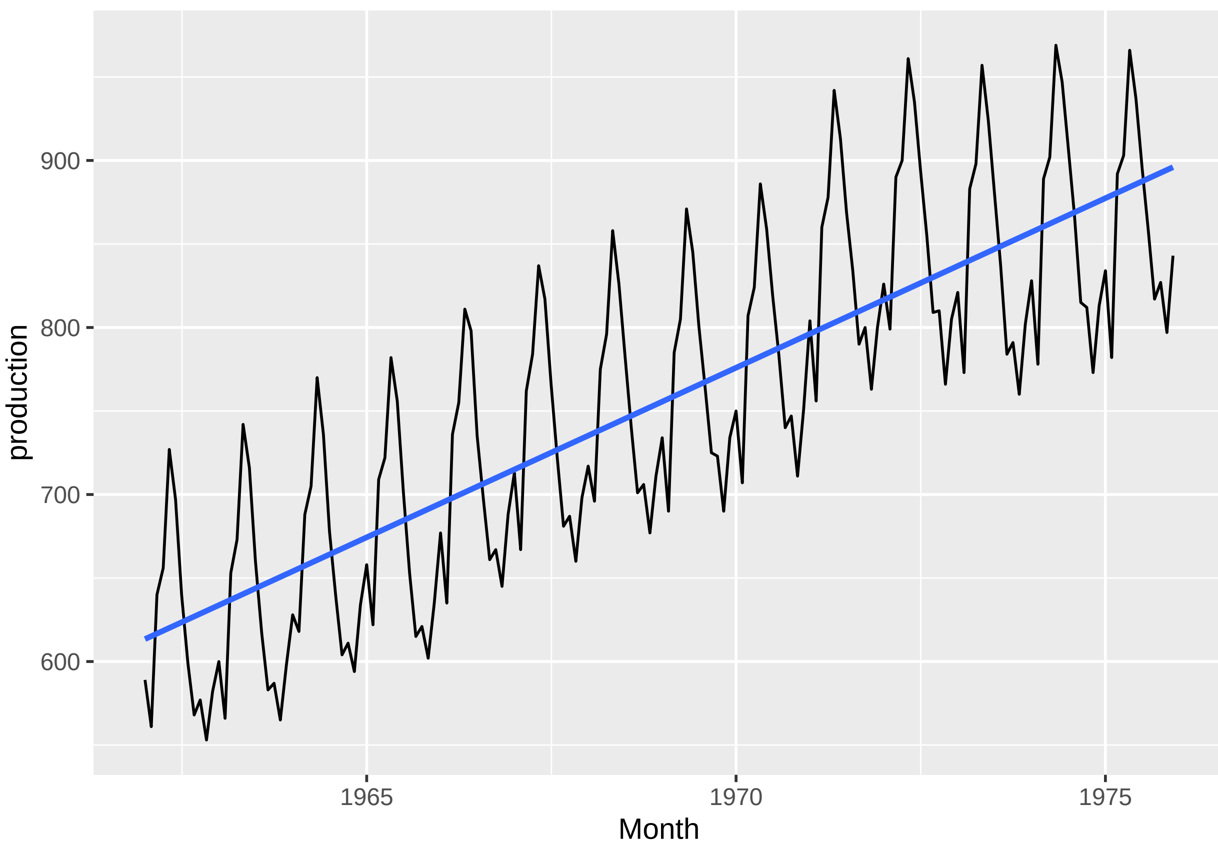
STAT5703 HW3 Exercise 1

Wen Fan(wf2255), Banruo Xie(bx2168), Hanjun Li(hl3339)

Exercise 1.

Question 1.

```
df_milk <- read.table(file ="milk.txt",skip = 15,col.names=c("Month", "production"))
df_milk$Month<-ymd(df_milk$Month,truncated = 1)
df_milk <- df_milk %>%
  mutate(num_month=row_number())
df_milk %>% ggplot(aes(x=Month, y=production)) +
  geom_line() +
  geom_smooth(method = "lm", se = FALSE)
```



