Analysis of data sets

1.

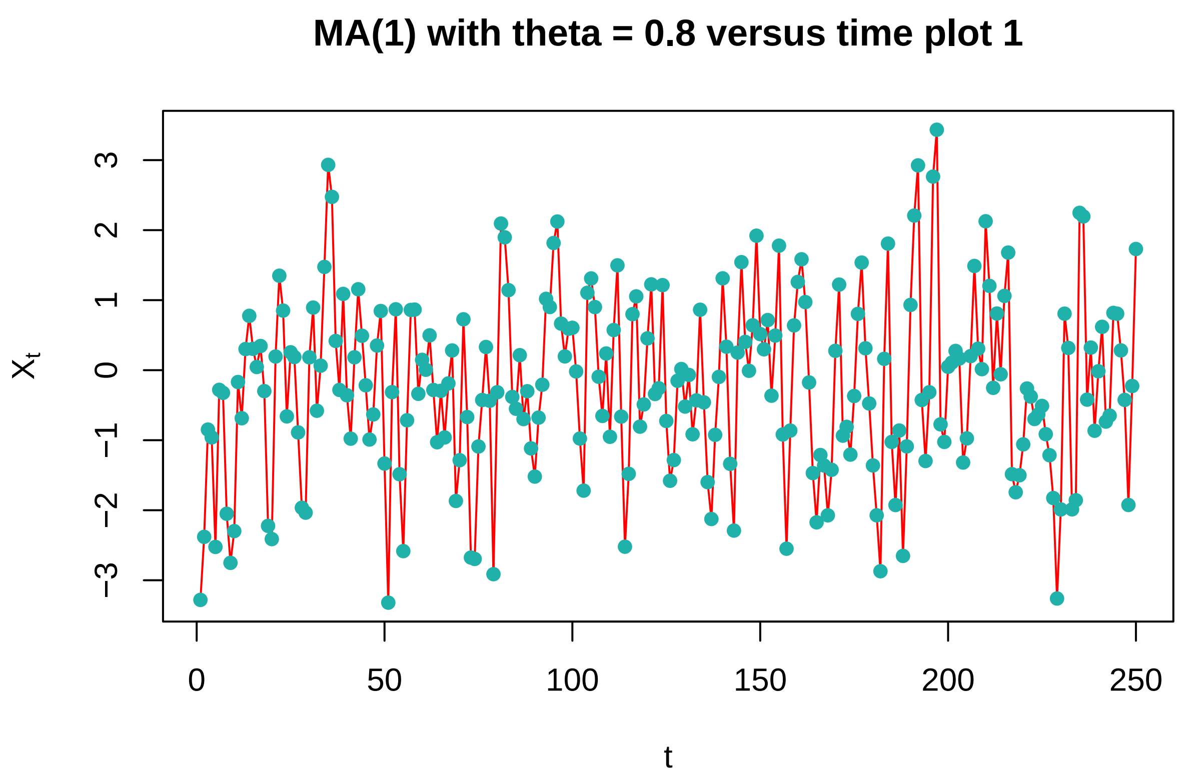
(a)

```
set.seed(485)
MA1 = arima.sim(model = list(ma = 0.8), n = 250)
MA2 = arima.sim(model = list(ma = 0.8), n = 250)
MA3 = arima.sim(model = list(ma = 0.8), n = 250)
```

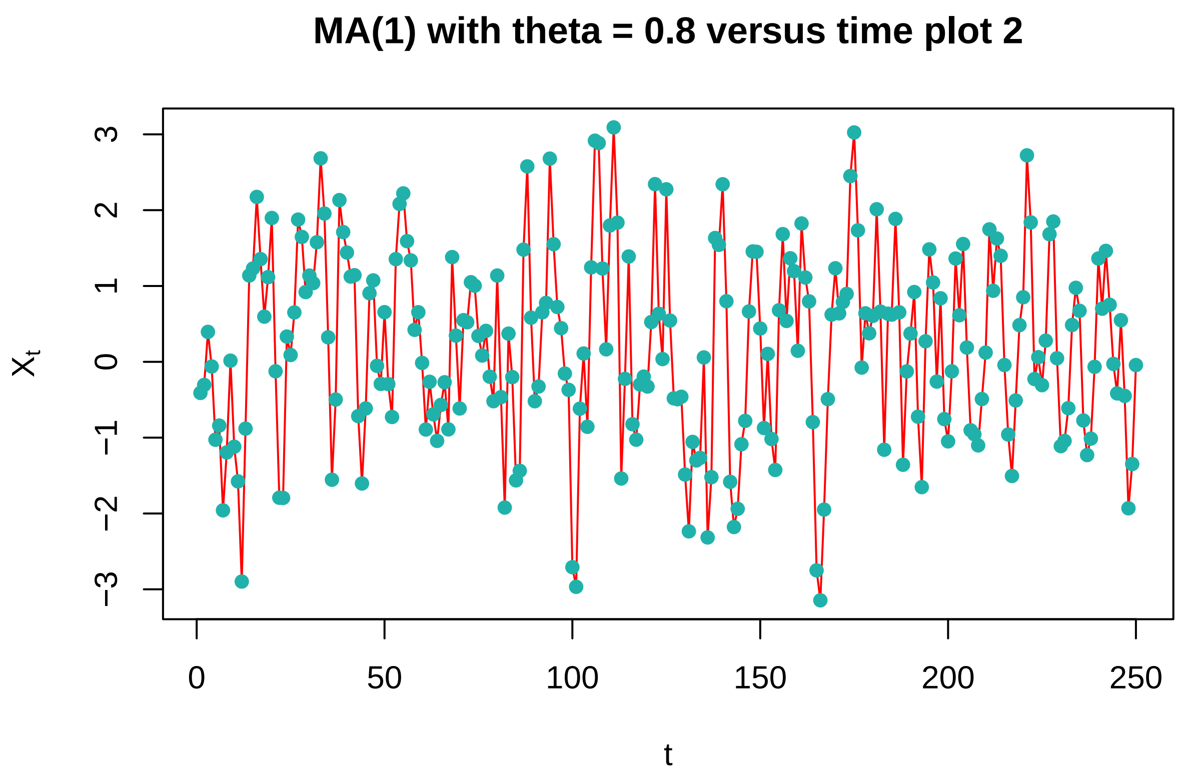
(b)

