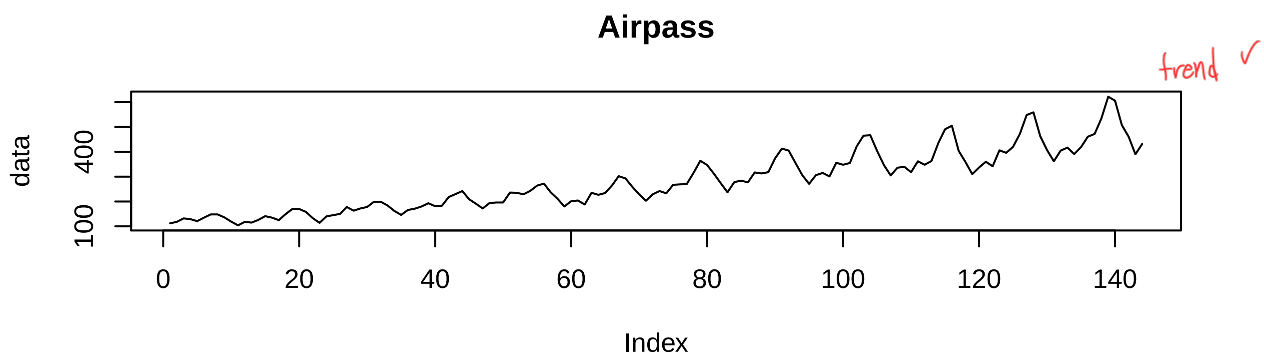
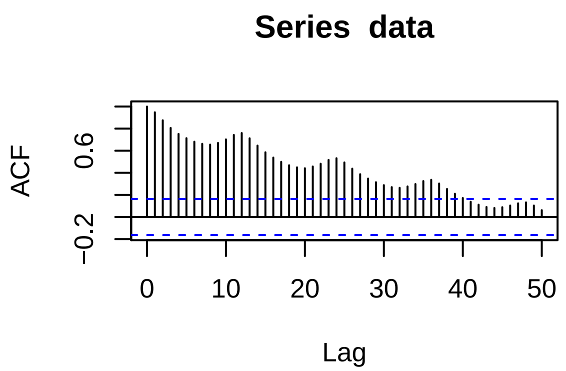
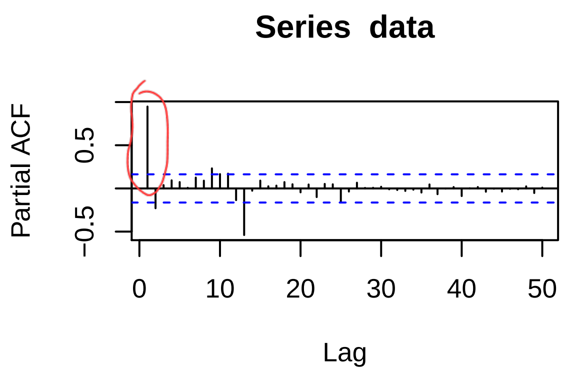
BoxCox Transformation

Here we illustrate how to apply Box-Cox Transformation. The data set is a airpassenger data in US.

```
data = scan("airpass.txt");
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
plot(data, type="l")
title("Airpass")
acf(data, lag=50);
pacf(data, lag=50);
```

