Lab 10 - Time Series

Julieth Santamaria

April 12, 2019

Data Source

This lab is based on a dataset that you already have. This lab uses the following sources of information:

- http://r-statistics.co/Time-Series-Analysis-With-R.html
- https://libguides.rutgers.edu/data/data_R

Application

We will use data about air passengers from 1948 to 1960

1. Import the dataset

```
data("AirPassengers")
```

2. What type of data are you dealing with

```
class(AirPassengers) # Time series
```

```
## [1] "ts"
```

```
# To know when does the data start and end
start(AirPassengers)
```

```
## [1] 1949    1
```

```
end(AirPassengers)
```

```
## [1] 1960   12
```

```
# To know the frequency of the data
frequency(AirPassengers) # 12 means we're dealing with a monthly dataset
```

```
## [1] 12
```

```
summary(AirPassengers)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   104.0   180.0   265.5   280.3   360.5   622.0
```

```
cycle(AirPassengers) # Shows the time period of your dataset
```

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1949   1   2   3   4   5   6   7   8   9  10  11  12
## 1950   1   2   3   4   5   6   7   8   9  10  11  12
## 1951   1   2   3   4   5   6   7   8   9  10  11  12
## 1952   1   2   3   4   5   6   7   8   9  10  11  12
## 1953   1   2   3   4   5   6   7   8   9  10  11  12
## 1954   1   2   3   4   5   6   7   8   9  10  11  12
## 1955   1   2   3   4   5   6   7   8   9  10  11  12
## 1956   1   2   3   4   5   6   7   8   9  10  11  12
## 1957   1   2   3   4   5   6   7   8   9  10  11  12
## 1958   1   2   3   4   5   6   7   8   9  10  11  12
## 1959   1   2   3   4   5   6   7   8   9  10  11  12
```

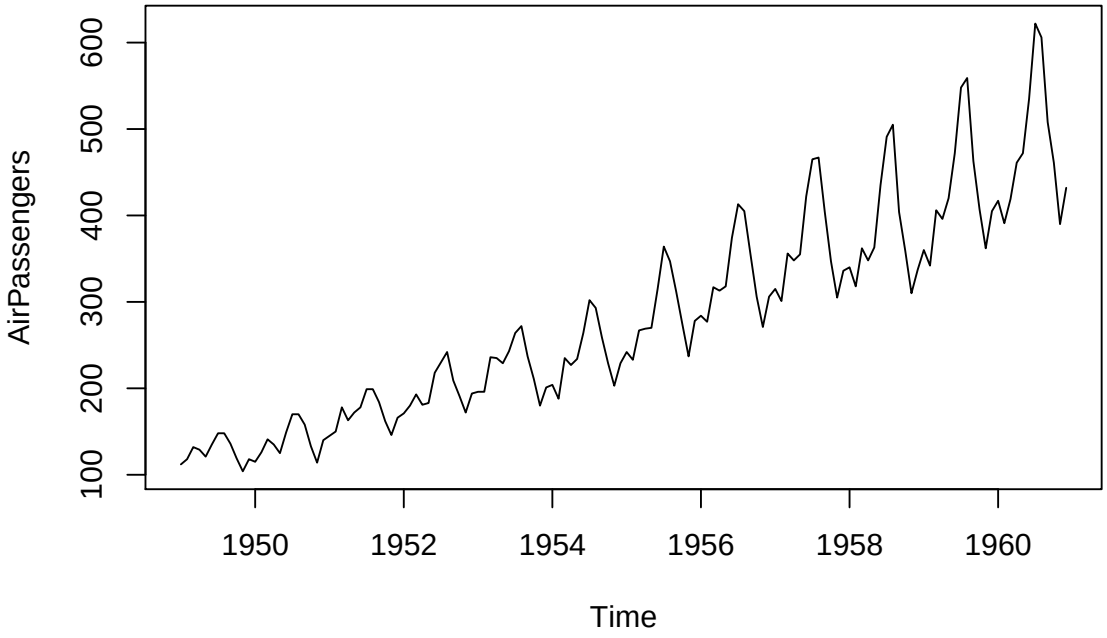
```
## 1960   1   2   3   4   5   6   7   8   9  10  11  12
```

3. Analyze the series

   Remember that time series usually have three components:

   - Trend
   - Seasonal component
   - Random component

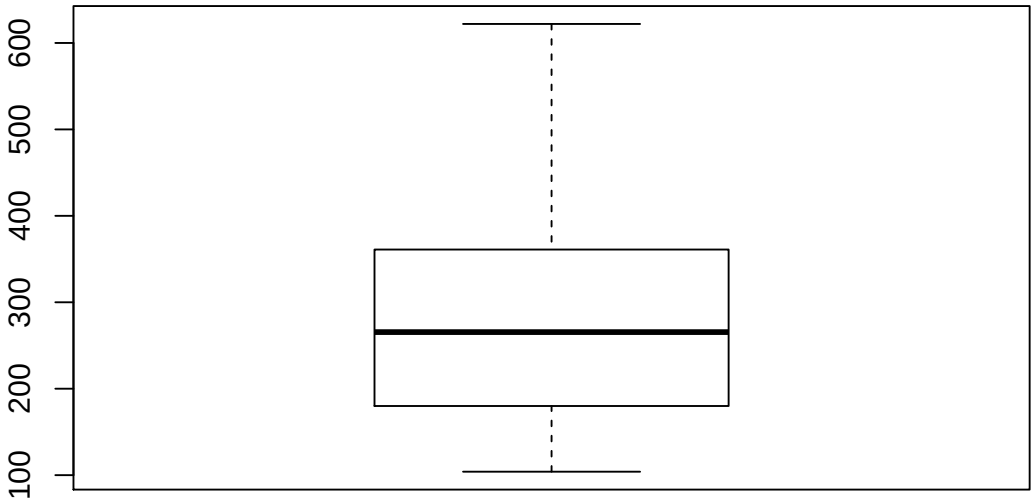
```
plot(AirPassengers)
```



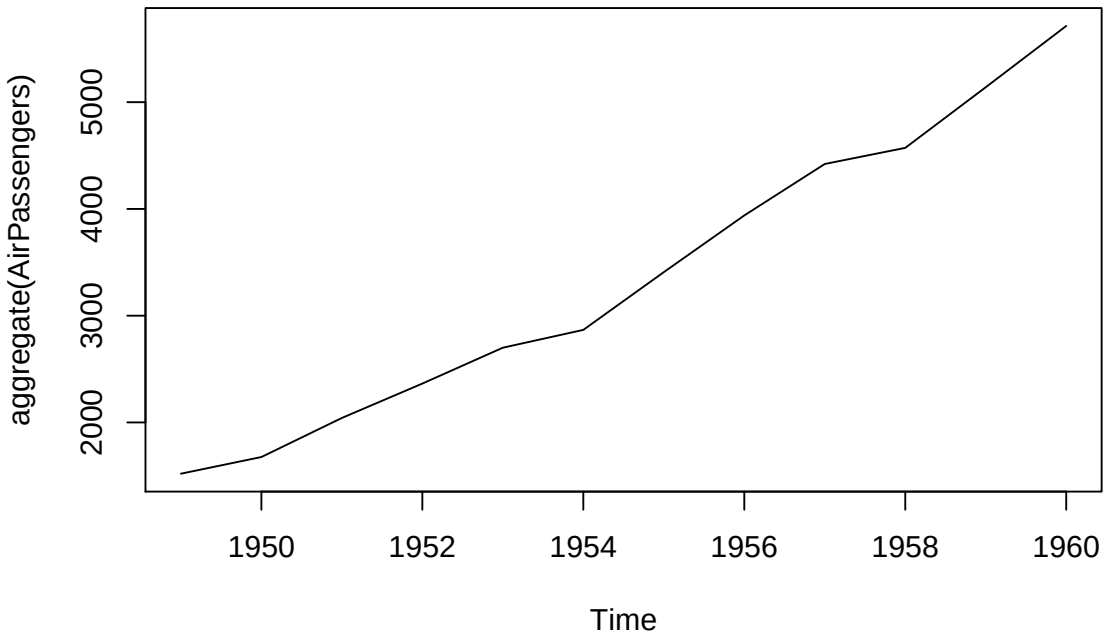