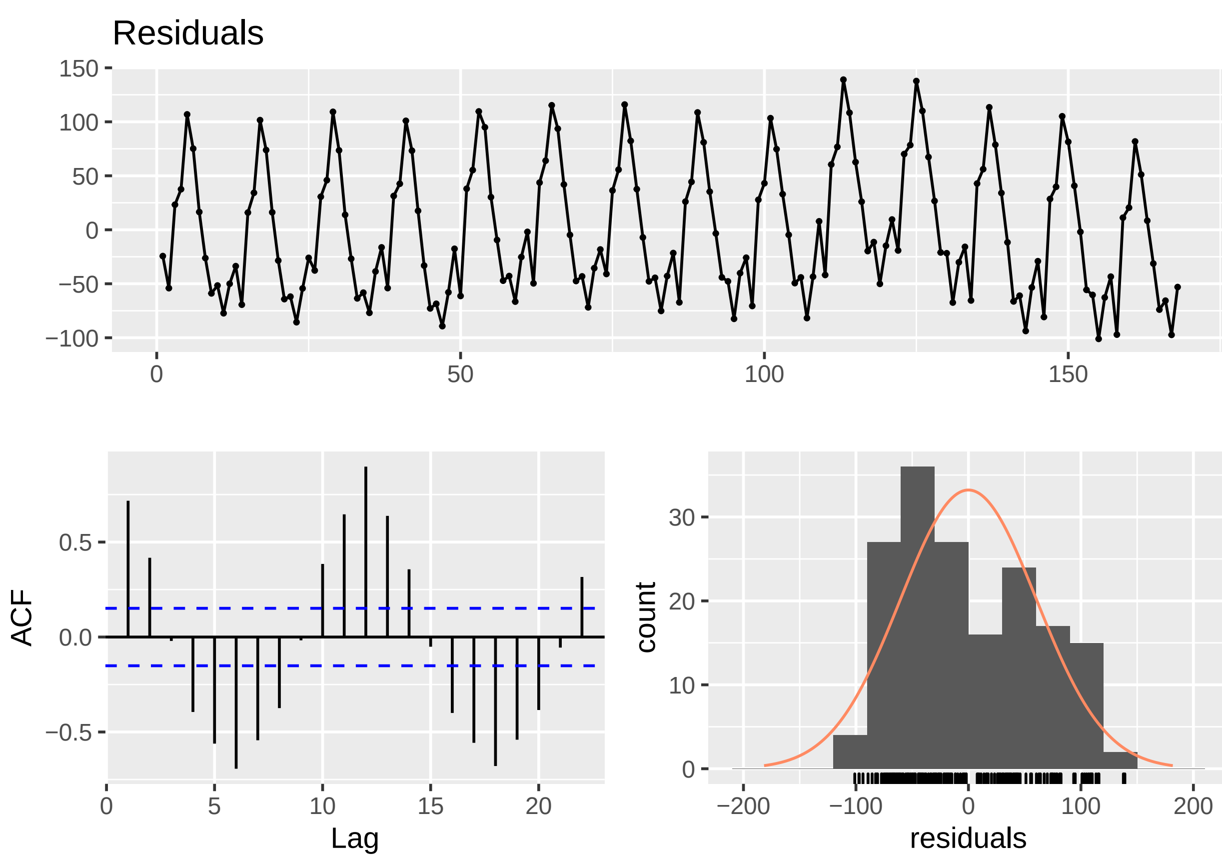

```
linearMod <- lm(production ~ num_month, data=df_milk)
summary(linearMod)
```

```
##
## Call:
## lm(formula = production ~ num_month, data = df_milk)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -101.04  -50.02  -15.30   42.88  139.05
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 611.68235    9.41444   64.97   <2e-16 ***
## num_month     1.69262    0.09663   17.52   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.74 on 166 degrees of freedom
## Multiple R-squared:  0.6489, Adjusted R-squared:  0.6468
## F-statistic: 306.8 on 1 and 166 DF,  p-value: < 2.2e-16
```

From linear model, the production is 611 pounds per cow for the first month and increase 1.69 pounds per cow for each month.

```r
checkresiduals(linearMod)
```

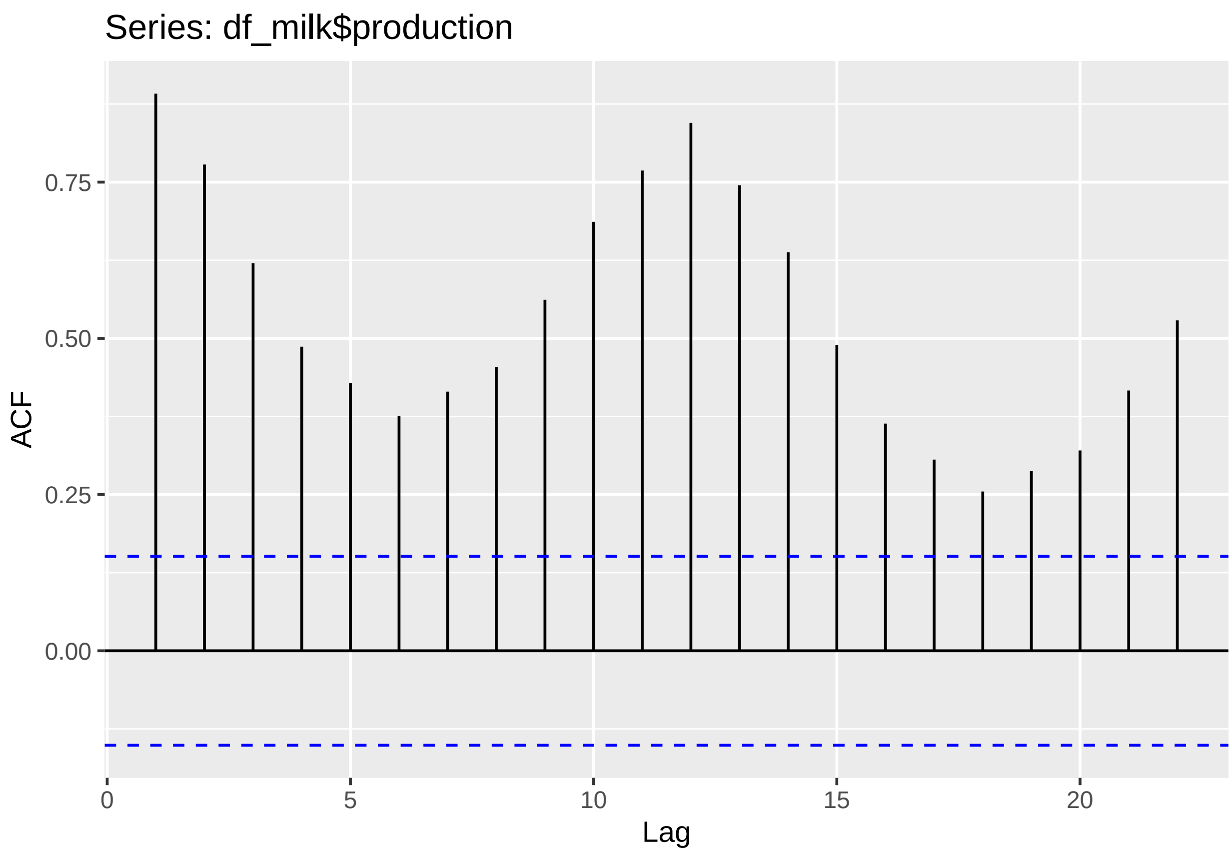



```
##
##  Breusch-Godfrey test for serial correlation of order up to 10
##
## data:  Residuals
## LM test = 135.24, df = 10, p-value < 2.2e-16
```