3 time series plots

```
plot.ts(MA1,
        col = "red",
        type = "l",
        main = "MA(1) with theta = 0.8 versus time plot 1",
        xlab = "t",
        ylab = expression(X[t]))

points(MA1,
       pch = 16, col = "lightseagreen",)
```

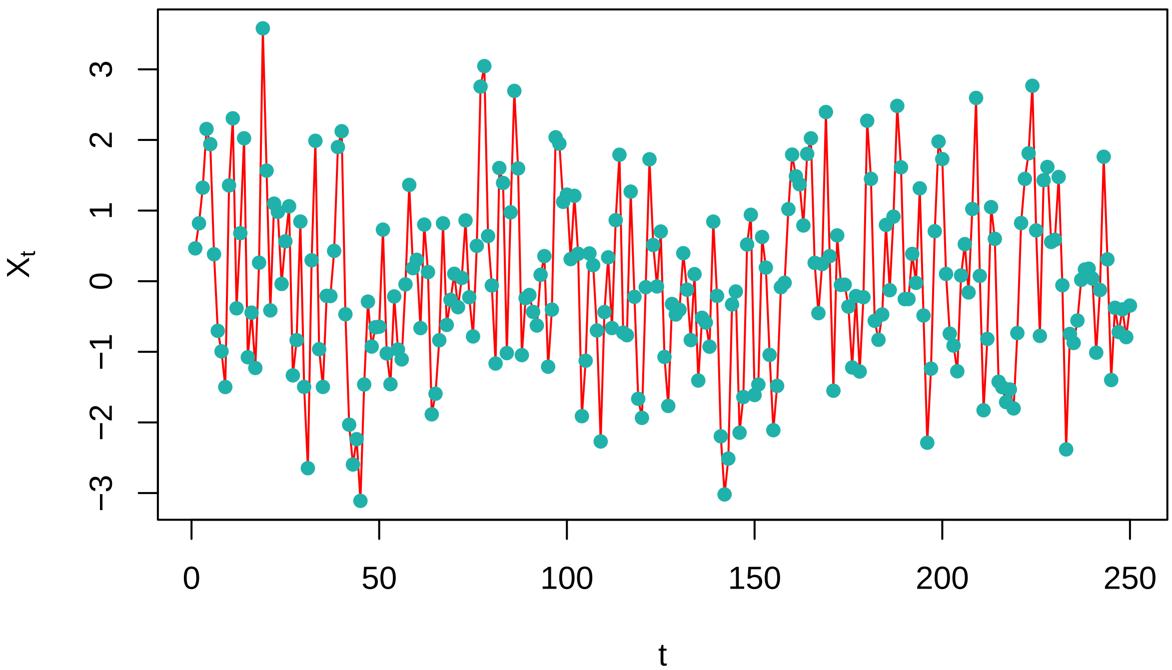


```
plot.ts(MA2,
        col = "red",
        type = "l",
        main = "MA(1) with theta = 0.8 versus time plot 2",
        xlab = "t",
        ylab = expression(X[t]))

points(MA2,
       pch = 16, col = "lightseagreen",)
```




```
plot.ts(MA3,
        col = "red",
        type = "l",
        main = "MA(1) with theta = 0.8 versus time plot 3",
        xlab = "t",
        ylab = expression(X[t]))

points(MA3,
       pch = 16, col = "lightseagreen",)
```

