Homework 3 Exercise 1

Chutian Chen cc4515; Congcheng Yan cy2550; Mingrui Liu ml4404

2020/4/16

Question 1

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
milk <- read.csv("milk.csv", sep = ",", header = TRUE,
                 col.names = c("Month", "Production"))
```

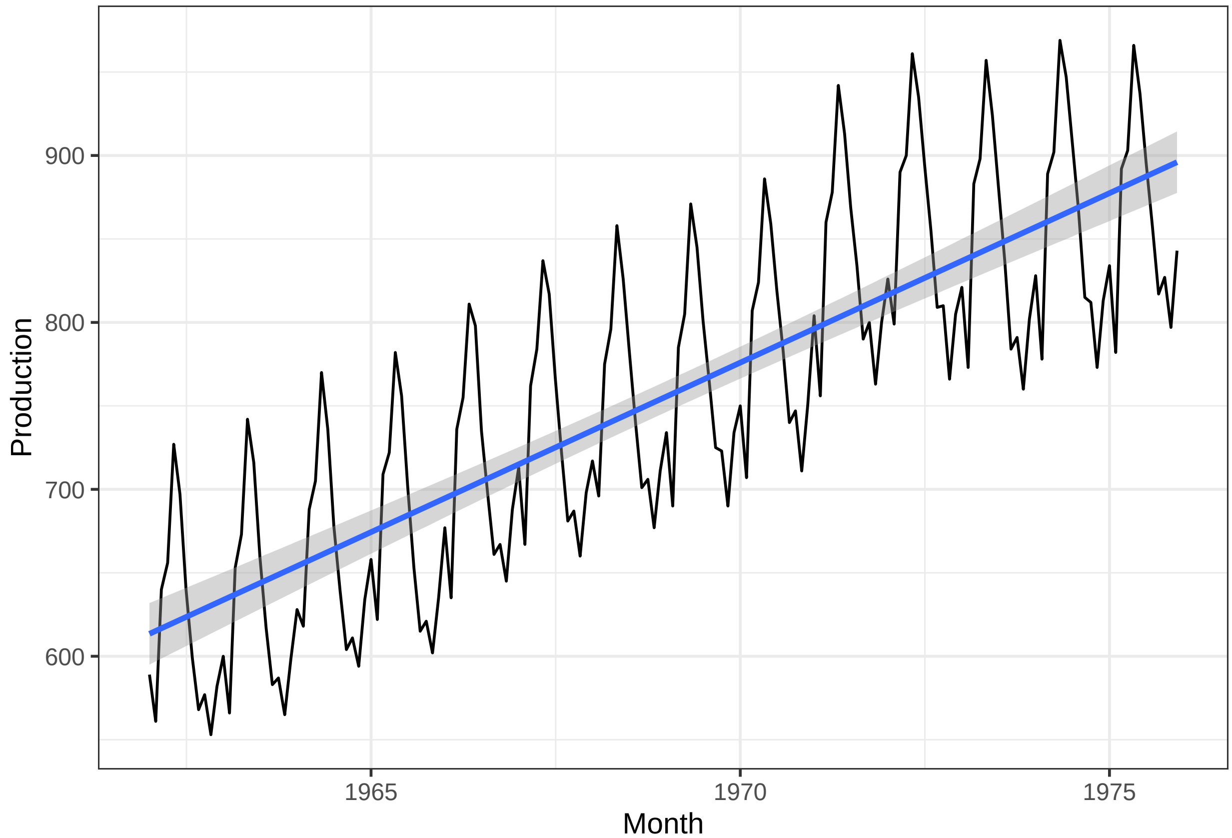
```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

```
milk <- milk %>%
  mutate(Month=ymd(Month, truncated = 1)) %>%
  mutate(Row=row_number())
```

