STAT 443: Lab 4

Flora Zhang 52135365

11 Feb, 2022

MAKE SURE TO NAME ALL THE GRAPHS

Question 1: Without using any mathematical notation, describe in words what it means for a time series to be stationary.

A time series is stationary when its properties do not depend on the time at which the series is observed. It does not have any predictable patterns in the long-run. Therefore, its mean is constant, variance is finite, and ACF will only depend on the lag. For example, a white noise series is stationary, since it looks the same at any point in time, aka, it does not depend on time. Therefore, if a time series has a trend or seasonality, it will not be stationary. The trend/ seasonality will affect the value at different times, aka, its properties depends on time.

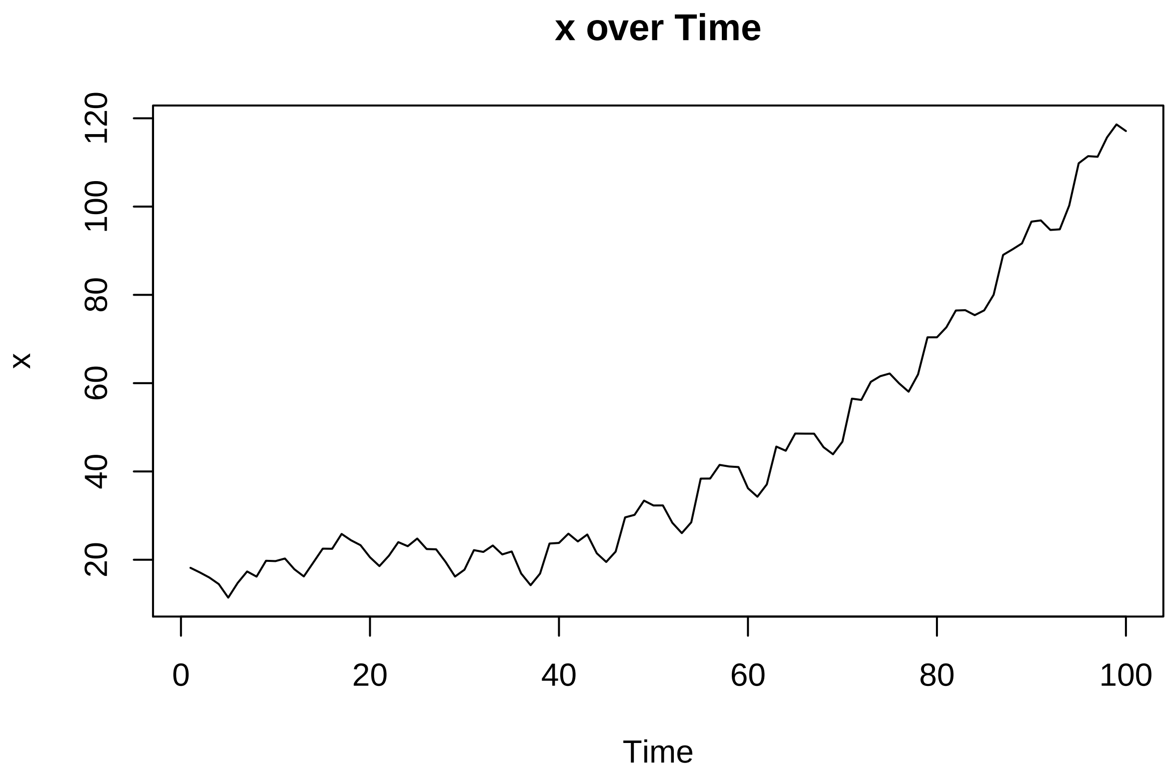
Question 2:

```
lab4data = read_csv("lab4data.csv")
```

