HM#1 rmarkdown

Steven Indrayana Istimur 20554526

Coding Howework (submmited to Canvas)

Use the dataset Demographics_State.csv, and conduct the following data analysis in R.

```
dat <- read.csv(file = 'Demographics_State.csv')
head(dat)
```

```
##       region total_population percent_white percent_black percent_asian
## 1    alabama          4799277            67            26             1
## 2     alaska           720316            63             3             5
## 3    arizona          6479703            57             4             3
## 4   arkansas          2933369            74            15             1
## 5 california         37659181            40             6            13
## 6   colorado          5119329            70             4             3
##   percent_hispanic per_capita_income median_rent median_age
## 1                4             23680         501       38.1
## 2                6             32651         978       33.6
## 3               30             25358         747       36.3
## 4                7             22170         480       37.5
## 5               38             29527        1119       35.4
## 6               21             31109         825       36.1
```

(1) Compute the average, median, range, standard deviation and quartiles of total_population.

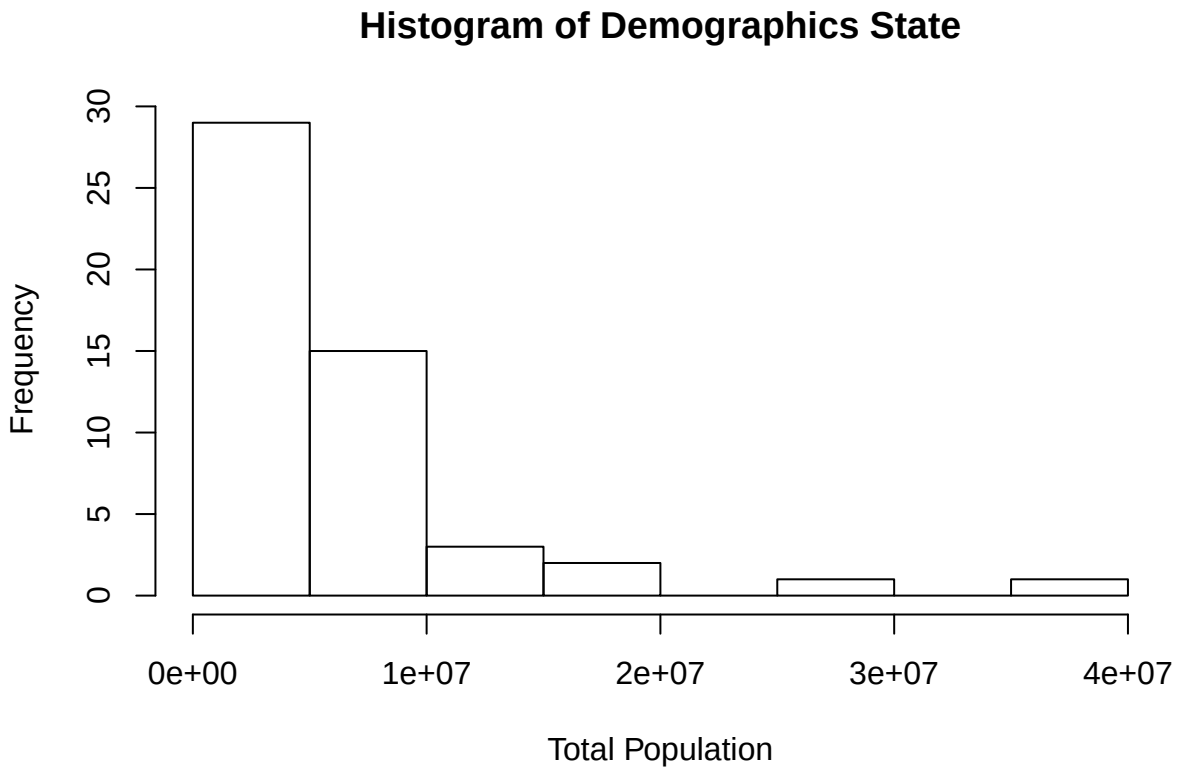
```
#input your r code here
ave = mean(dat$total_population)
med = median(dat$total_population)
ran = max(dat$total_population) - min(dat$total_population)
std = sd(dat$total_population)
qrt = quantile(dat$total_population)

cat(" Average \t: ", ave, '\n',
    "Median \t \t: ", med, '\n',
    "Range  \t \t: ", ran, '\n',
    "Standard Deviation \t: ", std, '\n',
    "Quartile (Q1)\t \t: ", qrt[2],'\n',
    "Quartile (Q2)\t \t: ", qrt[3],'\n',
    "Quartile (Q3)\t \t: ", qrt[4])
```

```
##  Average     :  6108561
##  Median      :  4361333
##  Range       :  37089047
```