```
aggregate(AirPassengers) # Adds all of the months per year, Year total
```

```
## Time Series:
## Start = 1949
## End = 1960
## Frequency = 1
##  [1] 1520 1676 2042 2364 2700 2867 3408 3939 4421 4572 5140 5714
```

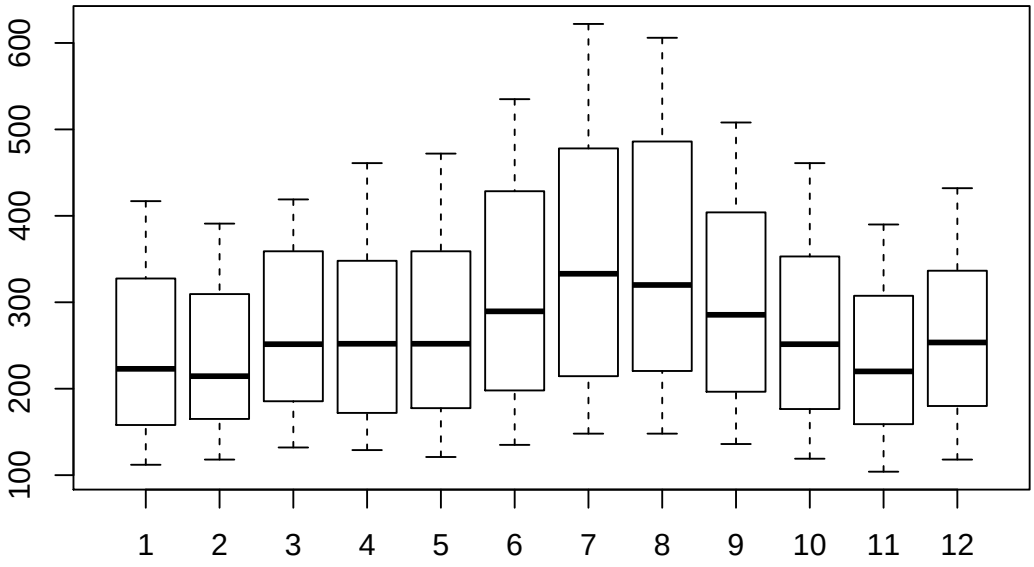
```
boxplot(AirPassengers)
```



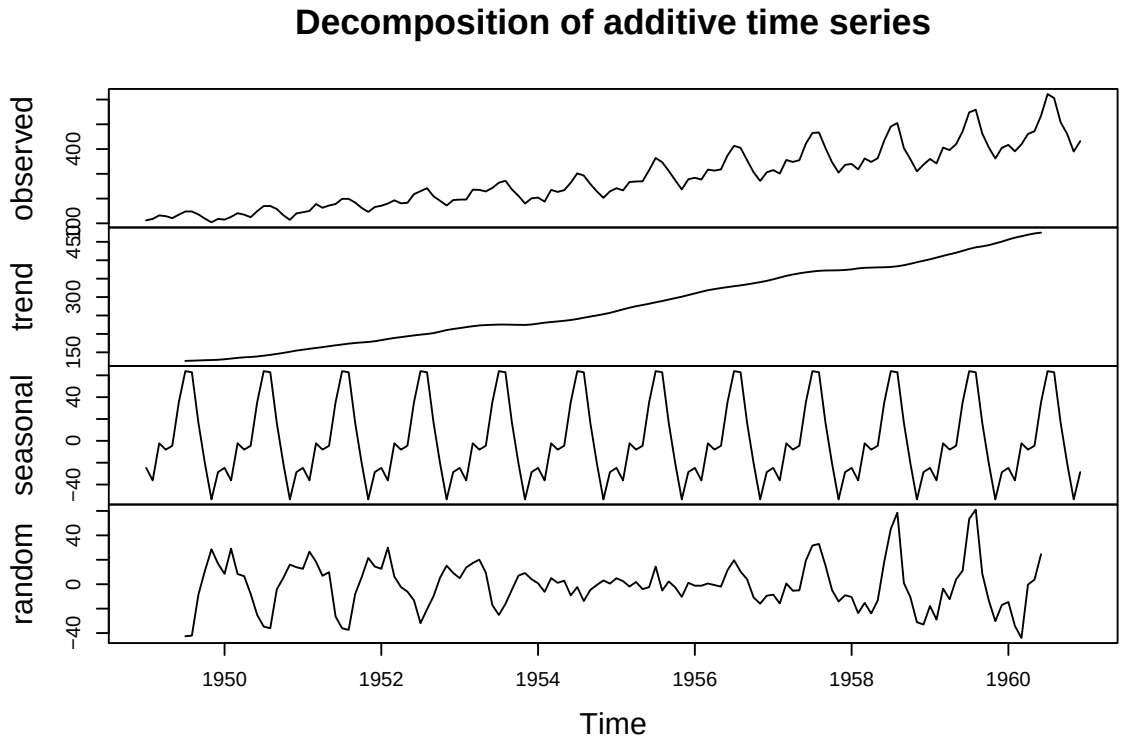

