Basic Time Series Analysis

Joseph Sepich (jps6444)

09/24/2020

```
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
```

```
library(lubridate)
```

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

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

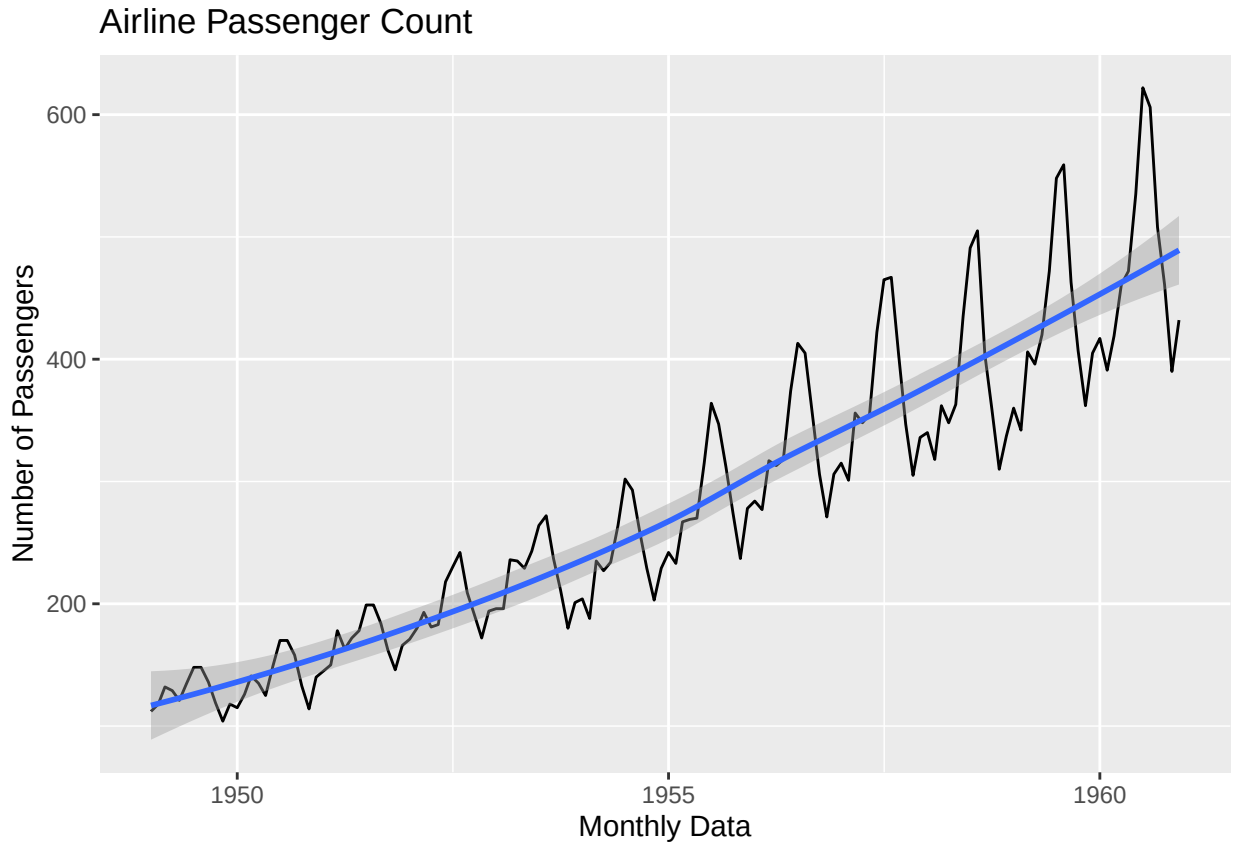
```
library(ggplot2)
```

```
# read airline data
airlines <- read.table('./airpassenger.txt')
airlines <- airlines %>%
  rename(year = V1, month = V2, count = V3) %>%
  mutate(time = ymd(paste0(year, "-", month,"-01"))) %>%
  arrange(time)
airlines %>%
  head()
```

```
##   year month count       time
## 1 1949     1   112 1949-01-01
## 2 1949     2   118 1949-02-01
## 3 1949     3   132 1949-03-01
## 4 1949     4   129 1949-04-01
## 5 1949     5   121 1949-05-01
## 6 1949     6   135 1949-06-01
```

```
airlines %>%
  ggplot(aes(x = time, y = count)) +
  geom_line() +
  geom_smooth() +
  ggtitle("Airline Passenger Count") +
  xlab("Monthly Data") +
  ylab("Number of Passengers")
```

