# Module 3 - Assignment 1

# Subramanian, Subbu

# Data Visulaization

Will be using the datasets containing candy rankings and production that can be found on the Canvas course website.

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.7 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

candy\_data<-read\_csv("~/Documents/MBA/Programming for Analytics/Module 3/Assignment 1/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.

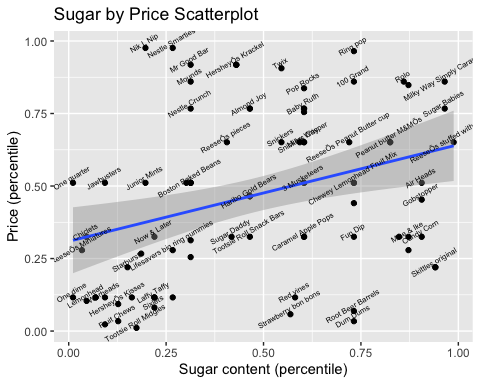
candy\_production <- read\_csv("~/Documents/MBA/Programming for Analytics/Module 3/Assignment 1/candy\_production.csv")

## Rows: 548 Columns: 2  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (1): IPG3113N  
## date (1): observation\_date  
##   
## 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.

# Visualization with Scatterplots (geom\_point)

library(tidyverse)  
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'

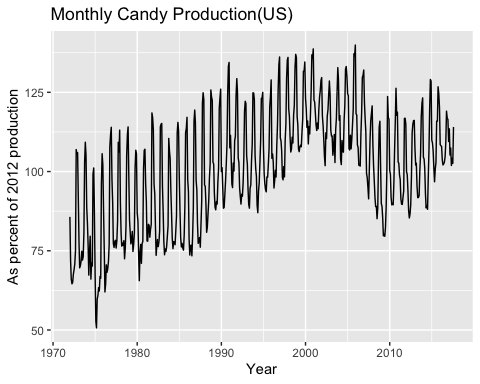


On analysis of the scattered plot of the candy data, “Skittles Original” has the most sugarpercent and lowest pricepercent.

# Line Chart of Candy Production

Will be using the datasets containing candy production dataset which will display a specific date and how production during that month is compared to 2012

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



# Bar Chart of Ingredients

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