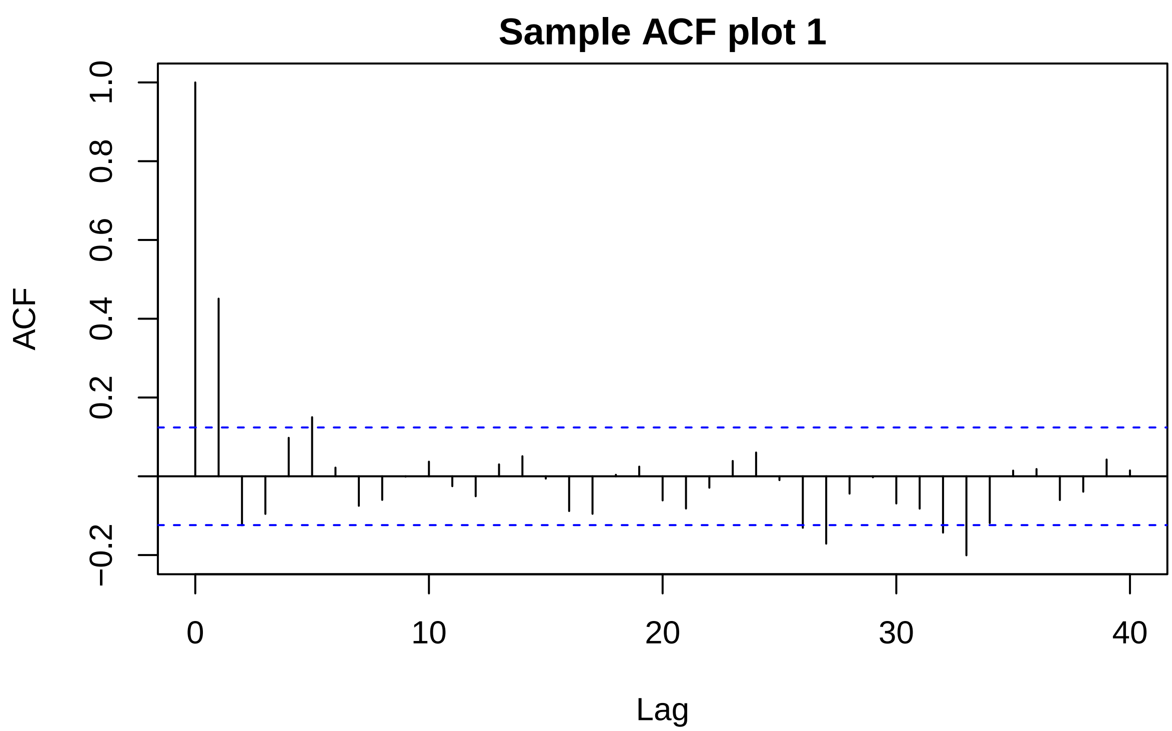
MA(1) with theta = 0.8 versus time plot 3



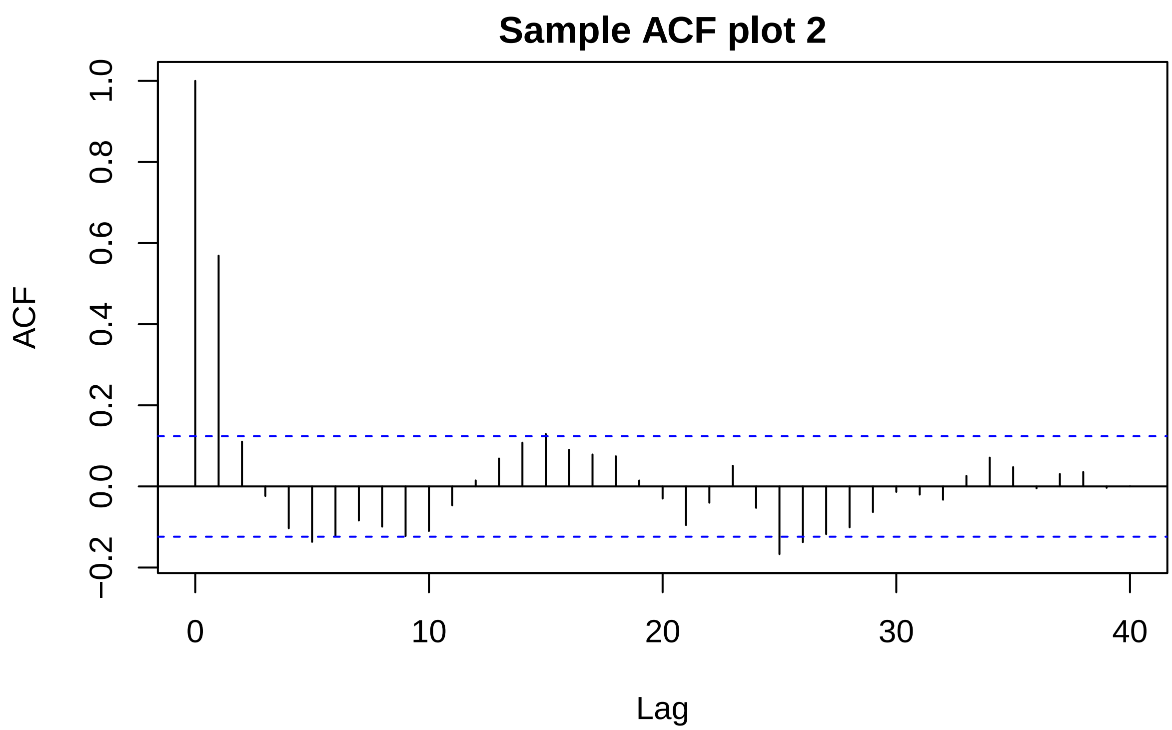
(c)