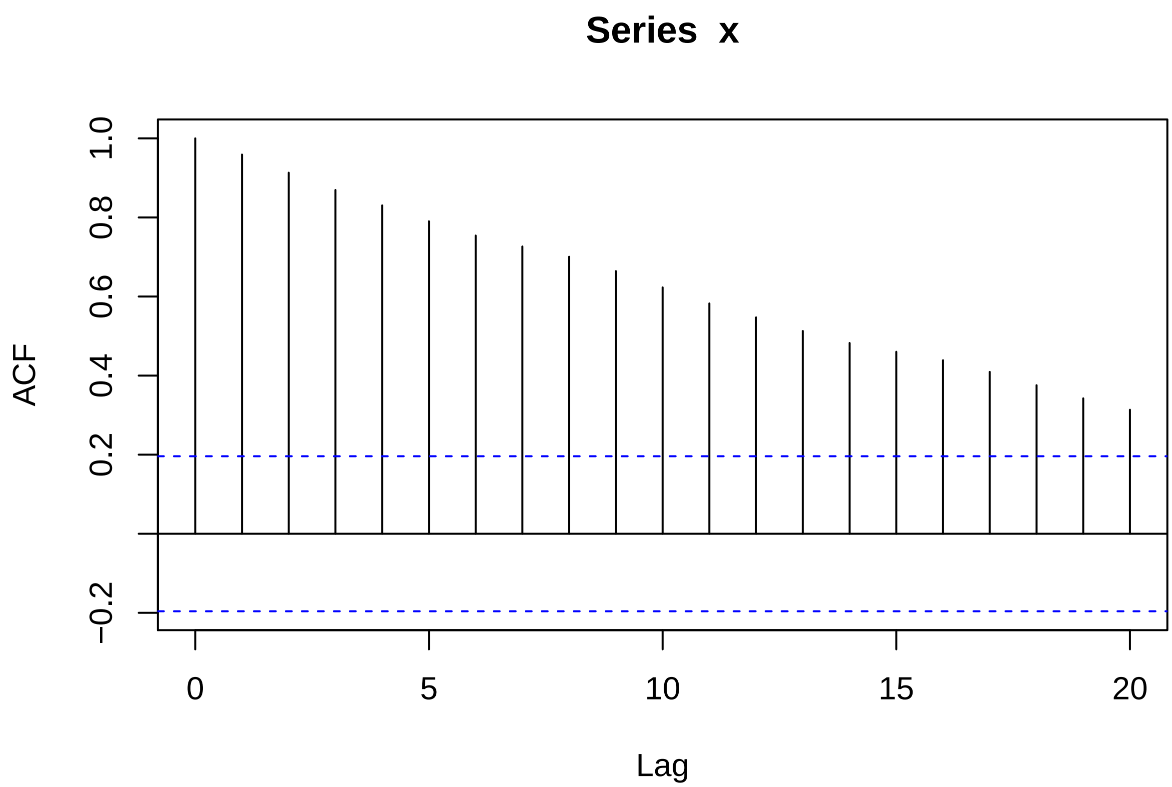
```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   X1 = col_double(),
##   x = col_double()
## )
```

```
x = ts(data=lab4data$x, start=c(1), end=c(100))
plot(x, main="x over Time")
```



