# Report on econ Data Set

Saikat Bera, Presidency University, KOLKATA

December 4, 2022

**Abstract**

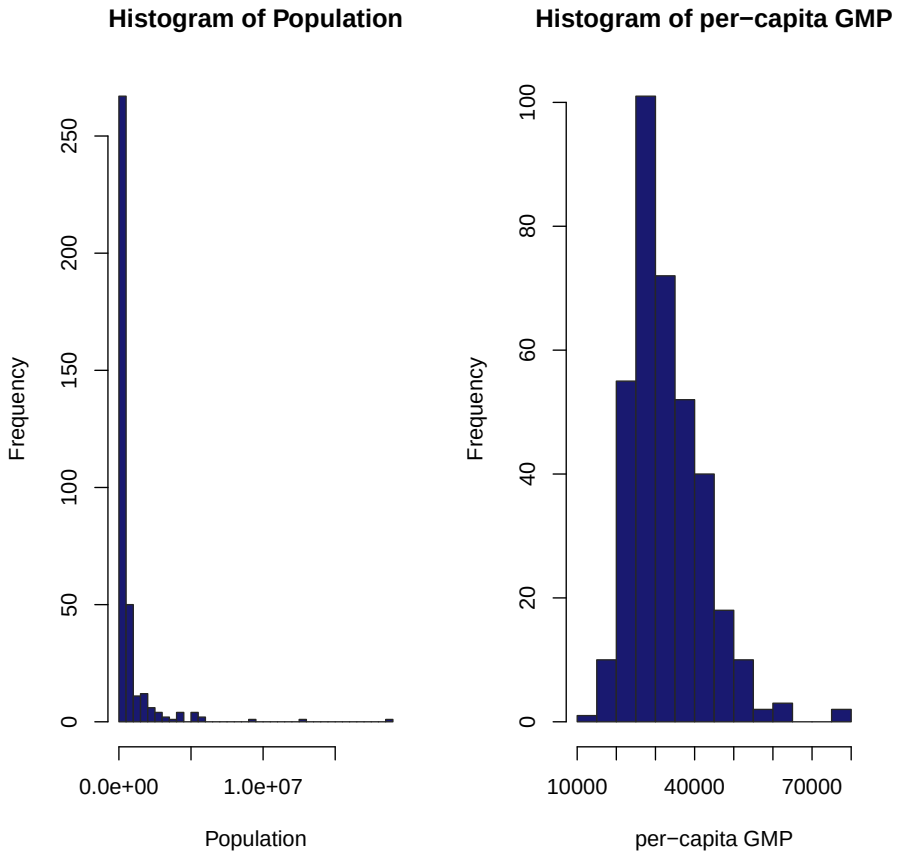
The data file econ.csv contains information about the economies of the 366 "metropolitan statistical areas" ( cities) of the US in 2006. In particular, it lists, for each city, the population, the total value of all goods and services produced for sale in the city that year per person ("per capita gross metropolitan product", pcgmp), and the share of economic output coming from four selected industries.

Loading the data file econ.csv which contains information about the economies of the 366 "metropolitan statistical areas" (≈ cities) of the US in 2006.

```
## [1] 366   7
```