```r
# Trend
plot(aggregate(AirPassengers))
```




```
# Cycle
boxplot(AirPassengers~cycle(AirPassengers))    # July and August have high travel volumes
```




```
# Decomposition
plot(decompose(AirPassengers))
```



4. Do a regression analysis

4.1. Trend

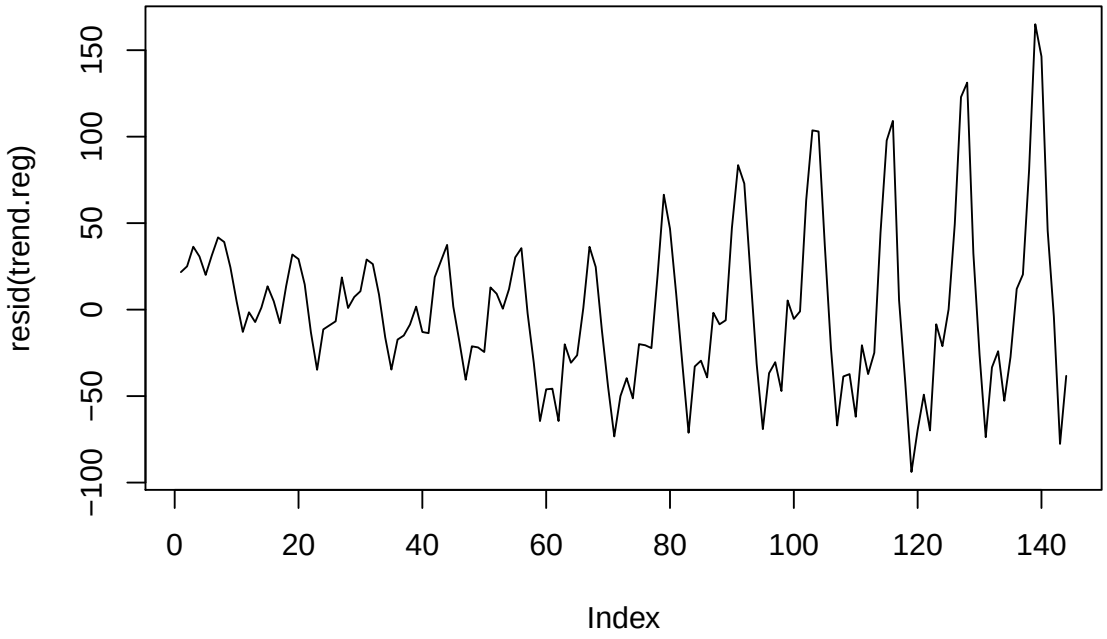
- Create a trend
- Run a regression of AirPassengers on the trend. Interpret the coefficient
- Write down the equation
- Plot the residuals. How does it compare to the original graph?