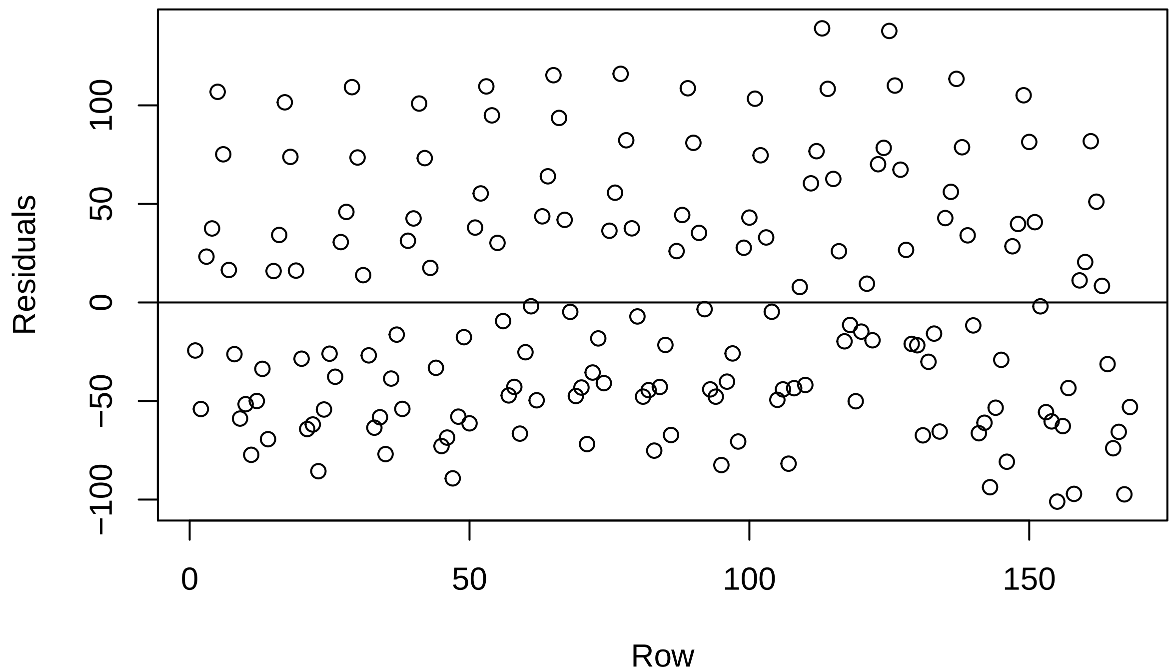
```
library(ggplot2)
ggplot(milk, aes(x=Month, y=Production)) +
  geom_line() +
  geom_smooth(method = "lm") +
  theme_bw()
```




```
milk.lm <- lm(Production ~ Row, milk)
milk.lm
```

```
##
## Call:
## lm(formula = Production ~ Row, data = milk)
##
## Coefficients:
## (Intercept)          Row
##     611.682        1.693
```