The residuals are normally distributed, we can say it is a stationary time series

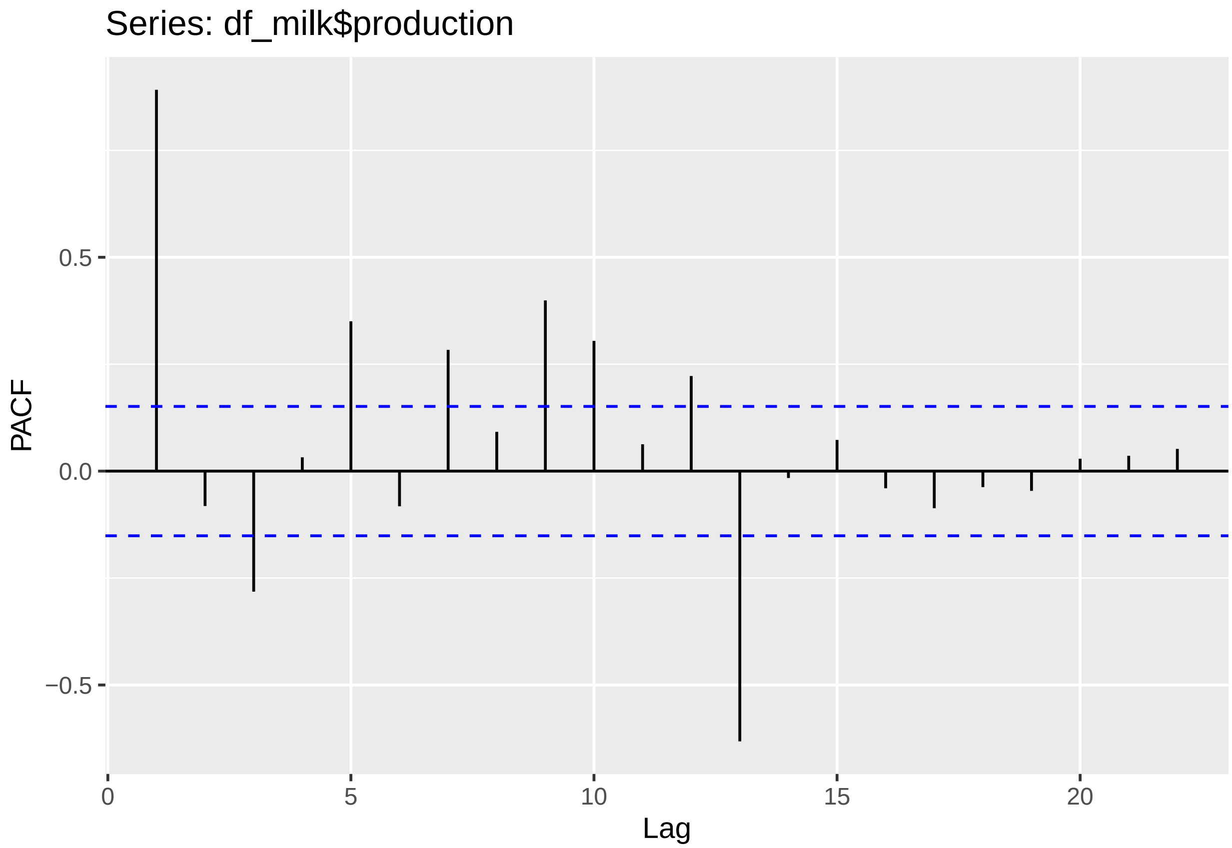
Question 2.

```
ggAcf(df_milk$production)
```