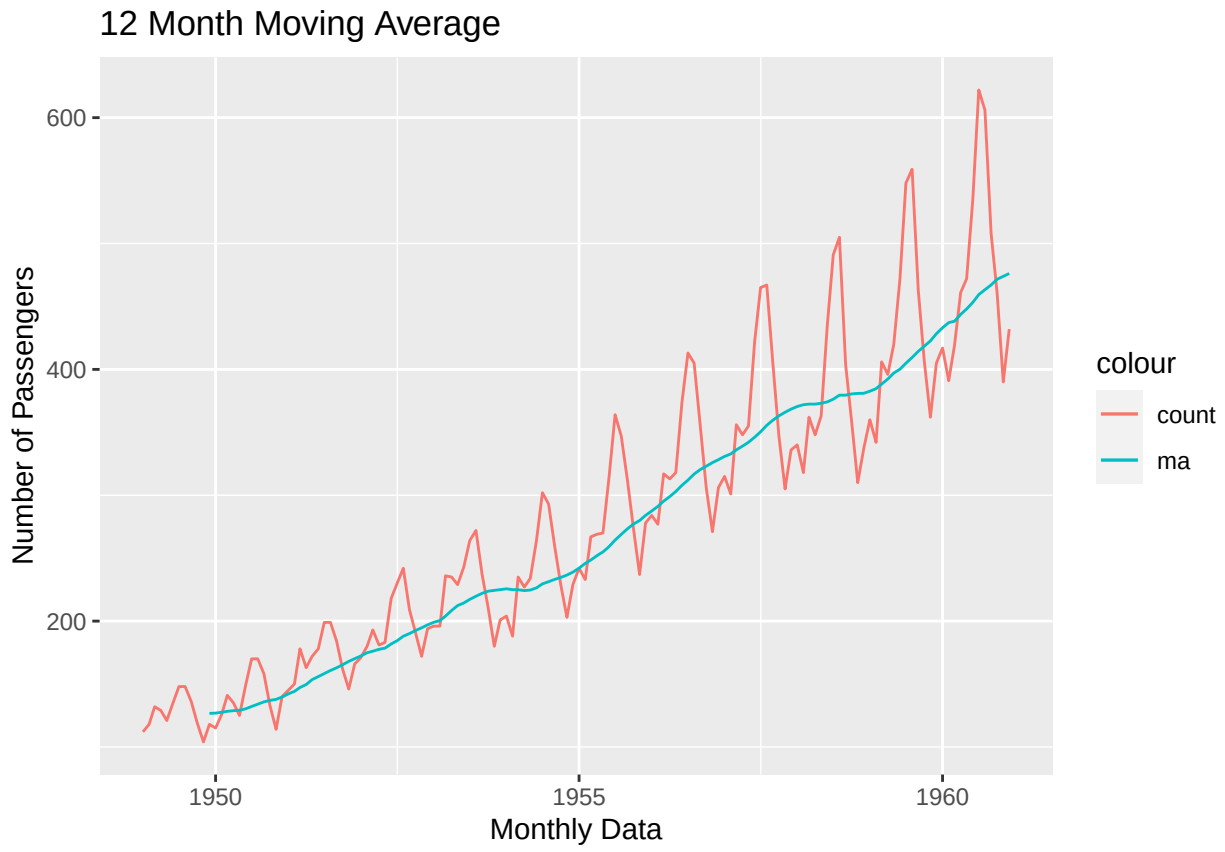
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Moving Average

```
sma <- function(x, n = 12){stats::filter(x, rep(1 / n, n), sides = 1)}
ma_count <- sma(airlines$count)
airlines %>%
  ggplot(aes(x = time, y = count, color="count")) +
  geom_line() +
  geom_line(aes(y=ma_count, color="ma")) +
  ggtitle("12 Month Moving Average") +
  xlab("Monthly Data") +
  ylab("Number of Passengers")
```

```
## Warning: Removed 11 row(s) containing missing values (geom_path).
```