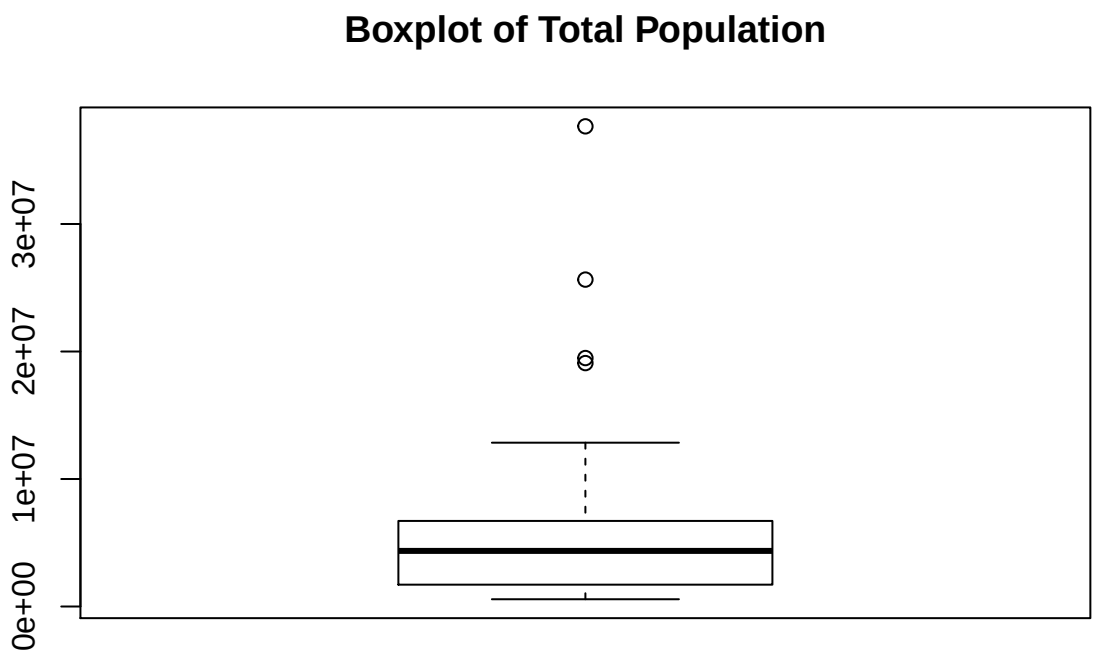
```
##  Standard Deviation  :  6904016
##  Quartile (Q1)       :  1712495
##  Quartile (Q2)       :  4361333
##  Quartile (Q3)       :  6712319
```

(2) Plot the histogram and the boxplot of total_population.

```r
#input your r code here
hist(dat$total_population,
     main = "Histogram of Demographics State",
     xlab = "Total Population")
```