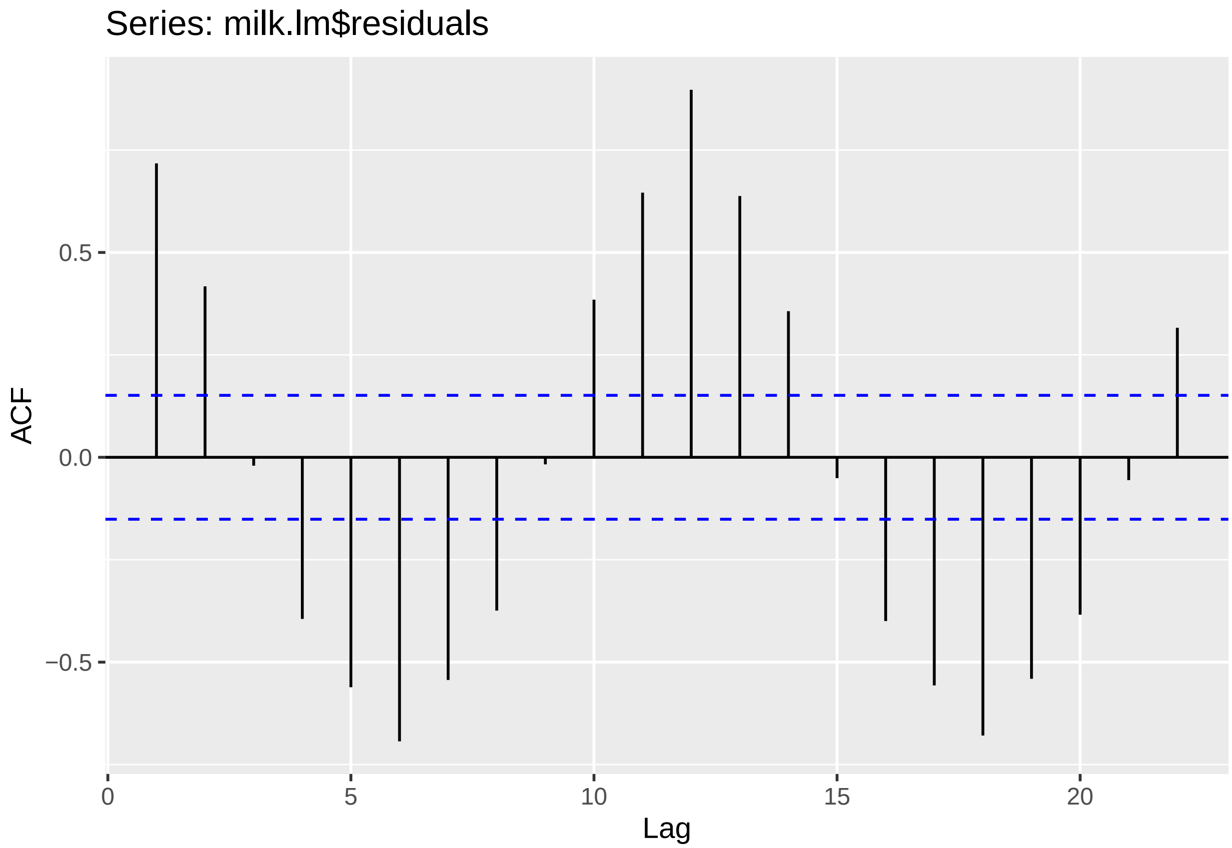
```
res = resid(milk.lm)
plot(milk$Row, res,
    ylab="Residuals", xlab="Row")
abline(0, 0)
```