From ACF plot above. the autocorrelation crosses the dashed blue line, it means that specific lag is significantly correlated with current series. The slow decrease in the ACF as the lags increase and due to the seasonality.

```
ggPacf(df_milk$production)
```



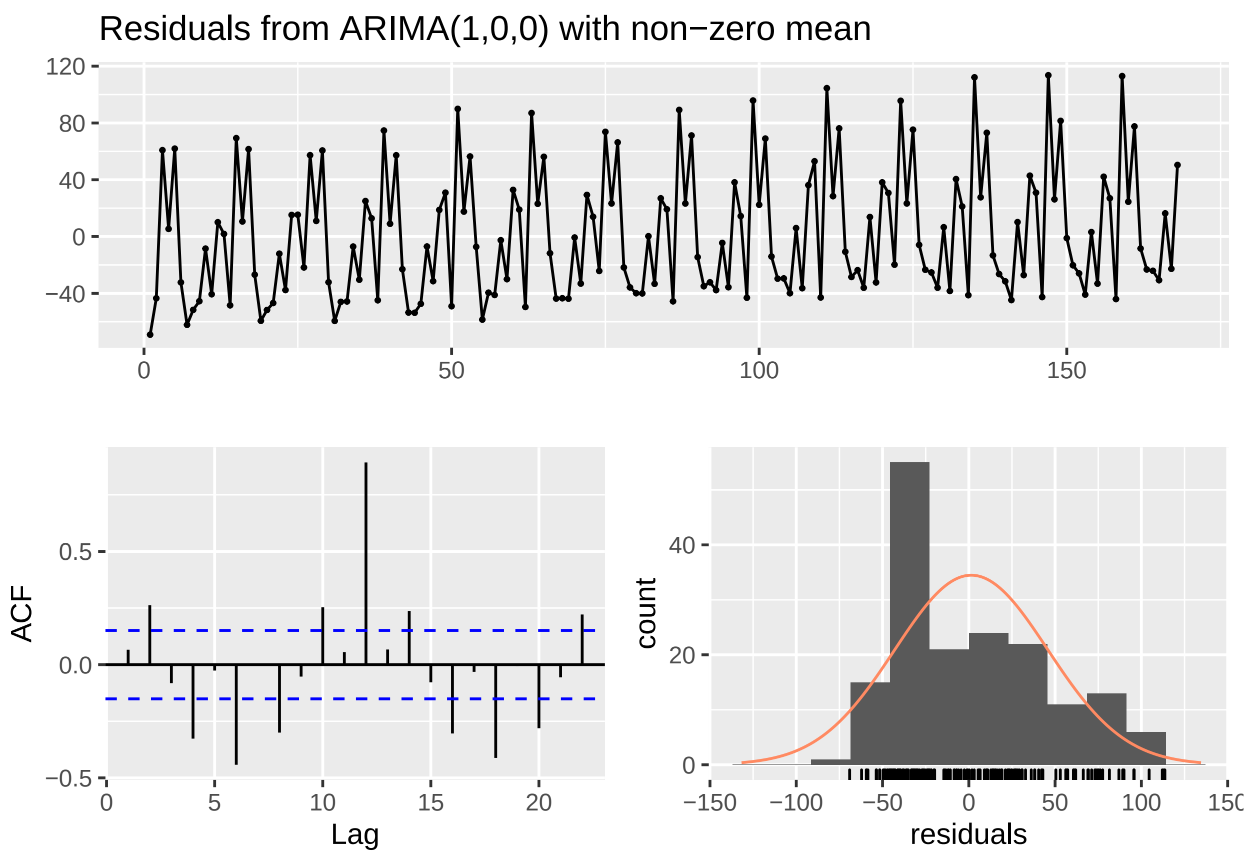
After remove linear trends in a timeseries, we can say the plot indicates a seasonal AR(1) component because the Pacf Cuts off after lag 1.

Question 3.

```
fitAR1 <- Arima(df_milk$production, order=c(1,0,0))
fitAR1
```

```
## Series: df_milk$production
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1      mean
##       0.9043  750.7911
## s.e.  0.0328   33.8590
##
## sigma^2 estimated as 1983:  log likelihood=-875.99
## AIC=1757.99   AICc=1758.13   BIC=1767.36
```

```
checkresiduals(fitAR1)
```

