Assignment 3

Frank Woodling

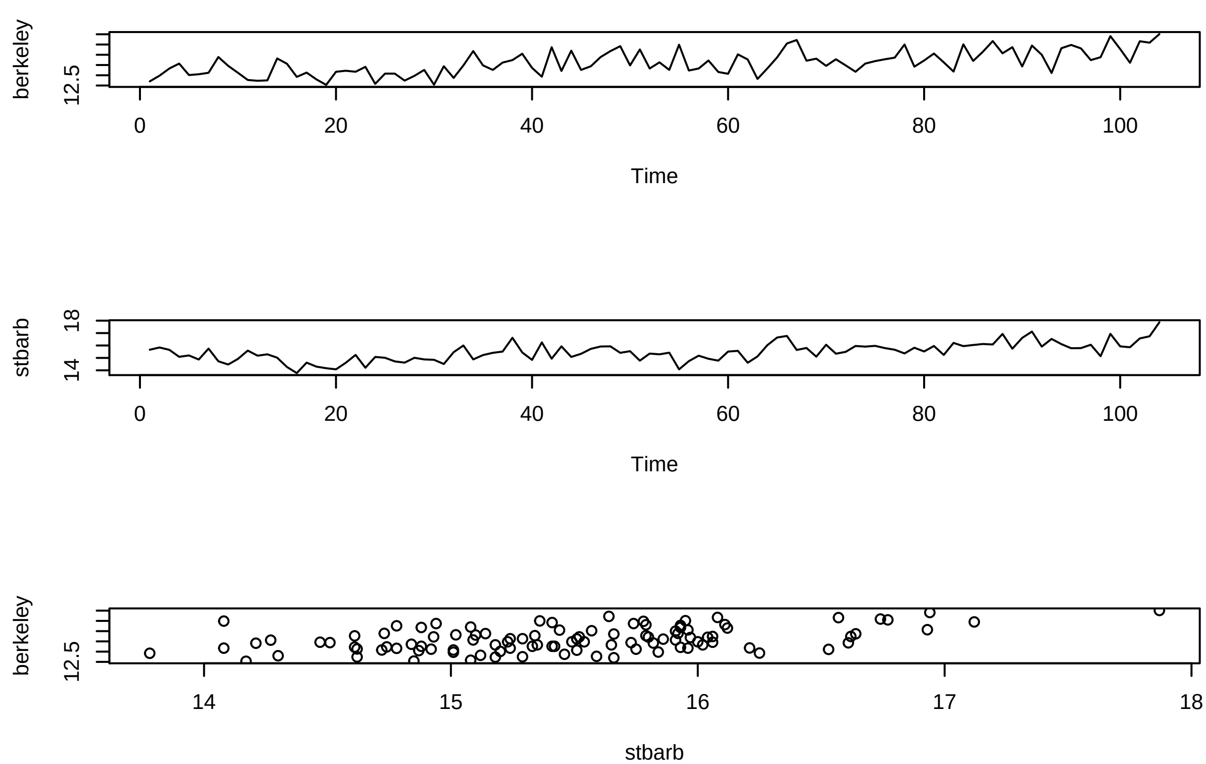
February 10, 2016

1.

a.

```
berk <- scan("berkeley.dat", what=list(double(0),double(0),double(0)))
time <- berk[[1]]
berkeley <- berk[[2]]
stbarb <- berk[[3]]
```

```
par(mfrow=c(3,1))
plot.ts(berkeley)
plot.ts(stbarb)
plot(berkeley~stbarb)
```