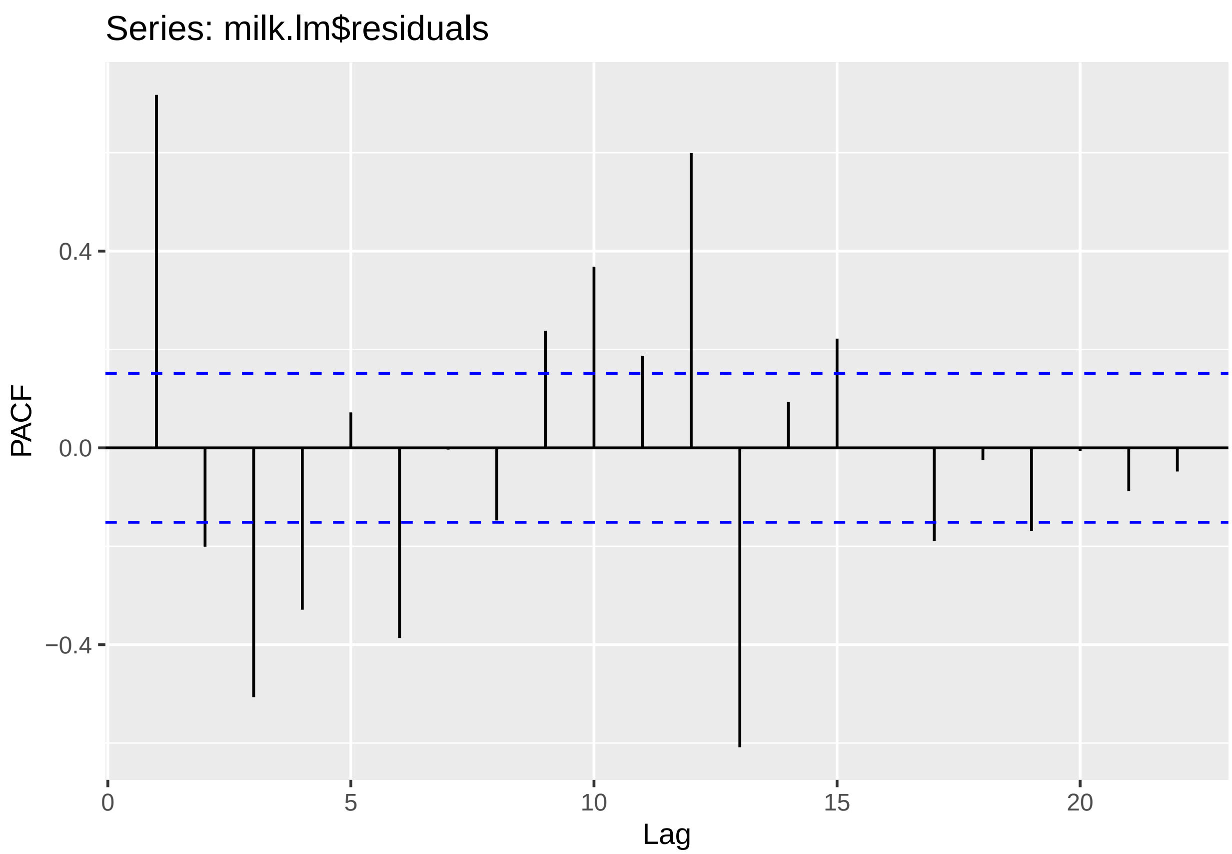
We make the regression with index instead of the timestamp. We can find a good inceasing trend through this linear regression. Also, the residuals seem to have a random value with mean = 0.

Question 2

```
library(forecast)
ggAcf(milk.lm$residuals)
```




```r
ggPacf(milk.lm$residuals)
```