```
trend<-c(1:length(AirPassengers))
trend.reg<-lm(AirPassengers~trend)
summary(trend.reg)
```

```
##
## Call:
## lm(formula = AirPassengers ~ trend)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -93.858 -30.727  -5.757  24.489 164.999
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 87.65278    7.71635   11.36   <2e-16 ***
## trend        2.65718    0.09233   28.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46.06 on 142 degrees of freedom
```

```
## Multiple R-squared:  0.8536, Adjusted R-squared:  0.8526
## F-statistic: 828.2 on 1 and 142 DF,  p-value: < 2.2e-16
```

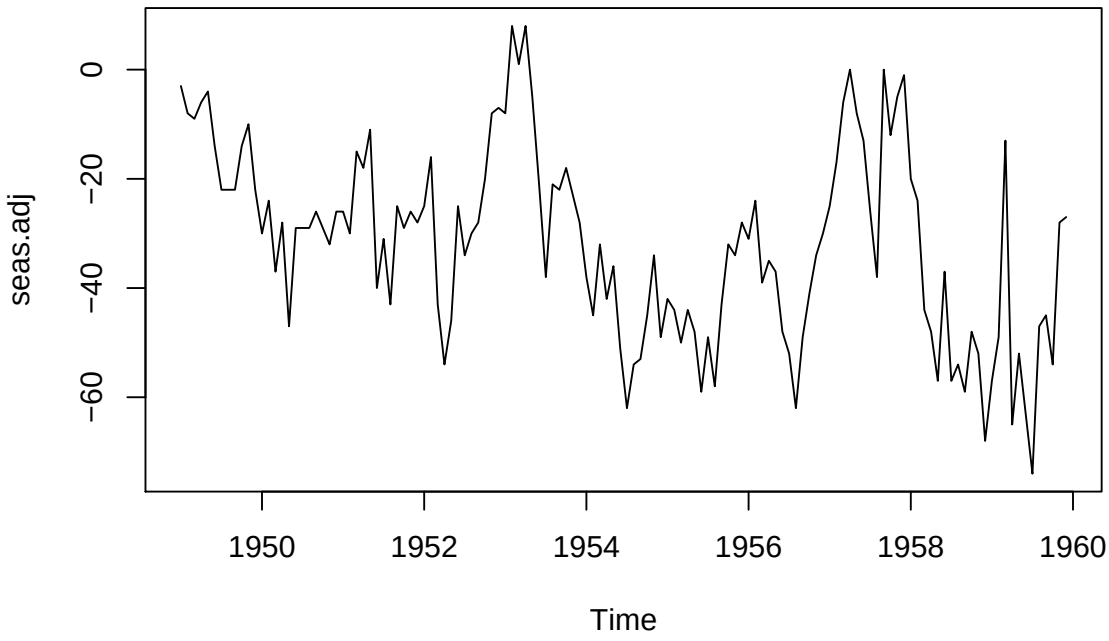
```
plot(resid(trend.reg),type = "l") # This means we de-trended out data
```



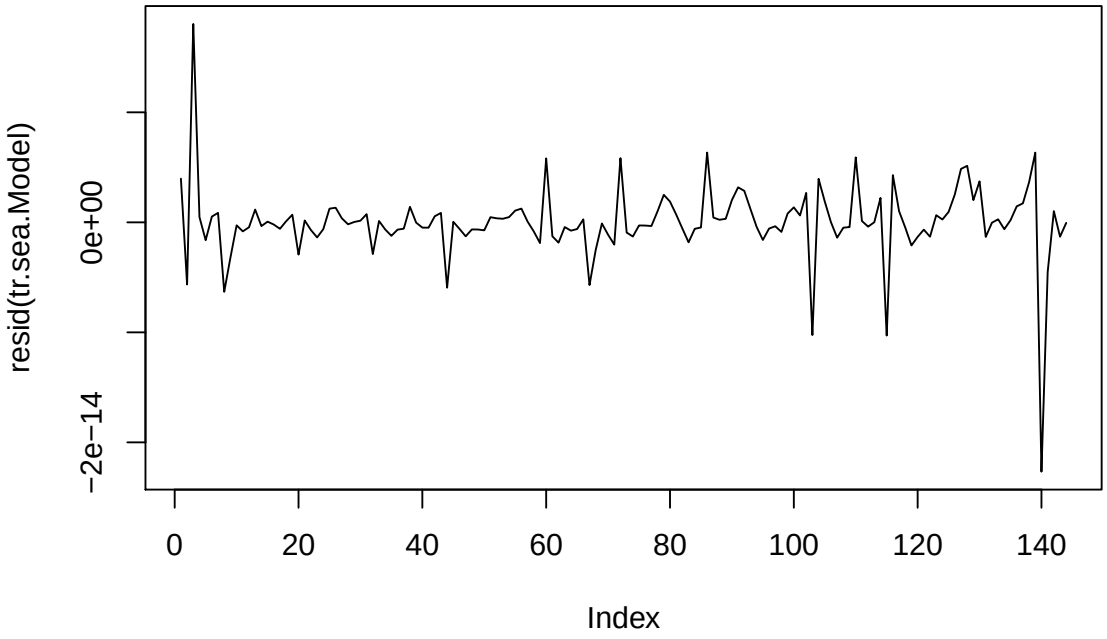
5.2 Seasonality

- What seasonal component makes sense?
- Create the seasonal component
- How is this model called?
- Run a regression of Air Passengers on a trend and the seasonal component.
- Write down the equation
- Plot the residuals