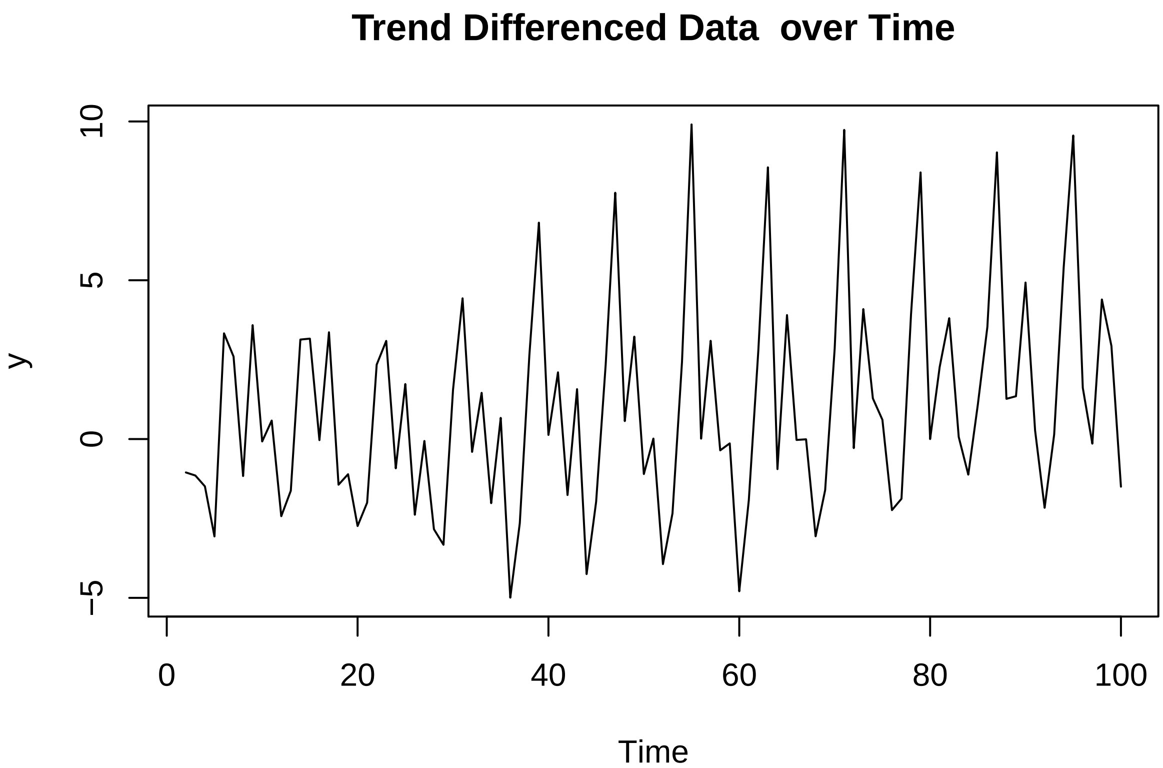

```
acf(x)
```



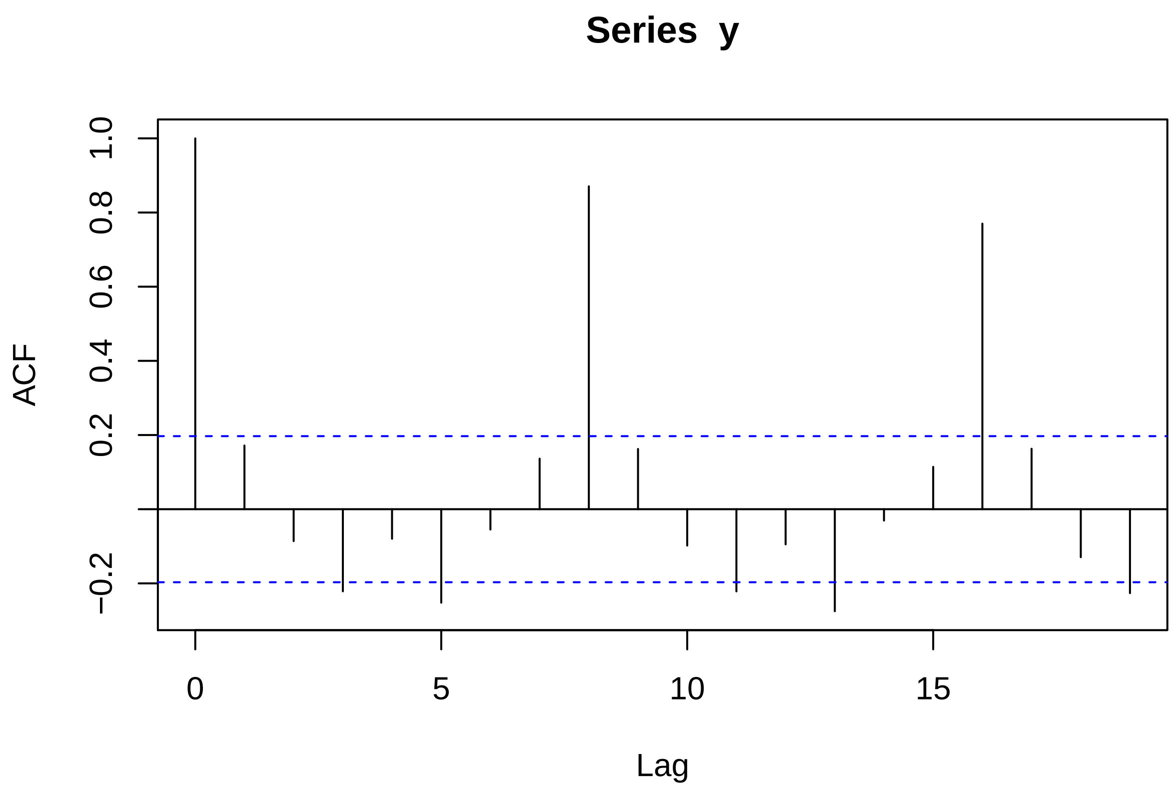
From the plot, we can see that there is an upwards trend in the data as time increases, meaning its mean is not constant. Additionally, the acf has a very slow decay, which further indicates that there is a trend in the data. There also appears to be seasonality. Therefore,this time series does not appear to satisfy the requirements of stationarity.

