Employment Analysis

Leala Darby

03/11/2020

First load all required packages:

```
library(car)
library(tseries)
library(astsa)
```

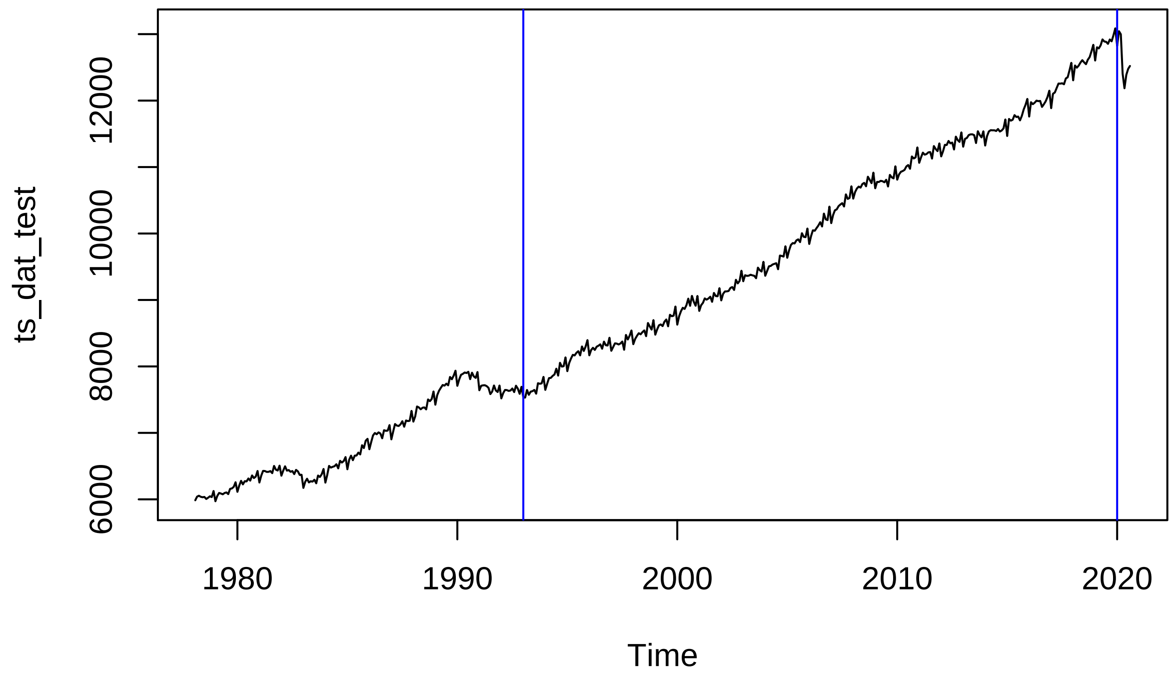
Load in the data:

```
dat <- read.csv("employment_data.csv", fileEncoding = 'UTF-8-BOM')
head(dat)
```

```
##   Observation.times Time.series.values
## 1            Feb-78             5985.7
## 2            Mar-78             6040.6
## 3            Apr-78             6054.2
## 4            May-78             6038.3
## 5            Jun-78             6031.3
## 6            Jul-78             6036.1
```

Create a time series object from the data and plot

```
ts_dat_test <- ts(dat[, 2], start = c(1978, 2), end = c(2020, 8), frequency = 12)
plot.ts(ts_dat_test)
abline(v = 1993, col = "blue")
abline(v = 2020, col = "blue")
```



Instructed to truncate data from January 1993 to December 2019 (inclusive)

```
dat[dat$Observation.times == "Jan-93",]
```

