Stat674 Final

Kotomi Oda

```r
library(dplyr)
library(fpp3)
library(tidyverse)
library(lmtest)
library(TSA)
library(forecast)
```

Section 7.10 Exercise 4

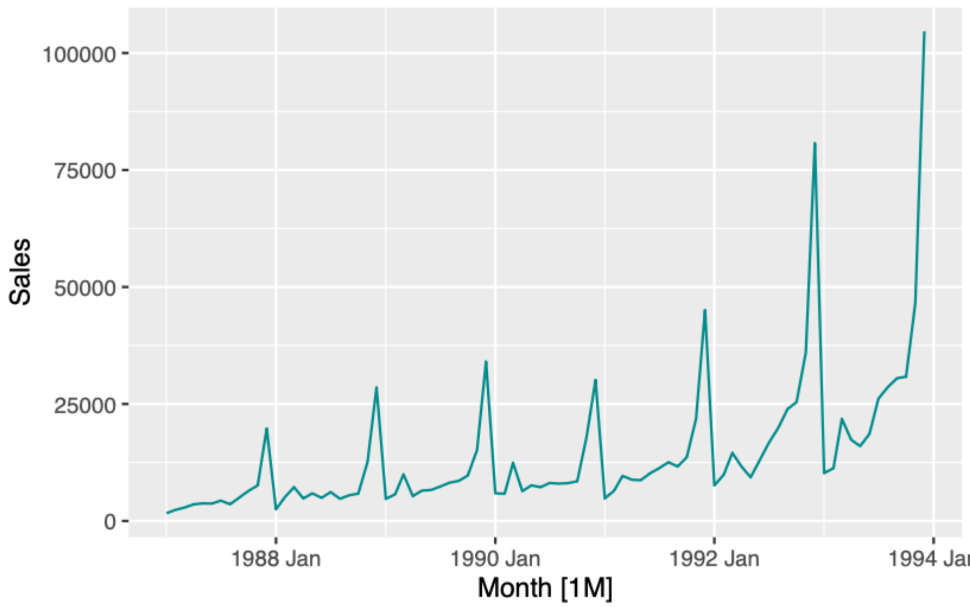
a.

```r
data(souvenirs)
head(souvenirs)
```

```
# A tsibble: 6 x 2 [1M]
     Month Sales
     <mth> <dbl>
1 1987 Jan 1665.
2 1987 Feb 2398.
3 1987 Mar 2841.
4 1987 Apr 3547.
5 1987 May 3753.
6 1987 Jun 3715.
```

```r
souvenirs %>%
  autoplot(color = "cyan4")
```

```
Plot variable not specified, automatically selected `.vars = Sales`
```



b.

ANS variability increases with the time, therefore, by applying a logarithmic transformation can give great results when trying to control this type of variation and normalize them.

c.

