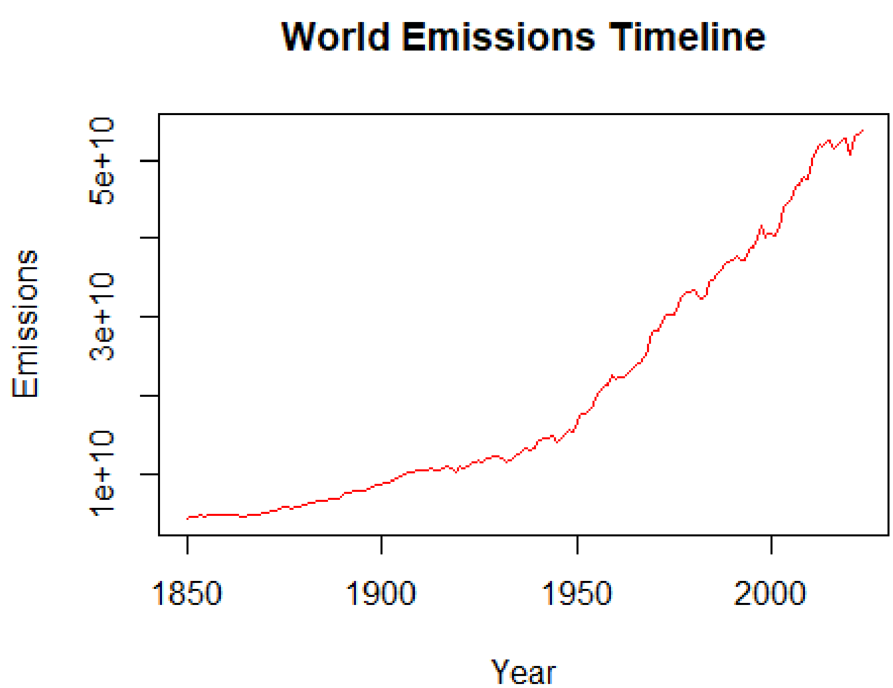
Environmental Data Analysis Code

Darshita Guckhool

2025-06-12

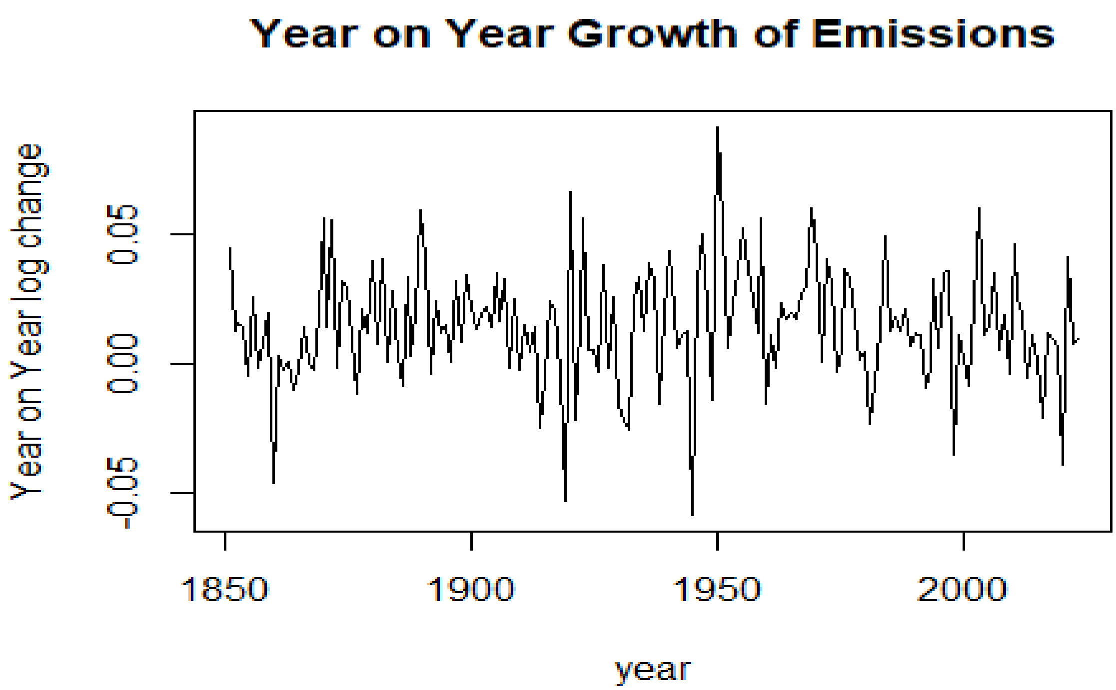
```r
rm(list = ls())
library(rio)

library(tseries)

library(forecast)

library(vars)

library(dplyr)

library(ggplot2)

data.emissions <- import("total-ghg-emissions.csv")
split_data <- split(data.emissions, data.emissions$Entity)
data.world<- subset(split_data$World, select = c(Year,Emissions))
years <- data.world$Year
plot(data.world, type="l", col= "red", main="World Emissions Timeline" )
```




```r
The Line graph above represents the timeline as to how global GHG emissions
```

