# Module 3 Assignment 1

## Graybill, Gil

### Data Visualization

This document will be using candy\_data.csv and candy\_production.csv from the course website to create examples about how to visualize data.

# bring in the library  
library(tidyverse)

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

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

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

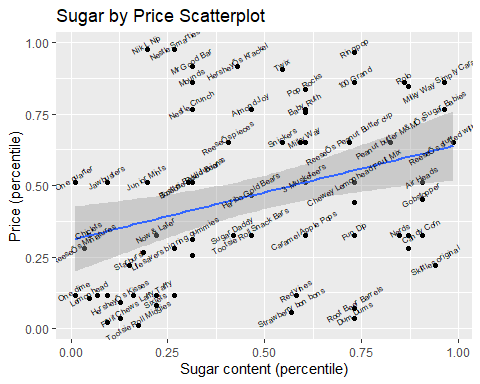
candy\_data <- read\_csv("candy\_data.csv")

## 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",   
 col\_types = cols(observation\_date = col\_date(format = "%m/%d/%Y")))

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

ggplot(data = candy\_data, 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   
 labs(title = "Sugar by Price Scatterplot", # plot title  
 x = "Sugar content (percentile)", # x axis label  
 y = "Price (percentile)" # y axis label  
 )

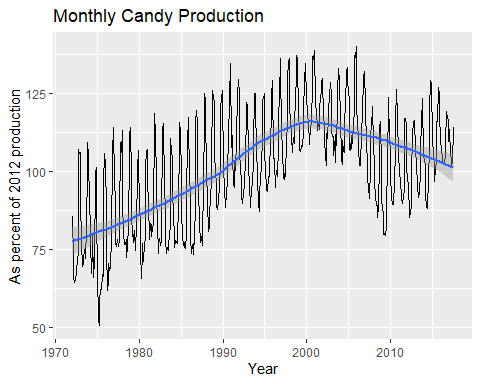


Based on this plot, Skittles Original has the highest sugar content for the lowest price. That makes sense, because Skittles are always the leftover candy at Halloween. Now if your looking for a sugar high but want to spend more money on it, get a Ring Pop. It’s the most expensive candy with the highest sugar content.

#### Line Chart of Candy Production

We are now going to use the candy production dataset to see how monthly candy production has changed over the years compared with 2012, which is known in the candy industry as “The year to compare all others against.”

p <- ggplot(data = candy\_production, aes(x = observation\_date, y = IPG3113N))+  
 geom\_line()+  
 geom\_smooth()+   
 labs(title = "Monthly Candy Production", # plot title  
 x = "Year", # x axis label  
 y = "As percent of 2012 production" # y axis label  
 )  
  
suppressMessages(print(p))



#### Bar Chart of Ingredients

Now let’s look at how many of these candies have chocolate, using a logical features table

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