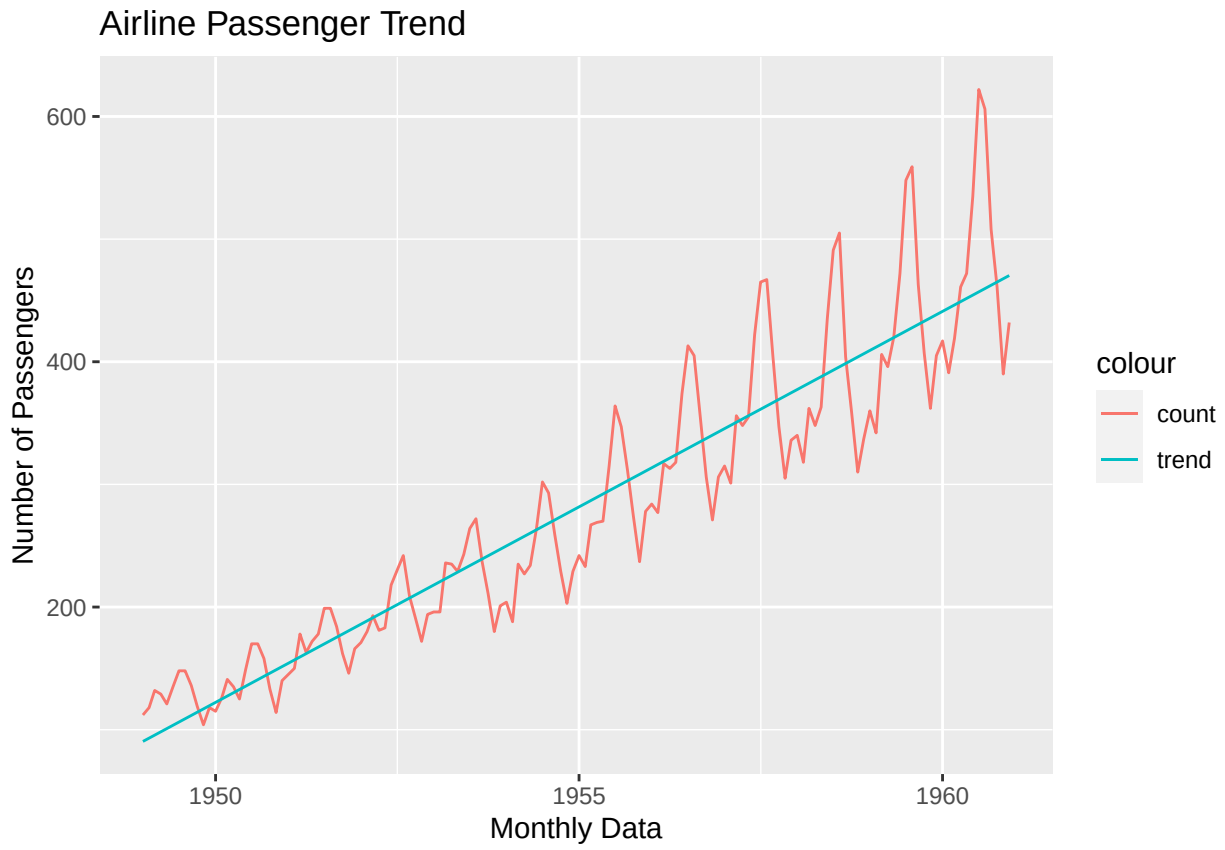
Trendline

```
model <- lm(count~time, data=airlines)
summary(model)
```

```
##
## Call:
## lm(formula = count ~ time, data = airlines)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -93.887 -30.700  -5.723  24.447 165.039
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.600e+02  1.711e+01   44.41   <2e-16 ***
