# Module 3 - Assignment 1

## Angelo, Elizabeth

### Data Visualization

I will be using datasets containing candy rankings and production.

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.5 ✔ purrr 0.3.4  
## ✔ tibble 3.1.6 ✔ dplyr 1.0.7  
## ✔ tidyr 1.1.4 ✔ stringr 1.4.0  
## ✔ readr 2.1.1 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

candy\_data <- read\_csv("candy\_data.csv")

## Rows: 85 Columns: 13

## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): competitorname  
## dbl (12): chocolate, fruity, caramel, peanutyalmondy, nougat, crispedricewaf...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

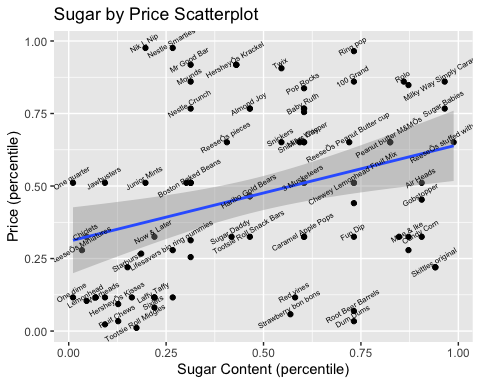
candy\_production <- read\_csv("candy\_production.csv")

## Rows: 548 Columns: 2  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (1): IPG3113N  
## date (1): observation\_date  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#### Visualization with Scatterplots (geom\_point)

library(ggplot2)  
ggplot(candy\_data, aes(x = sugarpercent, y = pricepercent, label = competitorname)) +   
 geom\_point() +   
 labs(title = "Scatterplot of Sugar Content vs Price of Candy",   
 x = "Sugar Content",   
 y = "Price Percent") +   
 geom\_smooth(method = "lm") +   
 geom\_text(check\_overlap = T,   
 vjust = "bottom",   
 nudge\_y = 0.01,   
 angle = 30,   
 size = 2) +   
 labs(title = "Sugar by Price Scatterplot",   
 x = "Sugar Content (percentile)",   
 y = "Price (percentile)")

## `geom\_smooth()` using formula 'y ~ x'



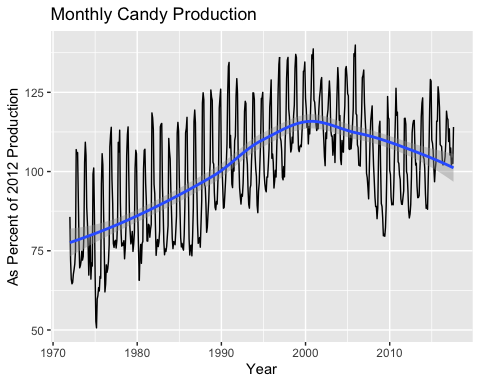
After reviewing the scatter-plot, Dum Dums have the most sugar content at the lowest price, although Skittles appear to be a strong competitor with more sugar content at a slightly higher price. Ring Pops have the most sugar content at the highest price, with Rolo and Milky Way Simply Caramel being strong competitors with higher sugar content at a slightly lower price.

#### Line Chart of Candy Production

I will be using the candy production dataset that will display a specific date and how production during that month is compared to 2012.

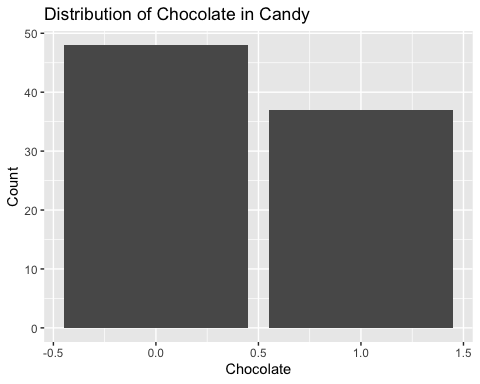
ggplot(candy\_production, aes(x = observation\_date, y = IPG3113N)) +   
 geom\_line() +   
 geom\_smooth() +   
 labs(title = "Monthly Candy Production",   
 x = "Year",   
 y = "As Percent of 2012 Production")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



#### Bar Chart of Ingredients

ggplot(candy\_data, aes(x = chocolate)) +   
 geom\_bar() +   
 labs(title = "Distribution of Chocolate in Candy",   
 x = "Chocolate",   
 y = "Count")



candyFeatures <- candy\_data %>% select(2:10)  
candyFeatures[] <- lapply(candyFeatures, as.logical)

ggplot(candyFeatures, aes(x = chocolate)) +   
 geom\_bar() +   
 labs(title = "Distribution of Chocolate in Candy",   
 x = "Chocolate",   
 y = "Count")