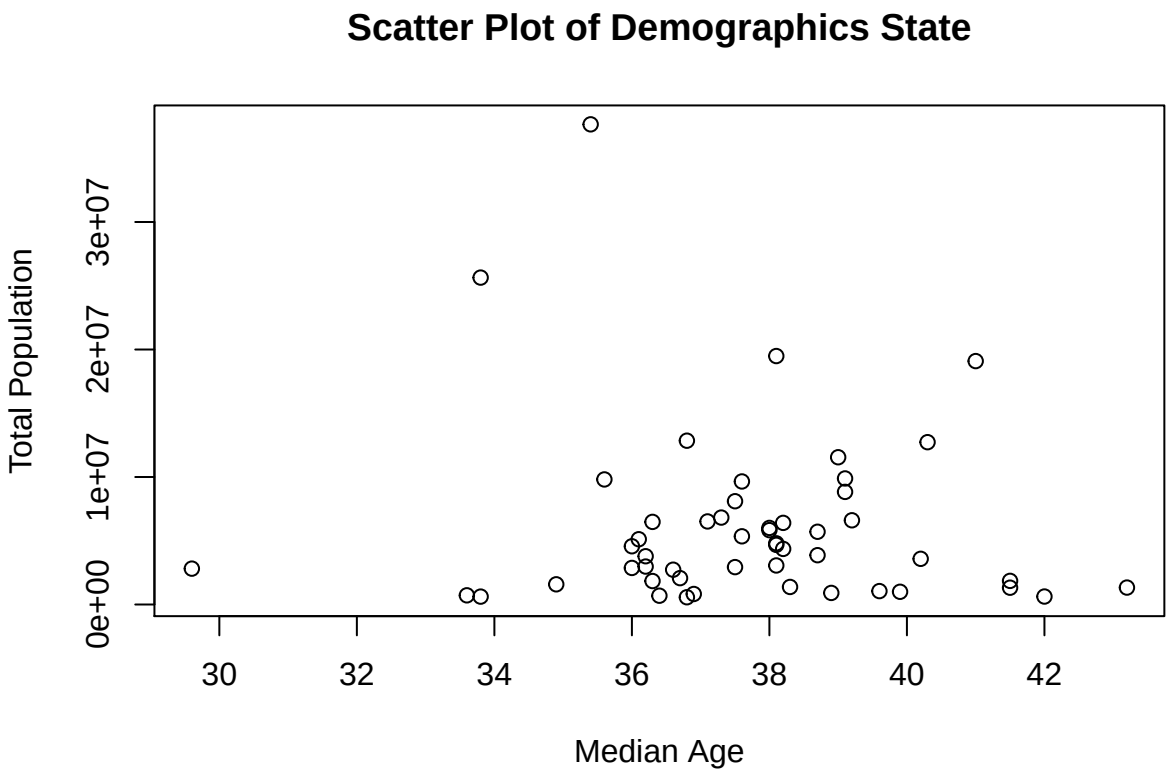



```r
boxplot(dat$total_population,
        main = "Boxplot of Total Population")
```



(3) Draw the scatter plot of total_population versus median_age. Compute the covariance of the two variables without and then with the built-in function; check if the results are the same.

```r
#input your r code here
#scatter plot
plot(dat$median_age,
     dat$total_population,
     main = "Scatter Plot of Demographics State",
     xlab = "Median Age",
     ylab = "Total Population")
```




```
#covariance without built-in function
cov_1 = 0
for (i in c(1:length(dat$total_population))) {
  cov_1 = cov_1 +
        ((dat$total_population[i] - mean(dat$total_population)) *
        (dat$median_age[i]- mean(dat$median_age)))
}
cov_1 = cov_1 / (length(dat$total_population)-1)



#covariance built-in function
cov_2 = cov(dat$total_population, dat$median_age)

cat("The Cov(Total Population, Median Age)\n",
    "without built-in function \t:",
    cov_1,
    "\n with built-in function \t \t:",
    cov_2)
```

```
## The Cov(Total Population, Median Age)
##  without built-in function   : -1537375
##  with built-in function      : -1537375
```