## time        8.729e-02  3.035e-03   28.76   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

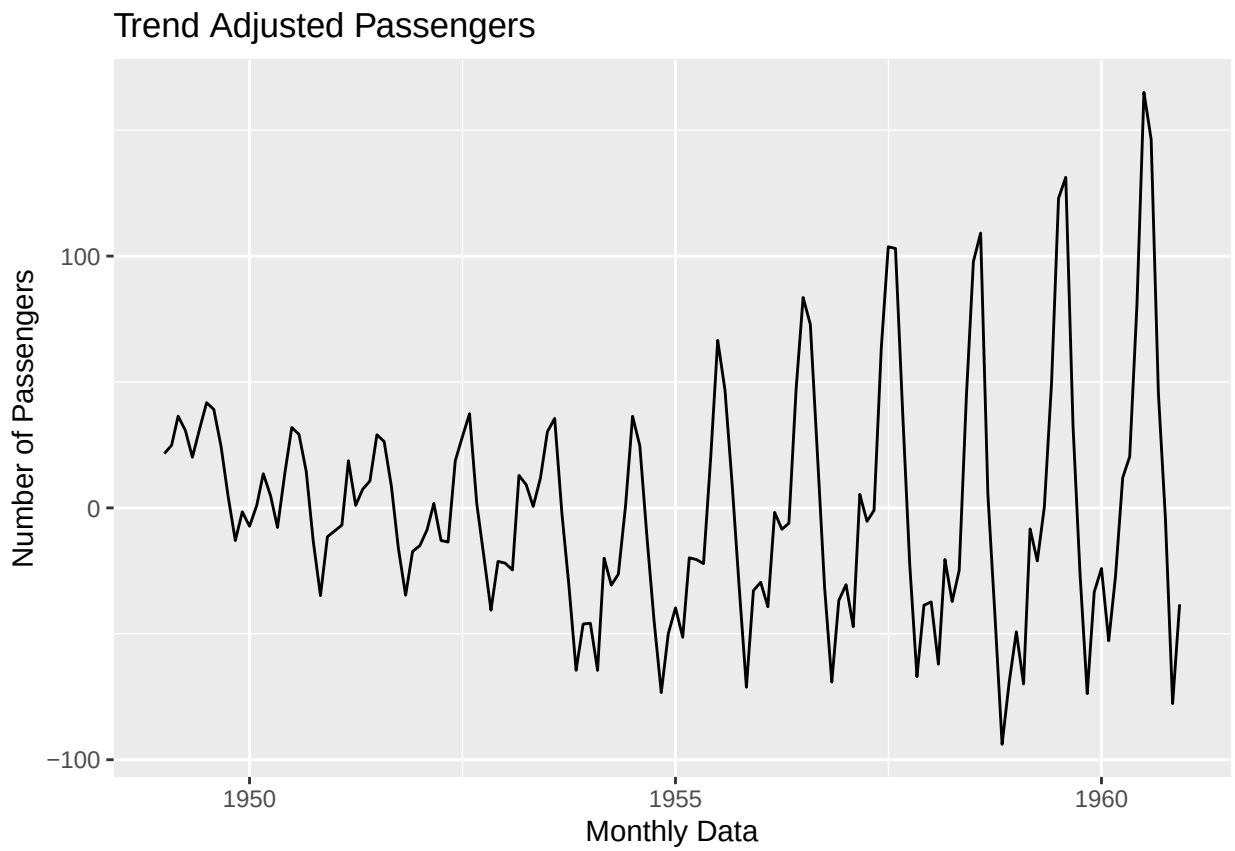
```
## Residual standard error: 46.08 on 142 degrees of freedom
## Multiple R-squared:  0.8535, Adjusted R-squared:  0.8525
## F-statistic: 827.3 on 1 and 142 DF,  p-value: < 2.2e-16
```

