```
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$X2021)
median(DataAir2$X2021)
var(DataAir2$X2021)
sd(DataAir2$X2021)

#Year 2018 column
mean(DataAir2$X2018)
median(DataAir2$X2018)
var(DataAir2$X2018)
sd(DataAir2$X2018)

#Scatterplot and correlation of column
plot(Population~X2021, data = DataAir2, ylab = "Population", xlab = "Air Quality Index",
main = "Air Quality Index 2021 vs Population", col ="dodgerblue")
cor(DataAir2$X2021, DataAir2$Population)

#Confidence Intervals
#Year 2021
model <- lm(X2021~1,DataAir2 )
confint(model, level=0.95)
#Year 2018
model2 <- lm(X2018~1, DataAir2)
confint(model2, level=0.95)
#fit of the model Year 2021
summary(model)
#fit of the model Year 2018
summary(model2)

#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)
```