b.

```r
timefit <- lm(berkeley~time)
anova(timefit)
```

```
## Analysis of Variance Table
##
## Response: berkeley
##            Df Sum Sq Mean Sq F value    Pr(>F)
## time        1 14.096  14.096  68.431 5.228e-13 ***
## Residuals 102 21.011   0.206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

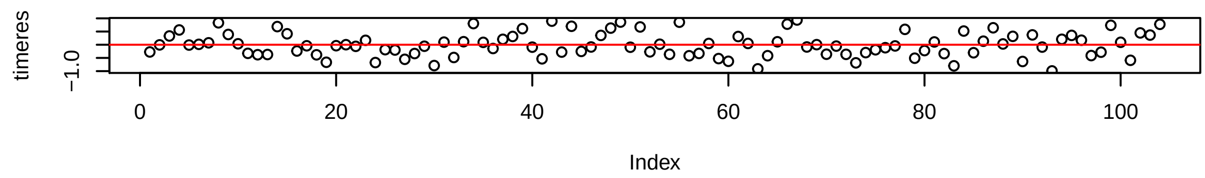
```r
summary(timefit)
```

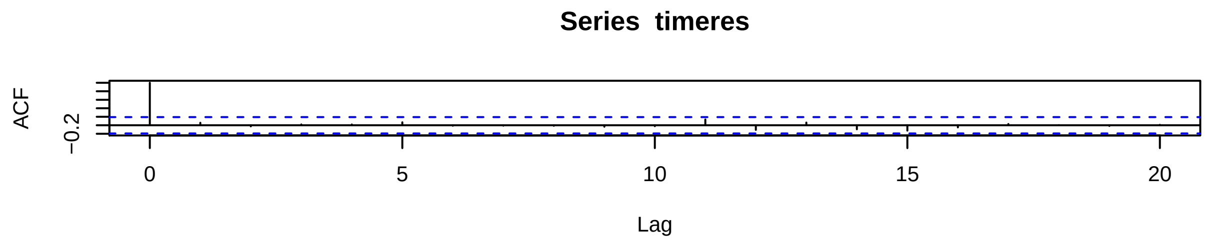
```
##
## Call:
## lm(formula = berkeley ~ time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.99195 -0.33156 -0.03834  0.32076  0.93689
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.252876   2.884447  -3.555 0.000575 ***
## time          0.012263   0.001482   8.272 5.23e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4539 on 102 degrees of freedom
## Multiple R-squared:  0.4015, Adjusted R-squared:  0.3956
## F-statistic: 68.43 on 1 and 102 DF,  p-value: 5.228e-13
```

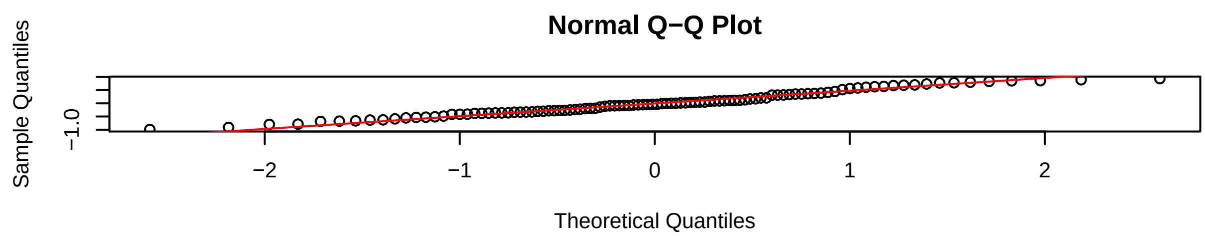
```r
AIC(timefit)
```

```
## [1] 134.8053
```

```r
par(mfrow=c(3,1))
timeres <- residuals(timefit)
plot(timeres)
abline(h=0, col="red")
acf(timeres)
qqnorm(timeres)
qqline(timeres, col="red")
```







The residuals seem very uniform, but may have a trend in them. The ACF decays quickly and looks to be stationary The Q-Q plot shows that the data is mostly normal. The tails have a couple of outliers.

c.

```
timefit <- lm(berkeley~stbarb)
anova(timefit)
```

```
## Analysis of Variance Table
##
## Response: berkeley
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## stbarb      1  7.2972  7.2972  26.765 1.153e-06 ***
## Residuals 102 27.8092  0.2726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(timefit)
```

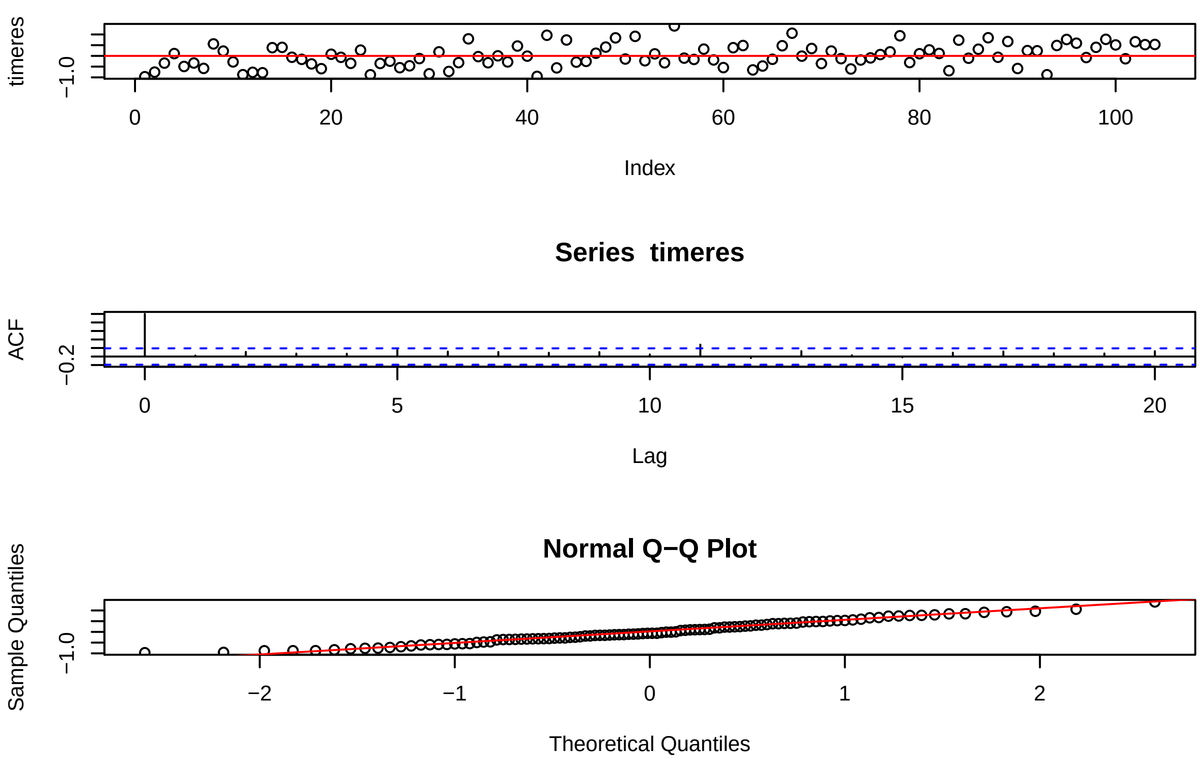
```
##
## Call:
## lm(formula = berkeley ~ stbarb)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97037 -0.33928 -0.06555  0.38689  1.39676
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.9502     1.0943   7.265 7.67e-11 ***
## stbarb        0.3653     0.0706   5.173 1.15e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5221 on 102 degrees of freedom
## Multiple R-squared:  0.2079, Adjusted R-squared:  0.2001
## F-statistic: 26.77 on 1 and 102 DF,  p-value: 1.153e-06
```

