Chart Crafters

**Exploratory analysis codes**

# Load necessary libraries  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.2

library(readr)  
  
# Read the CSV file  
data <- read\_csv("C:/Depaul/2 quater/Data Visualization/projects/master.csv", show\_col\_types = FALSE)

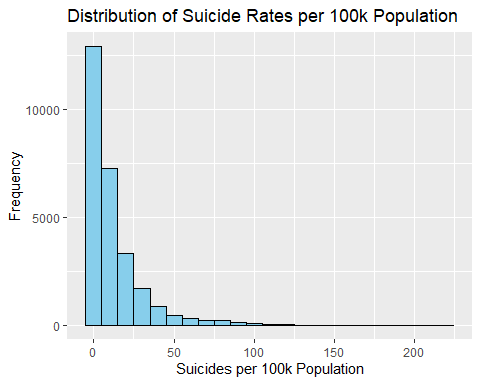
# View the structure of the dataset  
str(data)

## spc\_tbl\_ [27,820 × 12] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ country : chr [1:27820] "Albania" "Albania" "Albania" "Albania" ...  
## $ year : num [1:27820] 1987 1987 1987 1987 1987 ...  
## $ sex : chr [1:27820] "male" "male" "female" "male" ...  
## $ age : chr [1:27820] "15-24 years" "35-54 years" "15-24 years" "75+ years" ...  
## $ suicides\_no : num [1:27820] 21 16 14 1 9 1 6 4 1 0 ...  
## $ population : num [1:27820] 312900 308000 289700 21800 274300 ...  
## $ suicides\_100k\_pop: num [1:27820] 6.71 5.19 4.83 4.59 3.28 2.81 2.15 1.56 0.73 0 ...  
## $ country-year : chr [1:27820] "Albania1987" "Albania1987" "Albania1987" "Albania1987" ...  
## $ HDI for year : num [1:27820] NA NA NA NA NA NA NA NA NA NA ...  
## $ gdp\_for\_year ($) : num [1:27820] 2.16e+09 2.16e+09 2.16e+09 2.16e+09 2.16e+09 ...  
## $ gdp\_per\_capita : num [1:27820] 796 796 796 796 796 796 796 796 796 796 ...  
## $ generation : chr [1:27820] "Generation X" "Silent" "Generation X" "G.I. Generation" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. country = col\_character(),  
## .. year = col\_double(),  
## .. sex = col\_character(),  
## .. age = col\_character(),  
## .. suicides\_no = col\_double(),  
## .. population = col\_double(),  
## .. suicides\_100k\_pop = col\_double(),  
## .. `country-year` = col\_character(),  
## .. `HDI for year` = col\_double(),  
## .. `gdp\_for\_year ($)` = col\_number(),  
## .. gdp\_per\_capita = col\_double(),  
## .. generation = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

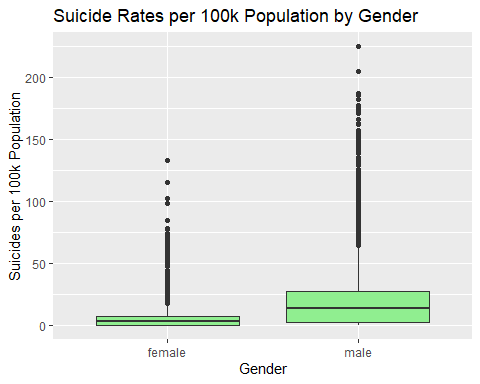
# Summary statistics  
summary(data)

## country year sex age   
## Length:27820 Min. :1985 Length:27820 Length:27820   
## Class :character 1st Qu.:1995 Class :character Class :character   
## Mode :character Median :2002 Mode :character Mode :character   
## Mean :2001   
## 3rd Qu.:2008   
## Max. :2016   
##   
## suicides\_no population suicides\_100k\_pop country-year   
## Min. : 0.0 Min. : 278 Min. : 0.00 Length:27820   
## 1st Qu.: 3.0 1st Qu.: 97498 1st Qu.: 0.92 Class :character   
## Median : 25.0 Median : 430150 Median : 5.99 Mode :character   
## Mean : 242.6 Mean : 1844794 Mean : 12.82   
## 3rd Qu.: 131.0 3rd Qu.: 1486143 3rd Qu.: 16.62   
## Max. :22338.0 Max. :43805214 Max. :224.97   
##   
## HDI for year gdp\_for\_year ($) gdp\_per\_capita generation   
## Min. :0.483 Min. :4.692e+07 Min. : 251 Length:27820   
## 1st Qu.:0.713 1st Qu.:8.985e+09 1st Qu.: 3447 Class :character   
## Median :0.779 Median :4.811e+10 Median : 9372 Mode :character   
## Mean :0.777 Mean :4.456e+11 Mean : 16866   
## 3rd Qu.:0.855 3rd Qu.:2.602e+11 3rd Qu.: 24874   
## Max. :0.944 Max. :1.812e+13 Max. :126352   
## NA's :19456

