Lab 6

Ben Chu

April 10, 2018

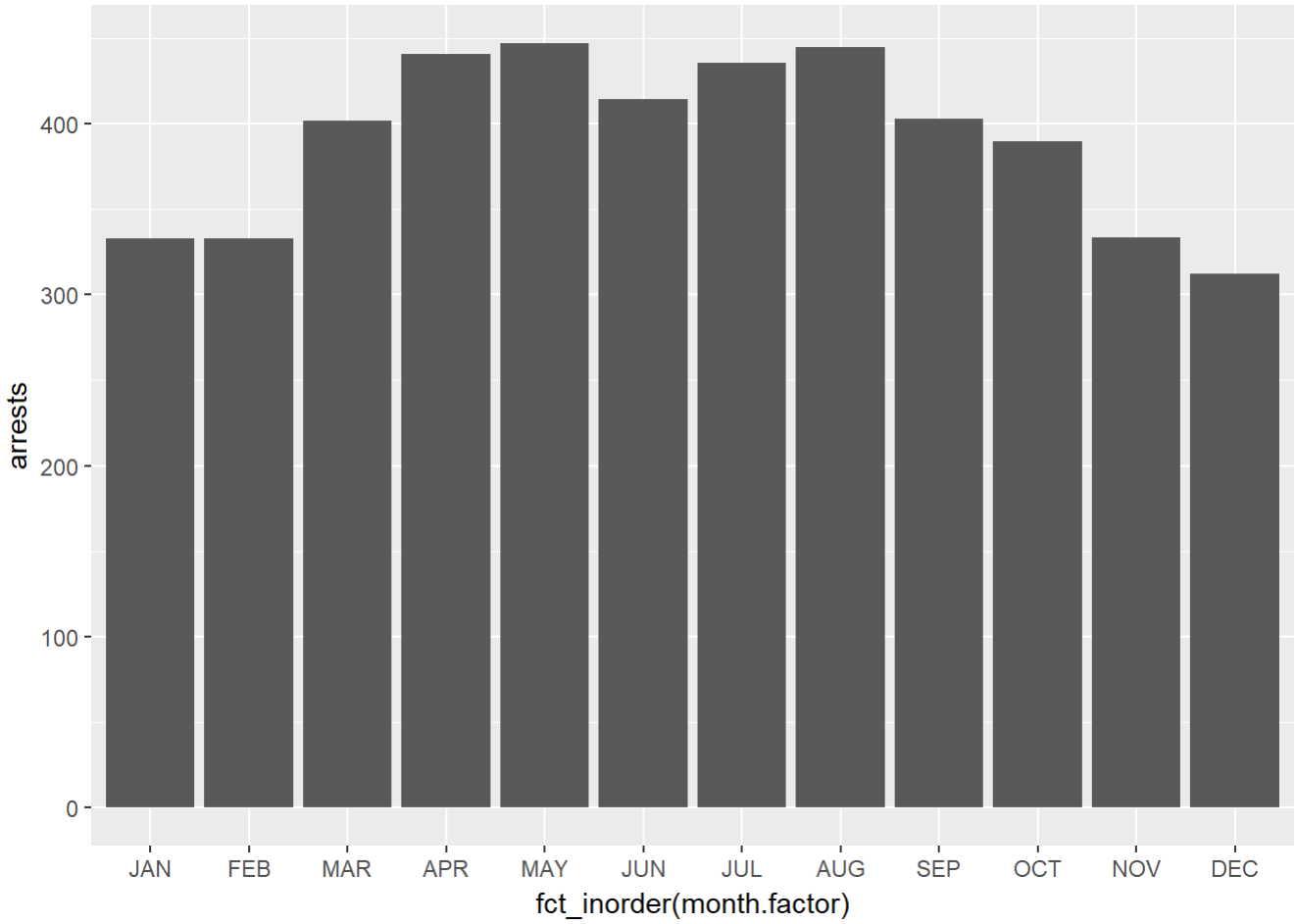
cleaning data

```
lab <- lab6 %>%
  separate(date_, c("month.factor","year1"), sep = " ") %>%
  mutate(law.numeric = as.numeric(lawprepo))
```

```
## Warning: package 'bindrcpp' was built under R version 3.3.3
```