have been progressing as time passes by. The graph shows that GHG emissions have rapidly increased, but we also notice that there is a shift around the 1950s where GHG emissions had a rapid increase. However, it is important to note that this graph simply shows the progression and not the year-on-year growth, meaning emissions could be decreasing compared to the previous years but it will not distinctively be seen on this graph since we are also comparing it to the 1850s.

```
emissions.yoy <- diff(log(data.world$Emissions), lag = 1)
years <- data.world$Year[-1]
plot(years,emissions.yoy, type="l", xlab= "year",ylab="Year on Year log change"
     ,main = "Year on Year Growth of Emissions")
```



To better show if emissions are actually rapidly increasing or decreasing we can consider the year-on-year growth, this allows us to see if the rate in emissions is increasing or decreasing.

As the graph below shows, the emissions rates have constantly been fluctuating, however notice as we go past the 2000s, we see that the peaks and troughs are shortening towards the center indicating fewer fluctuation and a decreasing rate in GHG emissions. However, a lot more decrease is necessary to achieve net zero emissions.

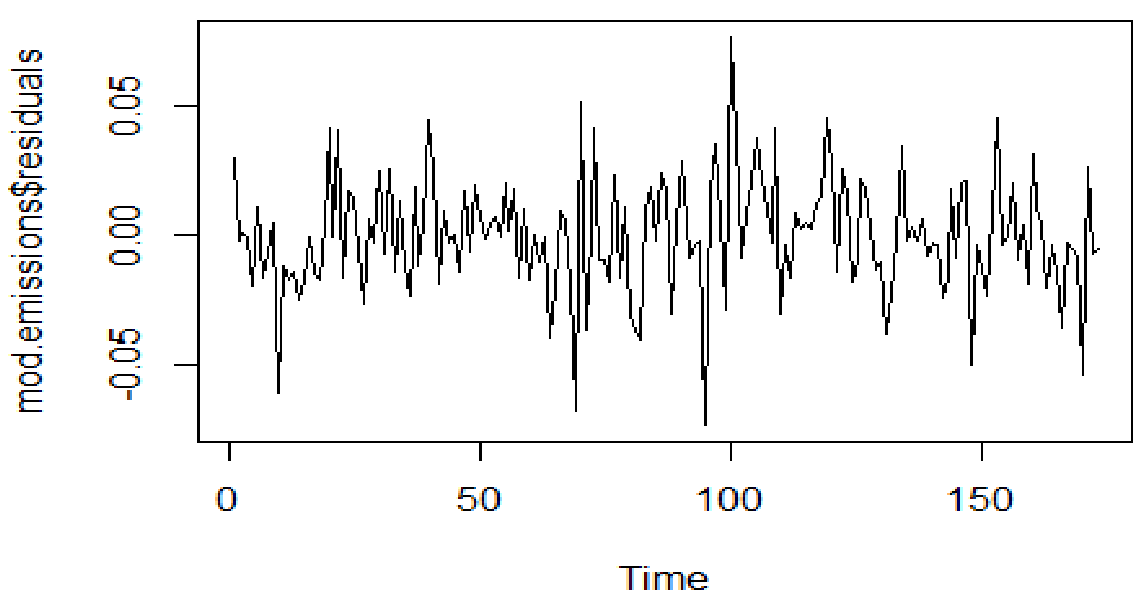
```
mod.emissions <- auto.arima(emissions.yoy, seasonal = FALSE, stationary = TRUE,
```

```
                                ic = "bic")
summary(mod.emissions)
```