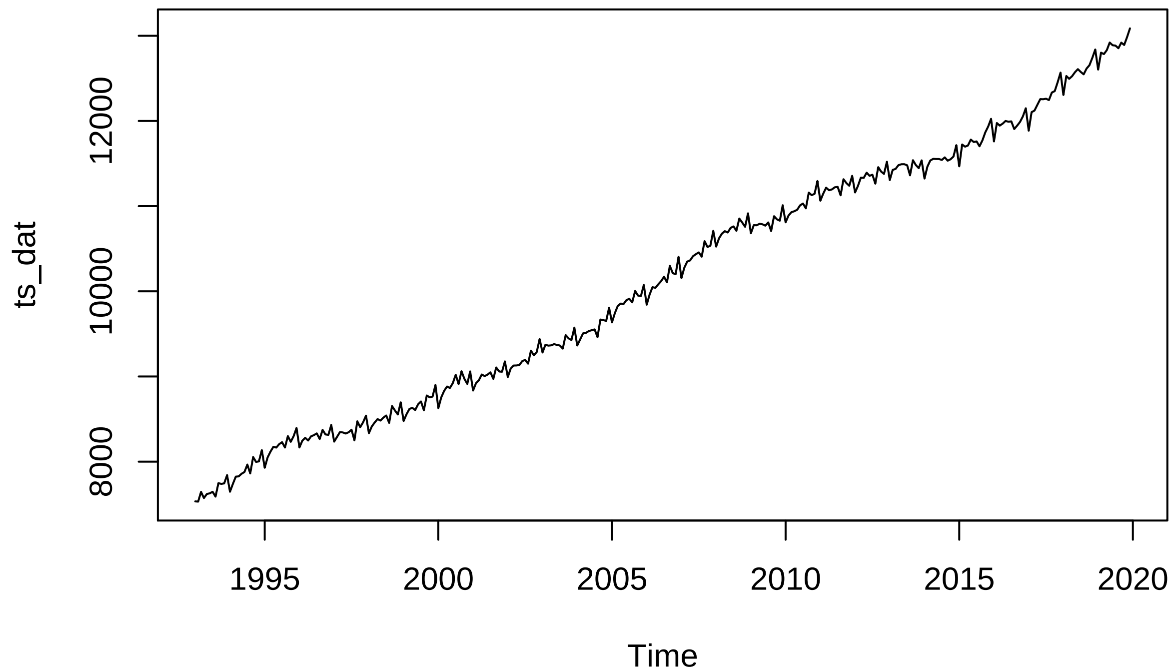
```
##     Observation.times Time.series.values
## 180            Jan-93             7533.7
```

```
dat[dat$Observation.times == "Dec-19",]
```

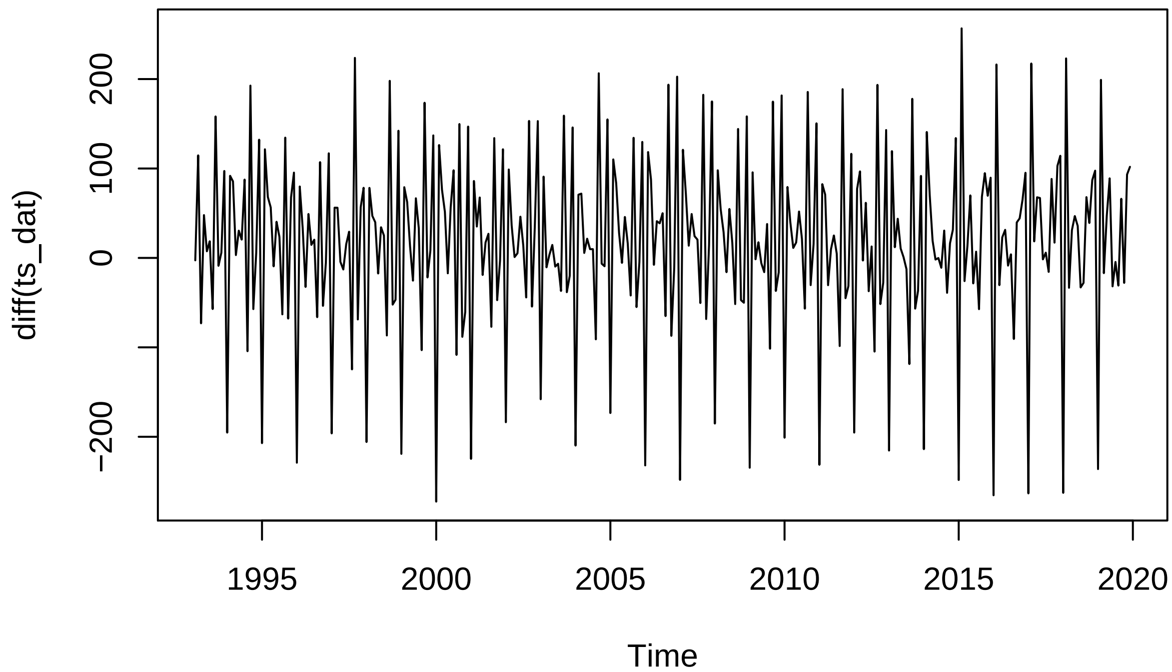
```
##     Observation.times Time.series.values
## 503            Dec-19            13087.1
```

