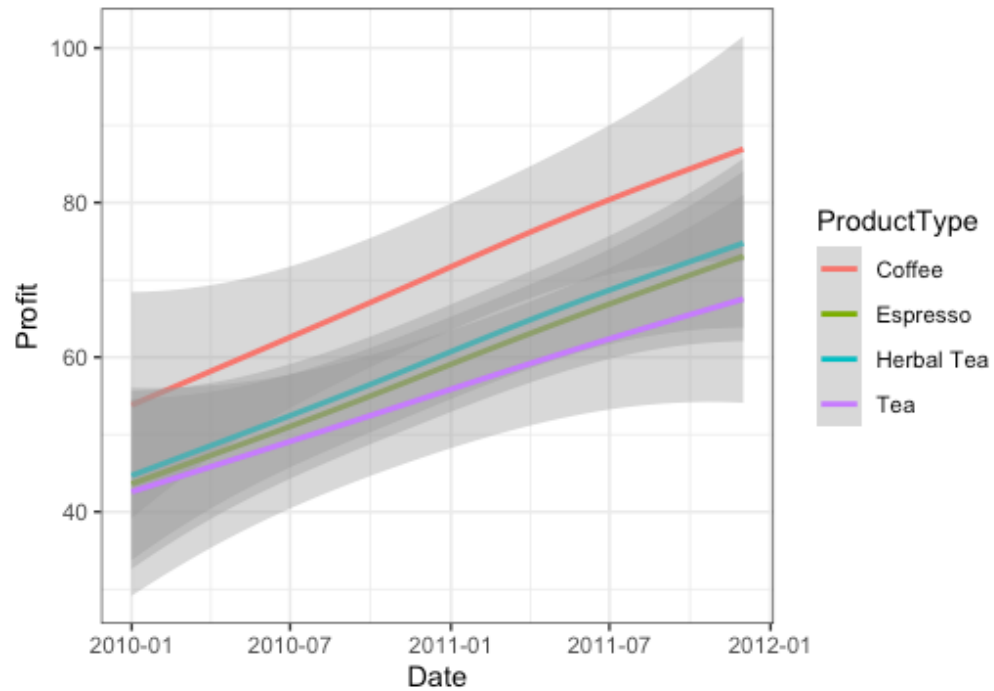
Label data directly and leverage consistent color