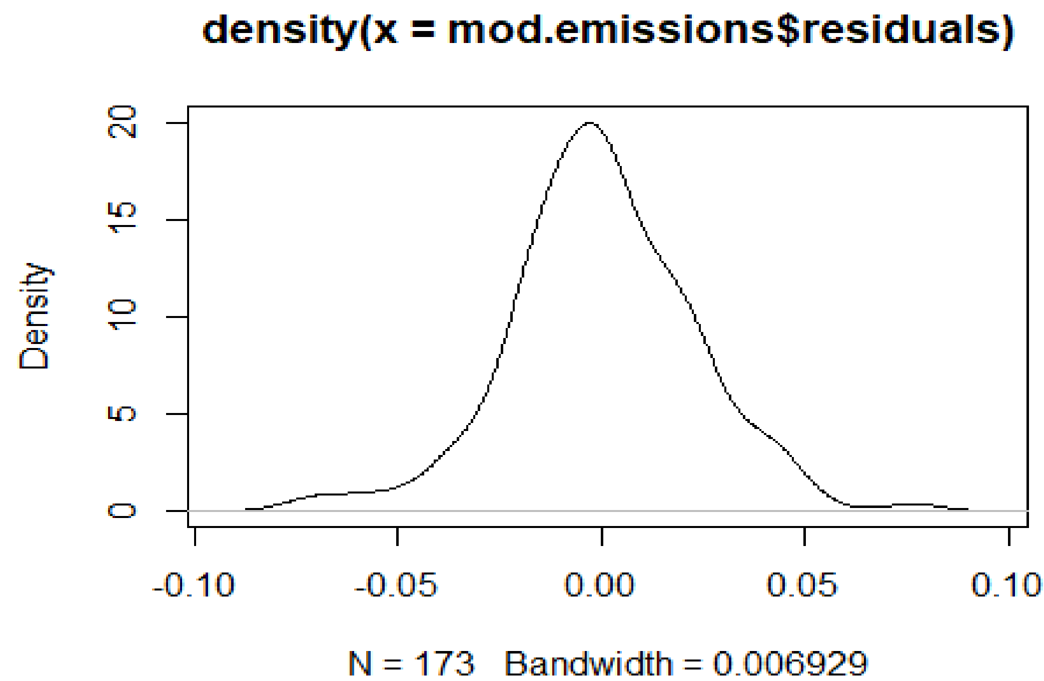
```
## Series: emissions.yoy
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##         mean
##       0.0147
## s.e.  0.0017
##
## sigma^2 = 0.0005069:  log likelihood = 411.31
## AIC=-818.62   AICc=-818.55   BIC=-812.32
##
## Training set error measures:
##                          ME       RMSE         MAE       MPE     MAPE
MASE
## Training set -1.886127e-18 0.02245003 0.01695938 -4.018194 237.5455
0.6848337
##                     ACF1
## Training set -0.01025181
```

Below is a simple ARIMA calculation needed to be able to create the neccessary forecast

```
plot(mod.emissions$residuals)
```