1

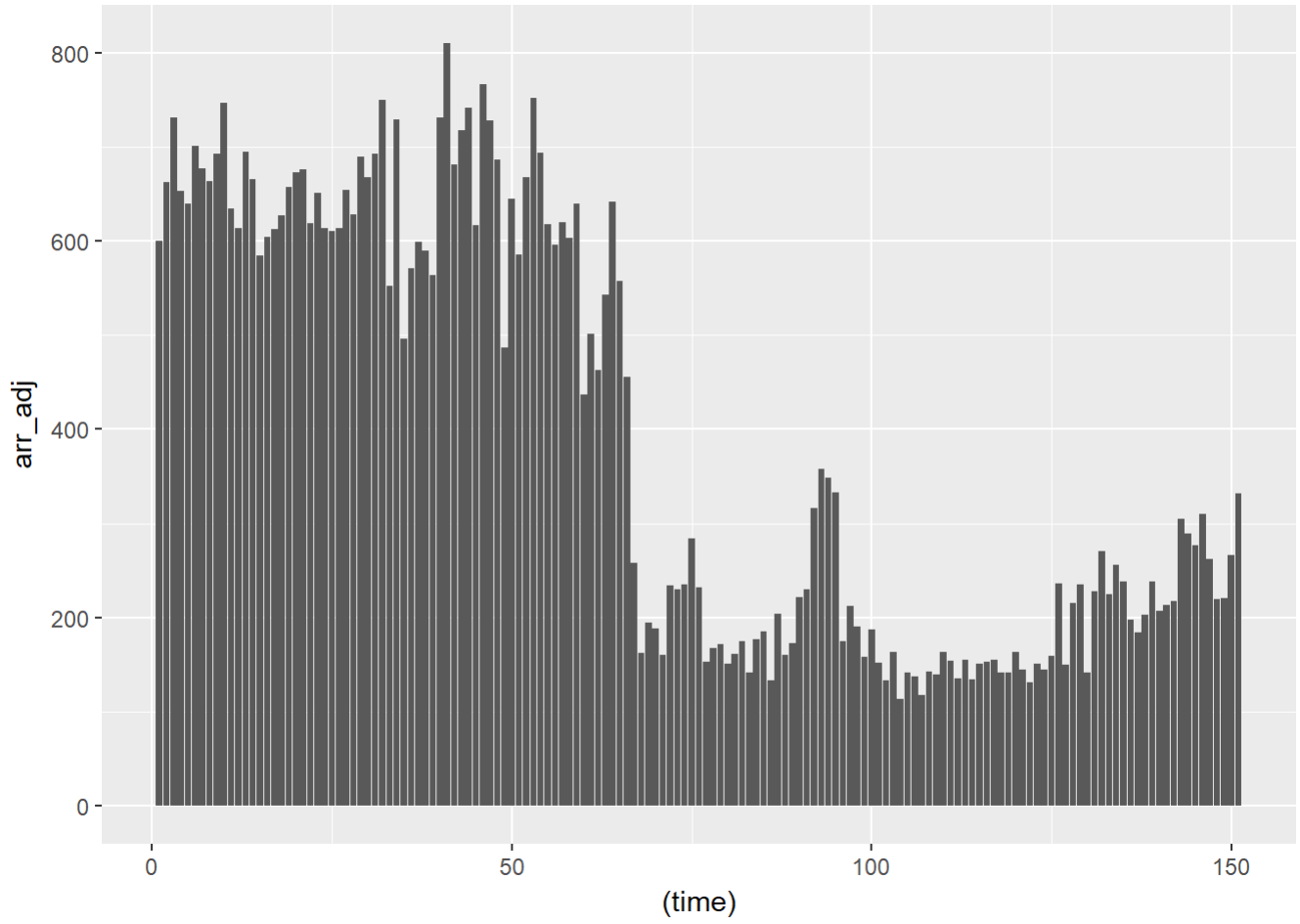
```
b <- ggplot(lab,aes(x=fct_inorder(month.factor),y = arrests))
b+stat_summary(fun.y=mean,geom="bar",position="dodge")
```



I'm not going to lie, it looks like there are more arrests happening from April to August, this is probably because ice cream sales are going up and apparently there is a correlation between ice cream sales and crime.

2

```
b <- ggplot(lab,aes(x=(time),y = arr_adj))
b+stat_summary(fun.y=mean,geom="bar",position="dodge")
```



Yes, Yes there does seem to be a reduction in arrests at that point.