```
AIC(timefit)
```

```
## [1] 163.9607
```

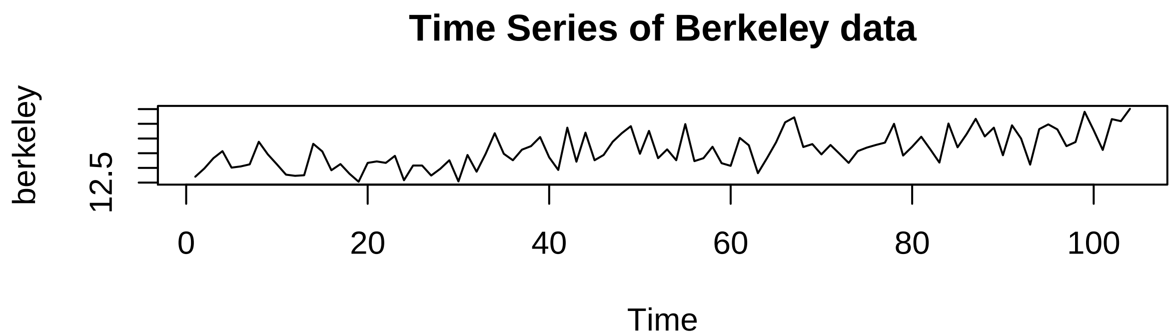
```
par(mfrow=c(3,1))
timeres <- residuals(timefit)
plot(timeres)
abline(h=0, col="red")
acf(timeres)
qqnorm(timeres)
qqline(timeres, col="red")
```

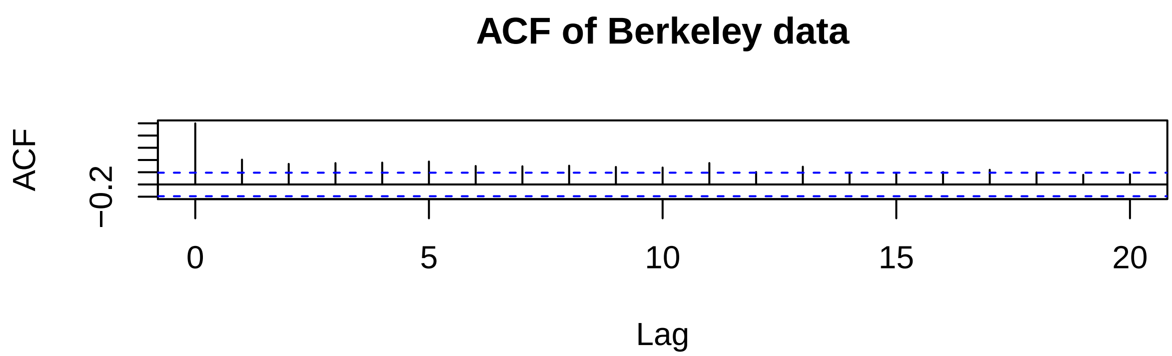


The residuals seem uniform in this as well. Again they do not seem to be completely random. There is an undulating trend. The R-squared is .4015, so there is at least some correlation between the two variables.

d.