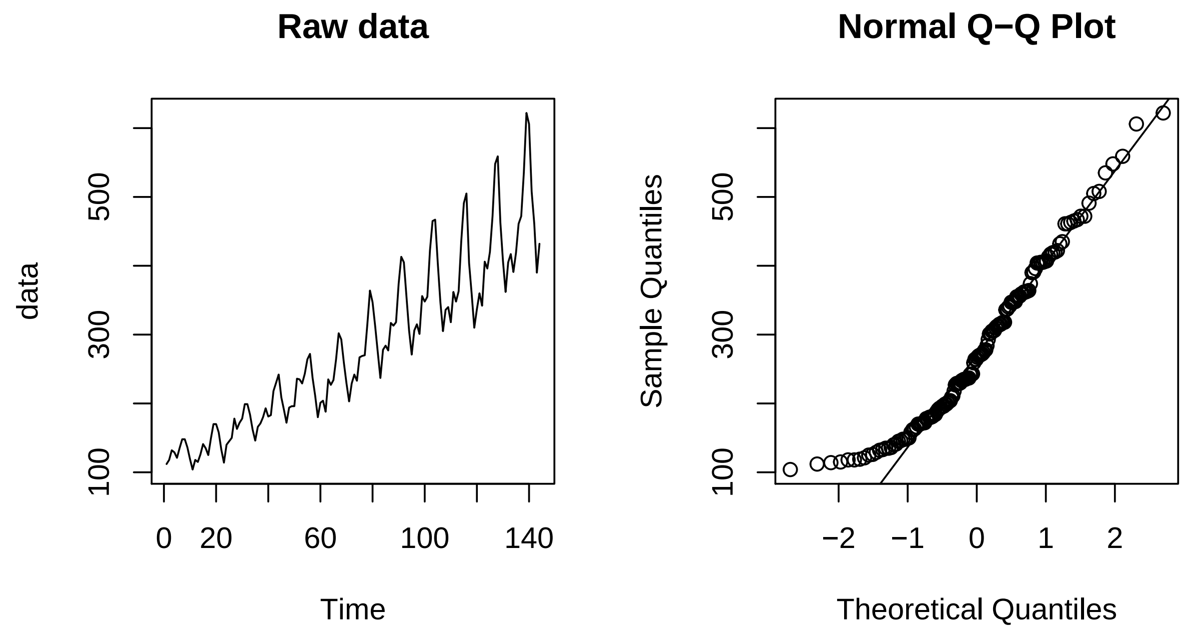




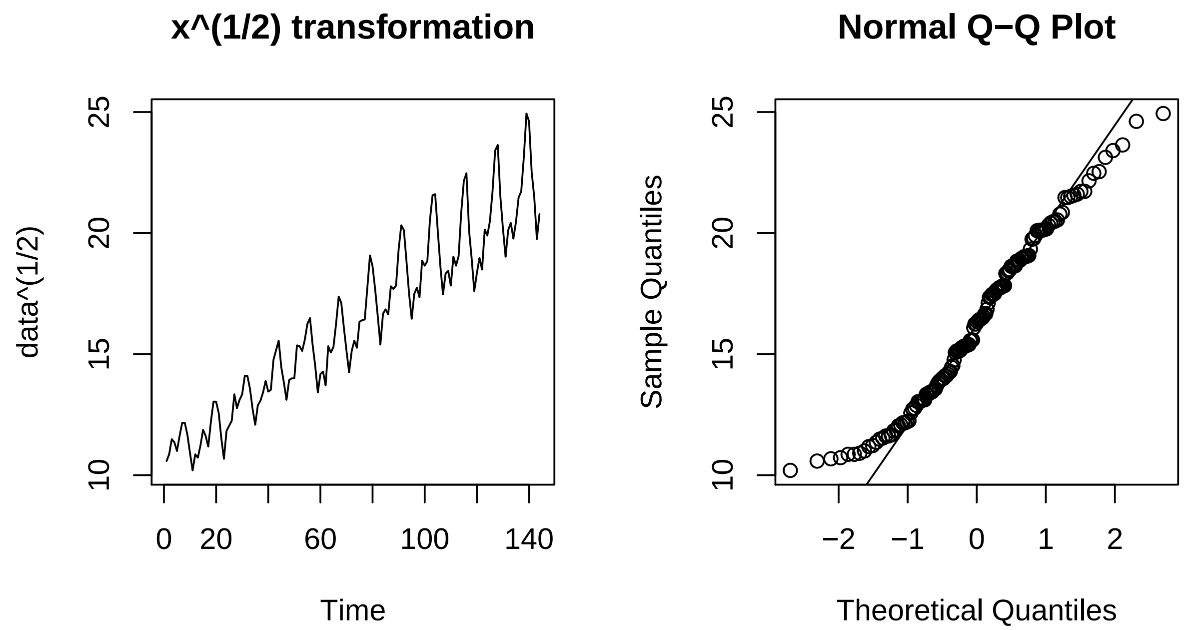


Albeit clear trend and seasonal component, we can observe some increasing variance, so we may need to do some transformation. The first method is ad-hoc graphical method. If we assume Gaussianity on the innovations, it is equivalent to find the transformation makes the data closest to Normal. Hence, we can use QQplot to detect it.