3 correlograms

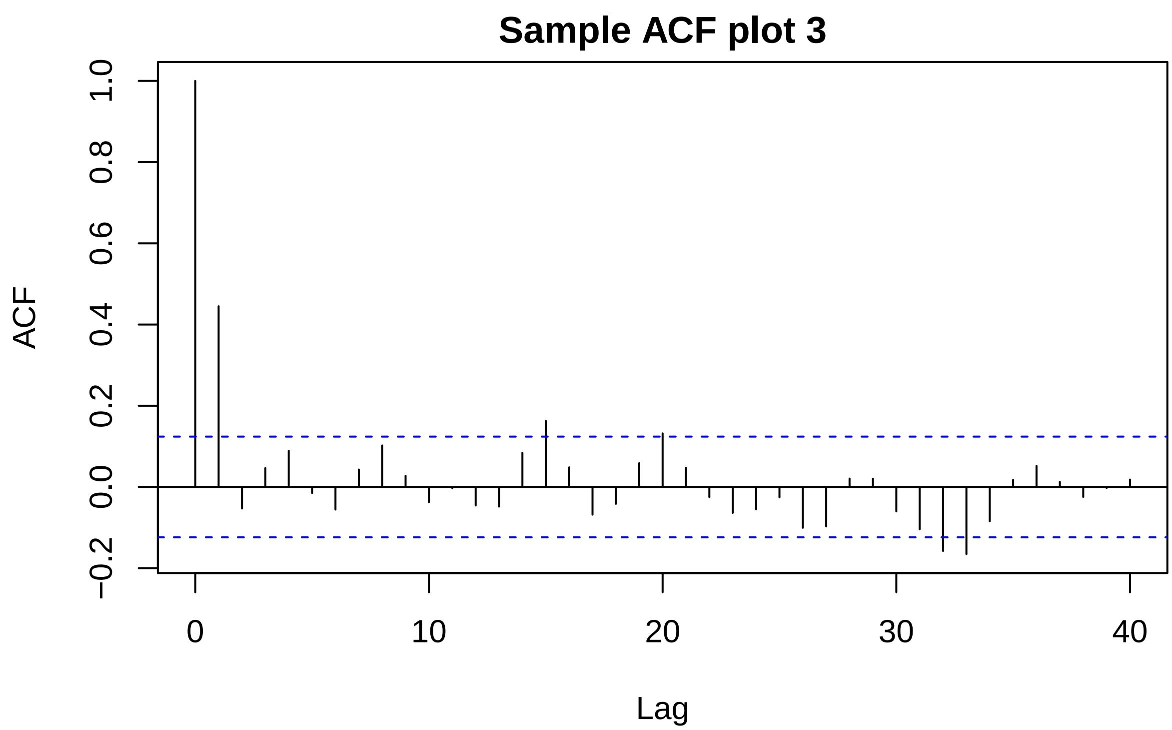
```
MA1.acf = acf(MA1, lag.max = 40,
               plot = F)
plot(MA1.acf, main = "")
title( "Sample ACF plot 1",line=0.5)
```




```
MA2.acf = acf(MA2, lag.max = 40,
               plot = F)
plot(MA2.acf, main = "")
title( "Sample ACF plot 2",line=0.5)
```




```
MA3.acf = acf(MA3, lag.max = 40,
               plot = F)
plot(MA3.acf, main = "")
title( "Sample ACF plot 3",line=0.5)
```


Sample ACF plot 3
ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
Lag

(d)

$$\rho_X(h) = \begin{cases} 1, & h = 0, \\ \theta/(1+\theta^2), & |h| = 1, \\ 0, & |h| \geq 2. \end{cases}$$

$h = 1,\ \theta = 0.8$

We have $\rho(1) = \frac{0.8}{1+0.8^2} \approx 0.4878$.

(e)

For plot 1 and plot 3, Estimated $\hat{\rho}(1)$ is equal to 0.4878, but $\hat{\rho}(1)$ from plot 2 is almost 0.6 which is different from what we calculated in question (d), so plot 2 is not consistent with the expected plot of $\rho(h)$.

Moreover, for $|\mathrm{h}| \geq 2$, $\rho(h) = 0$. We can expect approximately $40 \times 5\% = 2$ values to fall outside of the horizontal dash line. For plot 1, we have 5 values of $\hat{\rho}$ fall outside these bounds, and for plot 3, we have 4 values fall outside the bounds. The values of $\hat{\rho}$ are correlated, so if one values outside the bounds, the neighboring ones tend to do so as well. I would still consider plot 1 and plot 3 are consistent with the the expected plot since there are not too many values fall outside of the bounds and some of the points are neighbors.