```
lag12AP<-lag(AirPassengers,12)
seas.adj<-AirPassengers-lag12AP
plot(seas.adj)
```



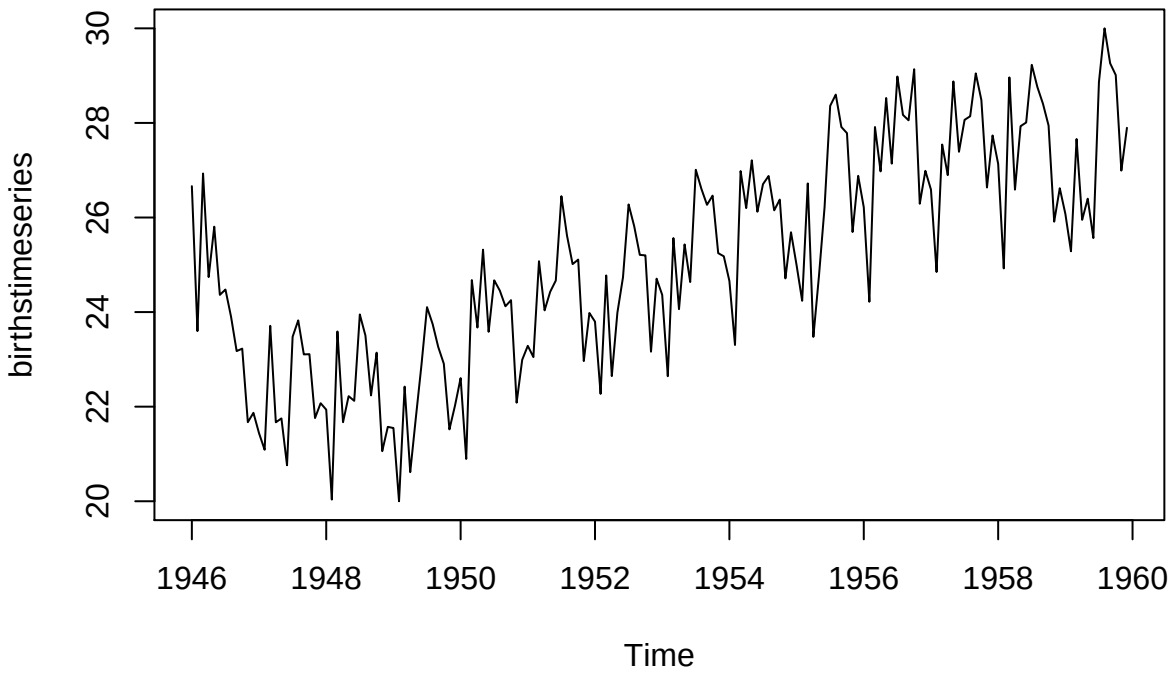

