Assignment\_1

2022-09-14

#Data source: <https://www.kaggle.com/datasets/iamsouravbanerjee/world-population-dataset> #Loading the data

WorldData = read.csv("~/Downloads/Soccer/world\_population.csv")

#Summary of descriptive statistics in this dataset:

summary(WorldData)

## Rank CCA3 Country Capital   
## Min. : 1.00 Length:234 Length:234 Length:234   
## 1st Qu.: 59.25 Class :character Class :character Class :character   
## Median :117.50 Mode :character Mode :character Mode :character   
## Mean :117.50   
## 3rd Qu.:175.75   
## Max. :234.00   
## Continent X2022.Population X2020.Population X2015.Population   
## Length:234 Min. :5.100e+02 Min. :5.200e+02 Min. :5.640e+02   
## Class :character 1st Qu.:4.197e+05 1st Qu.:4.153e+05 1st Qu.:4.047e+05   
## Mode :character Median :5.560e+06 Median :5.493e+06 Median :5.307e+06   
## Mean :3.407e+07 Mean :3.350e+07 Mean :3.173e+07   
## 3rd Qu.:2.248e+07 3rd Qu.:2.145e+07 3rd Qu.:1.973e+07   
## Max. :1.426e+09 Max. :1.425e+09 Max. :1.394e+09   
## X2010.Population X2000.Population X1990.Population   
## Min. :5.960e+02 Min. :6.510e+02 Min. :7.000e+02   
## 1st Qu.:3.931e+05 1st Qu.:3.272e+05 1st Qu.:2.641e+05   
## Median :4.943e+06 Median :4.293e+06 Median :3.825e+06   
## Mean :2.985e+07 Mean :2.627e+07 Mean :2.271e+07   
## 3rd Qu.:1.916e+07 3rd Qu.:1.576e+07 3rd Qu.:1.187e+07   
## Max. :1.348e+09 Max. :1.264e+09 Max. :1.154e+09   
## X1980.Population X1970.Population Area..km.. Density..per.km..   
## Min. : 733 Min. : 752 Min. : 1 Min. : 0.026   
## 1st Qu.: 229614 1st Qu.: 155997 1st Qu.: 2650 1st Qu.: 38.418   
## Median : 3141146 Median : 2604830 Median : 81200 Median : 95.347   
## Mean : 18984617 Mean : 15786909 Mean : 581449 Mean : 452.127   
## 3rd Qu.: 9826054 3rd Qu.: 8817329 3rd Qu.: 430426 3rd Qu.: 238.933   
## Max. :982372466 Max. :822534450 Max. :17098242 Max. :23172.267   
## Growth.Rate World.Population.Percentage  
## Min. :0.912 Min. : 0.0000   
## 1st Qu.:1.002 1st Qu.: 0.0100   
## Median :1.008 Median : 0.0700   
## Mean :1.010 Mean : 0.4271   
## 3rd Qu.:1.017 3rd Qu.: 0.2800   
## Max. :1.069 Max. :17.8800

#Tranformation of data is done below by calculating the percentage change in population from 1970 to 2022:

library(dplyr)

##   
## Attaching package: 'dplyr'

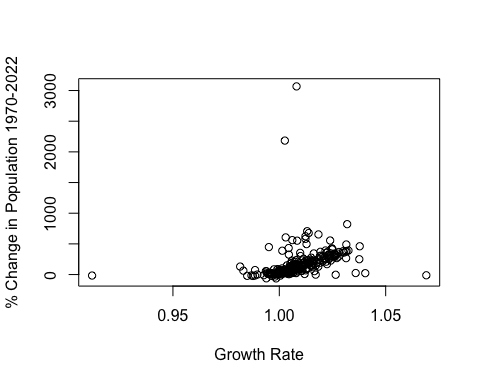
## The following objects are masked from 'package:stats':  
##   
## filter, lag

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

UpdatedWorldData = dplyr::mutate(WorldData, percentageChange\_1970\_2022 = ((WorldData$X2022.Population - WorldData$X1970.Population)/WorldData$X1970.Population)\*100)

#The following is a scatterplot of the population and growth rate:

plot(WorldData$Growth.Rate,UpdatedWorldData$percentageChange\_1970\_2022, xlab = "Growth Rate", ylab ="% Change in Population 1970-2022")

 #The follwoing is histogram of the population in 2022:

hist(WorldData$X2022.Population)