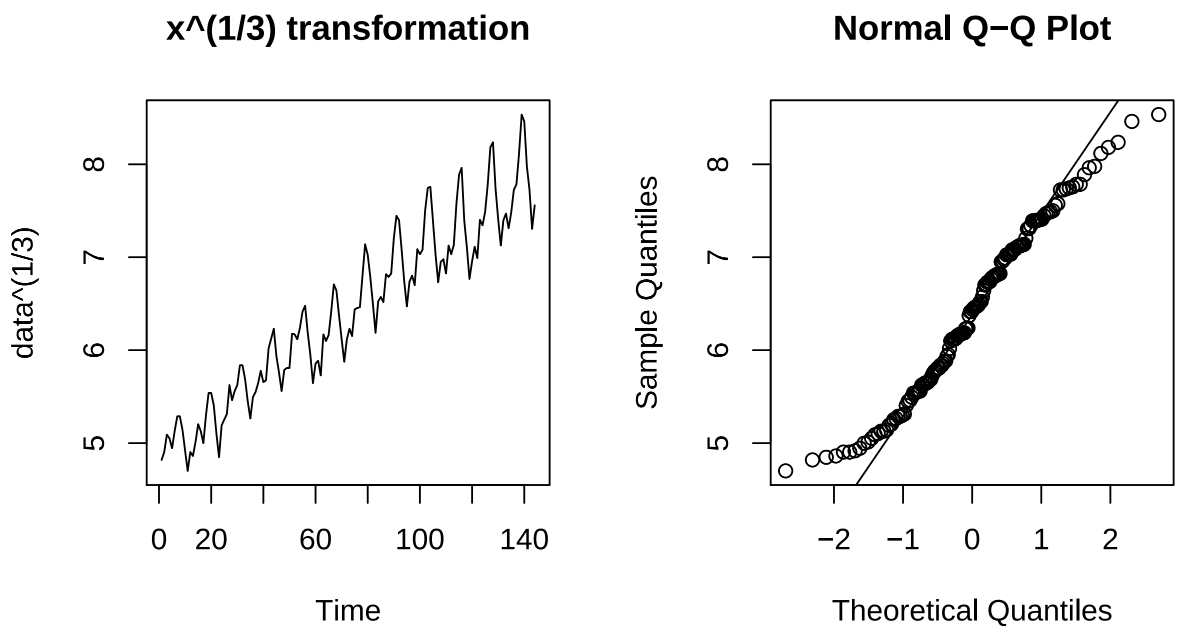
```
par(mfrow=c(1,2))
plot.ts(data); title("Raw data")
qqnorm(data);qqline(data);
```



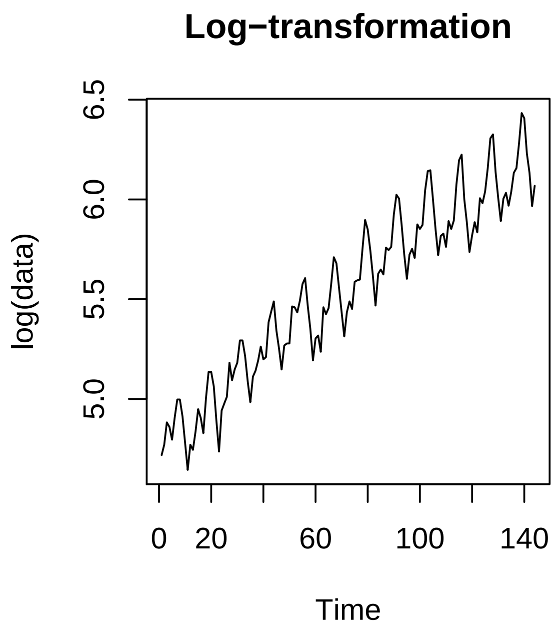

```
par(mfrow=c(1,2))
par(mfrow=c(1,2))
plot.ts(data^(1/2)); title("x^(1/2) transformation")
qqnorm(data^(1/2)); qqline(data^(1/2));
```