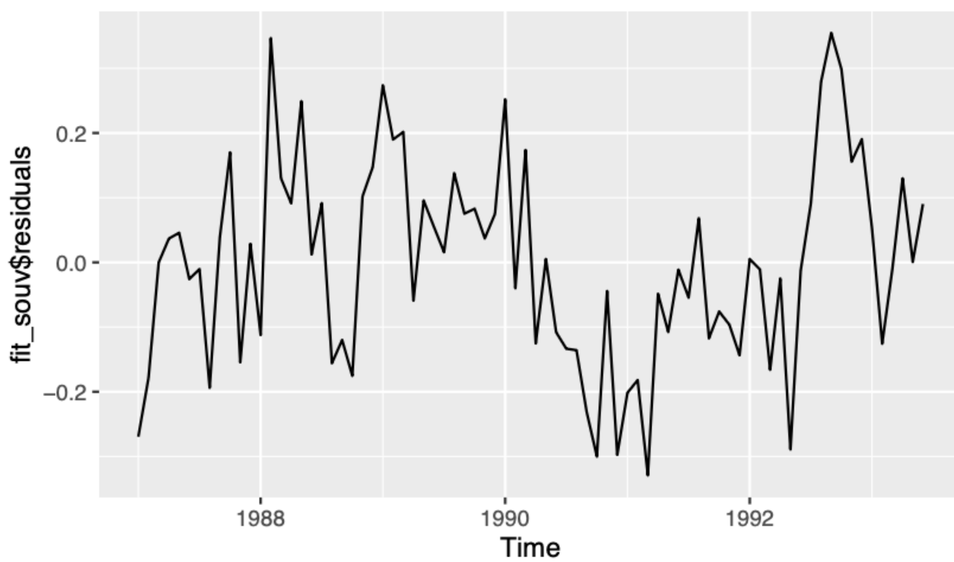
```
new_souvenirs <- ts(souvenirs$Sales,
                    start = c(1987, 1), end = c(1993, 6),
                    frequency = 12)

log_souvenirs <- log(new_souvenirs)

festival_dummy <- rep(0, length(log_souvenirs)) #festival dummy variable
festival_dummy[seq_along(festival_dummy) %% 12 == 3] = 1
festival_dummy[3] = 0
festival_dummy <- ts(festival_dummy, start = c(1987, 1), frequency = 12)

new_souvenirs_data <- data.frame(log_souvenirs, festival_dummy)
```

```r
fit_souv <- tslm(log_souvenirs ~ trend  + festival_dummy + season, data = new_souvenirs_da

summary(fit_souv)
```

```
Call:
tslm(formula = log_souvenirs ~ trend + festival_dummy + season,
    data = new_souvenirs_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.32917 -0.11936  0.00042  0.09144  0.35460 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    7.6662400  0.0729252 105.125  < 2e-16 ***
trend          0.0207611  0.0008864  23.423  < 2e-16 ***
festival_dummy 0.5543818  0.1898240   2.921  0.00482 ** 
season2        0.2526756  0.0921186   2.743  0.00789 ** 
season3        0.2232860  0.1866773   1.196  0.23607    
season4        0.3878297  0.0921527   4.209 8.18e-05 ***
season5        0.4145219  0.0921825   4.497 2.97e-05 ***
season6        0.4551219  0.0922209   4.935 6.02e-06 ***
season7        0.5767690  0.0958756   6.016 9.52e-08 ***
season8        0.5406444  0.0958797   5.639 4.16e-07 ***
season9        0.6296713  0.0958920   6.566 1.07e-08 ***
season10       0.7239552  0.0959125   7.548 2.02e-10 ***
season11       1.1957774  0.0959412  12.464  < 2e-16 ***
season12       1.9473841  0.0959780  20.290  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1723 on 64 degrees of freedom
Multiple R-squared:  0.9509,    Adjusted R-squared:  0.9409 
F-statistic: 95.34 on 13 and 64 DF,  p-value: < 2.2e-16
```

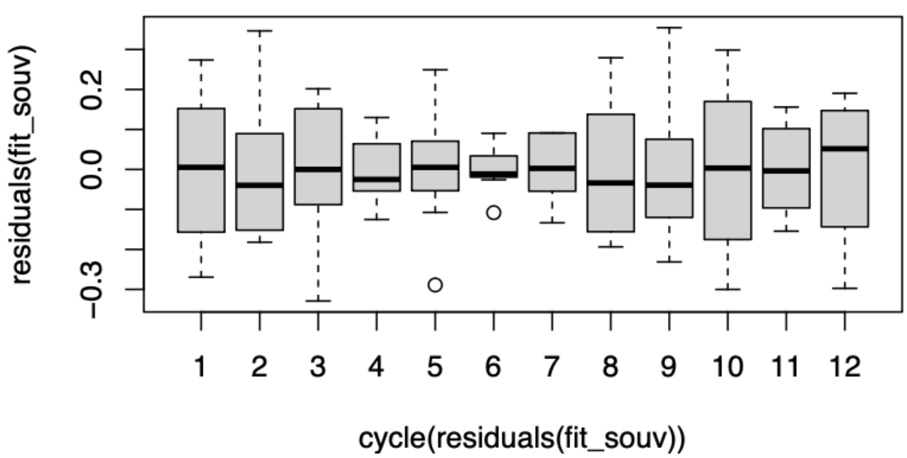
d.

```
autoplot(fit_souv$residuals)
```