```
tr.sea.Model<-lm(AirPassengers~trend+lag12AP)
plot(resid(tr.sea.Model),type = "l")
```



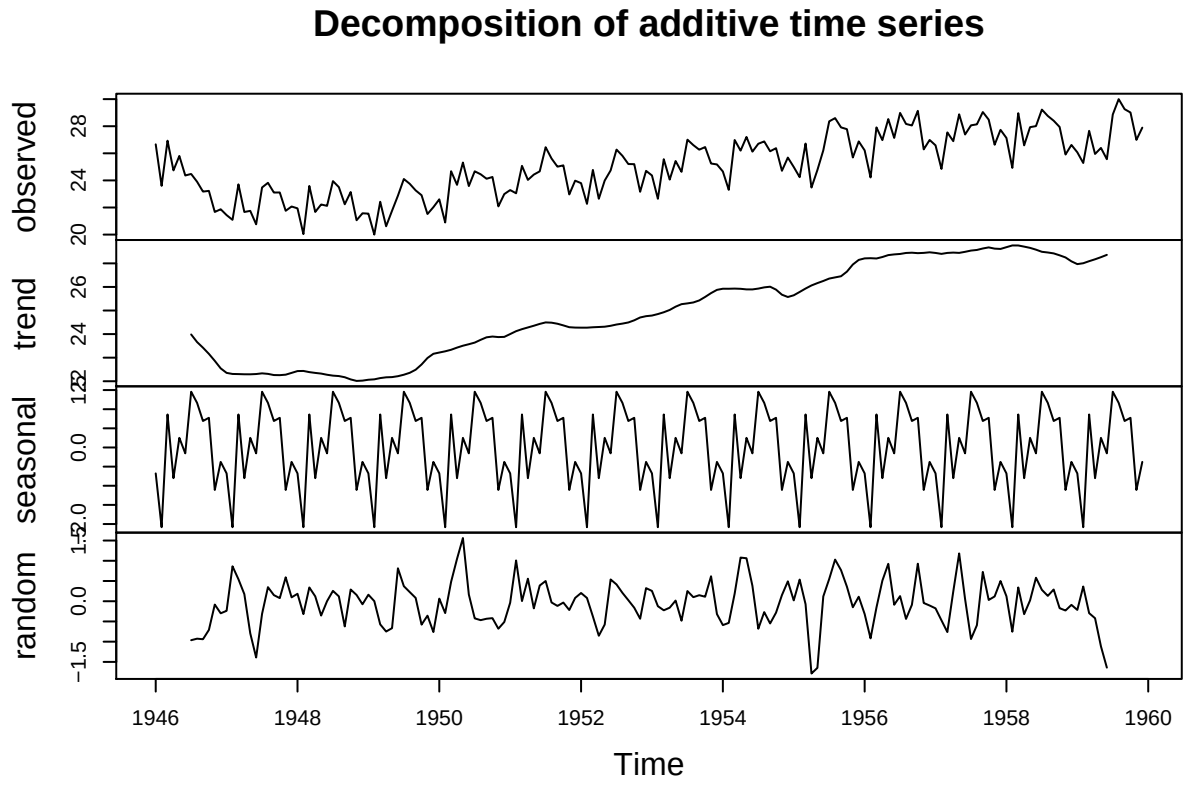
Other series

Analyze this series in the same manner as we did with air passengers. Now use the convenienve.r function that allows you to create differences and lags