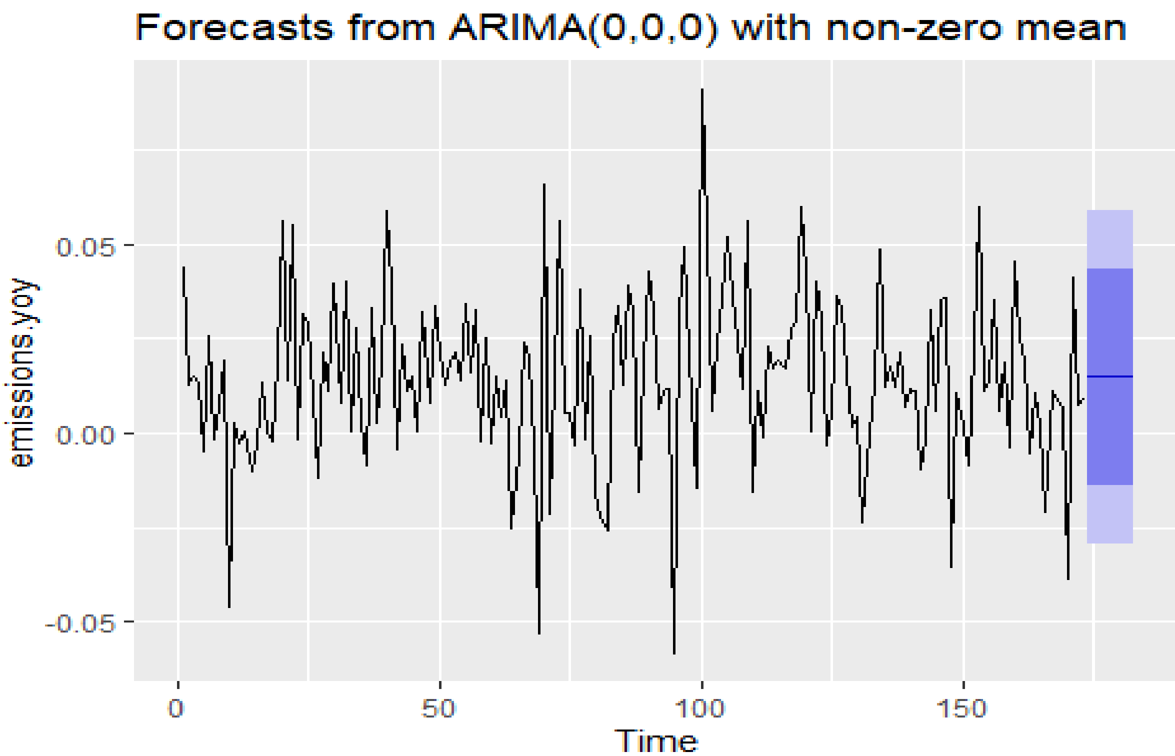

```
plot(density(mod.emissions$residuals))
```




```
ehat <- mod.emissions$residuals
summary(Arima(ehat, order = c(1,0,0), include.mean = FALSE))

## Series: ehat
## ARIMA(1,0,0) with zero mean
##
## Coefficients:
##           ar1
##       -0.0103
## s.e.   0.0762
##
## sigma^2 = 0.0005069:  log likelihood = 411.32
## AIC=-818.64   AICc=-818.57   BIC=-812.34
##
## Training set error measures:
##                          ME       RMSE        MAE      MPE     MAPE
MASE
## Training set 3.229499e-07 0.02244884 0.01694645 96.59378 100.0967
0.6843114
##                       ACF1
## Training set 9.54118e-05

adf.test(emissions.yoy, alternative = c("stationary"))

## Warning in adf.test(emissions.yoy, alternative = c("stationary")): p-value
## smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  emissions.yoy
## Dickey-Fuller = -5.1178, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

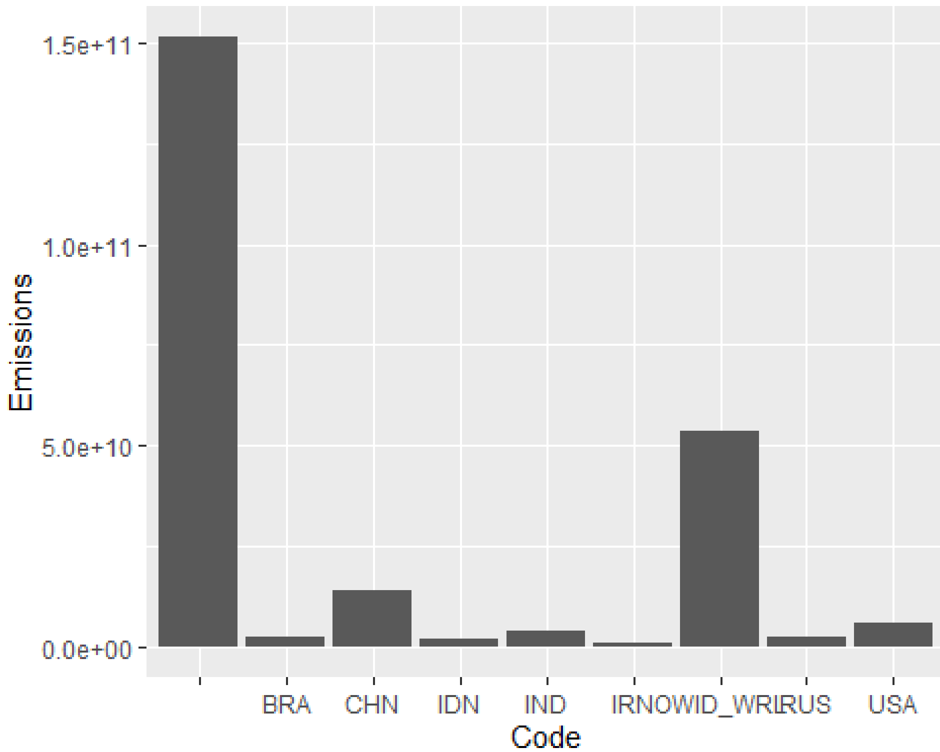
```r
start.year<- min(data.world$Year)+1
yoy_ts <- ts(emissions.yoy, start = start.year, frequency = 1)
emissions_model <-Arima(yoy_ts, order = c(0,0,0), include.mean = TRUE)
fc.emissions <- forecast(mod.emissions, h=10)
autoplot(fc.emissions)
```