Not easily embedded in ggplot2. If you are interested and willing to use another package, look at **directlabels**
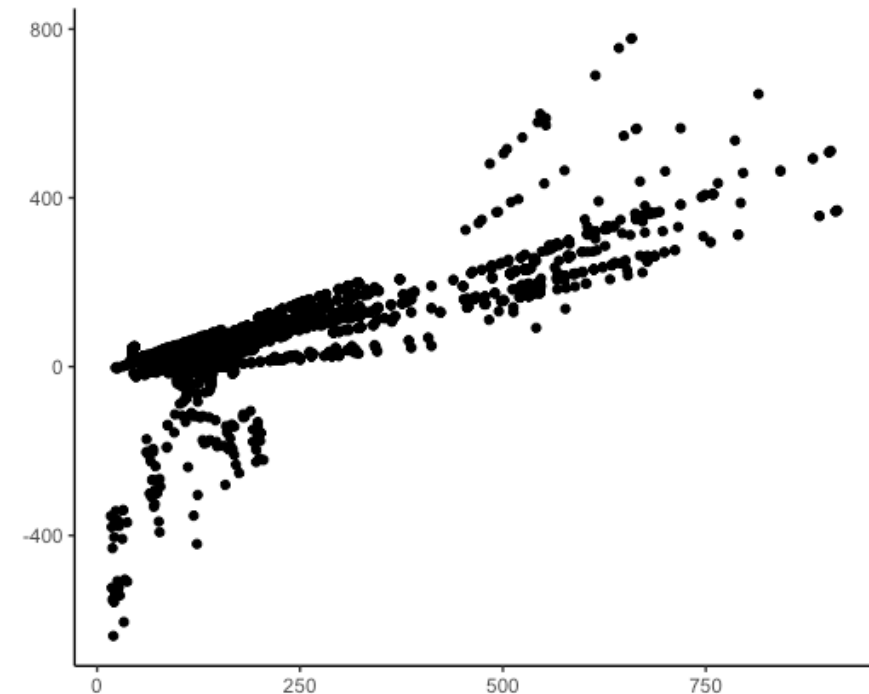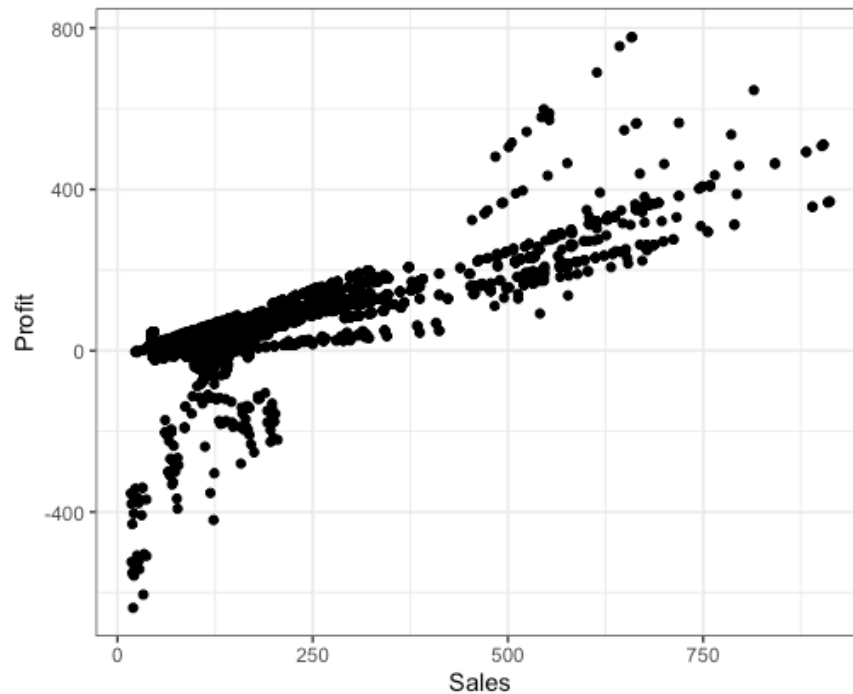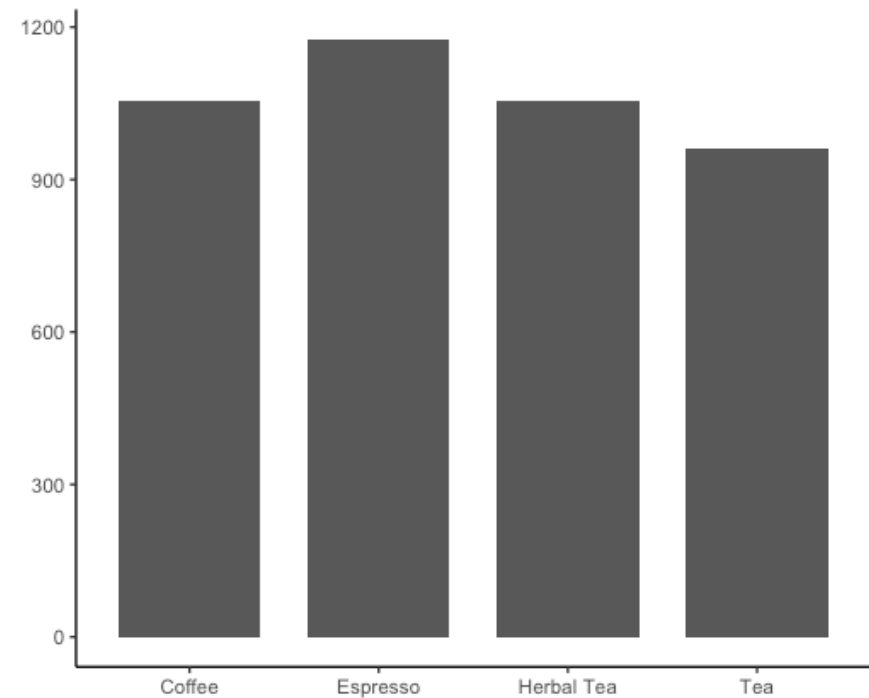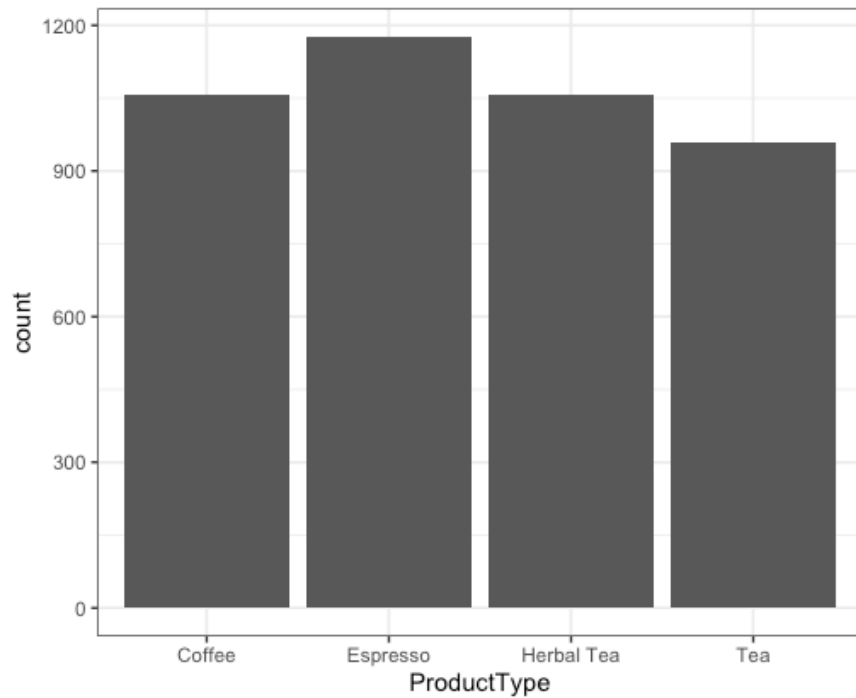
http://directlabels.r-forge.r-project.org