e.

```
boxplot(residuals(fit_souv) ~ cycle(residuals(fit_souv)))
```



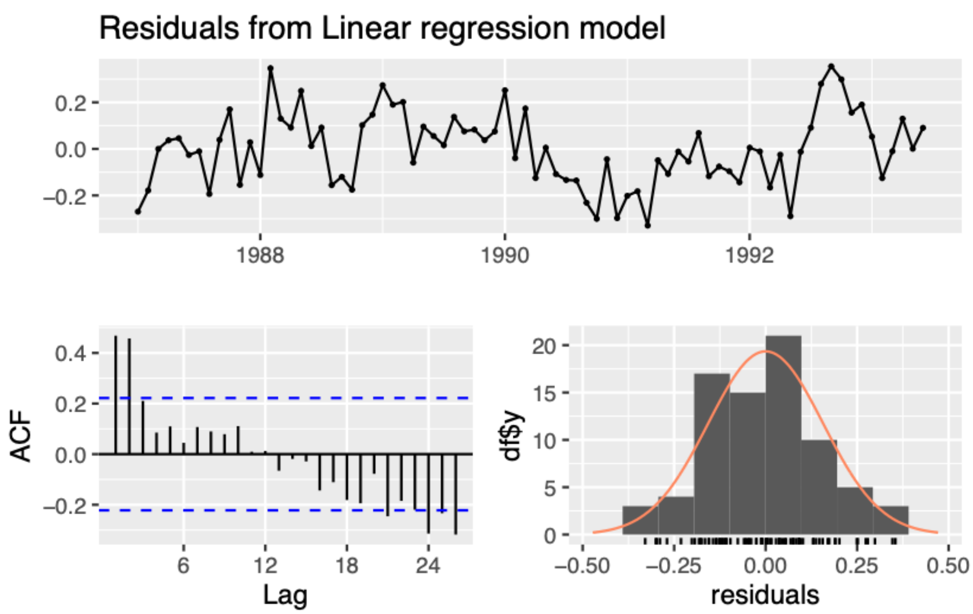
f.

```
fit_souv$coefficients
```

```
   (Intercept)          trend festival_dummy        season2        season3
     7.6662400      0.0207611      0.5543818      0.2526756      0.2232860
       season4        season5        season6        season7        season8
     0.3878297      0.4145219      0.4551219      0.5767690      0.5406444
       season9       season10       season11       season12
     0.6296713      0.7239552      1.1957774      1.9473841
```

g.

```
checkresiduals(fit_souv)
```




```
	Breusch-Godfrey test for serial correlation of order up to 17

data:  Residuals from Linear regression model
LM test = 31.535, df = 17, p-value = 0.01718
```

By Breusch-Godfrey test, the $p - value = 0.017 < 0.05$, which indicates that the data has a better fit with a logarithmic transformation.