So we only need rows 180-503.

```
trunc_dat <- dat[180:503,]
ts_dat <- ts(trunc_dat[, 2], start = c(1993, 1), end = c(2019, 12), frequency = 12)
plot.ts(ts_dat)
```




```r
plot.ts(diff(ts_dat))
```



The trend in mean is readily observable. Difficult to determine a trend in variance - there appears to be frequent changes, which are easier to see after incorporating lags of 1. Check statistically for stationarity using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, which has the following hypotheses [. . .]:

```
kpss.test(ts_dat)
```

```
## Warning in kpss.test(ts_dat): p-value smaller than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  ts_dat
## KPSS Level = 5.4937, Truncation lag parameter = 5, p-value = 0.01
```

The small p-value indicates that we should reject the null and conclude that the ts is not stationary.

As a rough test of constant variance (Levene's isn't really valid because time series data isn't independent)

```
length(ts_dat)
```

```
## [1] 324
```

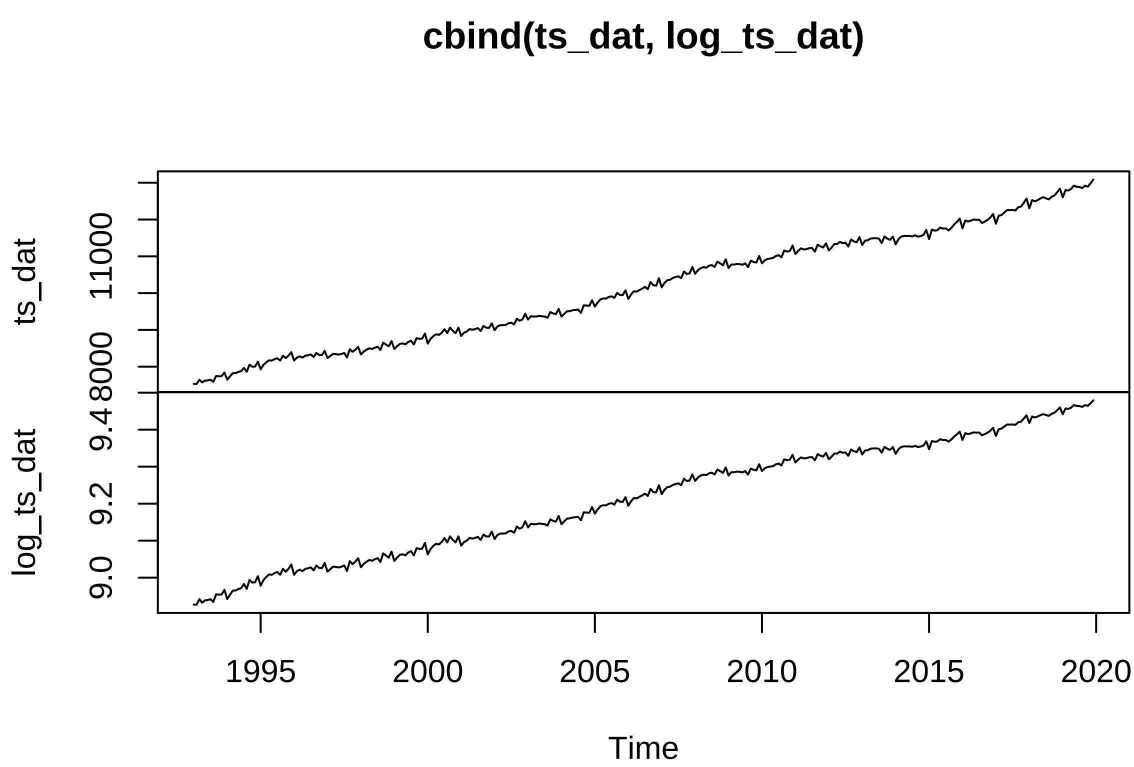
```
Group <- c(rep(1,81), rep(2, 81), rep(3, 81), rep(4, 81))
leveneTest(ts_dat, Group)
```

```
## Warning in leveneTest.default(ts_dat, Group): Group coerced to factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value    Pr(>F)
## group   3  7.2516 0.0001013 ***
##       320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The small p-value of 0.0001013 confirms that the data exhibits heteroscedasticity. Therefore we will perform a log transformation to attempt to reduce this:

```r
log_ts_dat <- log(ts_dat)
plot.ts(cbind(ts_dat, log_ts_dat))
```




```r
leveneTest(log_ts_dat, Group)
```

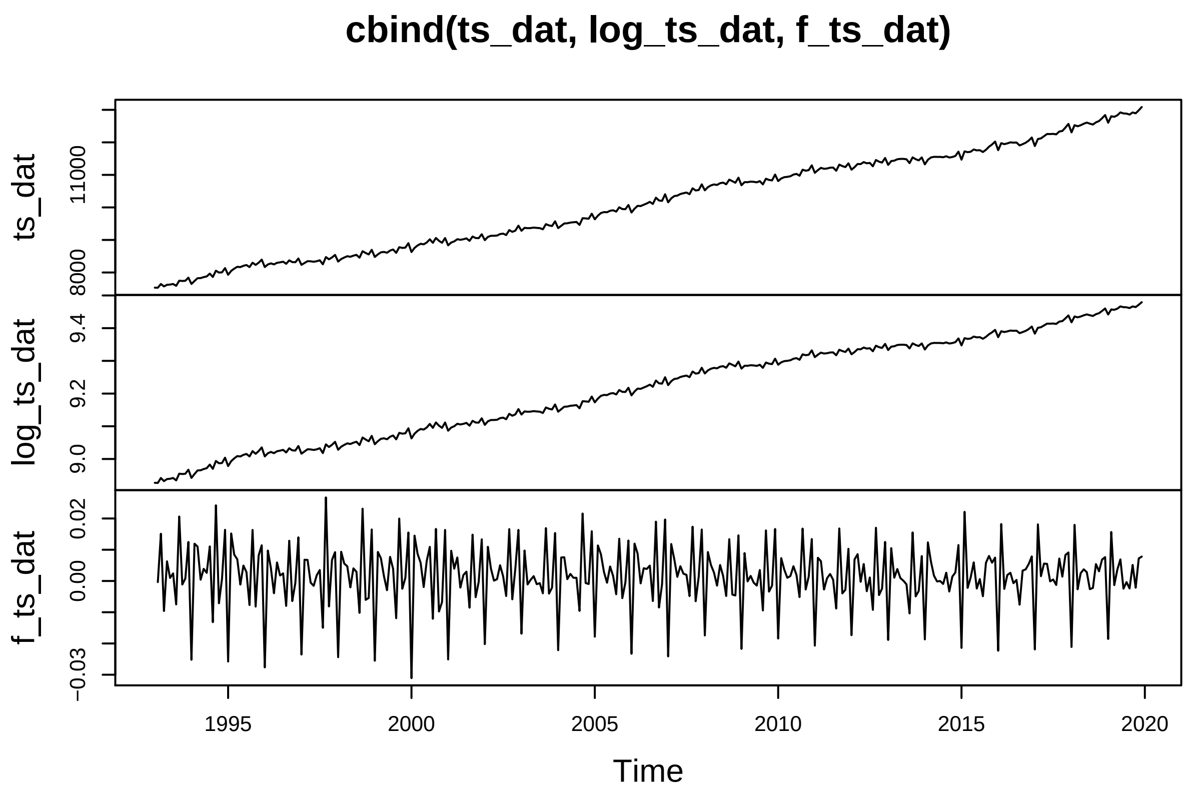
```
## Warning in leveneTest.default(log_ts_dat, Group): Group coerced to factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   3  1.4631 0.2245
##       320
```