```
plot.new()
par(mfrow=c(2,1))
plot.ts(berkeley, main = "Time Series of Berkeley data")
acf(berkeley, main = "ACF of Berkeley data")
```

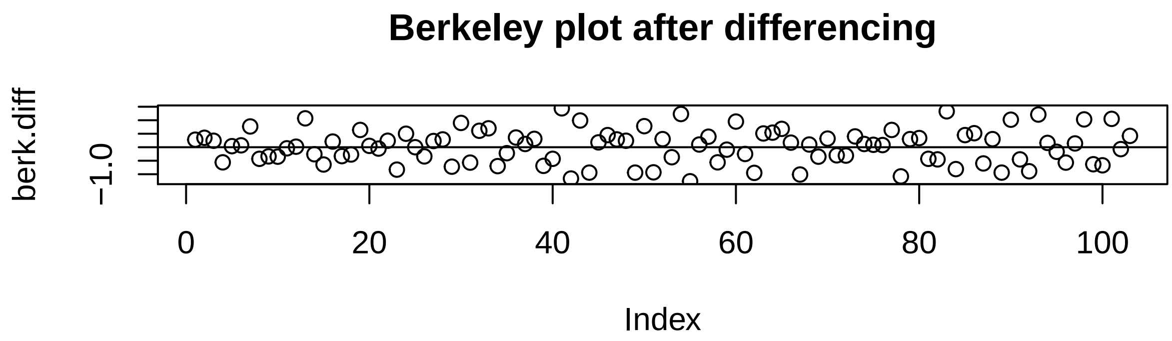


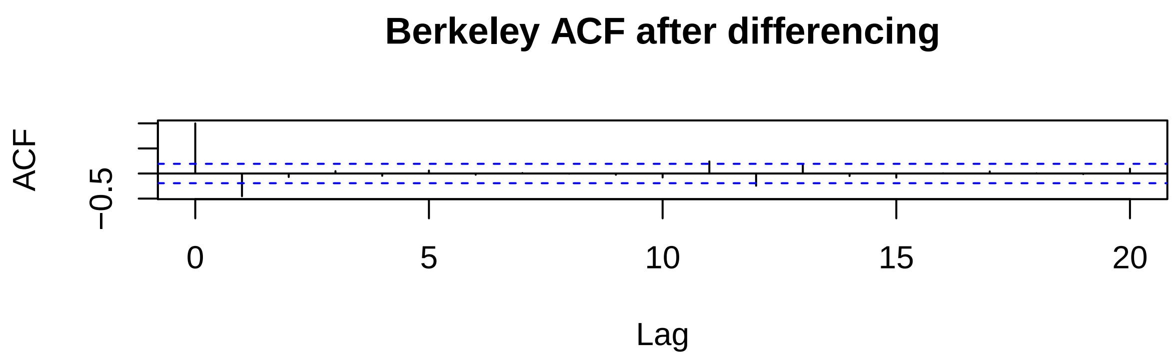


Many sections of the ACF plot are crossing the significance line. It seems to be decaying, but it does not look stationary.

e.

```
berk.diff <- diff(berkeley)
par(mfrow=c(2,1))
plot(berk.diff, main = "Berkeley plot after differencing")
abline(0, 0)
acf(berk.diff, main = "Berkeley ACF after differencing")
```





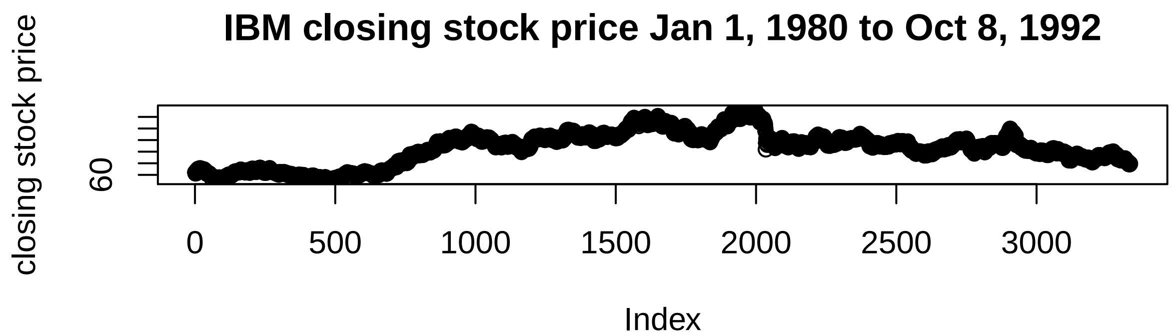
The data is much closer to being stationary after differencing, however, many of the lines still cross the dotted line. Even after differencing the series is still not stationary.

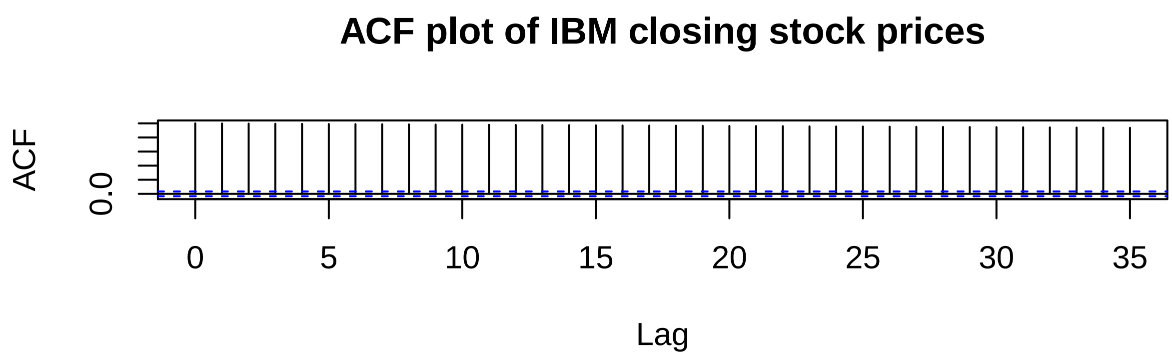
f.

2.

a.