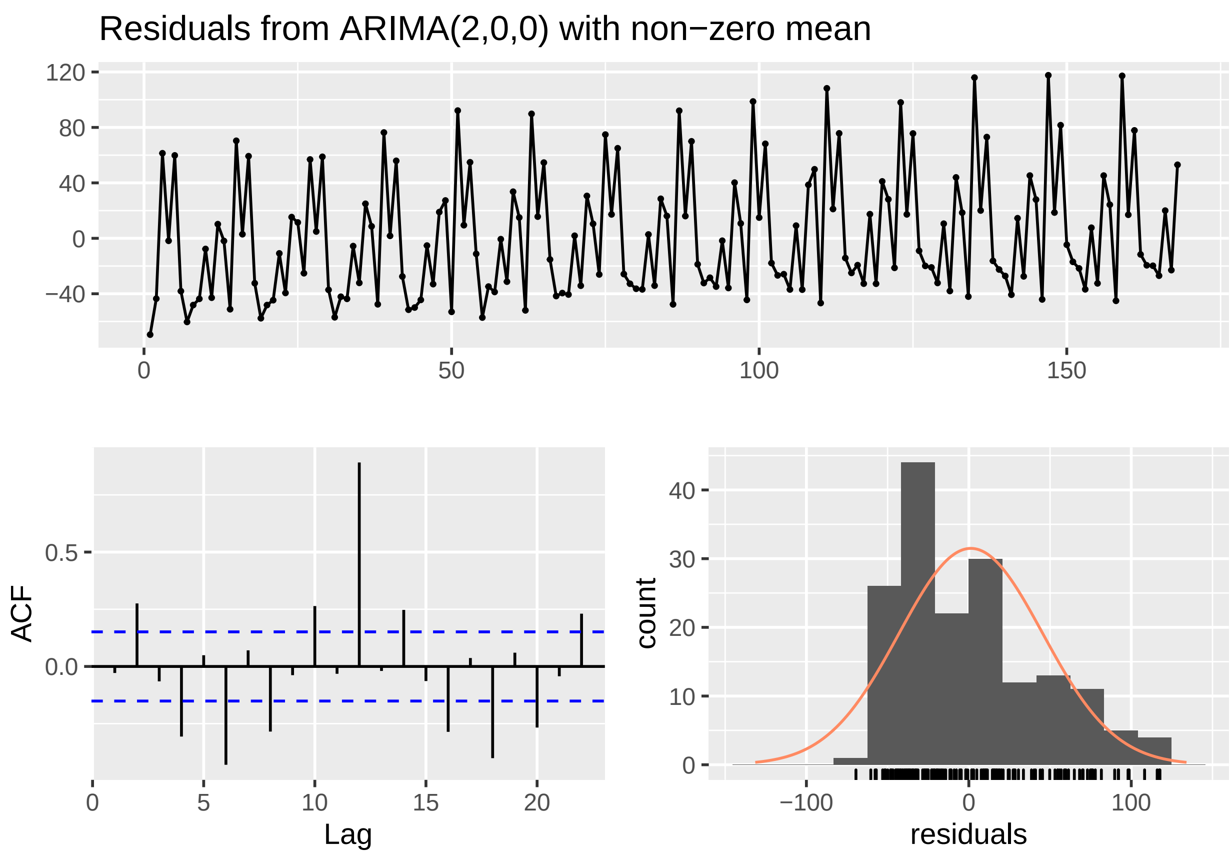



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0) with non-zero mean
## Q* = 95.079, df = 8, p-value < 2.2e-16
##
## Model df: 2.   Total lags used: 10
```

```
fitAR2 <- Arima(df_milk$production, order=c(2,0,0))
fitAR2
```

```
## Series: df_milk$production
## ARIMA(2,0,0) with non-zero mean
##
## Coefficients:
##          ar1      ar2      mean
##       0.9742  -0.0782  751.7093
## s.e.  0.0768   0.0776   31.3513
##
## sigma^2 estimated as 1983:  log likelihood=-875.49
## AIC=1758.97   AICc=1759.22   BIC=1771.47
```

