STAT 443: Lab 1

Alex Grinius 20712972

11 January, 2025

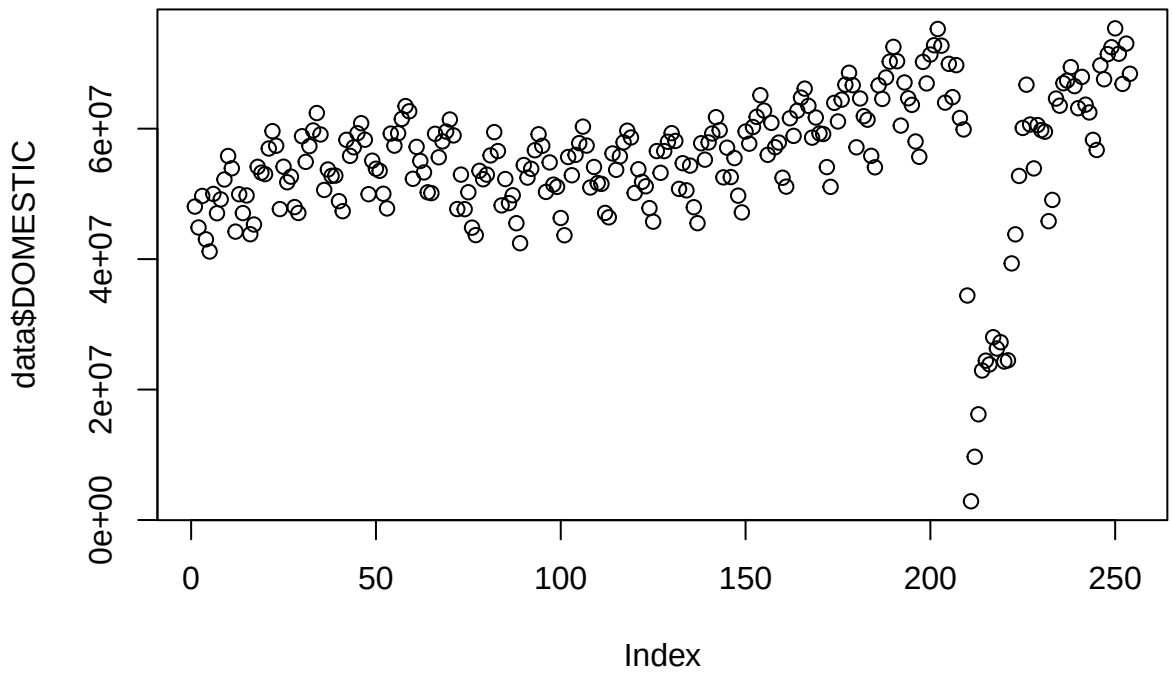
Question 1

(a)

```
data <- read.csv("dat_Passengers.csv", header = TRUE)
head(data)
```

```
##   Year Month DOMESTIC INTERNATIONAL    TOTAL
## 1 2002    10 48054917       9578435 57633352
## 2 2002    11 44850246       9016535 53866781
## 3 2002    12 49684353      10038794 59723147
## 4 2003     1 43032450       9726436 52758886
## 5 2003     2 41166780       8283372 49450152
## 6 2003     3 49992700       9538653 59531353
```

```
plot(data$DOMESTIC)
```