```
count_pred <- predict(model, data.frame(time=airlines$time))
airlines %>%
  ggplot(aes(x = time, y = count, color="count")) +
  geom_line() +
  geom_line(aes(y=count_pred, color="trend")) +
  ggtitle("Airline Passenger Trend") +
  xlab("Monthly Data") +
  ylab("Number of Passengers")
```



Adjust data using trend

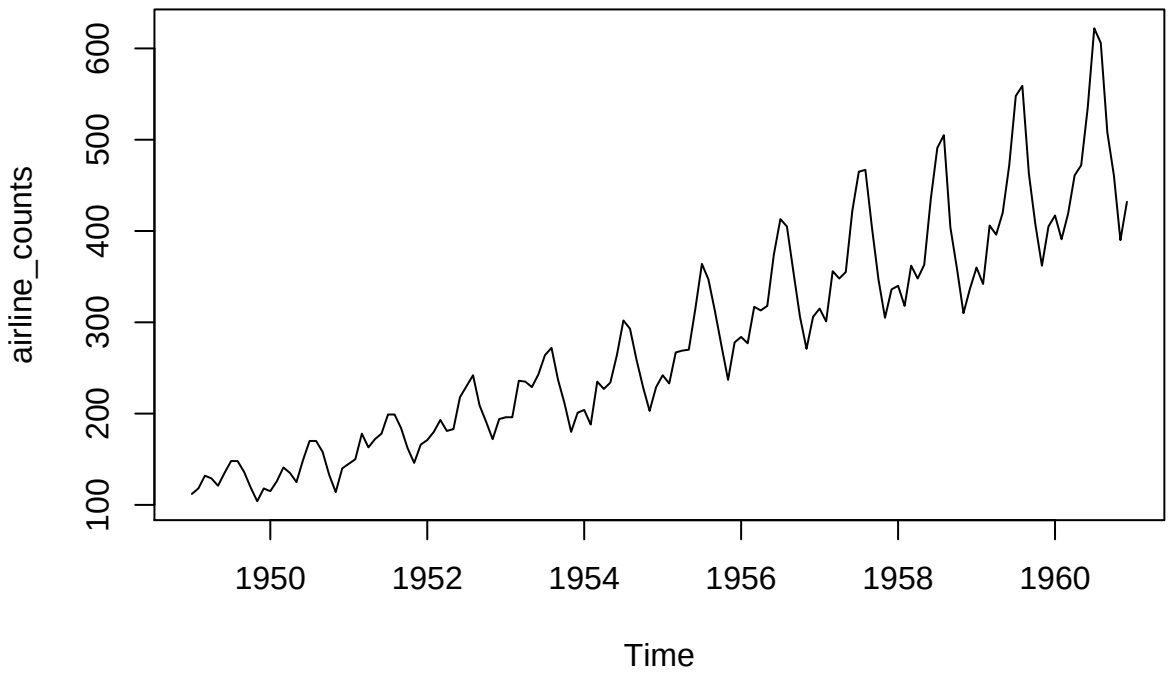
```
airlines %>%
  ggplot(aes(x = time, y = count - count_pred)) +
  geom_line() +
  ggtitle("Trend Adjusted Passengers") +
  xlab("Monthly Data") +
  ylab("Number of Passengers")
```



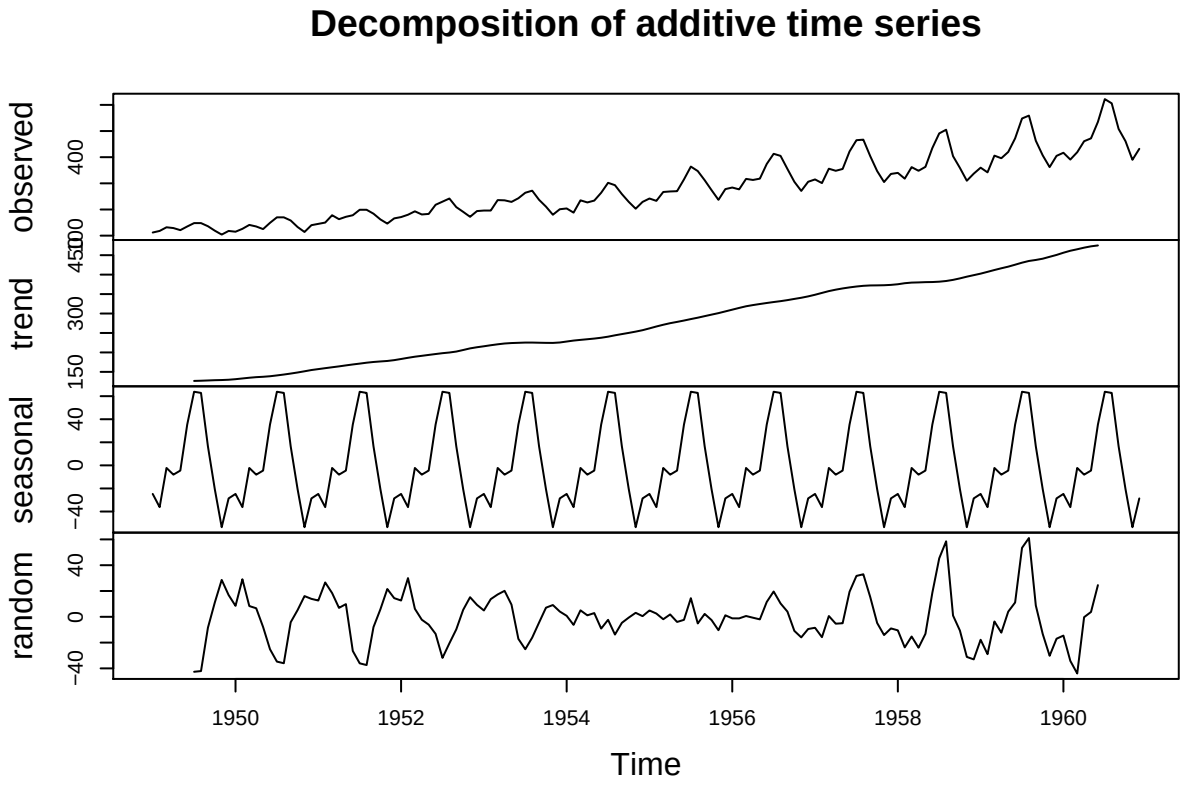
Using R ts