```
checkresiduals(fitAR2)
```

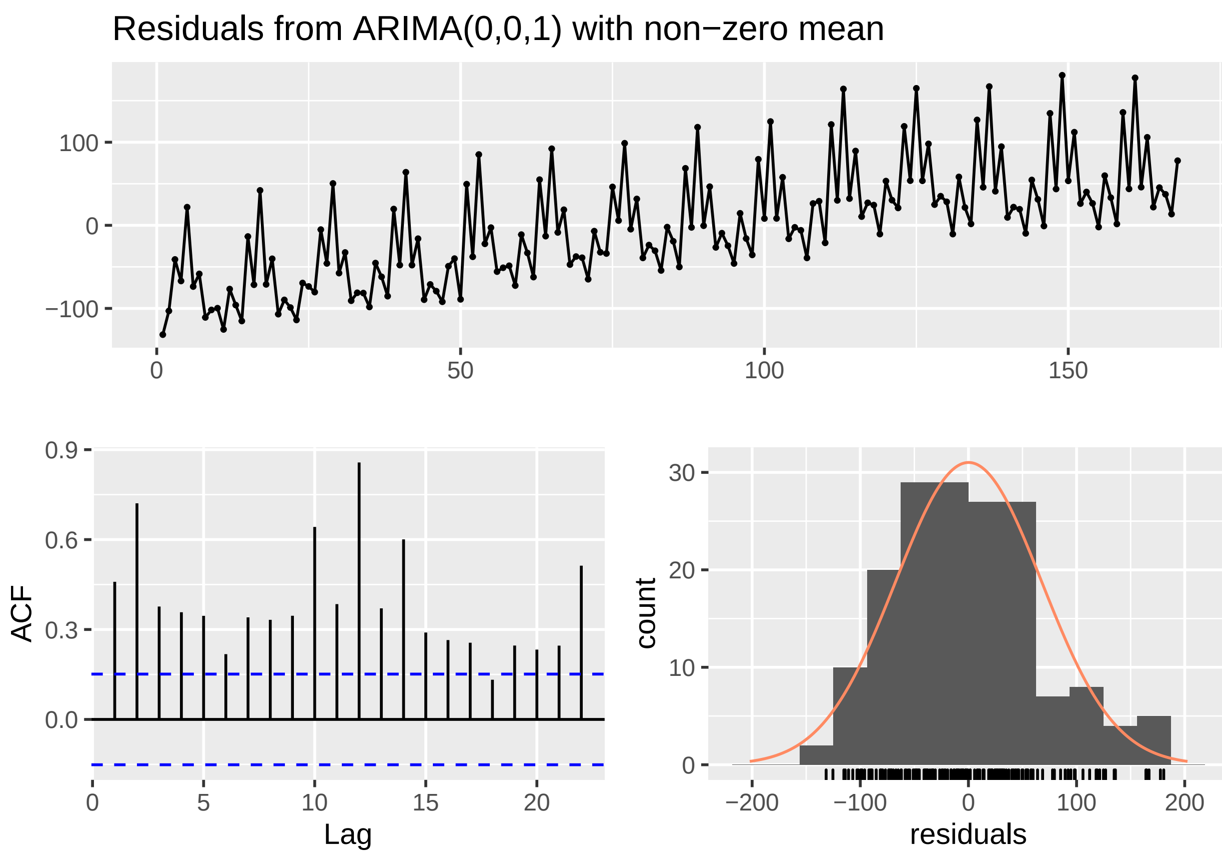



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,0,0) with non-zero mean
## Q* = 91.456, df = 7, p-value < 2.2e-16
##
## Model df: 3.   Total lags used: 10
```

```
fitMA1 <- Arima(df_milk$production, order=c(0,0,1))
fitMA1
```

```
## Series: df_milk$production
## ARIMA(0,0,1) with non-zero mean
##
## Coefficients:
##          ma1      mean
##       0.7676  754.8545
## s.e.  0.0360    9.1476
##
## sigma^2 estimated as 4577:  log likelihood=-945.84
## AIC=1897.69   AICc=1897.84   BIC=1907.06
```

```
checkresiduals(fitMA1)
```

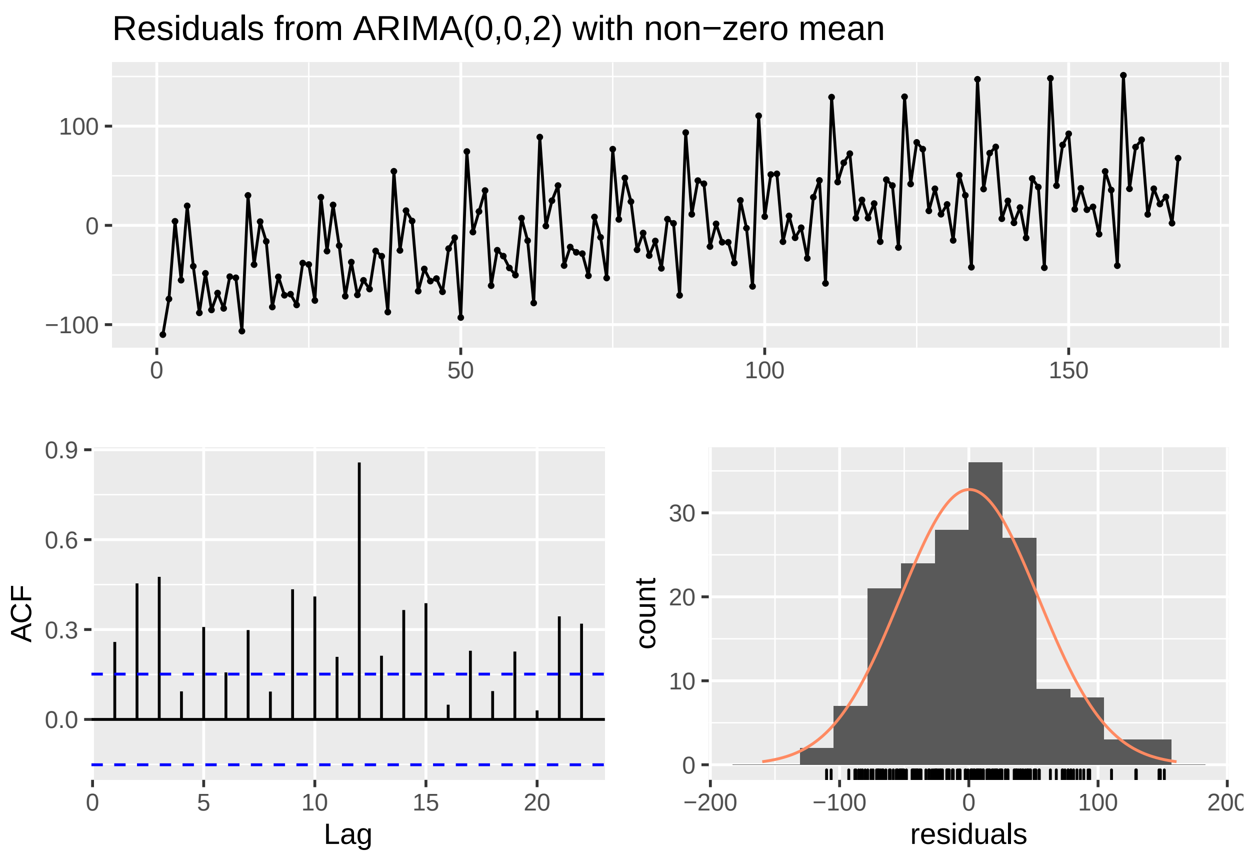



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,1) with non-zero mean
## Q* = 337.91, df = 8, p-value < 2.2e-16
##
## Model df: 2.   Total lags used: 10
```