At a significance level of 5%, the p-value above of 0.2245 provides very weak evidence and we fail to reject the null hypothesis of equal variance among groups. Thus the heteroscedasticity has been reduced.

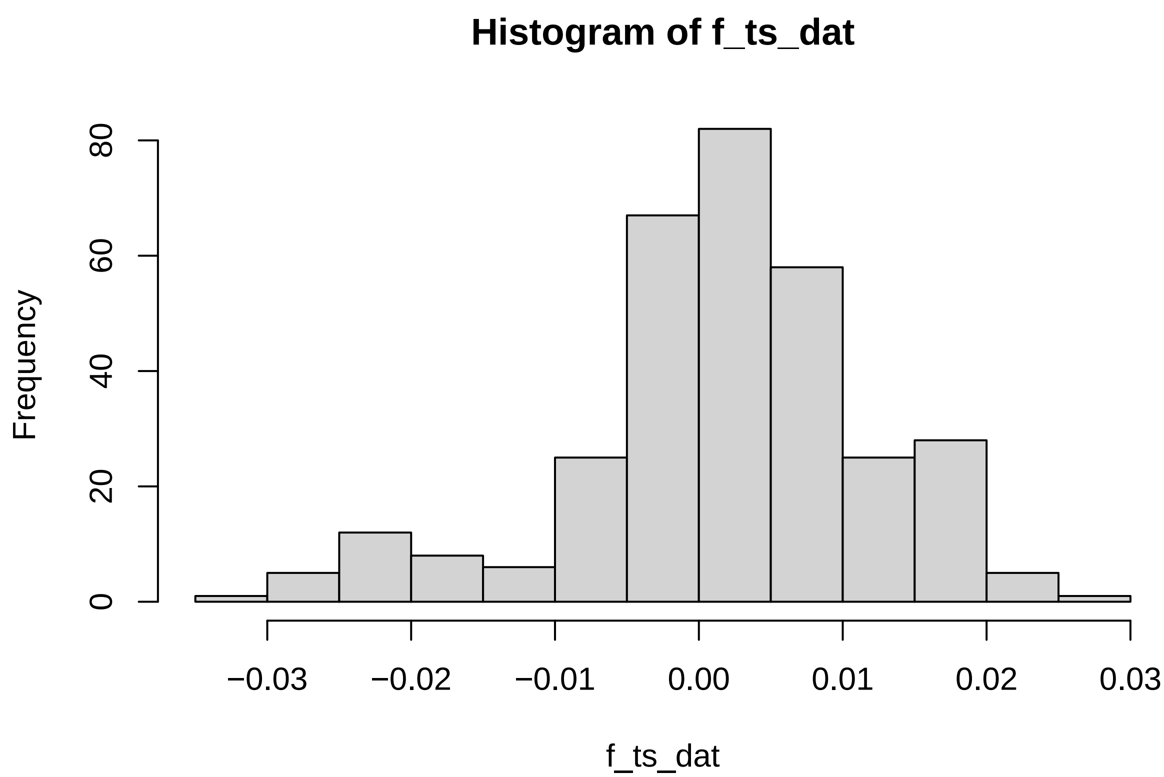
Next, to reduce the trend in mean, apply differencing of 1 lag to our TS with stabilised variance:

```
f_ts_dat <- diff(log_ts_dat, 1)
plot.ts(cbind(ts_dat, log_ts_dat, f_ts_dat))
```



To confirm constant mean and variance and a Gaussian distribution for the time series, a Shapiro-Wilk normality test is performed:

```
hist(f_ts_dat)
```




```r
shapiro.test(f_ts_dat)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  f_ts_dat
## W = 0.96138, p-value = 1.534e-07
```

The small p-value indicates likely non-normality, but this test isn't really valid for TS. Instead, check statistically for stationarity using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test:

```r
kpss.test(log_ts_dat)
```

```
## Warning in kpss.test(log_ts_dat): p-value smaller than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  log_ts_dat
## KPSS Level = 5.4933, Truncation lag parameter = 5, p-value = 0.01
```

```r
kpss.test(f_ts_dat)
```

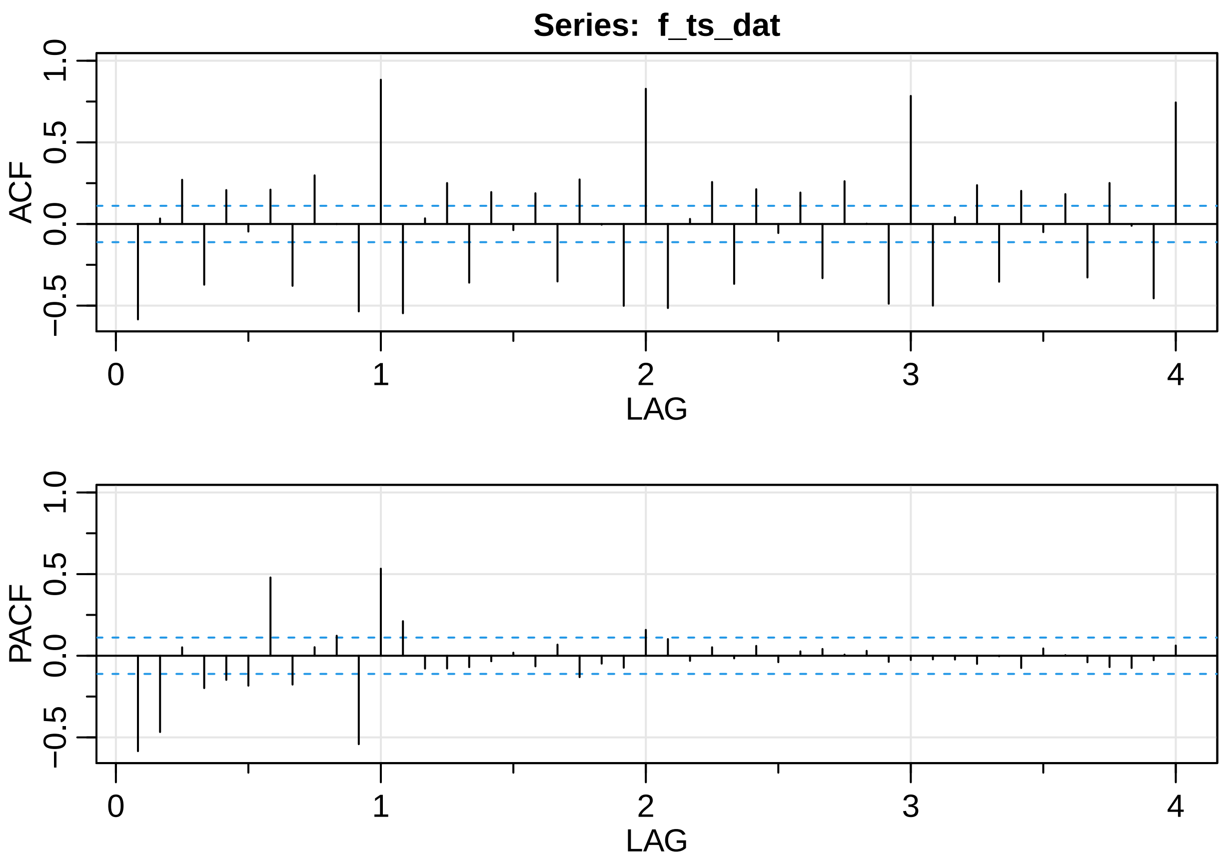
```
## Warning in kpss.test(f_ts_dat): p-value greater than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  f_ts_dat
## KPSS Level = 0.064047, Truncation lag parameter = 5, p-value = 0.1
```

The final ts has a high p-value of 0.1, which is statistically significant at a significance level of 5%. Therefore we fail to reject the null hypothesis, and have reasonable evidence that the final ts is stationary.

Next, the ACF and PACF of the differenced ts are plotted in order to estimate p and q.

```
acf2(f_ts_dat)
```




```
##       [,1]  [,2] [,3]  [,4]  [,5]  [,6] [,7]  [,8] [,9] [,10] [,11] [,12] [,13]
## ACF  -0.58  0.03 0.27 -0.37  0.21 -0.05 0.21 -0.38 0.30  0.00 -0.54  0.88 -0.55
## PACF -0.58 -0.47 0.05 -0.20 -0.15 -0.18 0.48 -0.18 0.05  0.12 -0.54  0.53  0.21
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF   0.03  0.25 -0.36  0.20 -0.04  0.19 -0.35  0.27  0.00 -0.50  0.83 -0.51
## PACF -0.08 -0.08 -0.07 -0.03  0.02 -0.06  0.07 -0.13 -0.05 -0.07  0.16  0.10
##      [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37]
## ACF   0.03  0.26 -0.37  0.21 -0.06  0.19 -0.33  0.26  0.00 -0.49  0.78 -0.50
## PACF -0.03  0.05 -0.02  0.06 -0.04  0.03  0.04  0.01  0.03 -0.04 -0.03 -0.02
##      [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48]
## ACF   0.04  0.24 -0.35  0.20 -0.05  0.18 -0.33  0.25 -0.01 -0.46  0.74
## PACF -0.02 -0.05  0.00 -0.08  0.04  0.00 -0.04 -0.07 -0.07 -0.03  0.06
```

Seasonal patterns are clear, more strongly in the ACF plot.