# Module 3 - Assignment 1

## Kiser, Stephen

### Data Visualization

For Assignment 1 I will be using datasets that contain the production and rankings of candy. These datasets are given to me on the MIS-503 course. I will now use these datasets to find out some data about different candy.

library(tidyverse)

## -- Attaching packages ------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.5  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts ---------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

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

## Parsed with column specification:  
## cols(  
## competitorname = col\_character(),  
## chocolate = col\_double(),  
## fruity = col\_double(),  
## caramel = col\_double(),  
## peanutyalmondy = col\_double(),  
## nougat = col\_double(),  
## crispedricewafer = col\_double(),  
## hard = col\_double(),  
## bar = col\_double(),  
## pluribus = col\_double(),  
## sugarpercent = col\_double(),  
## pricepercent = col\_double(),  
## winpercent = col\_double()  
## )

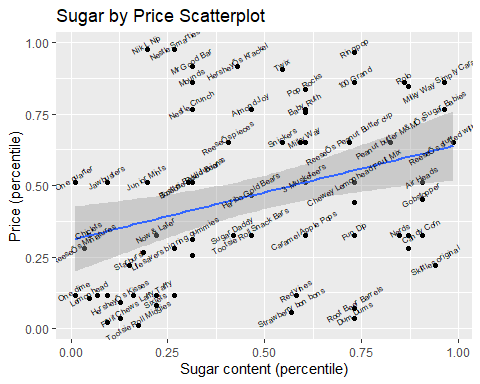
candy\_production <- read\_csv("candy\_production.csv")

## Parsed with column specification:  
## cols(  
## observation\_date = col\_date(format = ""),  
## IPG3113N = col\_double()  
## )

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

ggplot(data= candy\_data, aes(x= sugarpercent, y= pricepercent, label = competitorname)) +  
 geom\_point() +  
 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'



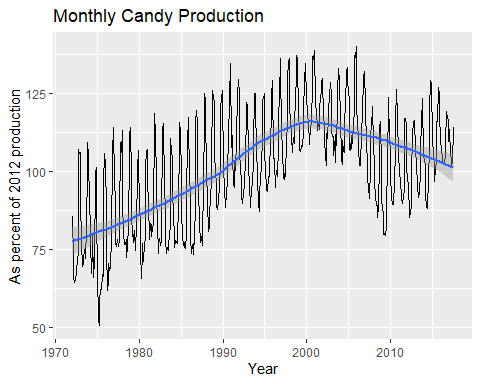
After the completion of the scatterplot I was able to find which candy has the most sugar and lowest price, and the most expensive candy with the highest sugar content. The candy that has the most sugar and lowest price is Dum Dums. They have almost a 75 % sugar content and less than 12.5% in price. The most expensive candy with the highest sugar content is Milky Way Simply Caramel. The sugar content is almost 100%, and the price is above the 75 percentile.

#### Line Chart of Candy Production

Now I will be using the Candy Production dataset to make a line chart. This data displays specific dates and how production was during that month. This data is compared to the 2012 results of the production of the candy.

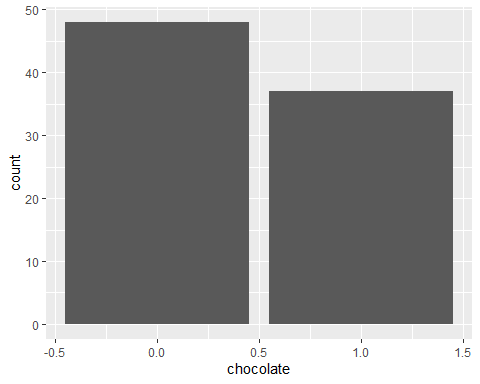
ggplot(data = candy\_production, aes(x= observation\_date, y= IPG3113N))+  
 geom\_line()+  
 geom\_smooth()+  
 labs(title = "Monthly Candy Production", x= "Year", y= "As percent of 2012 production")

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



#### Bar Chart of Ingredients

#Original Bar Chart  
ggplot(data = candy\_data, aes(x=chocolate)) +  
 geom\_bar()



#New Bar Chart  
candyFeatures <- candy\_data %>% select(2:10)  
candyFeatures[] <- lapply(candyFeatures, "as.logical")  
ggplot(data = candyFeatures, aes(x=chocolate)) +  
 geom\_bar()