3

```
mod <- lm(arr_adj~law.numeric,lab)
summary(mod)
```

```
##
## Call:
## lm(formula = arr_adj ~ law.numeric, data = lab)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -201.015  -44.570   -9.342   38.484  172.436
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1077.10      18.05   59.69   <2e-16 ***
## law.numeric  -439.63      11.00  -39.95   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67.08 on 149 degrees of freedom
## Multiple R-squared:  0.9146, Adjusted R-squared:  0.914
## F-statistic:  1596 on 1 and 149 DF,  p-value: < 2.2e-16
```

a Some durbins

There is not a autocorrelation so the assumptions are not met.

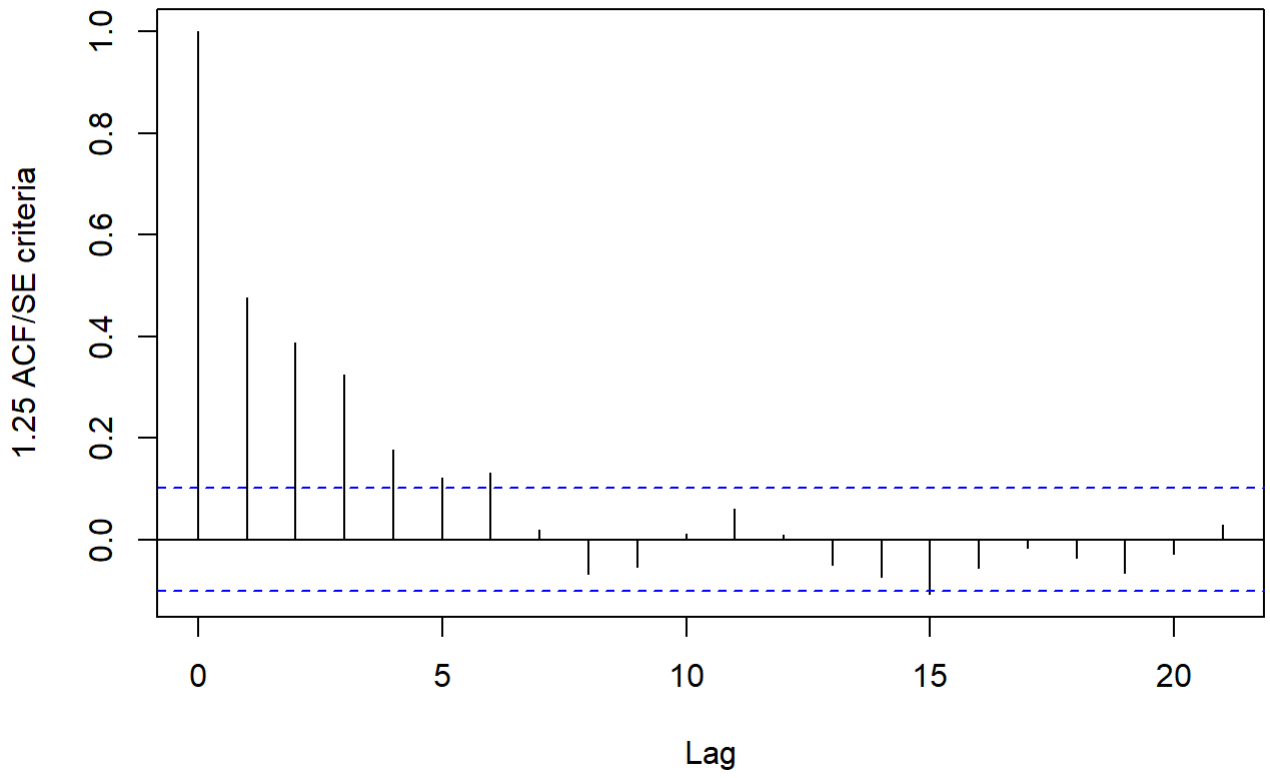
```
lmtest::dwtest(mod)
```

```
##
##  Durbin-Watson test
##
## data:  mod
## DW = 1.0176, p-value = 3.663e-10
## alternative hypothesis: true autocorrelation is greater than 0
```

