# Module 4 - Assignment 1

## Broach, Stuart

### Data Visualization

For this assignment, I will be using two different datasets: candy\_data.csv and candy\_production.csv.I have downloaded these datasets from Canvas and will apply statistical analysis and data visualizations on them.

library("tidyverse")

## -- Attaching packages --------------------------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.7  
## v tidyr 0.8.2 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

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

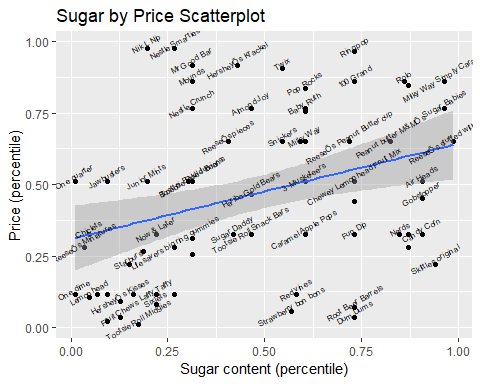
candy\_data <- read\_csv("candy\_data.csv")

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

candy\_production <- read\_csv("candy\_production.csv",   
 col\_types = cols(observation\_date = col\_date(format = "%Y-%m-%d")))

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

ggplot(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)")



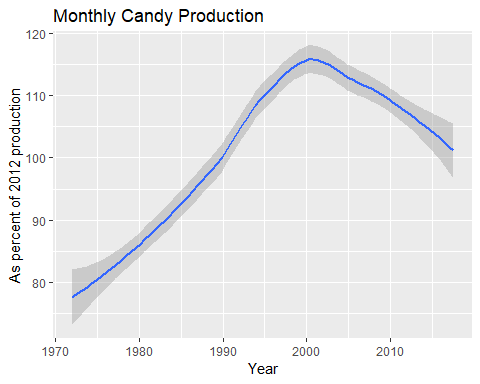
Based on these results, we can see that Skittles original has the most sugar with the lowest price while Milky Way Simply Caramel has the most sugar with the highest price.

#### Line Chart of Candy Production

Next, I will use the candy\_production dataset to display a specific date and how production during that month is compared to 2012.

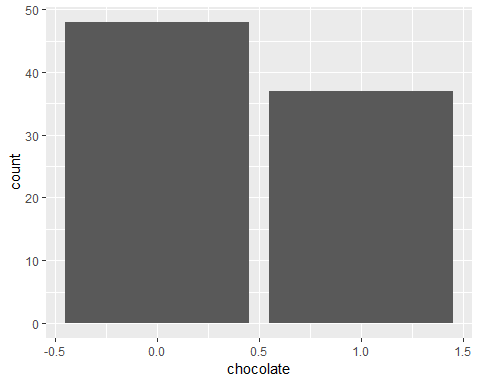
ggplot(candy\_production, aes(x=observation\_date,y=IPG3113N))+  
 geom\_smooth()+  
 labs(title = "Monthly Candy Production",  
 x = "Year",  
 y = "As percent of 2012 production")

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



#### Bar Chart of Ingredients

ggplot(candy\_data, aes(chocolate))+  
 geom\_bar()



candyFeatures <- candy\_data %>% select(2:10)  
 candyFeatures[] <- lapply(candyFeatures, as.logical)  
   
 ggplot(candyFeatures, aes(chocolate))+  
 geom\_bar()

