```
> library(readr)
> #DataAir <- read.csv("AIR QUALITY INDEX- top countries.csv")
> DataAir1 <- data.frame(AIR_QUALITY_INDEX_top_countries)
> #4 columns selected
> hist(DataAir1$X2021, main ="2021 Air Quality Index", ylab = "Frequency", xlab = "Quality Index")
> hist(DataAir1$X2020, main ="2020 Air Quality Index", ylab = "Frequency", xlab = "Quality Index")
> hist(DataAir1$X2019, main ="2019 Air Quality Index", ylab = "Frequency", xlab = "Quality Index")
> hist(DataAir1$X2018, main ="2018 Air Quality Index", ylab = "Frequency", xlab = "Quality Index")
> #removing outliers/missing data
> DataAir2 <-DataAir1[rowSums(is.na(DataAir1)) == 0, ]
> hist(DataAir2$X2021, main ="2021 Air Quality Index", ylab = "Frequency", xlab = "Quality Index")
> hist(DataAir2$X2020, main ="2020 Air Quality Index", ylab = "Frequency", xlab = "Quality Index")
> hist(DataAir2$X2019, main ="2019 Air Quality Index", ylab = "Frequency", xlab = "Quality Index")
> hist(DataAir2$X2018, main ="2018 Air Quality Index", ylab = "Frequency", xlab = "Quality Index")
> #Computing the mean, median, variance abd standard deviation
> #Year 2021 column
> mean(DataAir2$X2021)
[1] 20.31944
> median(DataAir2$X2021)
[1] 15.95
> var(DataAir2$X2021)
[1] 200.6072
> sd(DataAir2$X2021)
[1] 14.16359
> #Year 2018 column
> mean(DataAir2$X2018)
[1] 24.63472
> median(DataAir2$X2018)
[1] 18.6
> var(DataAir2$X2018)
[1] 337.8902
> sd(DataAir2$X2018)
[1] 18.38179
> #Scatterplot and correlation of column
> plot(Population~X2021, data = DataAir2, ylab = "Population", xlab = "Air Quality Index", main = "A
uality Index 2021 vs Population", col ="dodgerblue")
> cor(DataAir2$X2021, DataAir2$Population)
[1] 0.3571449
> #Confidence Intervals
> #Year 2021
> model <- lm(X2021~1,DataAir2 )
> confint(model, level=0.95)
               2.5 %   97.5 %
(Intercept) 16.99116 23.64772
> #Year 2018
> model2 <- lm(X2018~1, DataAir2)
> confint(model2, level=0.95)
               2.5 %   97.5 %
(Intercept) 20.31521 28.95423
> #fit of the model Year 2021
> summary(model)

Call:
```

```
lm(formula = X2021 ~ 1, data = DataAir2)

Residuals:
    Min      1Q  Median      3Q     Max
-15.519  -9.144  -4.369   5.106  56.581

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   20.319      1.669   12.17   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.16 on 71 degrees of freedom

> #fit of the model Year 2018
> summary(model2)

Call:
lm(formula = X2018 ~ 1, data = DataAir2)

Residuals:
    Min      1Q  Median      3Q     Max
-19.635 -12.710  -6.035   5.315  72.465

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   24.635      2.166   11.37   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.38 on 71 degrees of freedom

> #Histogram of the residual Year 2021
> residual1 <- residuals(model)
> hist(residual1, xlab = "x-axis", ylab = "y-axis", main = "Residual 2021")
> #Histogram of the residual Year 2018
> residual2 <- residuals(model2)
> hist(residual2 , xlab = "x-axis", ylab = "y-axis", main = "Residual 2018")
> plot(residual1)
> plot(residual2)
>

>
```