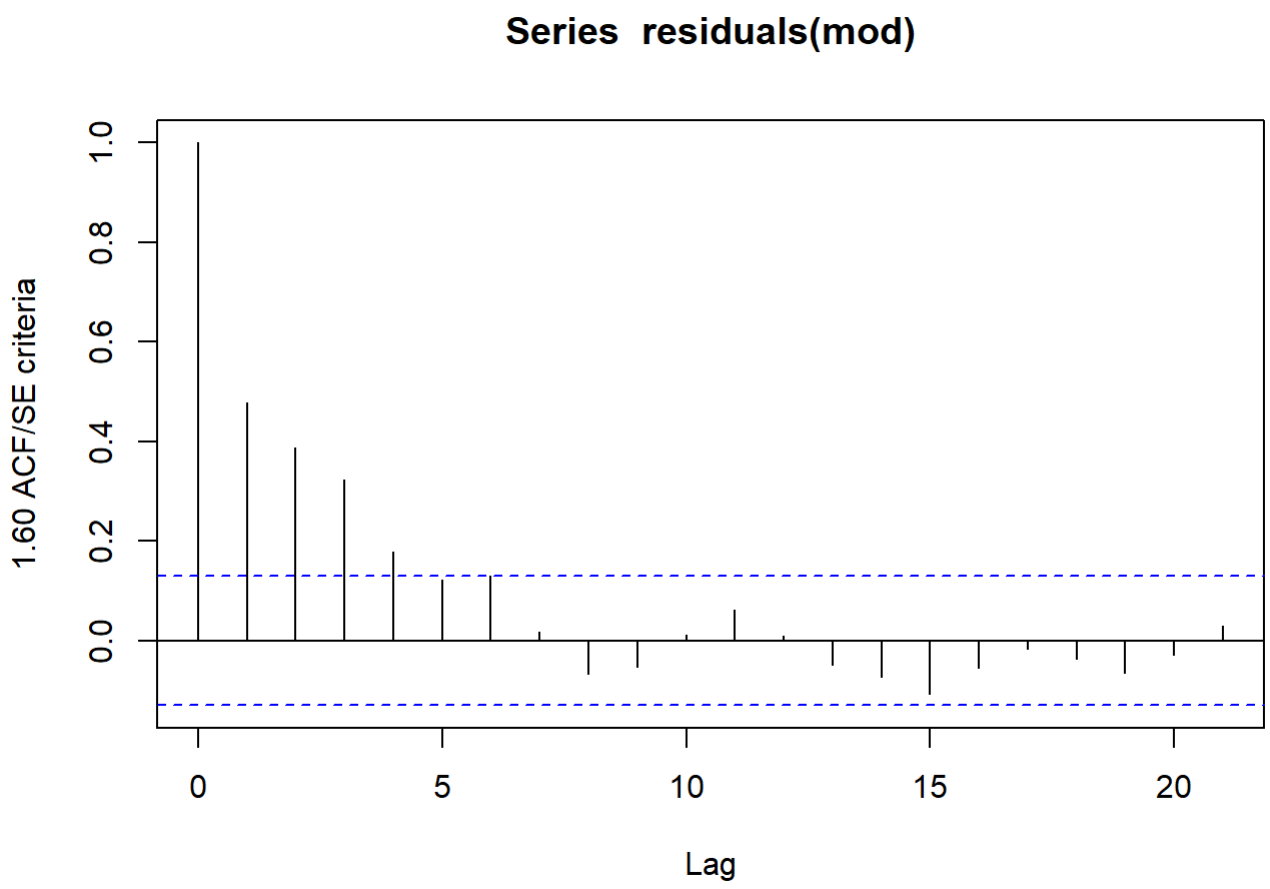
b Some Pankrantz

There are a series of lines spiking across the CI, so assumption is not met.

```
acf(residuals(mod), ci = .79, ylab="1.25 ACF/SE criteria")
```




```
acf(residuals(mod), ci = .89, ylab="1.60 ACF/SE criteria")
```



c Some BoxLjuns

These are just significant at the 1 lag.

```
Box.test(residuals(mod), lag = 1, type = c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  residuals(mod)
## X-squared = 35.021, df = 1, p-value = 3.262e-09
```

```
Box.test(residuals(mod), lag = 2, type = c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  residuals(mod)
## X-squared = 58.243, df = 2, p-value = 2.253e-13
```

```
Box.test(residuals(mod), lag = 3, type = c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  residuals(mod)
## X-squared = 74.612, df = 3, p-value = 4.441e-16
```

```
Box.test(residuals(mod), lag = 4, type = c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  residuals(mod)
## X-squared = 79.558, df = 4, p-value = 2.22e-16
```