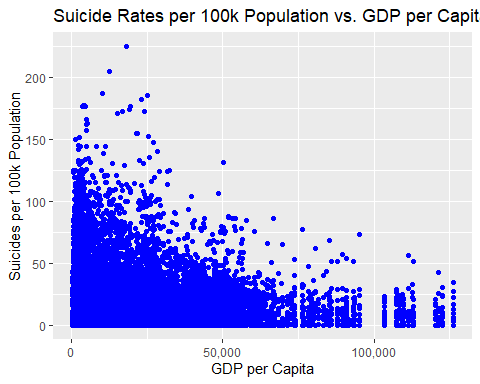
# Data visualization  
# Histogram for suicides per 100k population  
ggplot(data, aes(x = suicides\_100k\_pop)) +  
 geom\_histogram(binwidth = 10, fill = "skyblue", color = "black") +  
 labs(title = "Distribution of Suicide Rates per 100k Population", x = "Suicides per 100k Population", y = "Frequency")



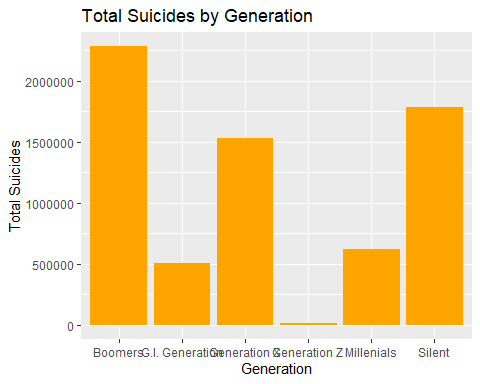
# Box plot for suicides per 100k population by sex  
ggplot(data, aes(x = sex, y = suicides\_100k\_pop)) +  
 geom\_boxplot(fill = "lightgreen") +  
 labs(title = "Suicide Rates per 100k Population by Gender", x = "Gender", y = "Suicides per 100k Population")



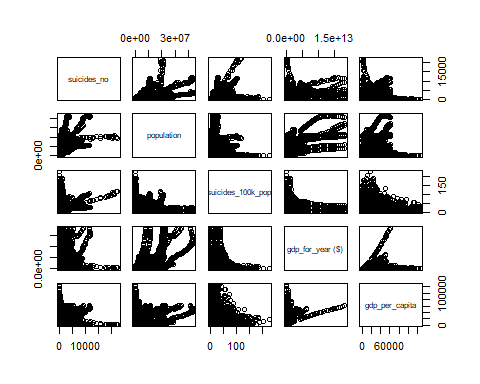
# Scatter plot for suicides per 100k population vs. GDP per capita  
ggplot(data, aes(x = gdp\_per\_capita, y = `suicides\_100k\_pop`)) +  
 geom\_point(color = "blue") +  
 labs(title = "Suicide Rates per 100k Population vs. GDP per Capita", x = "GDP per Capita", y = "Suicides per 100k Population") +  
 scale\_x\_continuous(labels = scales::comma) # Format x-axis labels with commas



# Bar plot for total suicides by generation  
data %>%  
 group\_by(generation) %>%  
 summarise(total\_suicides = sum(suicides\_no)) %>%  
 ggplot(aes(x = generation, y = total\_suicides)) +  
 geom\_bar(stat = "identity", fill = "orange") +  
 labs(title = "Total Suicides by Generation", x = "Generation", y = "Total Suicides")



# Scatter plot matrix for numerical variables  
pairs(data[, c(5:7, 10:11)])



# Correlation matrix  
cor(data[, c(5:7, 10:11)])

## suicides\_no population suicides\_100k\_pop gdp\_for\_year ($)  
## suicides\_no 1.00000000 0.616162268 0.306604451 0.43009585  
## population 0.61616227 1.000000000 0.008284973 0.71069732  
## suicides\_100k\_pop 0.30660445 0.008284973 1.000000000 0.02523964  
## gdp\_for\_year ($) 0.43009585 0.710697323 0.025239644 1.00000000  
## gdp\_per\_capita 0.06132975 0.081509858 0.001785134 0.30340454  
## gdp\_per\_capita  
## suicides\_no 0.061329749  
## population 0.081509858  
## suicides\_100k\_pop 0.001785134  
## gdp\_for\_year ($) 0.303404536  
## gdp\_per\_capita 1.000000000