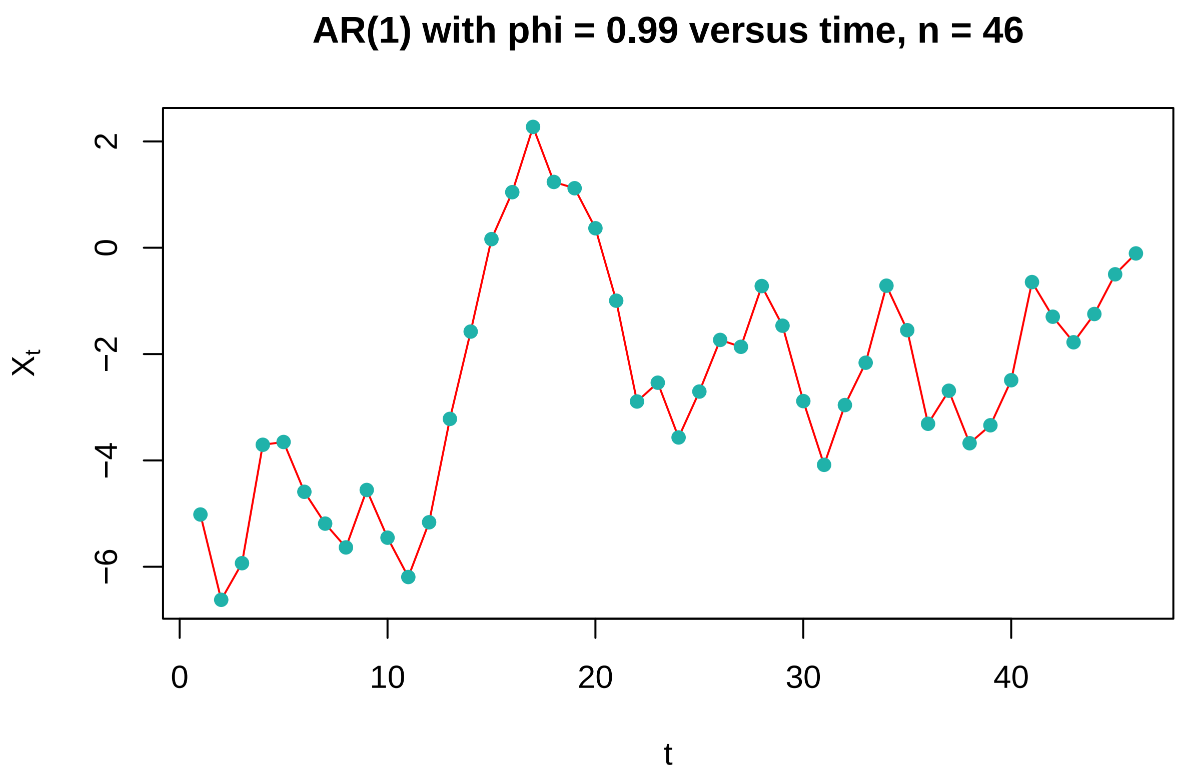
(f)

For different plots, the plots would vary. However, for different realizations we expect some general properties. For example, $\rho(0) = 1$, $\rho(1)$ should be the same as we calculated from our model, and there should not be too many points outside the horizontal dash line in the correlogram. We expect our plot consistent with the expected plot of $\rho(h)$.

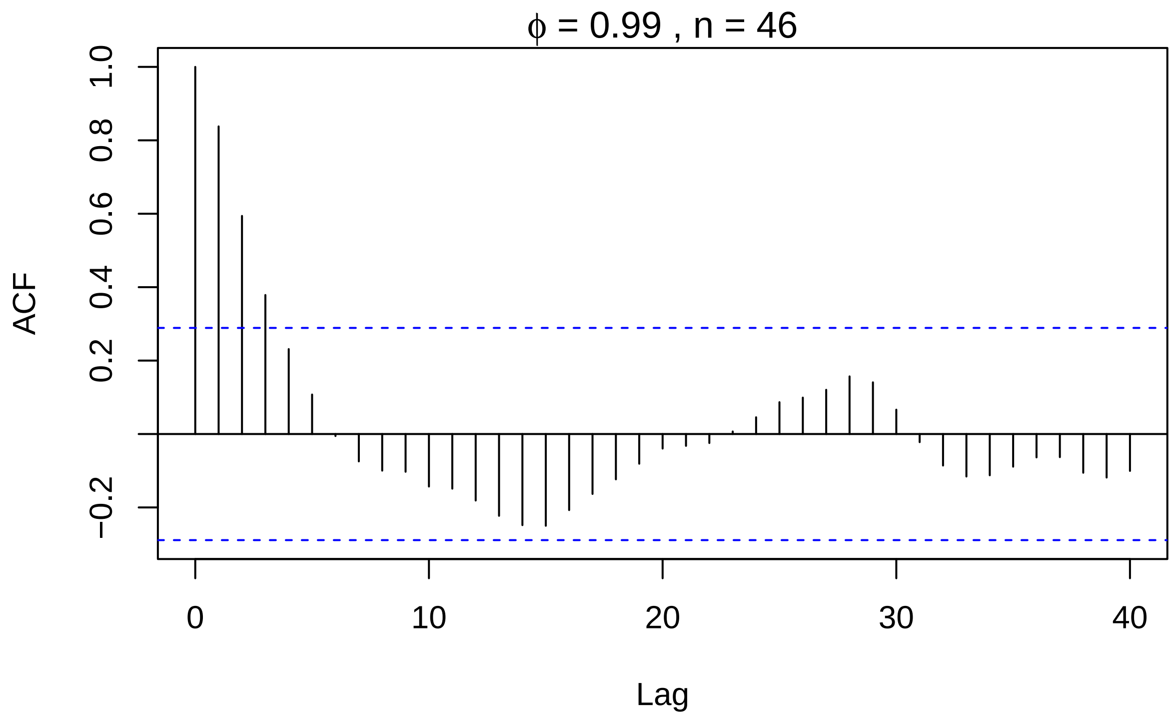
2.