h.

```
fc_data <- data.frame(festival_dummy = rep(0, 36))
pred <- forecast(fit_souv, newdata = fc_data)

head(pred)
```

```
$model

Call:
tslm(formula = log_souvenirs ~ trend + festival_dummy + season,
    data = new_souvenirs_data)

Coefficients:
   (Intercept)           trend  festival_dummy         season2         season3
       7.66624         0.02076         0.55438         0.25268         0.22329
       season4         season5         season6         season7         season8
       0.38783         0.41452         0.45512         0.57677         0.54064
       season9        season10        season11        season12
       0.62967         0.72396         1.19578         1.94738


$mean
           Jan       Feb       Mar       Apr       May       Jun       Jul
1993                                                               9.883136
1994  9.430933  9.704370  9.695742  9.881046  9.928500  9.989861 10.132269
1995  9.680067  9.953503  9.944875 10.130180 10.177633 10.238994 10.381402
1996  9.929200 10.202636 10.194008 10.379313 10.426766 10.488127
           Aug       Sep       Oct       Nov       Dec
1993  9.867772  9.977560 10.092605 10.585189 11.357556
1994 10.116905 10.226693 10.341738 10.834322 11.606690
1995 10.366039 10.475827 10.590872 11.083455 11.855823
1996

$lower
              [,1]      [,2]
Jul 1993  9.637329  9.503919
Aug 1993  9.621965  9.488555
```

```
Sep 1993  9.731753  9.598343
Oct 1993  9.846798  9.713388
Nov 1993 10.339382 10.205971
Dec 1993 11.111749 10.978339
Jan 1994  9.186093  9.053207
Feb 1994  9.459530  9.326644
Mar 1994  9.365756  9.186658
Apr 1994  9.636206  9.503320
May 1994  9.683659  9.550774
Jun 1994  9.745020  9.612135
Jul 1994  9.883394  9.748319
Aug 1994  9.868031  9.732955
Sep 1994  9.977819  9.842743
Oct 1994 10.092864  9.957788
Nov 1994 10.585447 10.450372
Dec 1994 11.357815 11.222739
Jan 1995  9.431764  9.296999
Feb 1995  9.705201  9.570436
Mar 1995  9.610605  9.429182
Apr 1995  9.881877  9.747112
May 1995  9.929330  9.794566
Jun 1995  9.990691  9.855927
Jul 1995 10.128745  9.991617
Aug 1995 10.113381  9.976253
Sep 1995 10.223169 10.086041
Oct 1995 10.338214 10.201086
Nov 1995 10.830798 10.693669
Dec 1995 11.603165 11.466037
Jan 1996  9.676730  9.539704
Feb 1996  9.950167  9.813141
Mar 1996  9.854949  9.670926
Apr 1996 10.126843  9.989817
May 1996 10.174296 10.037270
Jun 1996 10.235658 10.098631

$upper
              [,1]      [,2]
Jul 1993 10.128943 10.262353
Aug 1993 10.113579 10.246989
Sep 1993 10.223367 10.356777
Oct 1993 10.338412 10.471822
Nov 1993 10.830995 10.964406
Dec 1993 11.603363 11.736773
```

```
Jan 1994  9.675774  9.808659
Feb 1994  9.949210 10.082096
Mar 1994 10.025727 10.204825
Apr 1994 10.125887 10.258772
May 1994 10.173340 10.306226
Jun 1994 10.234701 10.367587
Jul 1994 10.381144 10.516219
Aug 1994 10.365780 10.500856
Sep 1994 10.475568 10.610644
Oct 1994 10.590613 10.725689
Nov 1994 11.083197 11.218272
Dec 1994 11.855564 11.990640
Jan 1995  9.928369 10.063134
Feb 1995 10.201806 10.336570
Mar 1995 10.279144 10.460567
Apr 1995 10.378482 10.513247
May 1995 10.425935 10.560700
Jun 1995 10.487296 10.622061
Jul 1995 10.634059 10.771188
Aug 1995 10.618696 10.755824
Sep 1995 10.728484 10.865612
Oct 1995 10.843529 10.980657
Nov 1995 11.336112 11.473240
Dec 1995 12.108480 12.245608
Jan 1996 10.181669 10.318696
Feb 1996 10.455106 10.592132
Mar 1996 10.533067 10.717090
Apr 1996 10.631782 10.768809
May 1996 10.679236 10.816262
Jun 1996 10.740597 10.877623

$level
[1] 80 95

$x
          Jan      Feb      Mar      Apr      May      Jun      Jul
1987 7.417466 7.782194 7.951809 8.173939 8.230300 8.220064 8.377841
1988 7.823970 8.556075 8.885322 8.477627 8.682857 8.507414 8.728931
1989 8.458933 8.648683 9.206089 8.576364 8.778392 8.799481 8.902404
1990 8.686278 8.668124 9.427164 8.759319 8.937103 8.885268 9.002236
1991 8.481906 8.774967 9.173549 9.084910 9.073646 9.231072 9.330481
1992 8.937879 9.195195 9.585923 9.357668 9.141265 9.478999 9.725125
1993 9.234373 9.329623 9.990896 9.761770 9.680206 9.830999
```

	Aug	Sep	Oct	Nov	Dec
1987	8.179295	8.521548	8.767715	8.935982	9.891223
1988	8.466352	8.611854	8.671647	9.441458	10.259122
1989	9.009034	9.056393	9.178901	9.625877	10.435909
1990	8.984600	8.998762	9.045077	9.793375	10.312759
1991	9.437653	9.361978	9.518332	9.990679	10.715766
1992	9.897902	10.083029	10.142164	10.491963	11.298763
1993					

i.