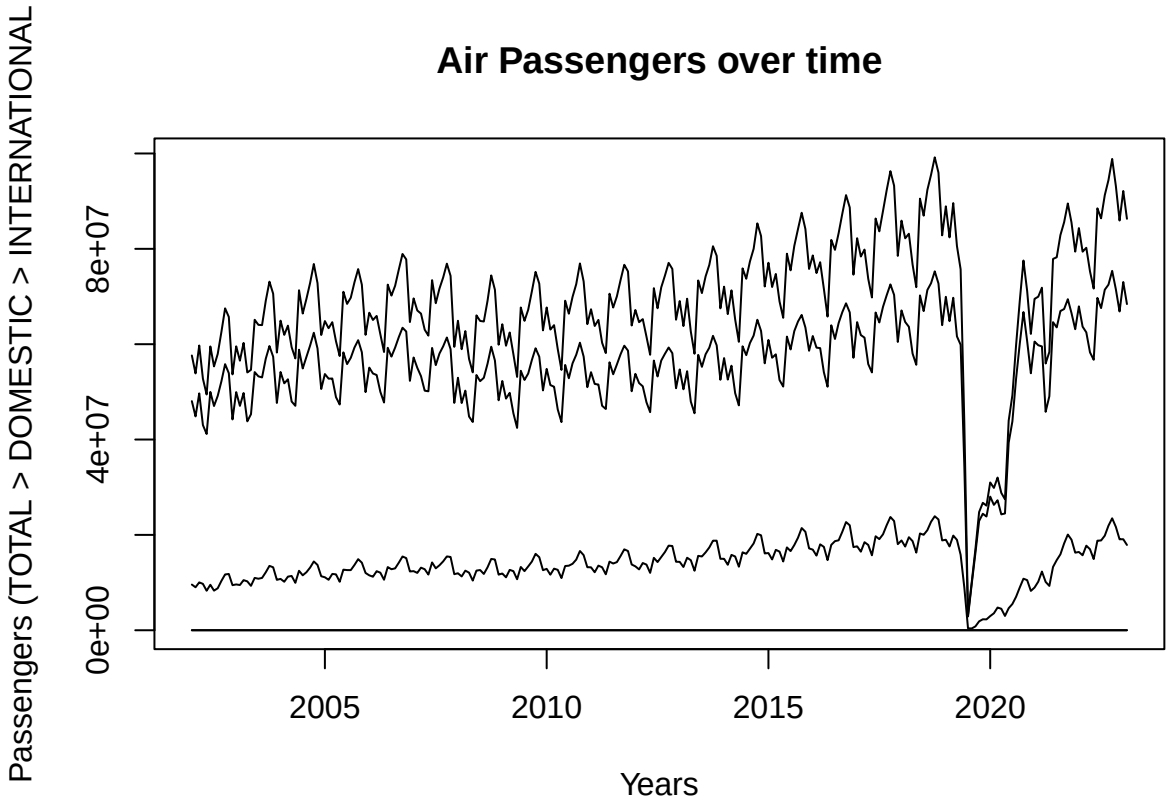
There is no clean way to show the date accocated with the data point as the month and year columns are split in 2

(b)

```
time_series <- ts(data = data, start = 2002, frequency = 12)
```

(c)

```
plot(time_series,plot.type = "single", xlab = "Years", main = "Air Passengers over time", ylab = "Passe
```



Here we can see the pandemic had a large effect on the number of passengers, much larger than the 2008 financial crisis. Number of passengers eems to be increasing over time, retuning to the trendline quickly after covid. We can also see periodicity in the number of passengers each year, having spikes over the summer and holiday season.