(b)

```
AR1 = arima.sim(model = list(ar = 0.99), n = 46, rand.gen = rnorm)
plot.ts(AR1,
        col = "red",
        main = "AR(1) with phi = 0.99 versus time, n = 46",
        xlab = "t",
        ylab = expression(X[t]))
points(AR1,
       pch = 16, col = "lightseagreen",)
```

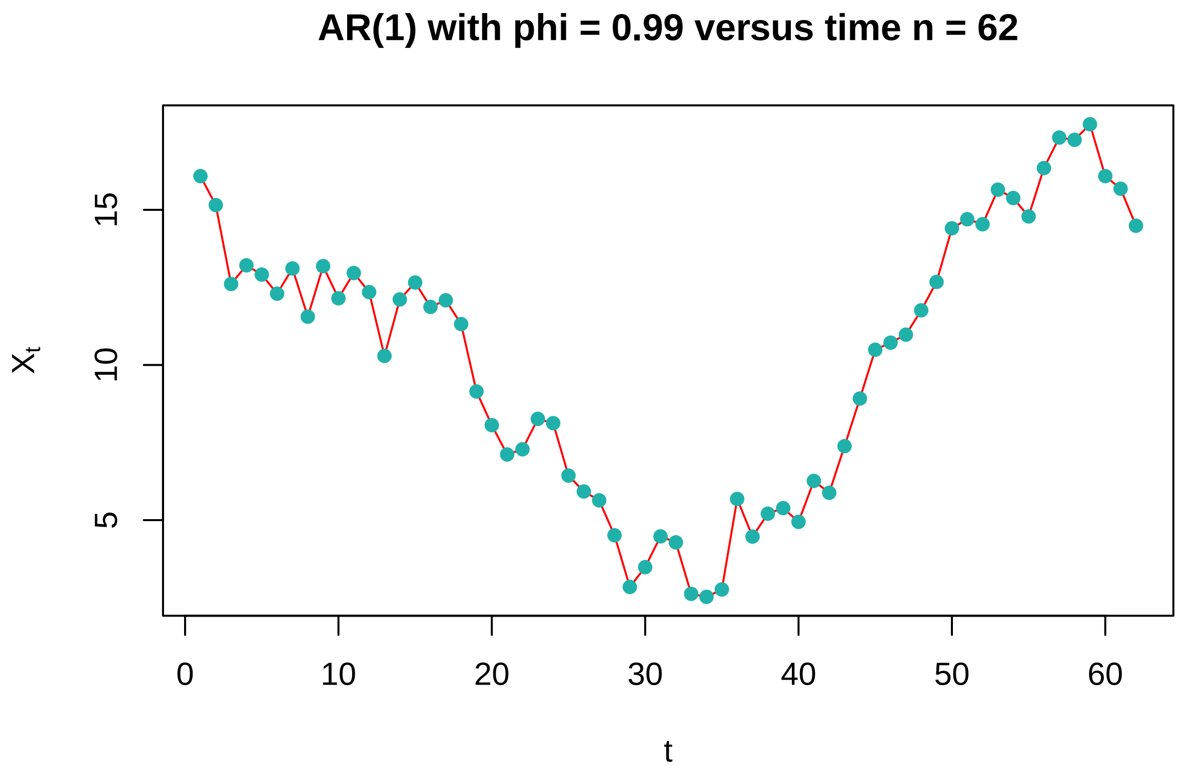



```
acf(AR1,main = " ",lag = 40)
title( expression(paste(phi, " = 0.99 , n = 46")),line=0.5)
```