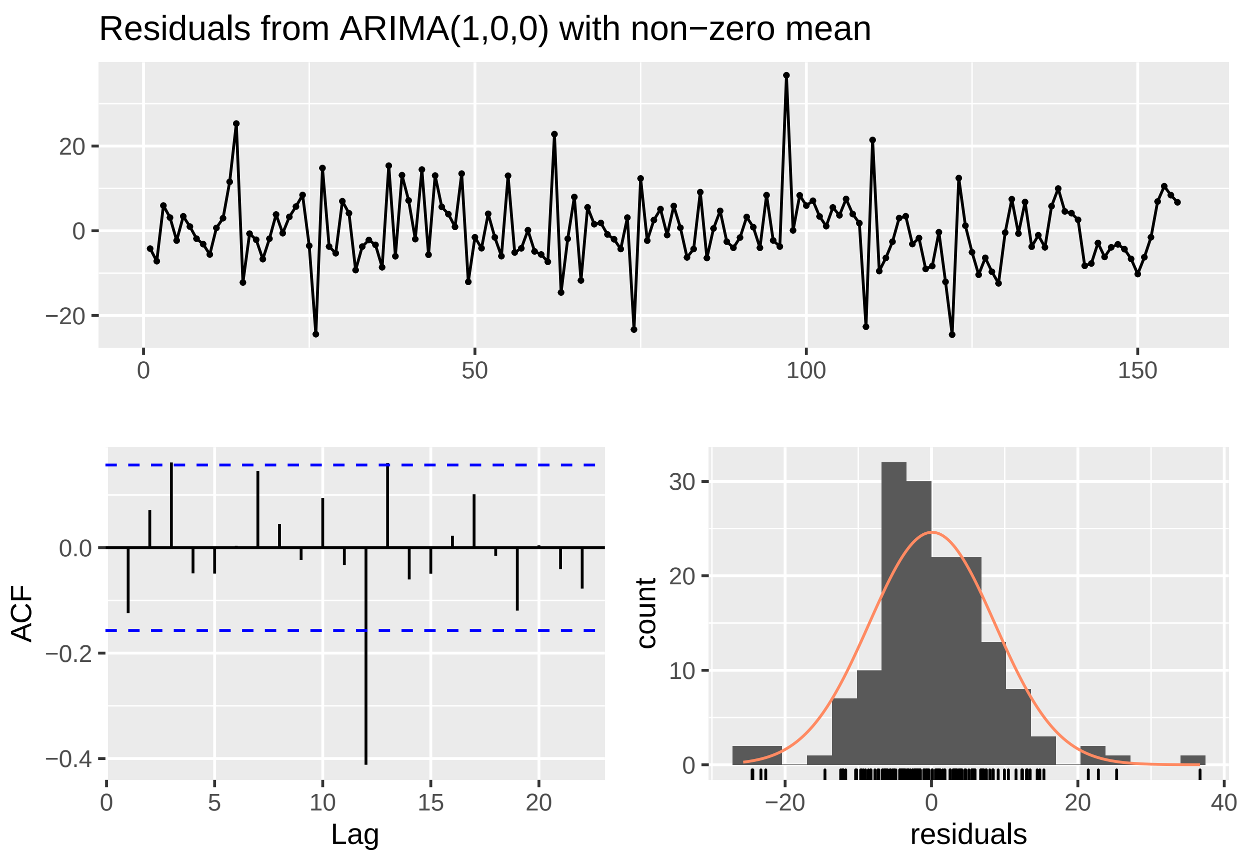
We can find a obvious seasonal trend in the graph. Maybe AR(1), AR(2) will have a better result.

Question 3

```
AR1 <- Arima(res %>% diff(12), order=c(1,0,0))
AR1
```

```
## Series: res %>% diff(12)
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1     mean
##       0.8543  -1.2121
## s.e.  0.0404   4.5427
##
## sigma^2 estimated as 74.42:  log likelihood=-557.17
## AIC=1120.33   AICc=1120.49   BIC=1129.48
```

```
checkresiduals(AR1)
```