In my opinion, this model can be improved by adding data to it - specifically average temperature by month. Considering given the location is in close to the beach, temperature could drastically explain the volatility in certain months.

Section 9.11 Exercise 7a 7b

```r
head(aus_airpassengers)
```

```
# A tsibble: 6 x 2 [1Y]
   Year Passengers
  <dbl>      <dbl>
1  1970       7.32
2  1971       7.33
3  1972       7.80
4  1973       9.38
5  1974      10.7 
6  1975      11.1 
```

7a.

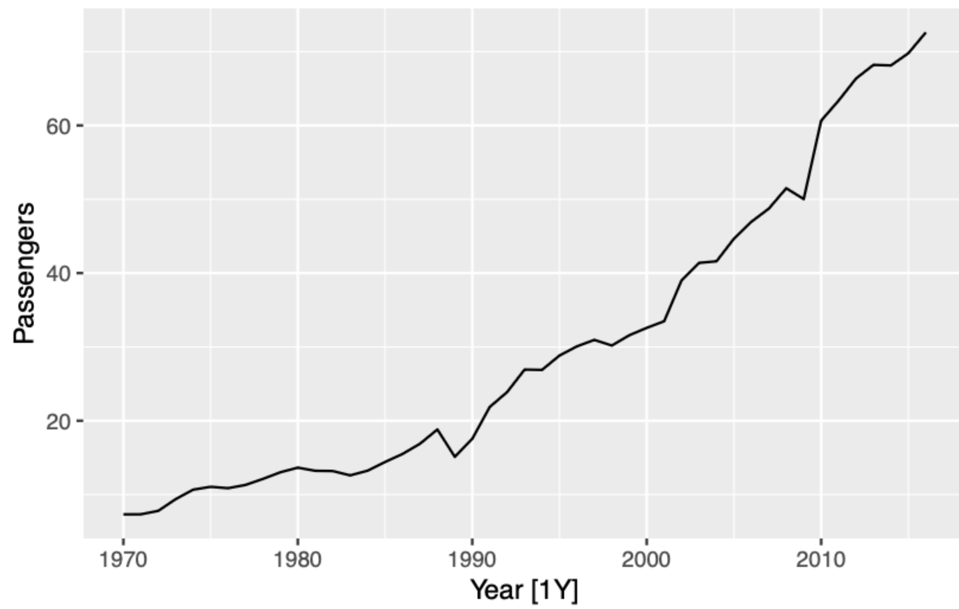
```r
fit_aus <- aus_airpassengers %>%
  filter(Year < 2012)

aus_model <- fit_aus %>%
  model(ARIMA(Passengers))

report(aus_model)
```

```
Series: Passengers
Model: ARIMA(0,2,1)

Coefficients:
          ma1
      -0.8756
s.e.   0.0722

sigma^2 estimated as 4.671:  log likelihood=-87.8
AIC=179.61   AICc=179.93   BIC=182.99
```

```r
autoplot(aus_airpassengers, h = 10)
```

```
Plot variable not specified, automatically selected `.vars = Passengers`
```

```
Warning in geom_line(...): Ignoring unknown parameters: `h`
```



By applying `ARIMA()`, ARIMA(0,2,1) was chosen to have an appropriate model. In addition, the residuals do resemble white noise. It is also clear to see that all the figures in the acf plot are between the boundaries.

7b.

$yt = -0.876\epsilon t - 1 + \epsilon t$

$(1 - B)2yt = (1 - 0.876B)\epsilon t$