```
setwd("G:/My Drive/U of M/TA/TA APEC3003/APEC 3003 - 2019/APEC 3003 R work/functions")
source("../functions/convenience.r")
births <- scan("http://robjhyndman.com/tsdldata/data/nybirths.dat")
birthstimeseries <- ts(births, frequency=12, start=c(1946,1))
plot.ts(birthstimeseries)
```




```
birthstimeseriescomponents <- decompose(birthstimeseries)
plot(birthstimeseriescomponents)
```

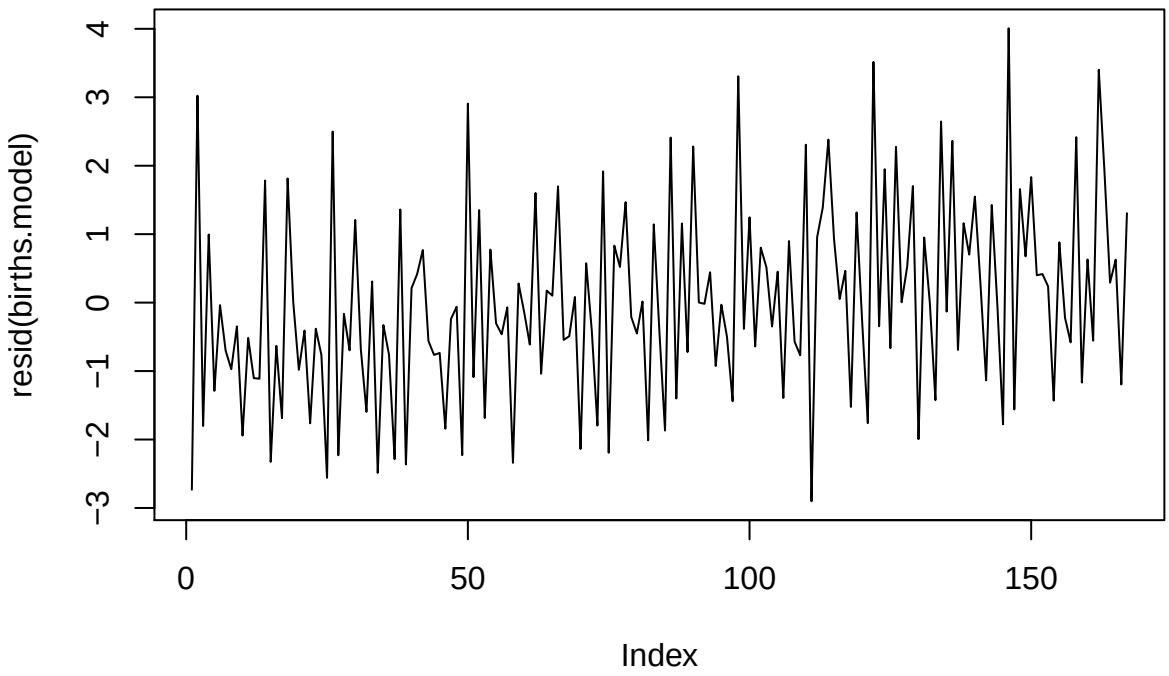


Now, we run a model the difference on the lag.

```
births.model<-lm(D(birthstimeseries)~L(birthstimeseries))
summary(births.model)
```

```
##
## Call:
## lm(formula = D(birthstimeseries) ~ L(birthstimeseries))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9017 -0.9492 -0.1305  0.9126  4.0068
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           5.25380    1.20895   4.346 2.42e-05 ***
## L(birthstimeseries) -0.20950    0.04807  -4.358 2.30e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.434 on 165 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1032, Adjusted R-squared:  0.09779
## F-statistic: 18.99 on 1 and 165 DF,  p-value: 2.301e-05
```

```
plot(resid(births.model),type = "l")
```



That model is equivalent to running a regression of the series today on a lag and a trend.