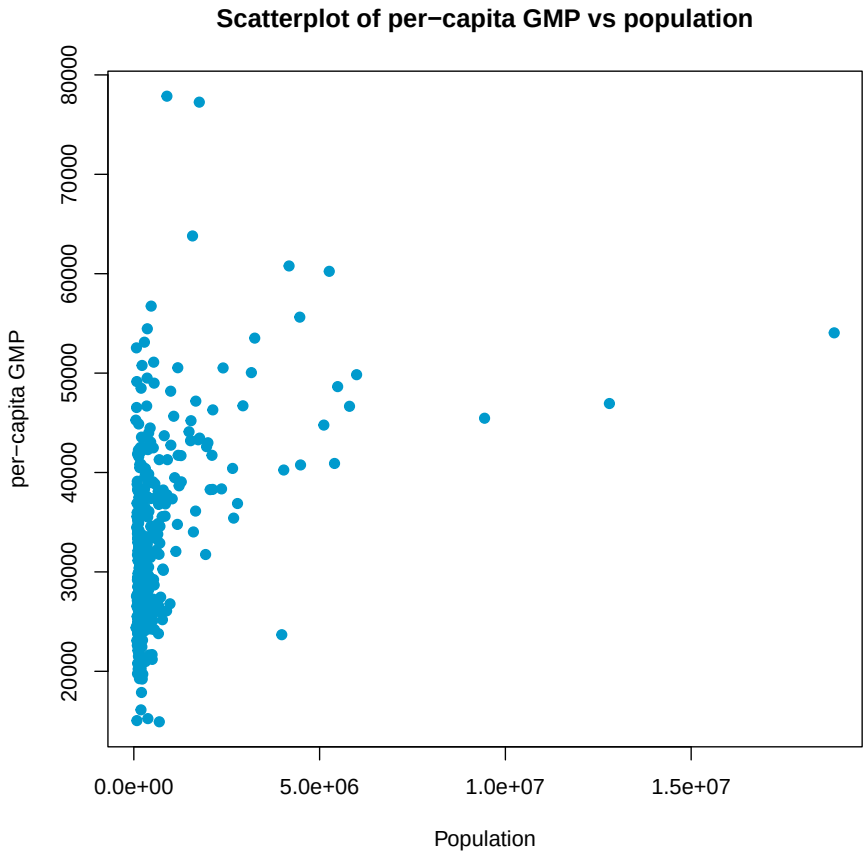
It contains the name of the cities (metropolitan statistical areas) in a column corresponding to each observation. Thats why, it should have seven columns.

```
##      pcgmp            pop              finance          prof.tech
##  Min.   :14920   Min.   :   54980   Min.   :0.03845   Min.   :0.01474
##  1st Qu.:26533   1st Qu.:  135625   1st Qu.:0.10403   1st Qu.:0.02932
##  Median :31615   Median :  231500   Median :0.14140   Median :0.04213
##  Mean   :32923   Mean   :  680898   Mean   :0.15082   Mean   :0.04905
##  3rd Qu.:38213   3rd Qu.:  530875   3rd Qu.:0.18122   3rd Qu.:0.05932
##  Max.   :77860   Max.   :18850000   Max.   :0.38480   Max.   :0.19080
##                                     NA's   :12        NA's   :112
##       ict            management
##  Min.   :0.00349   Min.   :0.00042
##  1st Qu.:0.01215   1st Qu.:0.00294
##  Median :0.02218   Median :0.00651
##  Mean   :0.03910   Mean   :0.00908
##  3rd Qu.:0.04072   3rd Qu.:0.01191
##  Max.   :0.58600   Max.   :0.05431
##  NA's   :76        NA's   :157
```




The distribution of population is highly positively skewed i.e. a large number of cities have a little amount of population, and very few cities have huge population.

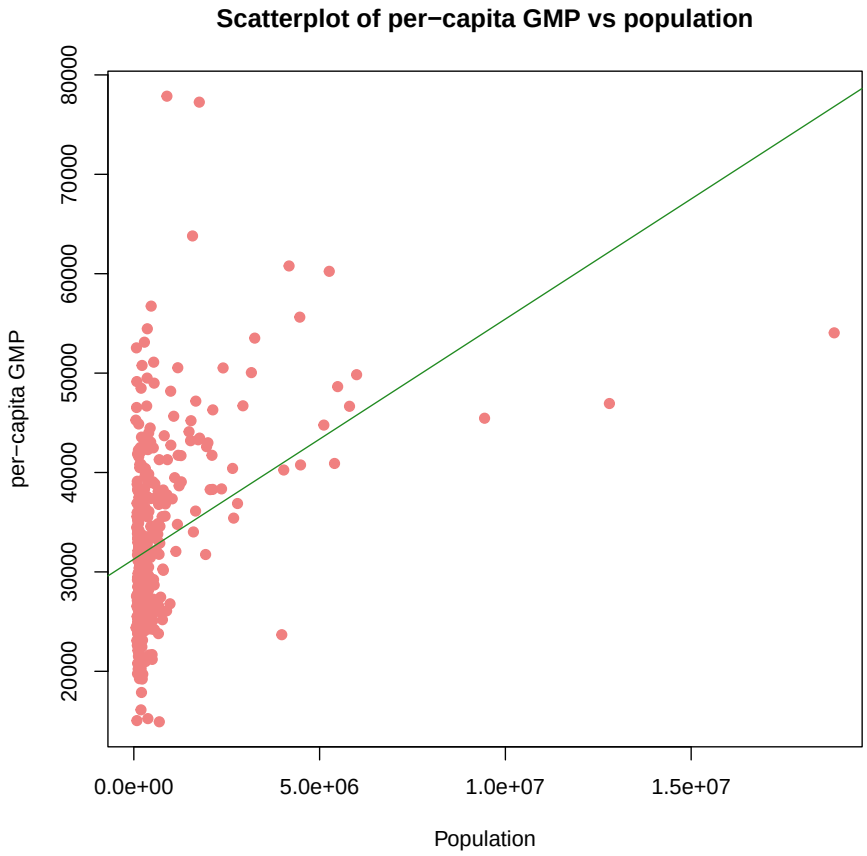
The distribution of per-capita GMP is slightly positively skewed.




```
## [1] 0.002416201
## [1] 31277.57
```