d

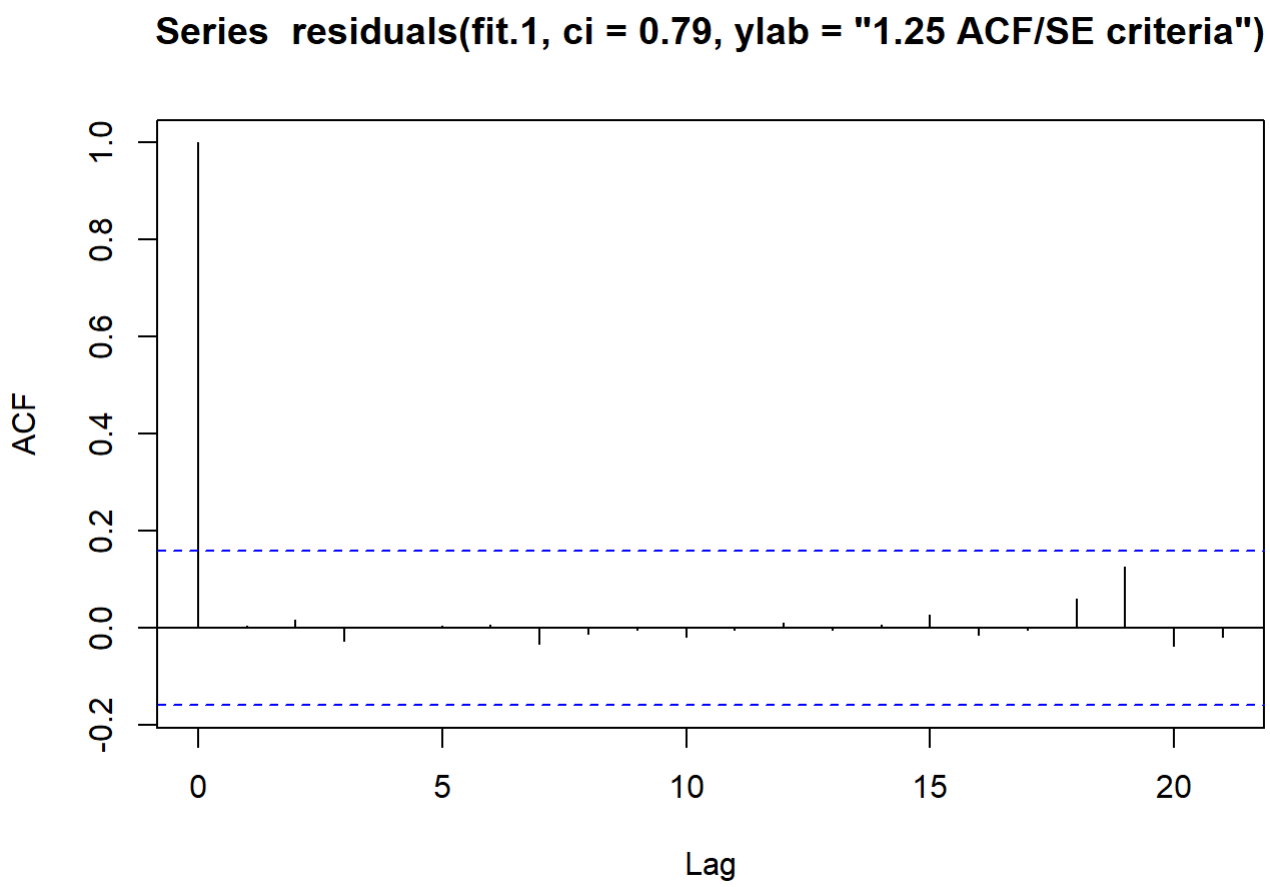
Quite Frankly, a LM is not appropriate to use.