```
ibm <- scan("dailyibm.dat", skip = 1)
par(mfrow=c(2,1))
plot(ibm, main = "IBM closing stock price Jan 1, 1980 to Oct 8, 1992", ylab = "closing stock price")
acf(ibm, main = "ACF plot of IBM closing stock prices")
```

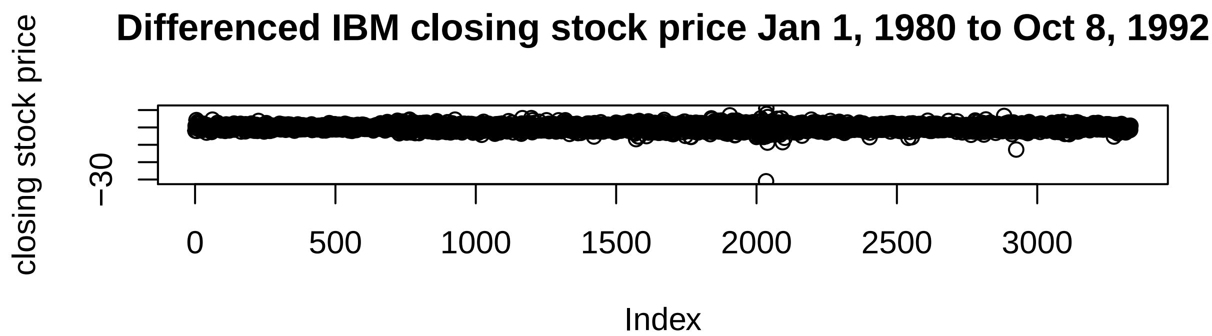


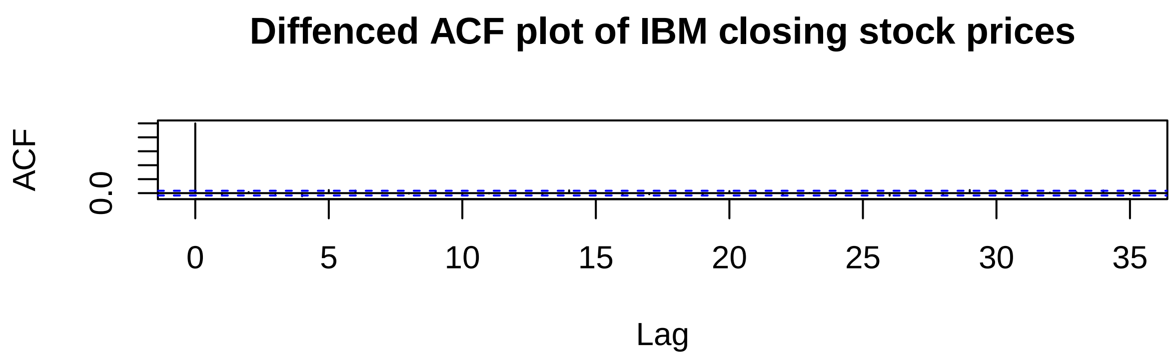


The time series is not stationary since the autocorrelations do not decay to zero. You can see in the first plot there is a constant upward trend.

b.

```
ibm.diff <- diff(ibm)
par(mfrow=c(2,1))
plot(ibm.diff, main = "Differenced IBM closing stock price Jan 1, 1980 to Oct 8, 1992", ylab = "closing
acf(ibm.diff, main = "Diffenced ACF plot of IBM closing stock prices")
```

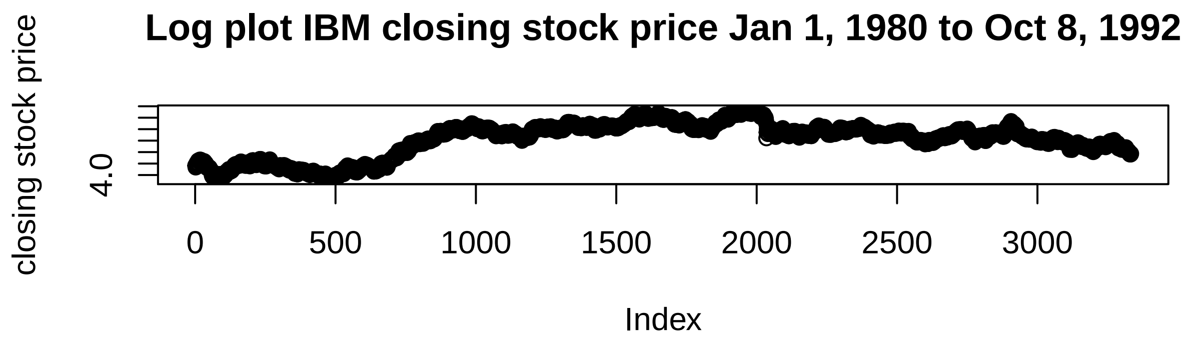


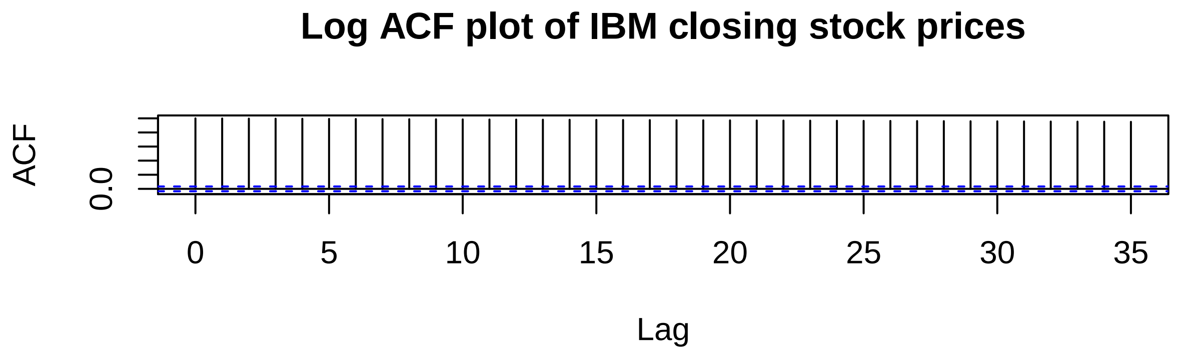


The data decayed to within the boundaries quickly. The differenced time series data is stationary.

c.