```
library(TTR)
```

```
airline_counts <- ts(airlines$count, frequency=12, start=c(1949,1))
plot(airline_counts)
```

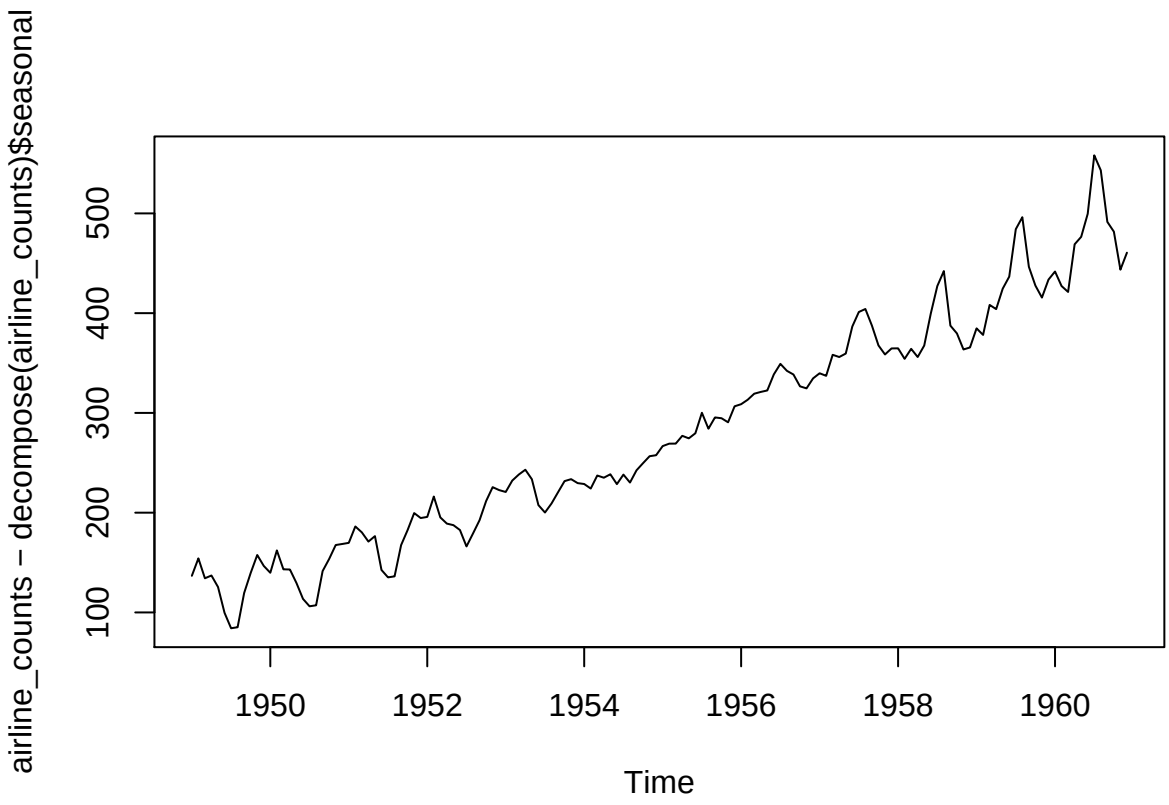