4

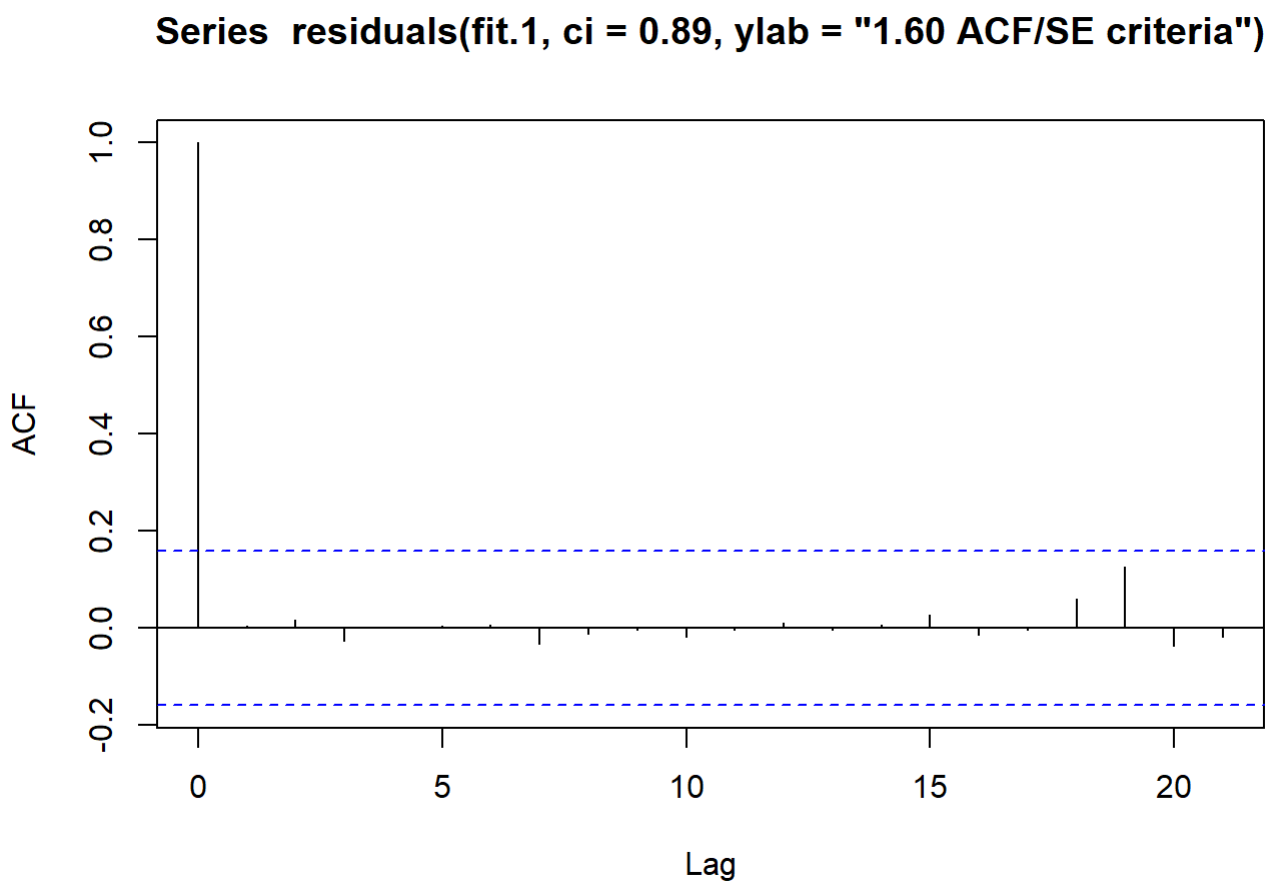
```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 3.3.3
```

```
fit.1 <- Arima(lab$arr_adj,lab$law.numeric,order=c(19,0,0))
fit.3 <- Arima(lab$arr_adj,lab$law.numeric,order=c(3,0,0))
fit.2 <- Arima(lab$arr_adj,lab$law.numeric,order=c(2,0,0))
acf(residuals(fit.1, ci=.79,ylab="1.25 ACF/SE criteria"))
```