The grpah below shows the predicted forecast of the year on year growth of emissions for the next 10 years (until 2033) we can see that emissions are predicted to almost flatline meaning the year on year change in emissions will not be changing, this doesnt mean that emissions are going to be zero, it simply means there is a high likely chance GHG emissions will not increase or decrease from the previous year. Of course this is a simple forecast, and the graph highlights other possibilities including peaks and troughs, the darker the shaded area the higher likely probability.

The predicition above does have its flaws, for one it is purely based on the historic distributiion of GHG emissions, meaning instances like the industrial boom, where emissions rose drastically and the covid-19 pandemic where emissions drastically reduced are reference points for this prediction. This forecast also does not account for political or economic changes that *may drastically affect how we as humans perceive emissions and prioritize enviromental conservation.*

```
split_data_year <- split(data.emissions, data.emissions$Year)
data_2023 <- split_data_year$`2023`
top_25<-data_2023 %>%
  top_n (25, data_2023$Emissions)
ggplot(top_25, aes(x=Code, y=Emissions)) + geom_bar(stat="Identity")
```



The bar chart above refelcts the highest emitters of GHG in the World where the first bar is the total emissions in the world and the remainders are the top 8 emitters.

References:

#Our World in Data – CO_2 and Greenhouse Gas Emissions Dataset
#World Bank Climate Data
#NASA Climate Reports
#IPCC Greenhouse Gas Emissions Inventory