```
fitMA2 <- Arima(df_milk$production, order=c(0,0,2))
fitMA2
```

```
## Series: df_milk$production
## ARIMA(0,0,2) with non-zero mean
##
## Coefficients:
##          ma1     ma2      mean
##       0.9132  0.6538  754.5311
## s.e.  0.1031  0.0510   10.4977
##
## sigma^2 estimated as 2890:  log likelihood=-907
## AIC=1822.01   AICc=1822.25   BIC=1834.5
```

```
checkresiduals(fitMA2)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,2) with non-zero mean
## Q* = 190.25, df = 7, p-value < 2.2e-16
##
## Model df: 3.   Total lags used: 10
```

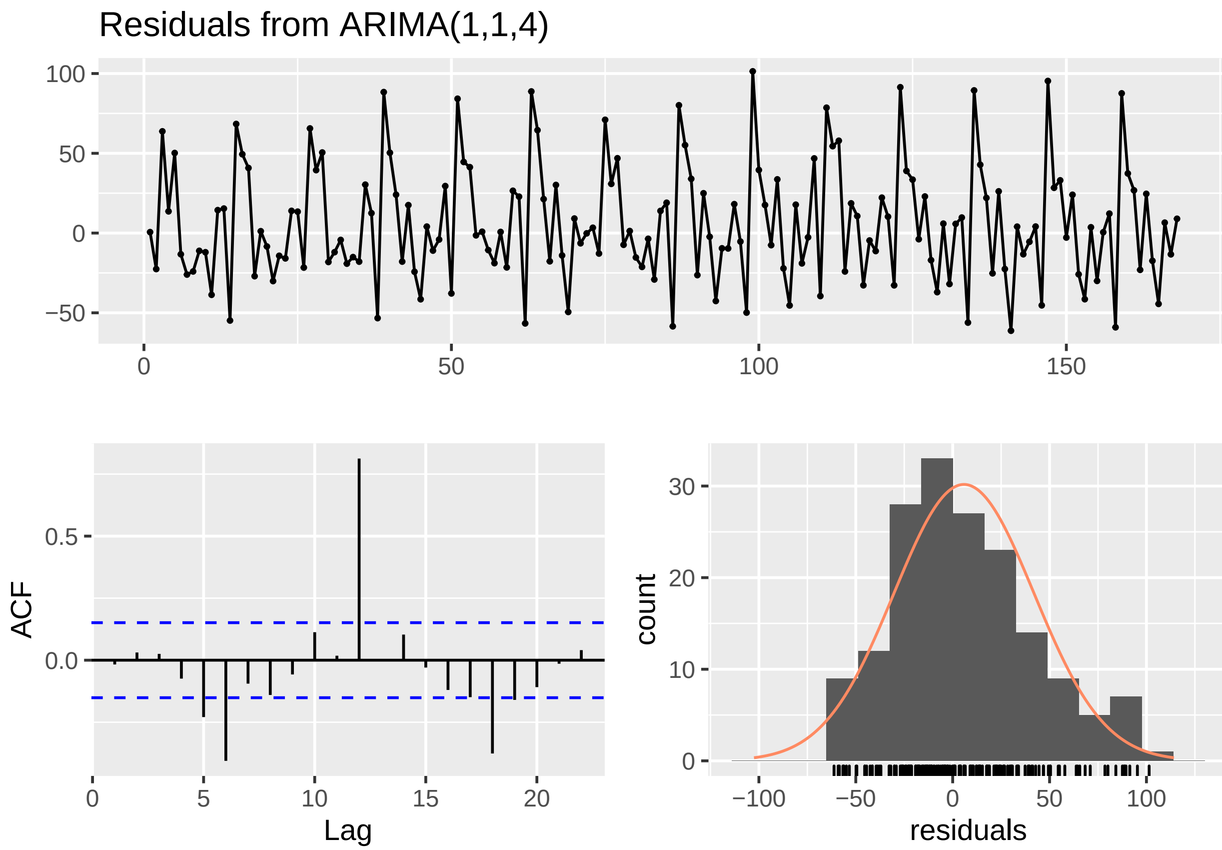
In all cases, residals are normally distributed white noise, AR(1) is better with lower AICc.

Question 4.