```
ibm.log <- log(ibm)
par(mfrow=c(2,1))
plot(ibm.log, main = "Log plot IBM closing stock price Jan 1, 1980 to Oct 8, 1992", ylab = "closing sto
acf(ibm.log, main = "Log ACF plot of IBM closing stock prices")
```

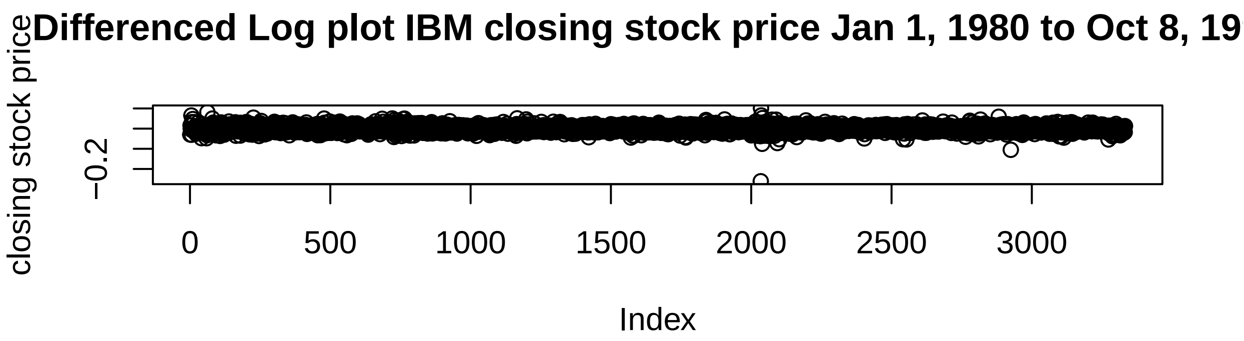


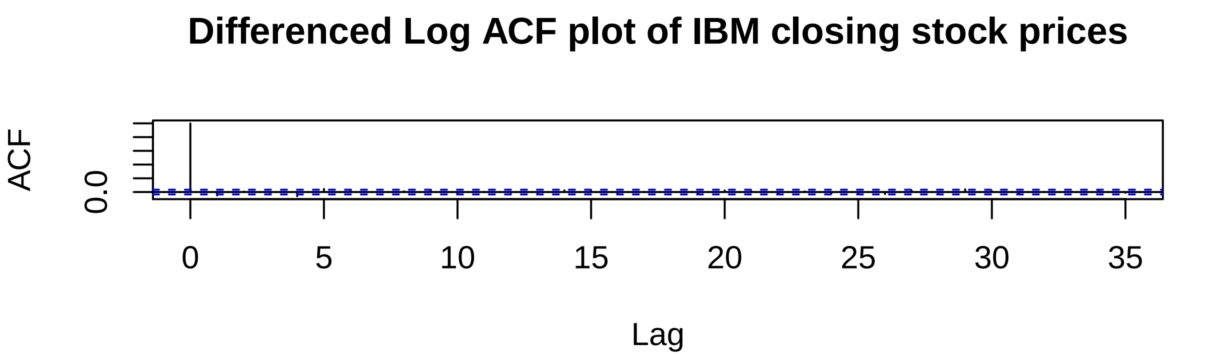


The data looks very similar to the initial plot, even though the new values are much smaller. The data is not stationary since it does not decay at all.

d.