Before and after

Decluttering (Points) in R
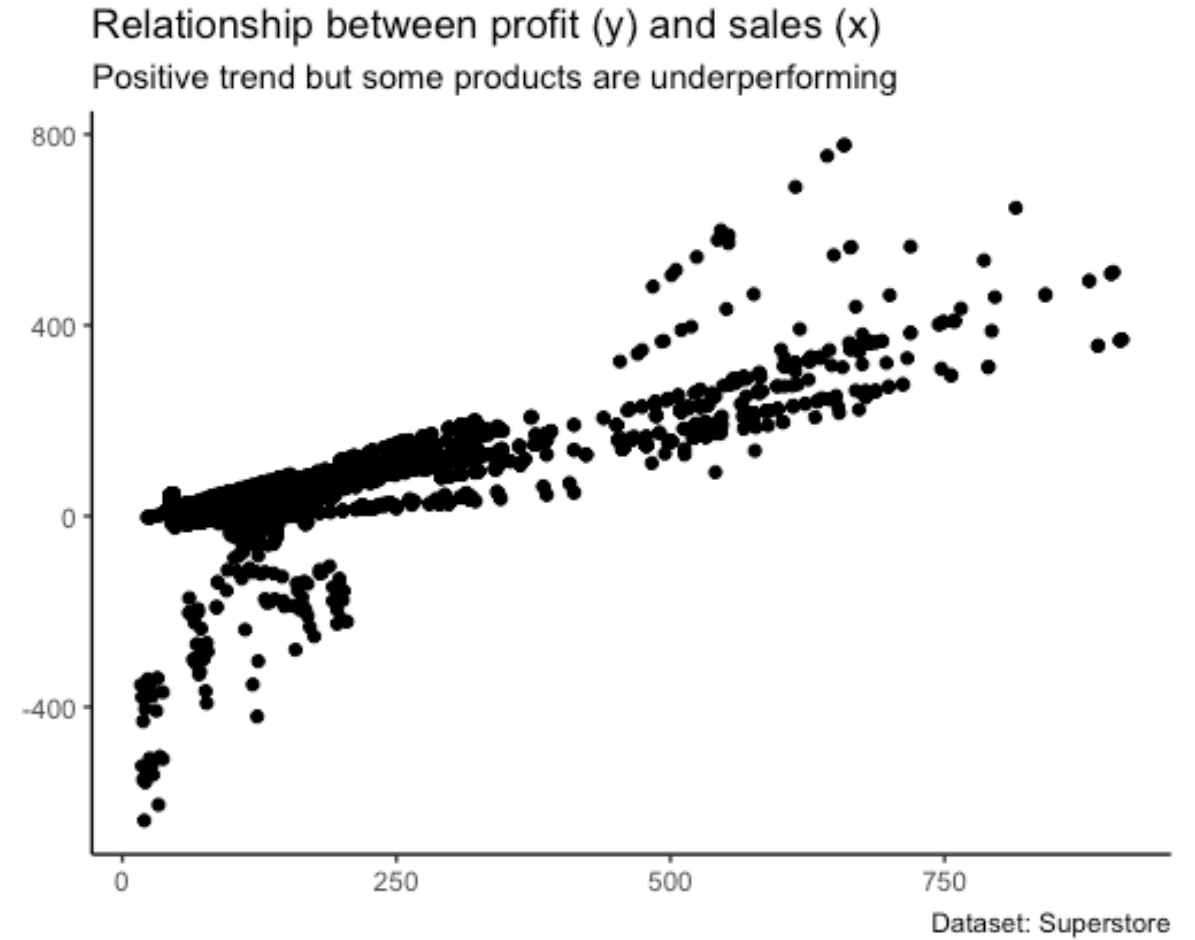
Decluttering (Bars) in R