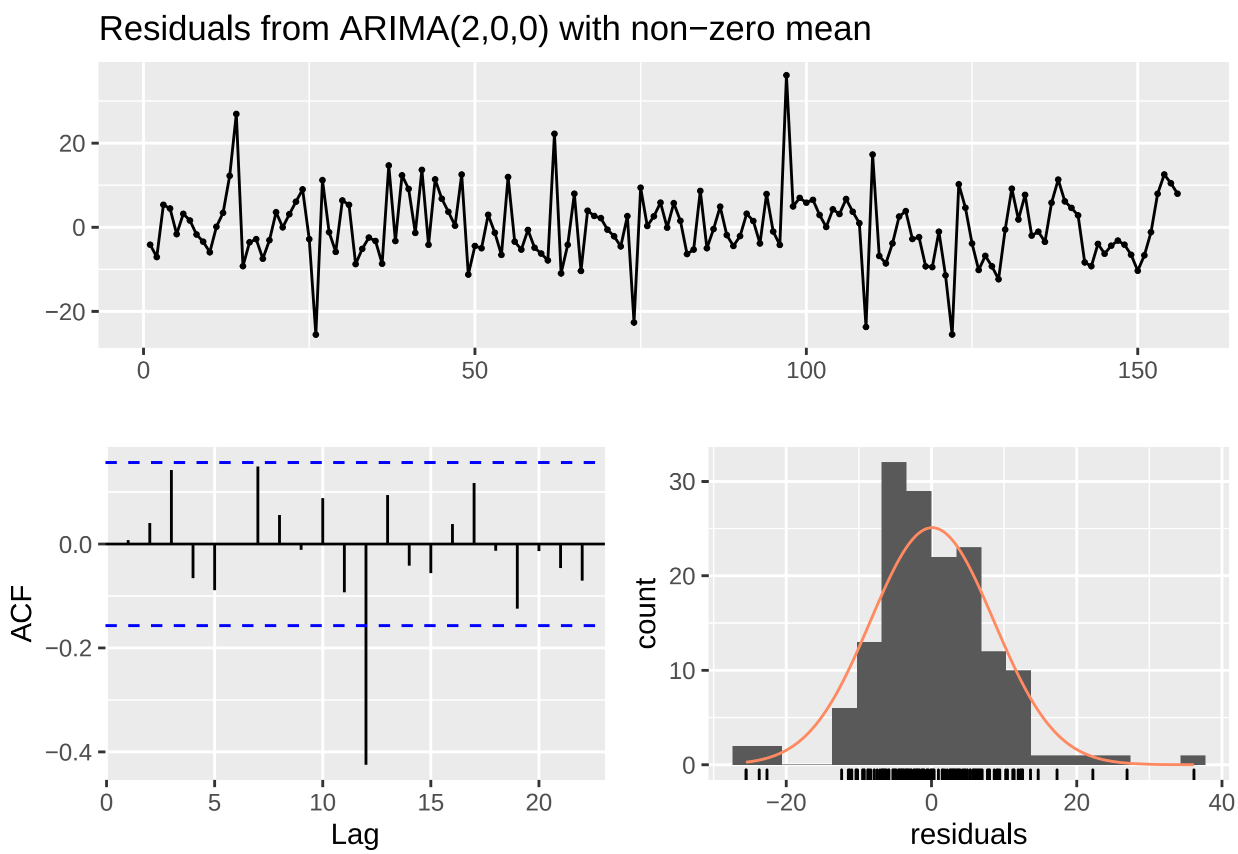



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0) with non-zero mean
## Q* = 13.707, df = 8, p-value = 0.08972
##
## Model df: 2.   Total lags used: 10
```

```r
AR2 <- Arima(res %>% diff(12), order=c(2,0,0))
AR2
```

```
## Series: res %>% diff(12)
## ARIMA(2,0,0) with non-zero mean
##
## Coefficients:
##          ar1     ar2     mean
##       0.7242  0.1512  -1.2623
## s.e.  0.0789  0.0790   5.1753
##
## sigma^2 estimated as 73.18:  log likelihood=-555.36
## AIC=1118.72   AICc=1118.98   BIC=1130.92
```

