Visualizing Data

Tabitha Hagen

2022-07-22

# Module 3 - Assignment 1

## Tabitha Hagen

### Data Visualization Part 1

The 2 datasets, candy\_data.csv (showing candy rankings) and candy\_production.csv (showing the production of candy) will be used to visualize data. Specifically, the tidyverse package will be used to manipulate how the data sets are viewed and compared after they are imported into R Studio.

library(tidyverse) # import the tidyverse package

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ 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()

# assign the dataset to a variable candy\_data  
candy\_data <- read.csv("candy\_data.csv")   
  
# assign the dataset to a variable candy\_production  
candy\_production <- read\_csv("candy\_production.csv",   
 col\_types = cols(observation\_date = col\_date(format = "%Y-%m-%d")))

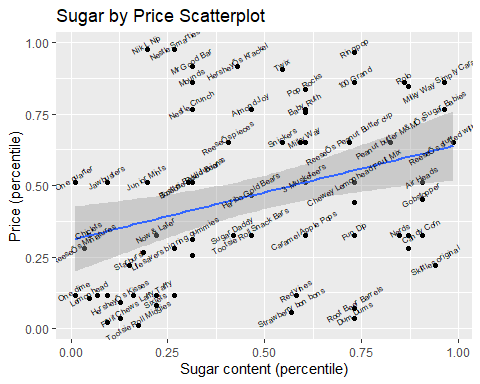
### Data Visualization Part 2 Scatterplots

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

A Scatterplot will be used to compare sugar content to the price of various types of candy on the market.

ggplot(data=candy\_data, aes(x=sugarpercent, y=pricepercent, label=competitorname)) +   
 geom\_point() + # Creates a Scatterplot on the graph  
 geom\_smooth(method = "lm") + # Adds a fitted line to the graph  
 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 on the results, it seems that ReeseÕs stuffed with pieces has the highest sugar content with 0.988 percent sugar.

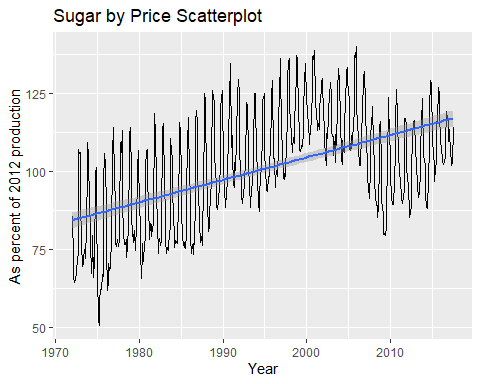
### Data Visualization Part 3 Line Charts

#### Line Chart of Candy Production

Using Tidyverse’s Line chart feature, the candy production dataset will be examined to compare prior date production to that same month in 2012.

ggplot(data=candy\_production, aes(x=observation\_date, y=IPG3113N)) +   
 geom\_line() + # Plot line chart  
 geom\_smooth(method = "lm") + # Plot fitted line  
 labs(title = "Sugar by Price Scatterplot", # Plot Title  
 x = "Year", # x axis label  
 y = "As percent of 2012 production" #y axis Label  
 )

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

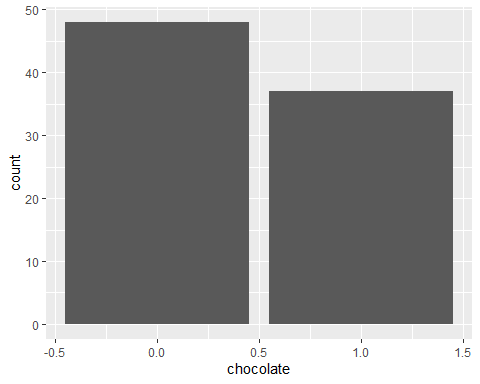


### Data Visualization Part 4 Bar Charts

#### Bar Chart Ingredients

Using Tidyverse’s Bar chart feature and the candy data, the category of Ingredients will be compared to each other.

ggplot(data=candy\_data, aes(x=chocolate )) +   
 geom\_bar() # Adds a barchart to the graph



# 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() # Adds a barchart to the graph

