# Module 3 - Assignment 1

## Tate, Levi

### Data Visualization

I will be using the data sets containing candy rankings and production. These data sets can be found on the Canvas course website under Module 3. This assignment will explore the most popular layers (plots) within ggplot2: geom\_point( ), geom\_bar( ) and geom\_line( ).

library(tidyverse)

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

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.0.6 v dplyr 1.0.4  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

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

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

##   
## -- 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()  
## )

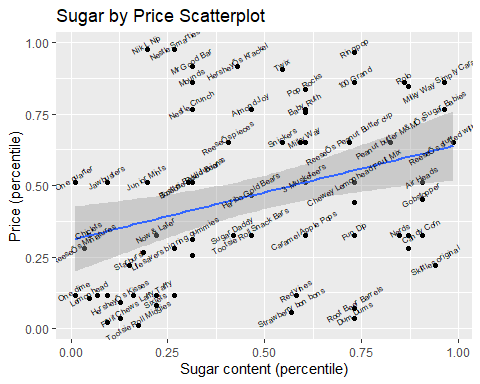
candy\_production <- read\_csv("candy\_production.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## observation\_date = col\_date(format = ""),  
## IPG3113N = col\_double()  
## )

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

ggplot(data = candy\_data, mapping = aes(x = sugarpercent, y = pricepercent, label = competitorname)) +   
 geom\_point() +   
 geom\_smooth(method = "lm") + # adds a fitted line  
 geom\_text(check\_overlap = T,# automatically reduce overlap (deletes some labels)   
 vjust = "bottom", # adjust the vertical orientation  
 nudge\_y = 0.01, # move the text up a bit so it doesn't touch the points   
 angle = 30,# tilt the text 30 degrees  
 size = 2 # make the text smaller (to reduce overlap more)  
 ) + # and then add labels to the points  
 labs(title = "Sugar by Price Scatterplot", # plot title  
 x = "Sugar content (percentile)", # x axis label   
 y = "Price (percentile)" # y axis label  
 )

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



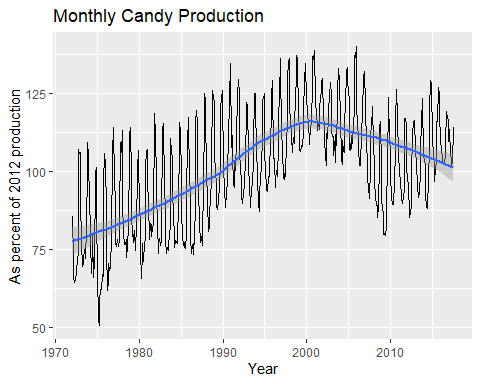
Based upon the results, the candy that has the most sugar and lowest price is Skittles original. Also, the most expensive candy with the highest sugar content is Milky Way Simply Caramel.

#### Line Chart of Candy Production

I am now using the candy production data set that will display a specific date and how production during that month is compared to 2012.

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

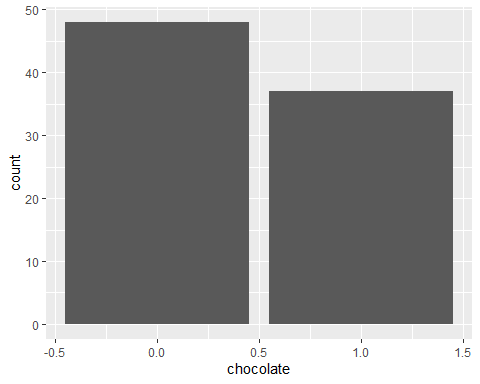
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



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

#### Bar Chart of Ingredients

ggplot(data = candy\_data) +  
 geom\_bar(mapping = aes(x = chocolate))



# select out the columns that have the features of the candy (chocolate, caramel, etc.)  
candyFeatures <- candy\_data %>% select(2:10)  
# make sure that these are booleans (logical)  
candyFeatures[] <- lapply(candyFeatures, as.logical)

ggplot(data = candyFeatures) +  
 geom\_bar( mapping = aes(x = chocolate))