```r
checkresiduals(AR2)
```

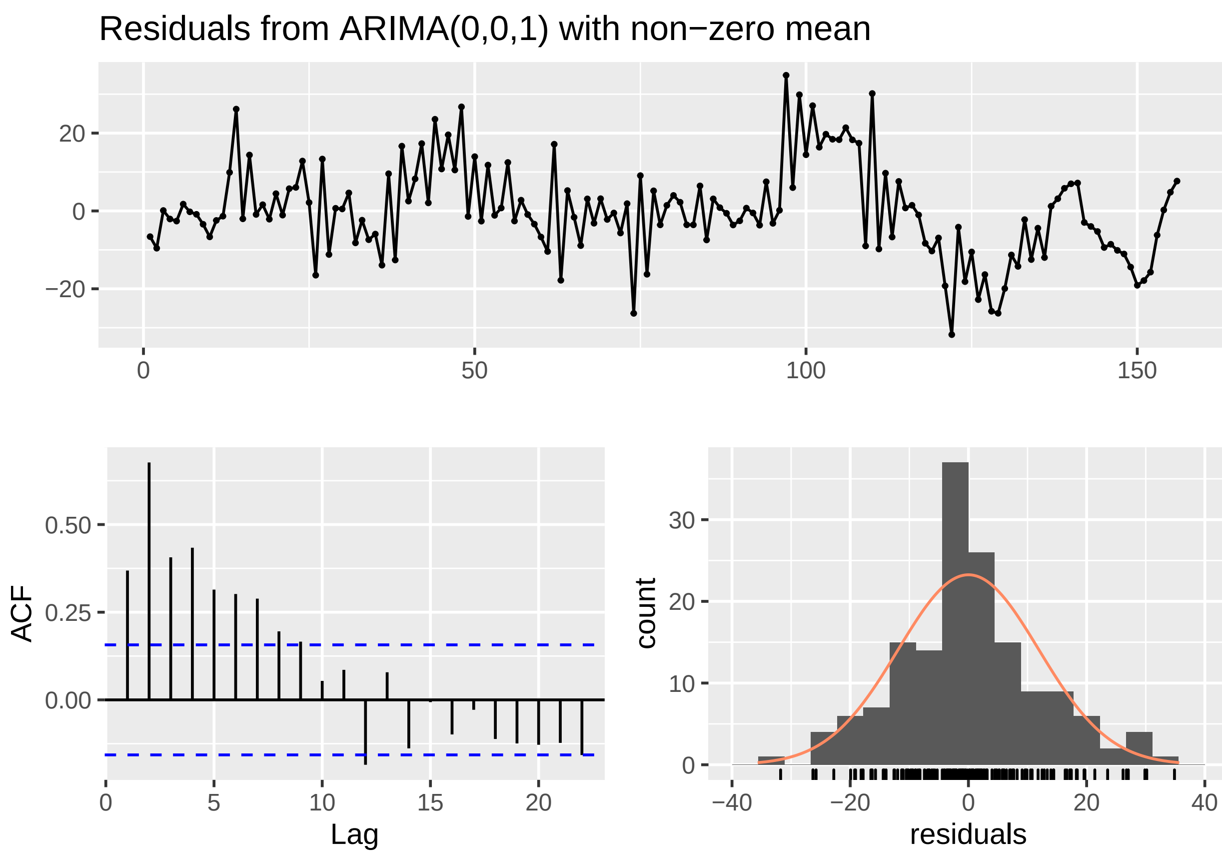



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,0,0) with non-zero mean
## Q* = 11.072, df = 7, p-value = 0.1355
##
## Model df: 3.   Total lags used: 10
```

```
MA1 <- Arima(res %>% diff(12), order=c(0,0,1))
MA1
```

```
## Series: res %>% diff(12)
## ARIMA(0,0,1) with non-zero mean
##
## Coefficients:
##          ma1     mean
##       0.6880  -1.2906
## s.e.  0.0496   1.5977
##
## sigma^2 estimated as 142.3:  log likelihood=-607.39
## AIC=1220.78   AICc=1220.94   BIC=1229.93
```

```
checkresiduals(MA1)
```