```
mol1 <- auto.arima(df_milk$production)
mol1
```

```
## Series: df_milk$production
## ARIMA(1,1,4)
##
## Coefficients:
##           ar1     ma1     ma2      ma3      ma4
##       -0.3045  0.2456  0.1500  -0.4257  -0.6493
## s.e.   0.1158  0.0816  0.0545   0.0486   0.0614
##
## sigma^2 estimated as 1380:  log likelihood=-839.88
## AIC=1691.77   AICc=1692.29   BIC=1710.48
```

```r
checkresiduals(mol1)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,4)
## Q* = 47.463, df = 5, p-value = 4.571e-09
##
## Model df: 5.   Total lags used: 10
```

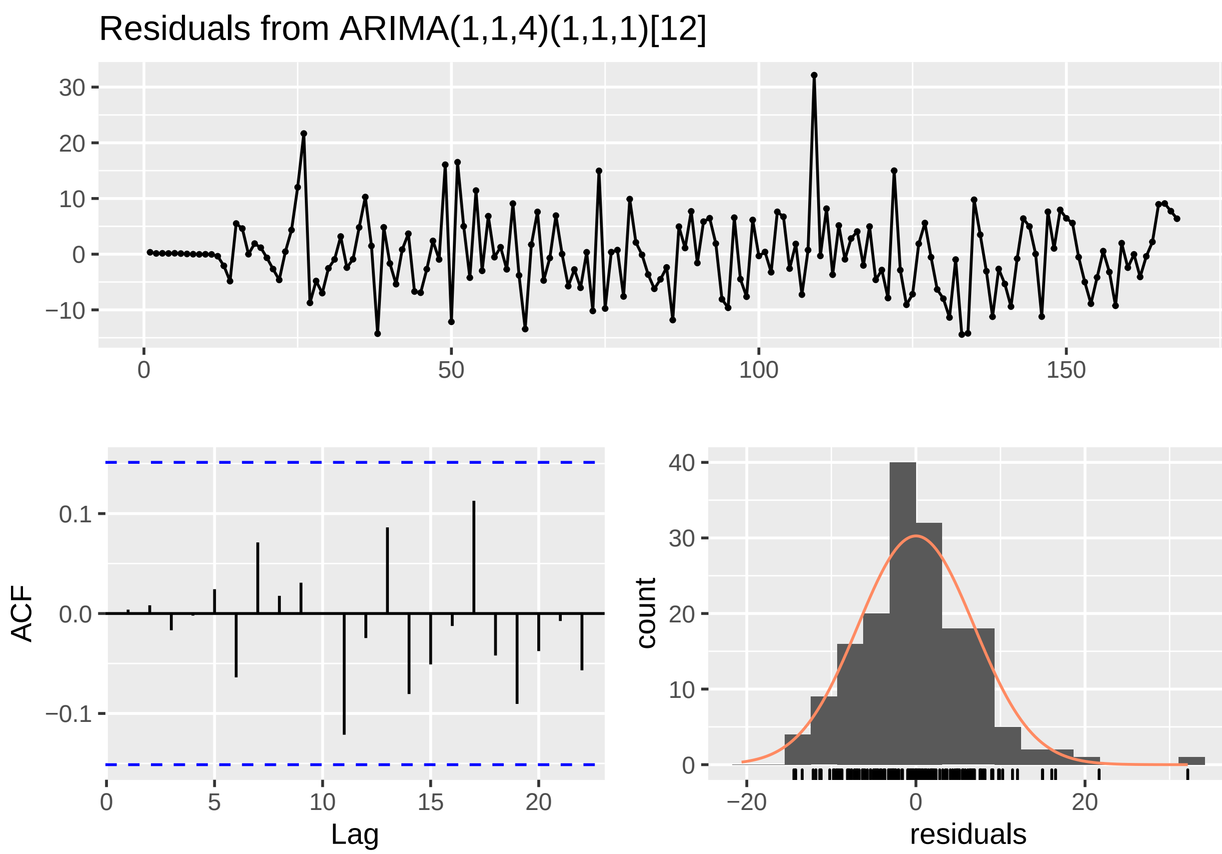
The moel with auto chosen gives us a lower AICc, which is better than AR(1) we get before. This model includes AR(1) and MA(4) with a first order difference.

```r
mol2 <- Arima(df_milk$production, order=c(1,1,4), seasonal = list(order = c(1, 1, 1), period = 12))
mol2
```

```
## Series: df_milk$production
## ARIMA(1,1,4)(1,1,1)[12]
##
## Coefficients:
##          ar1      ma1     ma2     ma3      ma4     sar1     sma1
##       0.3961  -0.6289  0.1273  0.1266  -0.1930  -0.0343  -0.5982
## s.e.  0.3990   0.3957  0.1384  0.1071   0.0948   0.1217   0.0965
##
```

```
## sigma^2 estimated as 53.34:  log likelihood=-527.53
## AIC=1071.05   AICc=1072.04   BIC=1095.4
```

```r
checkresiduals(mol2)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,4)(1,1,1)[12]
## Q* = 2.0126, df = 3, p-value = 0.5698
##
## Model df: 7.   Total lags used: 10
```

After manually changing the order, we get a better model ARIMA(1,1,4)(1,1,1)[12] has a much lower AICc.