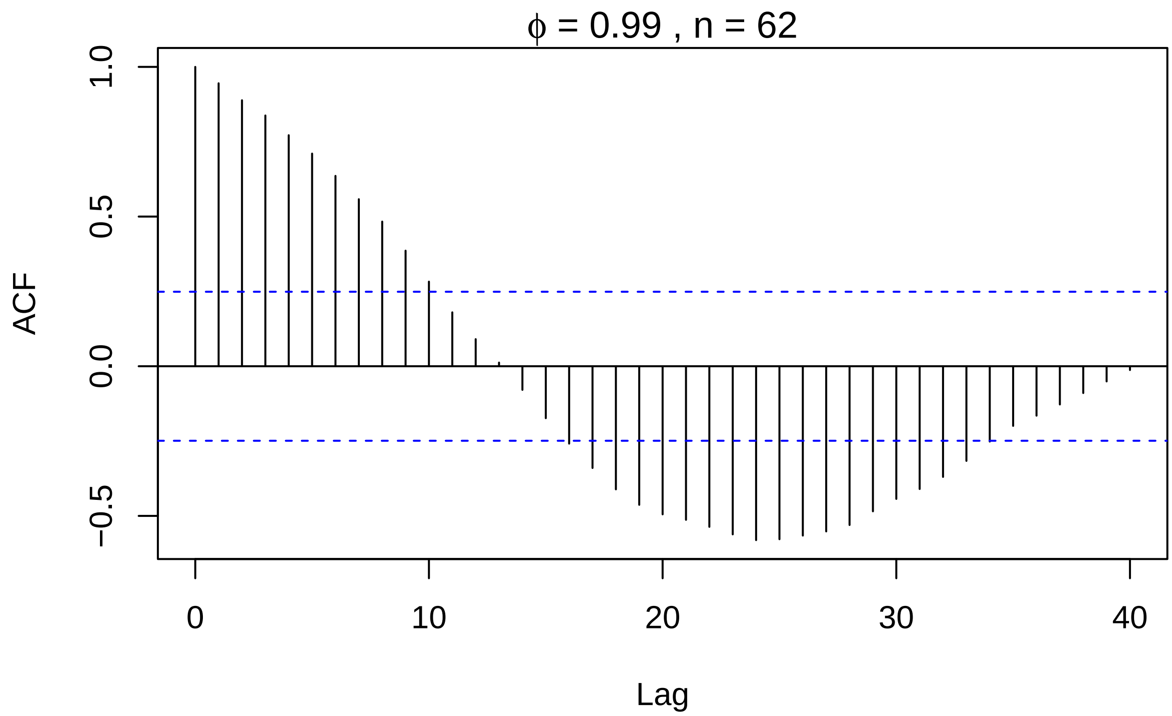


(c)

```
AR2 = arima.sim(model = list(ar = 0.99), n = 62, rand.gen = rnorm)
plot.ts(AR2,
        col = "red",
        main = "AR(1) with phi = 0.99 versus time n = 62",
        xlab = "t",
        ylab = expression(X[t]))
points(AR2,
       pch = 16, col = "lightseagreen",)
```

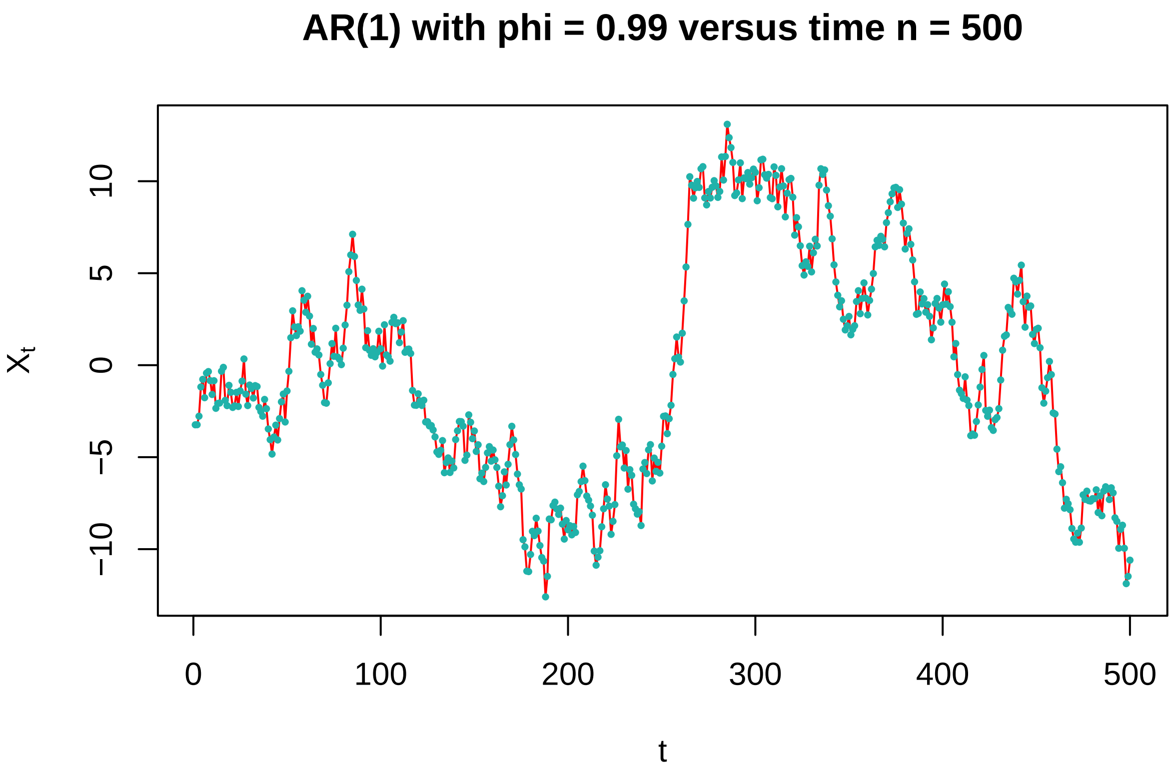



```
acf(AR2,main = " ",lag = 40)
title( expression(paste(phi, " = 0.99 , n = 62")),line=0.5)
```