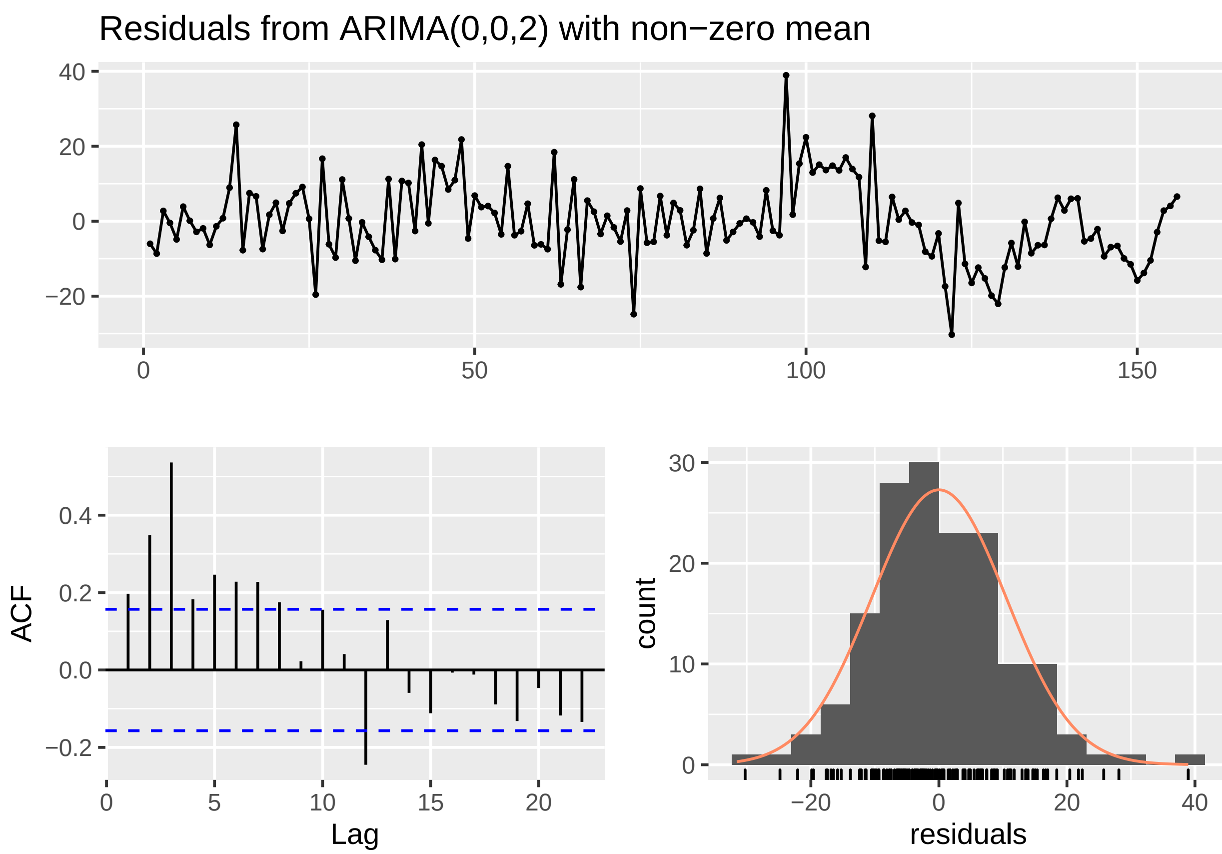



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,1) with non-zero mean
## Q* = 208.48, df = 8, p-value < 2.2e-16
##
## Model df: 2.   Total lags used: 10
```

```
MA2 <- Arima(res %>% diff(12), order=c(0,0,2))
MA2
```

```
## Series: res %>% diff(12)
## ARIMA(0,0,2) with non-zero mean
##
## Coefficients:
##          ma1     ma2     mean
##       0.7665  0.4426  -1.2676
## s.e.  0.0720  0.0684   1.8470
##
## sigma^2 estimated as 112.2:  log likelihood=-588.44
## AIC=1184.88   AICc=1185.15   BIC=1197.08
```

```
checkresiduals(MA2)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,2) with non-zero mean
## Q* = 113.62, df = 7, p-value < 2.2e-16
##
## Model df: 3.   Total lags used: 10
```

If we pick AIC as the evaluation, AR1 and AR2 are similar. AR2 is the best fit among the 4 fits.

Question 4

```r
AIC(Arima(res %>% diff(12), order=c(2,0,1)))
```

```
## [1] 1120.707
```

```r
AIC(Arima(res %>% diff(12), order=c(2,0,2)))
```

```
## [1] 1119.821
```

```r
AIC(Arima(res %>% diff(12), order=c(2,0,3)))
```

```
## [1] 1114.429
```

```r
AIC(Arima(res %>% diff(12), order=c(3,0,1)))
```

```
## [1] 1116.066
```

```r
AIC(Arima(res %>% diff(12), order=c(3,0,2)))
```

```
## [1] 1116.468
```

```r
AIC(Arima(res %>% diff(12), order=c(3,0,3)))
```

```
## [1] 1121.527
```

So, the top 2 fits are ARMA(2,3) and ARMA(3,1)