Question 3:

```
y = diff(x, lag = 1, difference=1)
plot(y, main="Trend Differenced Data  over Time")
```


Trend Differenced Data over Time
y
Time
10
5
0
−5
0
20
40
60
80
100

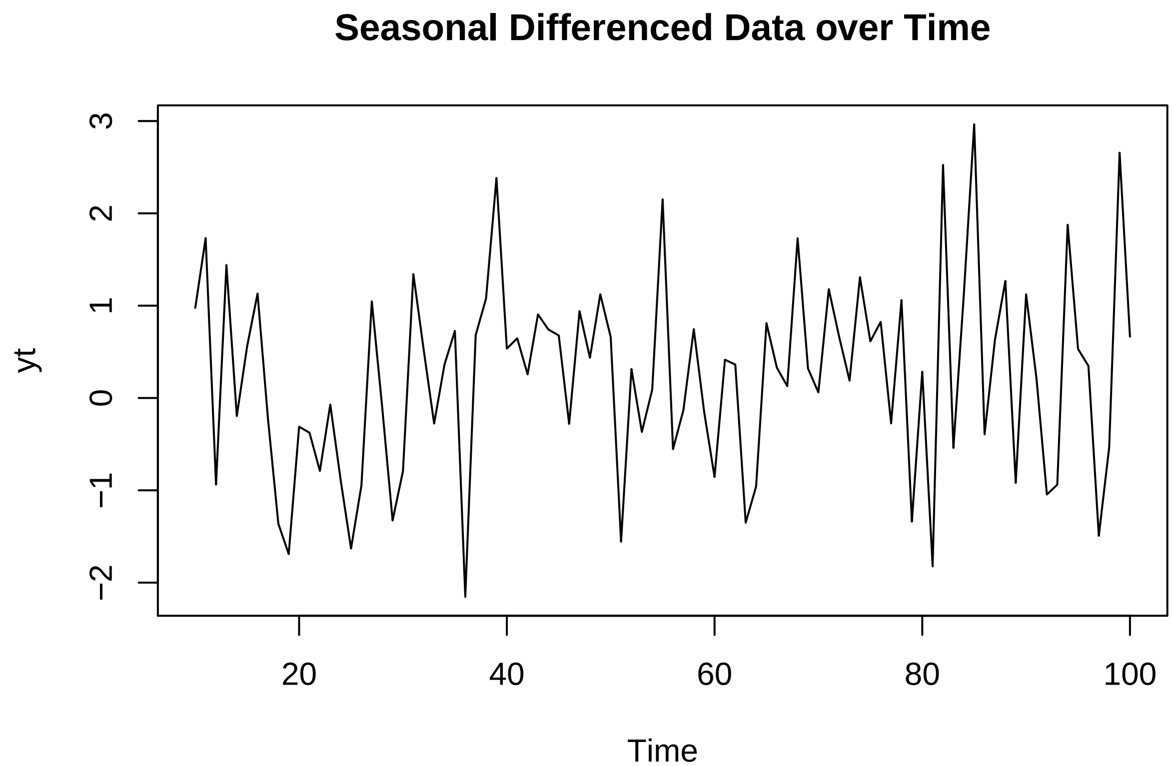
```
acf(y)
```