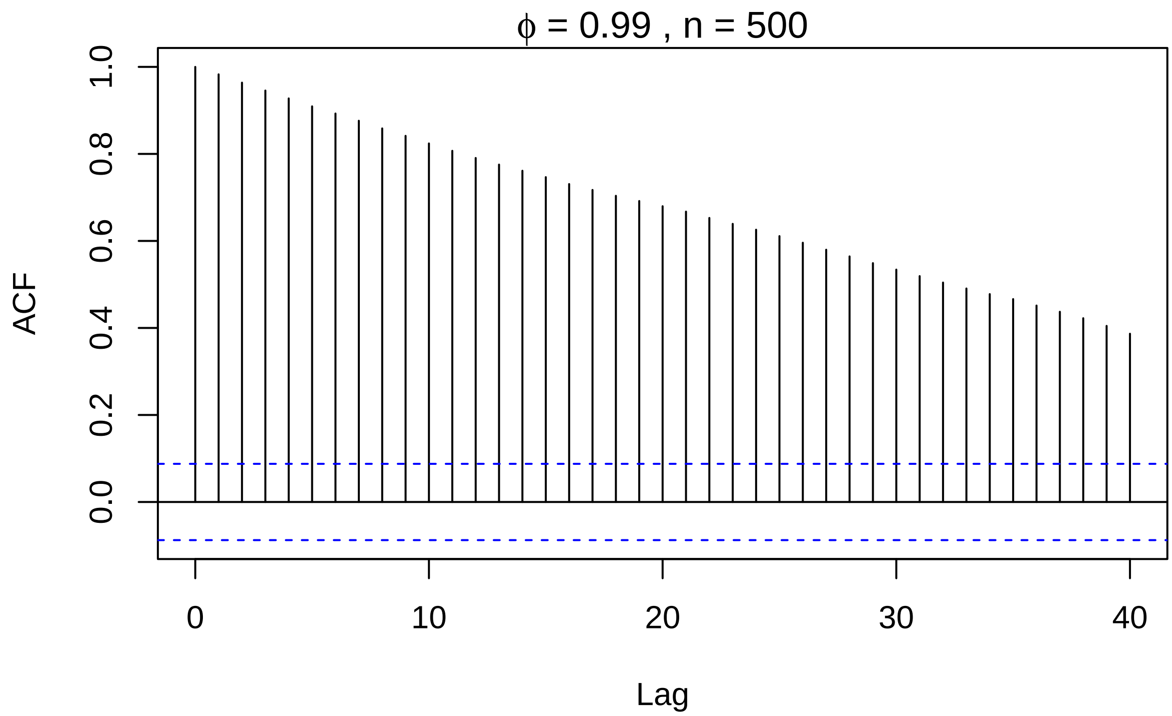


(d)

```
AR3 = arima.sim(model = list(ar = 0.99), n = 500, rand.gen = rnorm)
plot.ts(AR3,
        col = "red",
        main = "AR(1) with phi = 0.99 versus time n = 500",
        xlab = "t",
        ylab = expression(X[t]))
points(AR3,
       pch = 16, cex = .5, col = "lightseagreen",)
```




```
acf(AR3,main = " ",lag = 40)
title( expression(paste(phi, " = 0.99 , n = 500")),line=0.5)
```



(e)

$$\rho_X(h) = \phi^h \ , \ h > 0$$

(f)

There is correlation at any lag, though the strength decreases as the lag increases since $|\phi|^h \to 0$ as $h \to \infty$ if $|\phi| < 1$. Only when n is 500 the plot of correlogram is consistent with the expected plot of $\rho(h)$.

The smaller the $|\phi|$, the faster the correlations decrease as the lag increases. we have $\phi = 0.99$, the correlation should decrease slowly as the lag increases.

(g)

$\hat{\gamma}$ and $\hat{\rho}$ is baised; however, under general assumptions that they are nearly unbiased for large n. For h close to n, the estimates are unreliable. There is a rough rule thumb that says that n $\geq$ 50 and h $\leq$ n/4 will produce reasonable results.

For the first correlogram, we have n = 46 which is less than 50. So the correlogram do not consistent with the model.

For the second correlogram, we have n = 62, and we observed that the correlogram do not present the evidence of consistent with the model at h is 16 (62/4 = 15.5).

The third correlogram have n = 500 which is large than 50, and h is less or equal to 125 (500/4 = 125), so the plot of correlogram is consistent with the model.