Title, subtitle, and caption

Useful for showing relationship between two things

Is there a positive relationship between sales and profit?

```
ggplot(data = CoffeeChain,
       mapping = aes(x = Sales, y = Profit)) +
  geom_point() +
  theme_classic() +
  labs(x = NULL, y = NULL,
       title = "Relationship between profit (y) and sales (x)",
       subtitle = "Positive trend but some products are underperforming",
       caption = "Dataset: Superstore")
```


Relationship between profit (y) and sales (x)
Positive trend but some products are underperforming

Dataset: Superstore

Title, subtitle, caption (and faceting)

Relationship between profit (y) and sales (x)

Positive trend but some products are underperforming

Dataset: Superstore
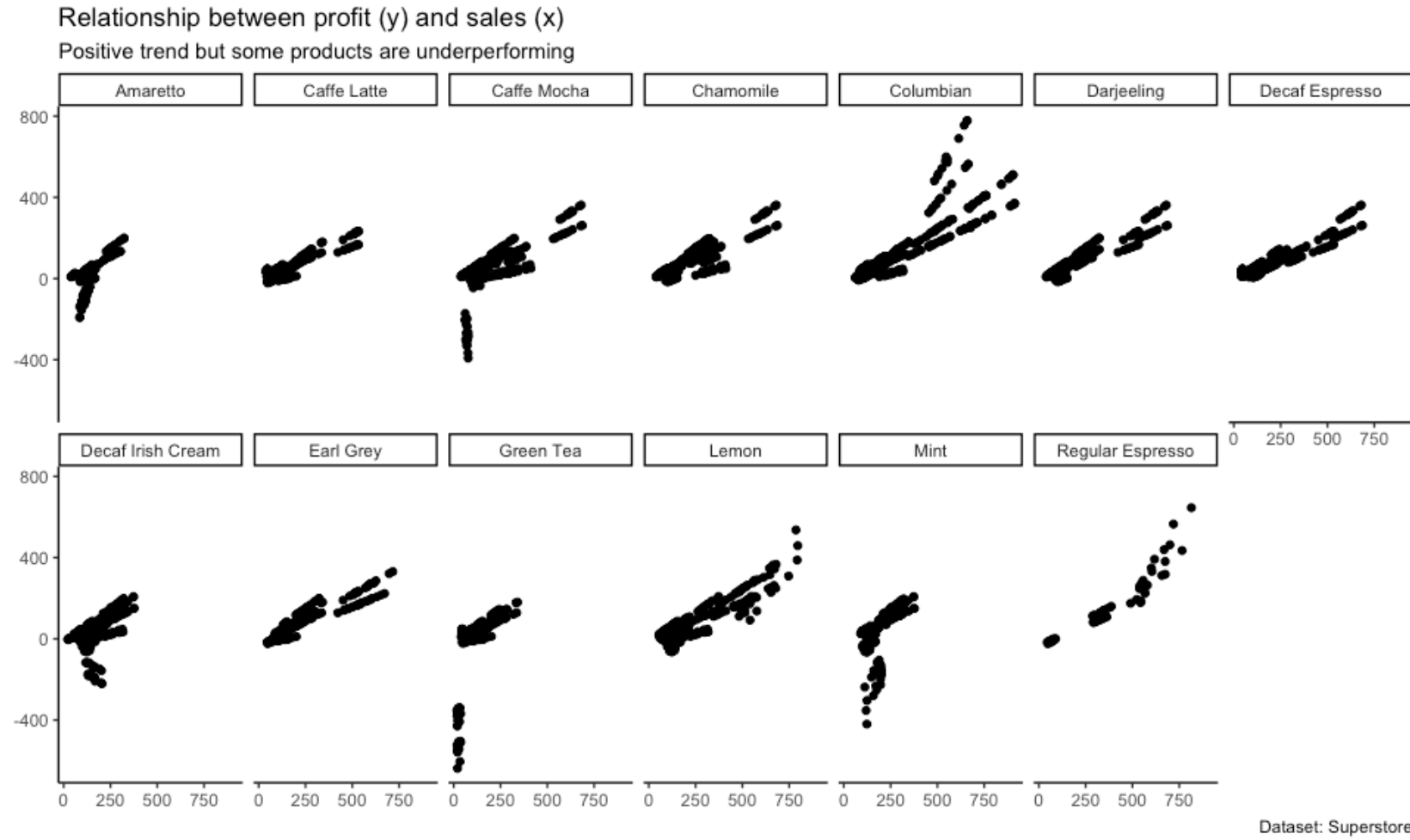
Main takeaways

Remove clutter

- Remove border, gridlines, markers
- Clean up axis
- Label directly
- Use consistent color

At-Home Exercises

Practice, practice, practice…

Open the CoffeeChain dataset in R. Next, de-clutter all graphs (points, lines, and bars) you created from last time
- Do the same for Tableau

Read chapter 3 ("clutter is your enemy") of Storytelling with data

Thank You!