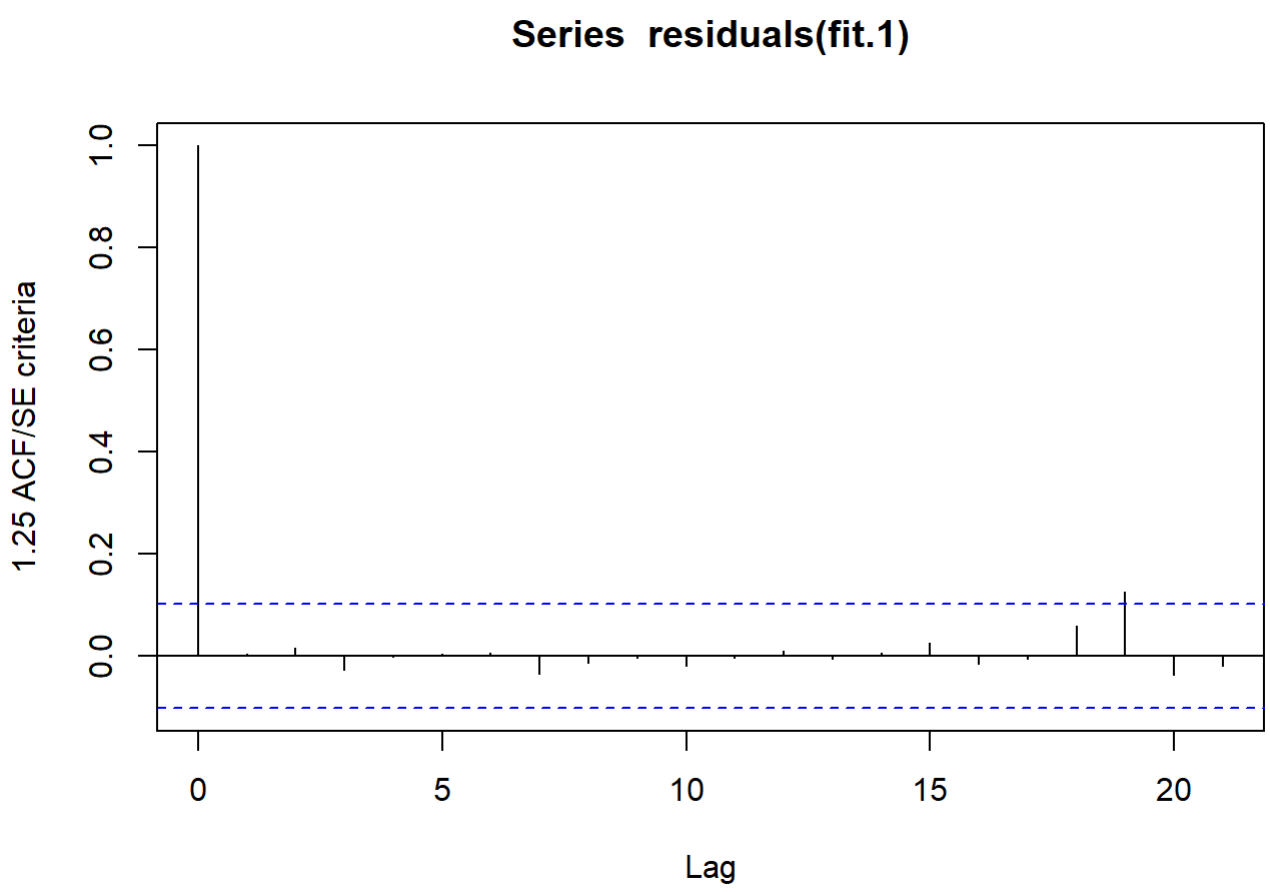

```
acf(residuals(fit.1, ci=.89,ylab="1.60 ACF/SE criteria"))
```