```
difflogibm <- diff(log(ibm))
par(mfrow=c(2,1))
plot(difflogibm , main = "Differenced Log plot IBM closing stock price Jan 1, 1980 to Oct 8, 1992", yla
acf(difflogibm , main = "Differenced Log ACF plot of IBM closing stock prices")
```

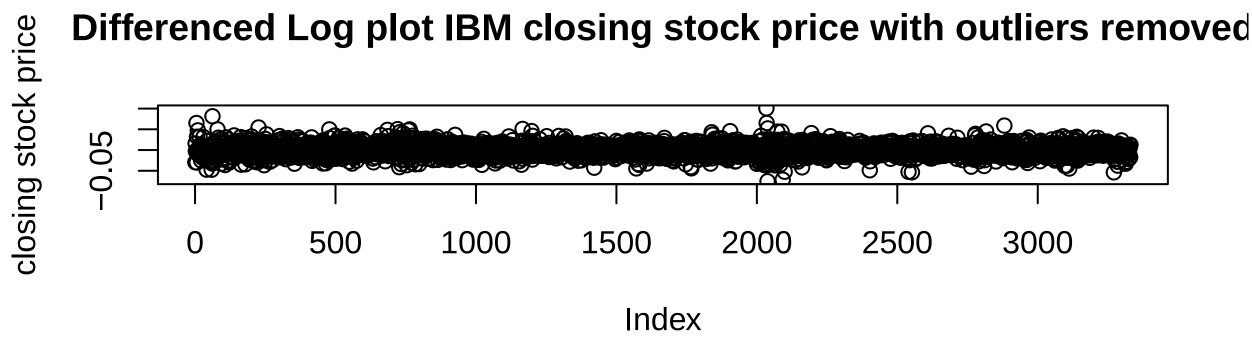


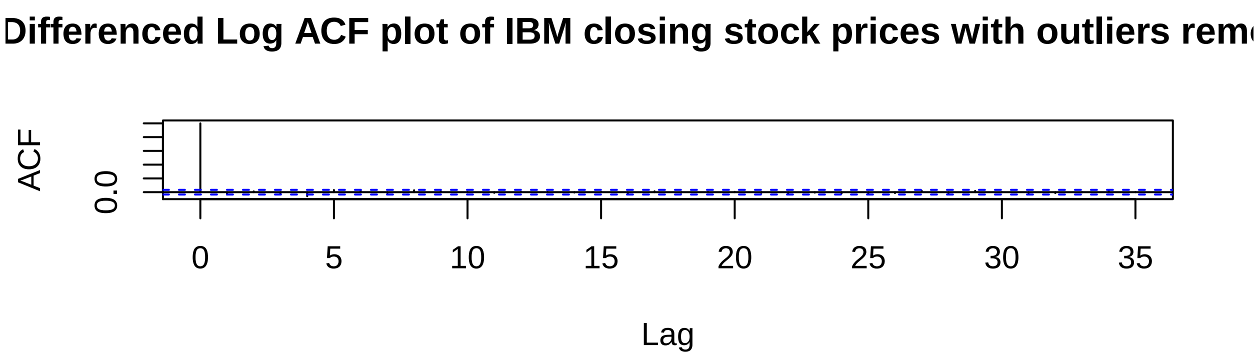


Taking the difference first removes the trend, so there would be many points with zero change from the previous. Since you cannot take the log of 0, it would be impossible to plot. Taking the log first the data is mostly stationary except from the few outliers mentioned. There is no discernable trend, and the ACF plot decays quickly.

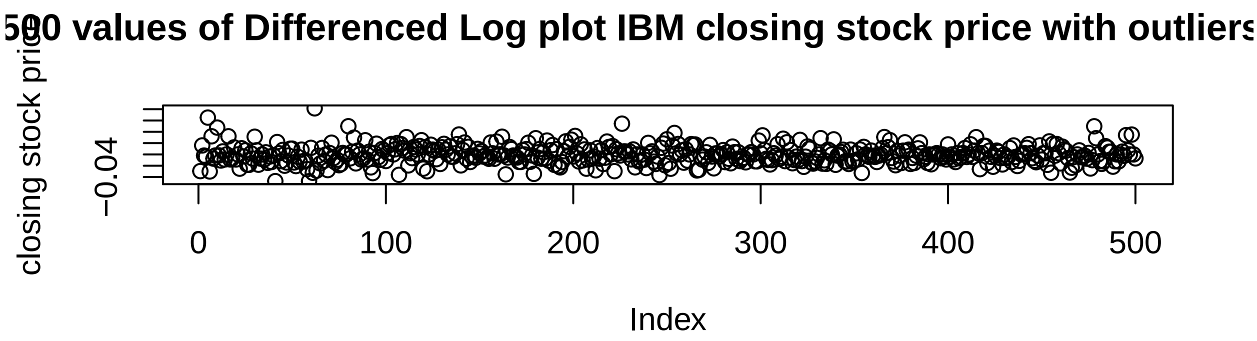
e.

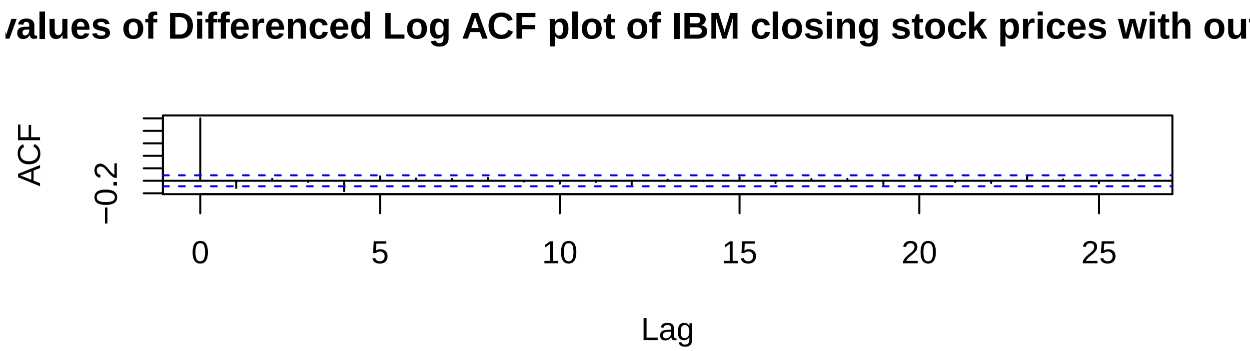
```
difflogibm <- difflogibm[difflogibm > -0.1]