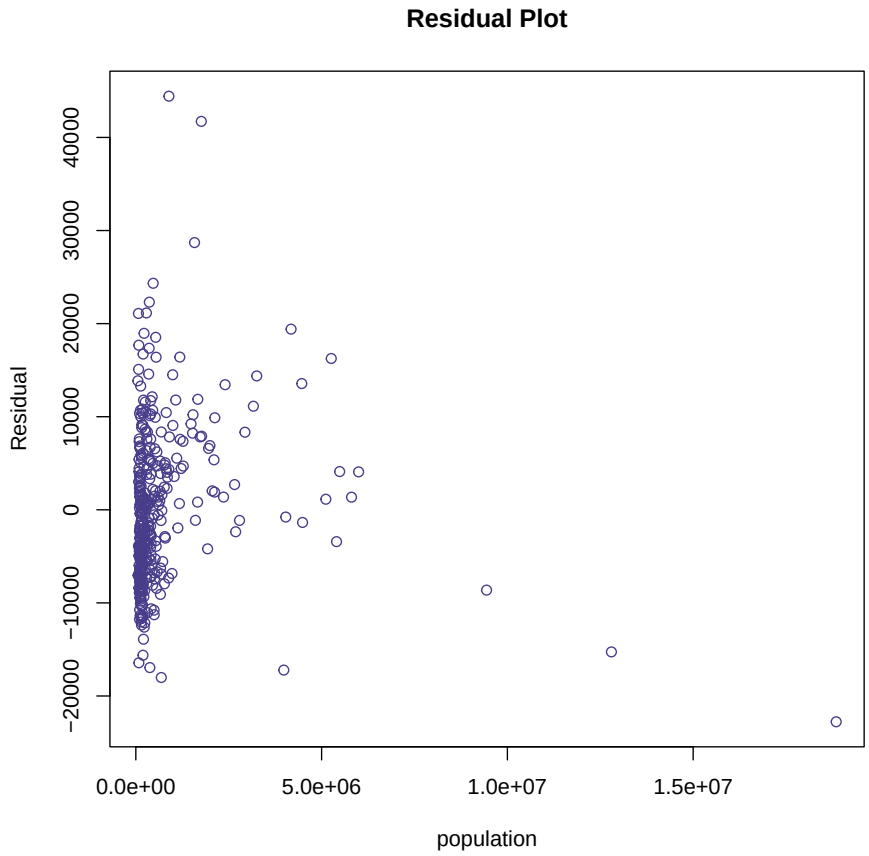
```
##         pop
## 0.002416201
## (Intercept)
##    31277.57
```

It agrees with the answer in the previous part. Yes, it should agree.




The line doesn't fit the data well.

Checking that if the assumptions of the simple linear regression model appear to hold or not.




From the above residual plot we can see that the variance of the residuals is decreasing with x. So, the homoscedasticity assumption of the error variables gets violated here.

```
## [1] 2361000
```

The population of Pittsburgh is 2361000.

```
## [1] 38350
```

The per-capita GMP of Pittsburgh is 38350.

```
## (Intercept)
##    36982.22
```

The per-capita GMP predicted by the model is 36982.22

```
## (Intercept)
##    1367.775
```

The residual for Pittsburgh is 1367.775

```
## [1] 70697145
```

The mean squared error of the regression is 70697145

Consider the ratio of Residual square for Pittsburgh to the Mean Residual Square or MSE:

```
## (Intercept)
##  0.02646231
```

Clearly, the residual square for Pittsburgh is only 2.6% (approx.) of the MSE. Now, as the residual for Pittsburgh is greater than 1, so we can say that the residual for Pittsburgh is quite small compared to the mean squared error.

Interpretation of the estimated slope:

If we select two sets of cases from the un-manipulated distribution where the population differs by 1, we expect per-capita GMP to differ by 0.002416201 (=estimated slope) unit.

```
## (Intercept)
##    37223.84
```

The predicted per-capita GMP for a city with $10^5$ more people than Pittsburgh is 37223.84

If $10^5$ people were added to the population, by a policy intervention, then the predicted Pittsburgh per-capita GMP would become more closer to the observed per-capita GMP i.e. the residual would decrease.