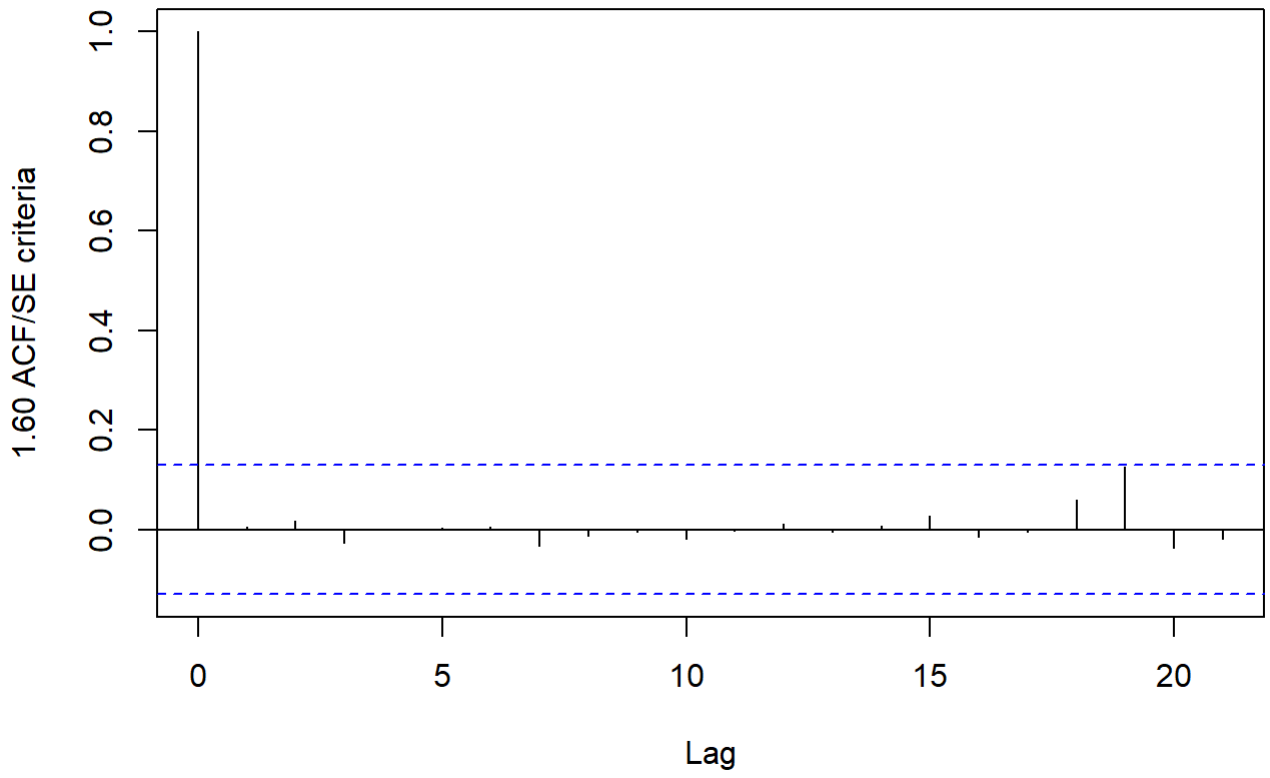
a

Assumptions

```
acf(residuals(fit.1), ci = .79, ylab="1.25 ACF/SE criteria")
```




```
acf(residuals(fit.1), ci = .89, ylab="1.60 ACF/SE criteria")
```



##b

```
Box.test(residuals(fit.1), lag = 1, type = c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  residuals(fit.1)
## X-squared = 0.0029018, df = 1, p-value = 0.957
```

```
Box.test(residuals(fit.1), lag = 2, type = c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  residuals(fit.1)
## X-squared = 0.04356, df = 2, p-value = 0.9785
```

```
Box.test(residuals(fit.1), lag = 3, type = c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  residuals(fit.1)
## X-squared = 0.15766, df = 3, p-value = 0.9841
```

```
Box.test(residuals(fit.1), lag = 4, type = c("Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  residuals(fit.1)
## X-squared = 0.1578, df = 4, p-value = 0.997
```

c

19 autoregressions seemed to work.

5

```
summary(fit.1)
```

```
## Series: lab$arr_adj
## ARIMA(19,0,0) with non-zero mean
##
## Coefficients:
##          ar1     ar2     ar3      ar4      ar5     ar6     ar7      ar8
##       0.6163  0.3701  0.0536  -0.3008  -0.0339  0.3613  0.1359  -0.3377
## s.e.  0.0820  0.0959  0.0997   0.1013   0.1029  0.1054  0.1081   0.1080
##           ar9    ar10    ar11     ar12     ar13   ar14    ar15    ar16
##       -0.0592  0.2899  0.2091  -0.1866  -0.1950  0.029  0.0256  0.0395
## s.e.   0.1118  0.1103  0.1140   0.1113   0.1136  0.109  0.1098  0.1049
##          ar17    ar18     ar19      mean
##       -0.1165  0.1241  -0.0529  426.1412
## s.e.   0.1050  0.0979   0.0843  130.6363
##
## sigma^2 estimated as 3576:  log likelihood=-823.79
## AIC=1689.57   AICc=1696.73   BIC=1752.93
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -2.63368 55.69828 40.28537 -3.650653 13.41538 0.8465304
##                      ACF1
## Training set 0.004340563
```

```
test <- fit.1$coef/sqrt(diag(fit.1$var.coef))
(1-pnorm(abs(fit.1$coef)/sqrt(diag(fit.1$var.coef))))*2
```

```
##          ar1          ar2          ar3          ar4          ar5
## 5.817569e-14 1.144321e-04 5.907871e-01 2.968852e-03 7.417666e-01
##          ar6          ar7          ar8          ar9         ar10
## 6.123558e-04 2.083964e-01 1.759638e-03 5.961706e-01 8.610909e-03
##         ar11         ar12         ar13         ar14         ar15
## 6.675294e-02 9.358785e-02 8.605766e-02 7.901098e-01 8.159861e-01
##         ar16         ar17         ar18         ar19    intercept
## 7.065813e-01 2.668534e-01 2.049352e-01 5.306182e-01 1.106122e-03
```

There was a significant decrease in arrests after a new law was instituted $b = 3.26$, $p < .05$