par(mfrow=c(2,1))
plot(difflogibm , main = "Differenced Log plot IBM closing stock price with outliers removed", ylab = "
acf(difflogibm , main = "Differenced Log ACF plot of IBM closing stock prices with outliers removed")
```



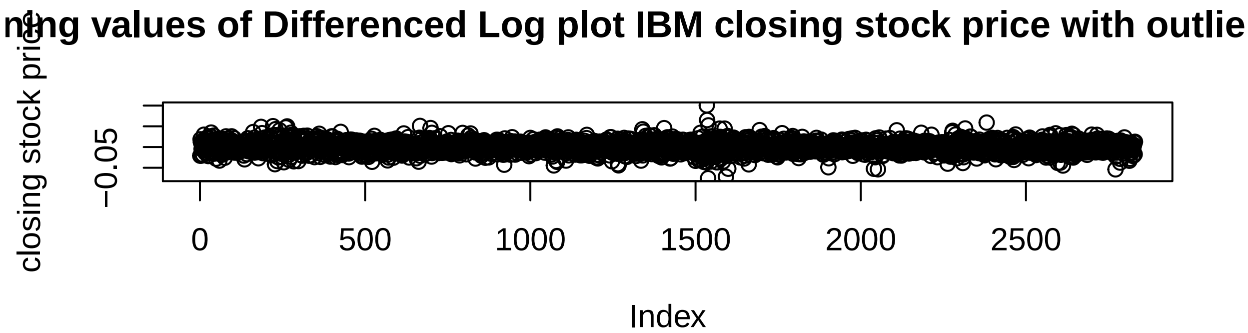


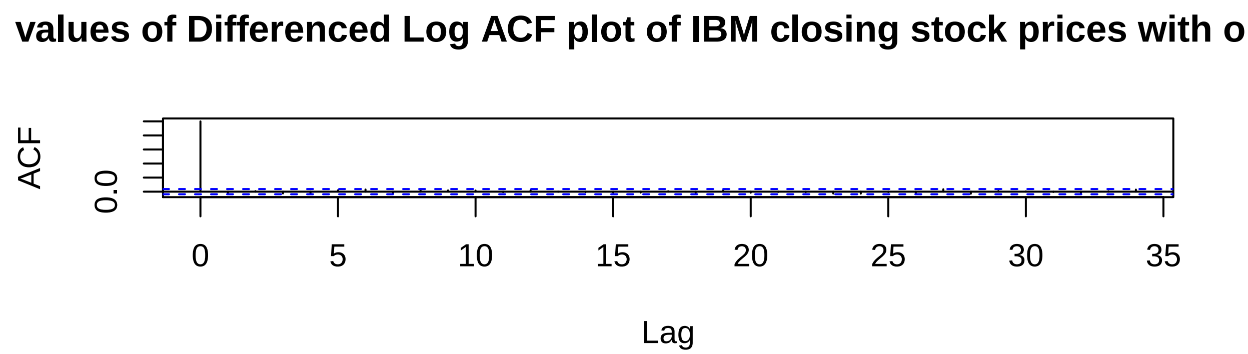

```
difflogibm1 <- difflogibm[1:500]

par(mfrow=c(2,1))
plot(difflogibm1, main = "First 500 values of Differenced Log plot IBM closing stock price with outliers
acf(difflogibm1, main = "First 500 values of Differenced Log ACF plot of IBM closing stock prices with
```






```
difflogibm2 <- difflogibm[501:length(difflogibm)]

par(mfrow=c(2,1))
plot(difflogibm2, main = "Remaining values of Differenced Log plot IBM closing stock price with outliers
acf(difflogibm2, main = "Remaining values of Differenced Log ACF plot of IBM closing stock prices with o
```





The price values of the second graph are double the first. The plots seem similar, except there are much fewer points in the first plot. Both ACFs seem to decay quickly, and are stationary. The ACFs seem slightly better than before we split them up.

f.