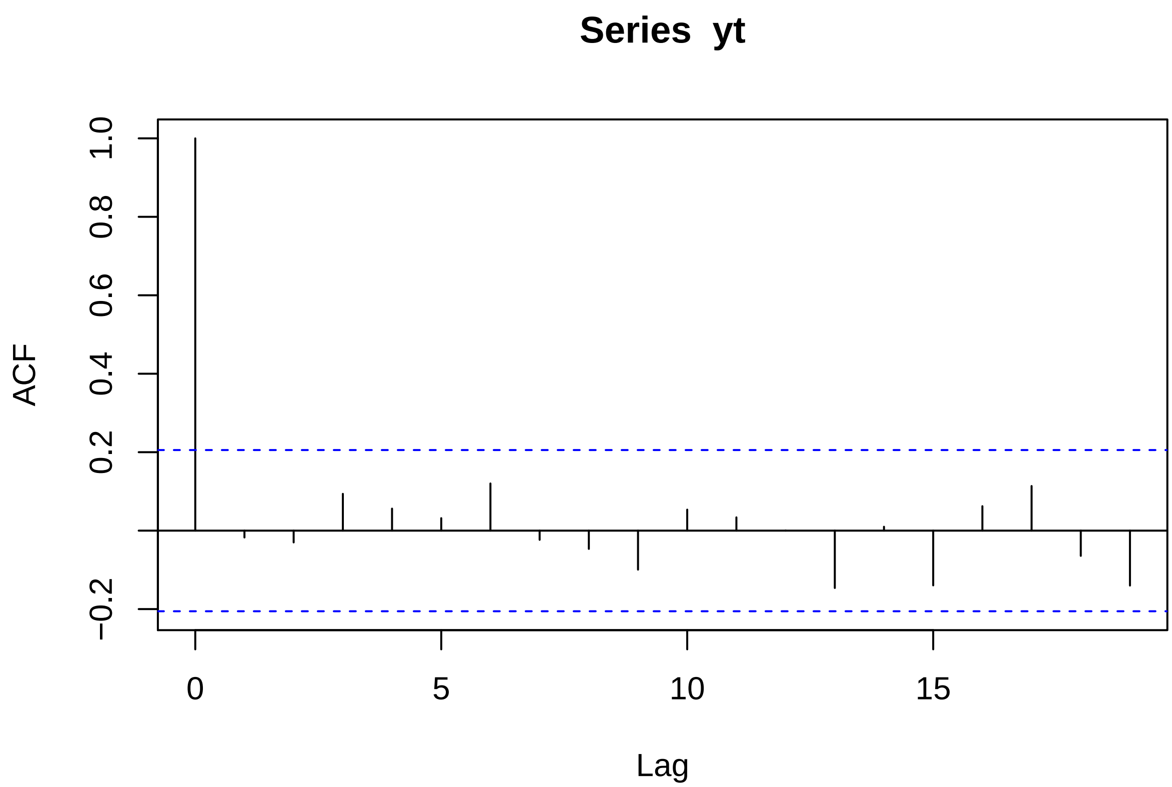
After differencing the data, we can see that the trend is removed. There looks to be seasonality, which can be seen both in the plot and the acf graph fluctuating with a period of around 8.

Question 4:

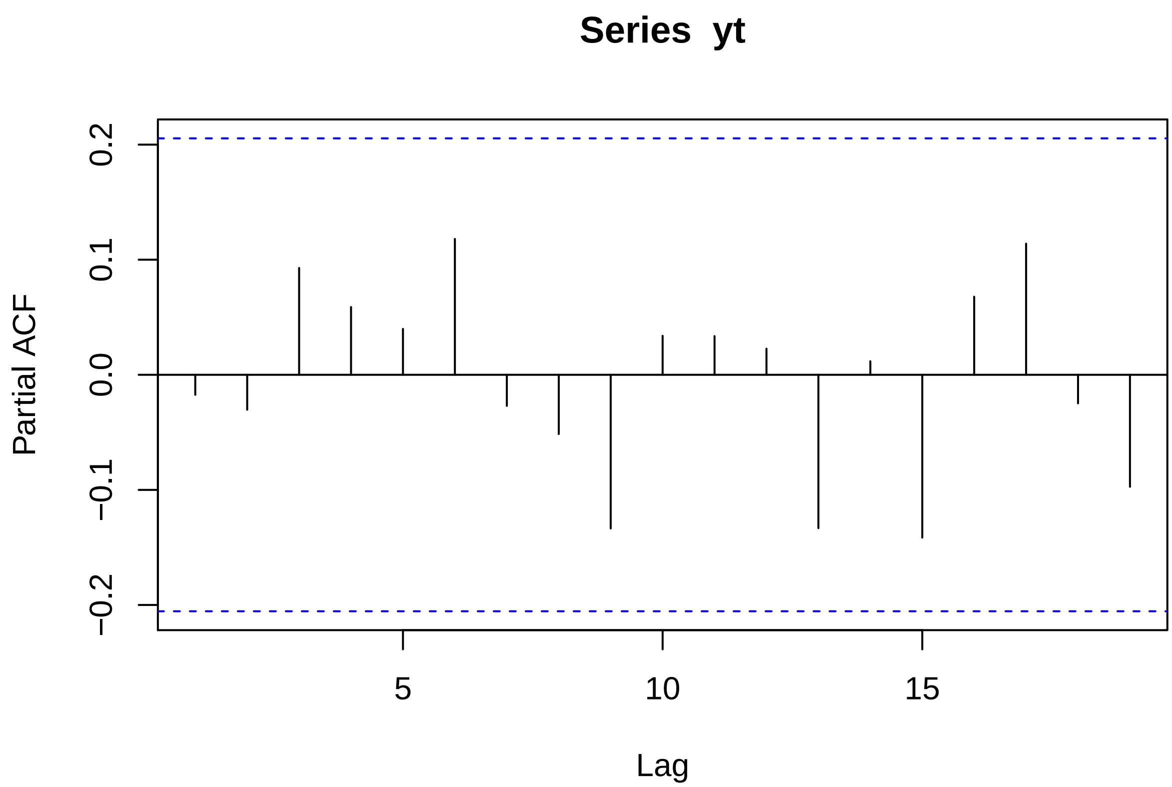
```
yt = diff(y, lag = 8, difference=1)
plot(yt, main="Seasonal Differenced Data over Time")
```


Seasonal Differenced Data over Time
yt
3
2
1
0
−1
−2
20
40
60
80
100
Time

```
acf(yt)
```




```
pacf(yt)
```



The deseasonalized time series demonstrates no significant autocorrelation as observed from the correlogram. We can see that about 95% of the spikes in the correlogram lies within (+/-)2/sqrt(n). There also appears to be no seasonality since the correlogram shows no obvious fluctuations. Thus this time series resembles white noise.

5. To determine which type of model from the SARIMA family to use, we need to determine the corresponding (p,d,q), (P, D, Q), and s terms from the acf and pacf graphs: s=8 - period of 8 d=1 - one time to remove the trend D=1 - one time to remove the seasonal effect q=0 - examining the autocorrelation at early lags, there is significant autocorrelation at lag 0, which is then cutoff Q=0 - examining the autocorrelation at seasonal lags, in this case s=8, so examining autocorrelation at lags 8, 16, 24, etc., and do not see any significant autocorrelation. p=0 - autocorrelation cuts off and does not decay P=0 - seasonal lag cuts off and does not decay Additionally, after differencing for trend and seasonality, the observed series resembles white noise, further proving that p=q=P=Q=0. Therefore, our final terms for the SARIMA model are: $(0, 1, 0), (0, 1, 0)_8$

6a) $Y_t = X_t - X_{t-1}$, $W_t = Y_t - Y_{t-s}$

$W_t = X_t - X_{t-1} - X_{t-s} + X_{t-(s+1)}$

b) $Y_t = X_t - BX_t = (1 - B)X_t$, where $BX_t = X_{t-1}$

c) $W_t = Y_t - B^s Y_t = (1 - B^s)Y_t = (1 - B^s)(1 - B)X_t$