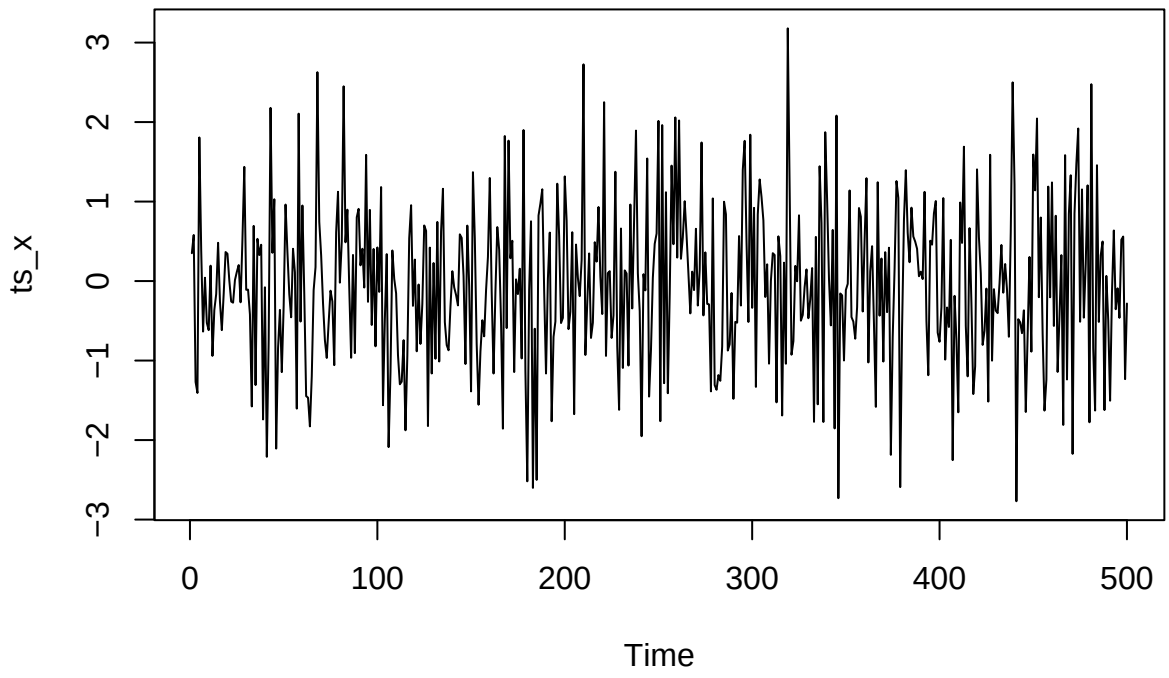
Question 2

(a)

```
set.seed(443)
X <- rnorm(500)
ts_x <- ts(data = X, start = 1, frequency = 1)
```

(b)

```
plot(ts_x)
```



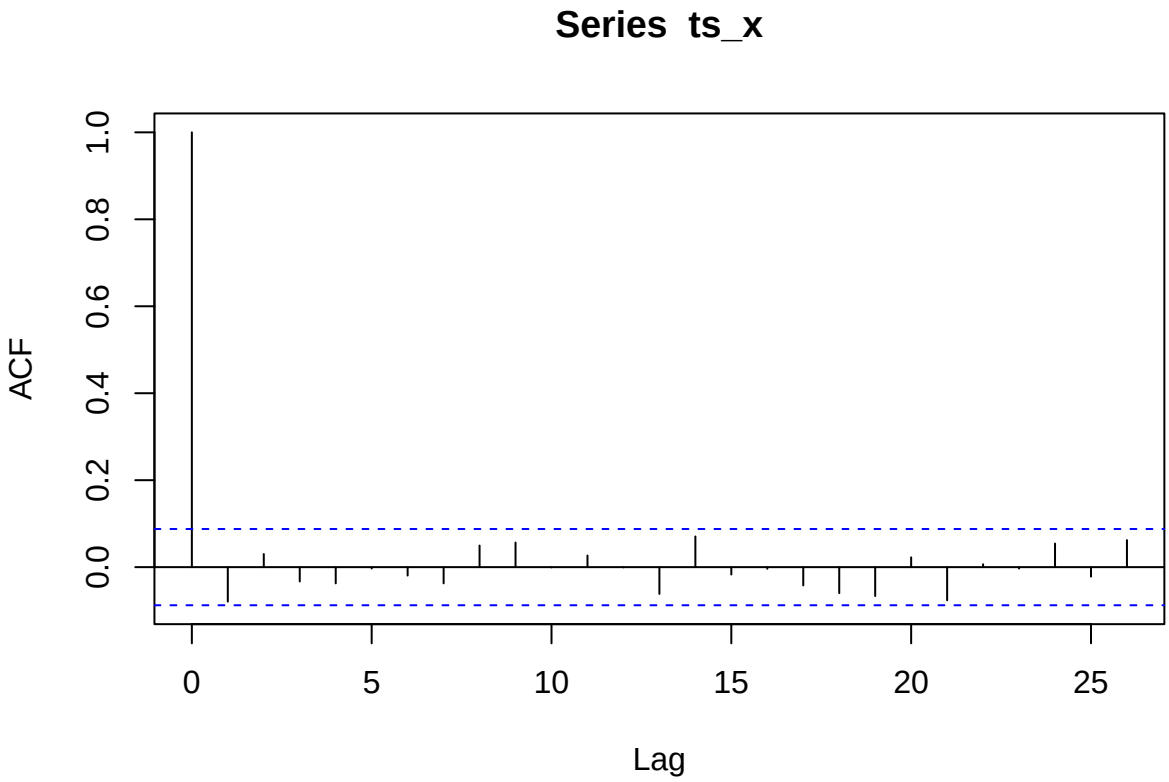

```
sum(abs(ts_x) > 1)
```

```
## [1] 157
```

There are 157 data points that are more than one standard deviation away from the mean, which is close to the 33.3% we are looking for

(c)

```
acf_x <- acf(ts_x)
```



This graph only has strong correlation at lag = 0, i.e. with itself. This is what we'd expect from a graph that was generated by purely random data.