```
plot(decompose(airline_counts))
```

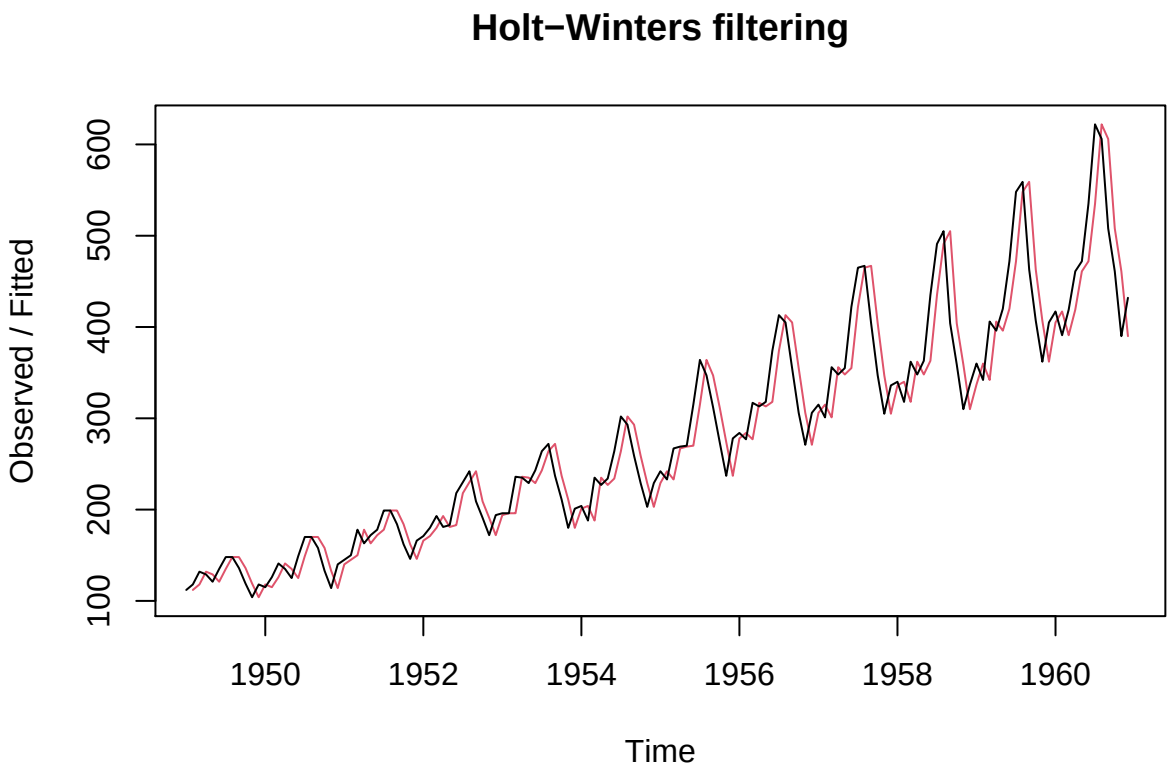


Seasonally Adjusted

```
plot(airline_counts - decompose(airline_counts)$seasonal)
```




```
airline_forecast <- HoltWinters(airline_counts, beta = FALSE, gamma = FALSE)
plot(airline_forecast)
```




```
airline_forecast$SSE
```

```
## [1] 162510.6
```