# Module 3 - Assignment 1

## Black, Tyler

### Data Visualization

In this assignment, we will be using the candy\_data and candy\_production data sets to produce different graphs, including a scatter plot, line graph, and bar graph.

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readr)  
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.

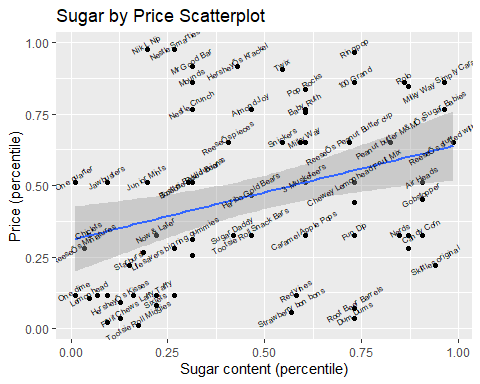
candy\_production <- read\_csv("candy\_production.csv",   
 col\_types = cols(observation\_date = col\_date(format = "%Y-%m-%d")))

#### 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,# 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'

## Warning: The following aesthetics were dropped during statistical transformation: label.  
## ℹ This can happen when ggplot fails to infer the correct grouping structure in  
## the data.  
## ℹ Did you forget to specify a `group` aesthetic or to convert a numerical  
## variable into a factor?

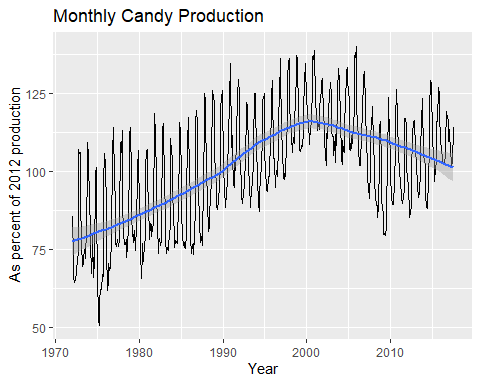


It looks like Reese’s stuffed with pieces is the most sugary candy with the lowest cost. Reese’s stuffed with pieces is the most sugary candy there is, but other candies are more expensive.  
Nestle Smarties is the most expensive candy with the most sugar. There are other candies that have more sugar but they cost less.

#### Line Chart of Candy Production

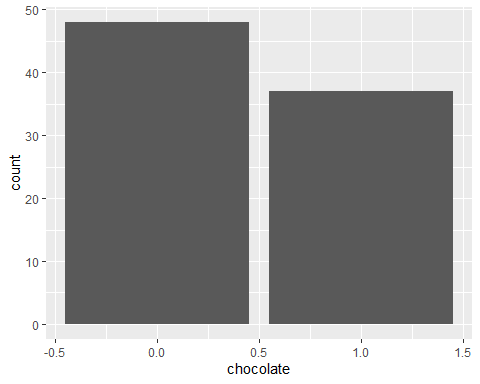
ggplot(data=candy\_production, aes(x=observation\_date,y=IPG3113N)) +  
 geom\_line() +  
 geom\_smooth() +  
 labs(title = "Monthly Candy Production", # plot title  
 x = "Year", # x axis label  
 y = " As percent of 2012 production" # y axis label  
 )

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



#### Bar Chart of Ingredients

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



# 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,aes(x=chocolate)) +  
 geom\_bar()