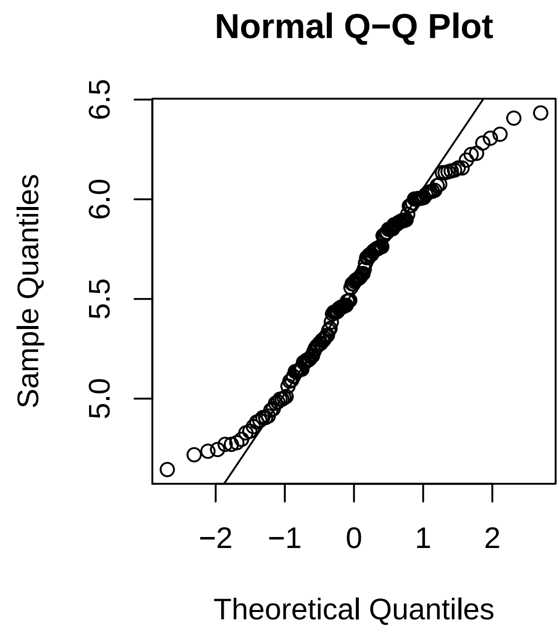


```
plot.ts(data^(1/3)); title("x^(1/3) transformation")
qqnorm(data^(1/3)); qqline(data^(1/3));
```




```
par(mfrow=c(1,2))
plot.ts(log(data)); title("Log-transformation")
qqnorm(log(data)); qqline(log(data));
```





log-transformation seems to best-stabilize the variance by this method.

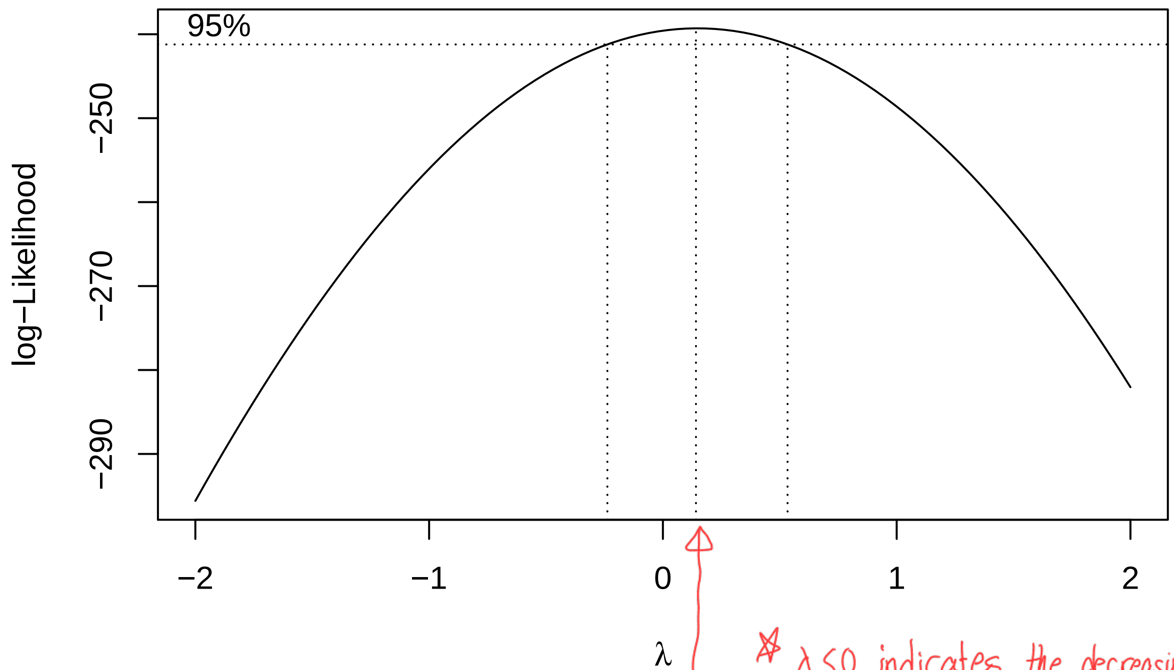
The second method is to estimate lambda by MLE. It is implemented in MASS() library.

```
library(MASS)
library(itsmr)
```

```
##
## Attaching package: 'itsmr'

## The following object is masked from 'package:MASS':
##
##     deaths
```

```
x = 1:length(data)
fit = boxcox(data~1, plotit=TRUE);
```



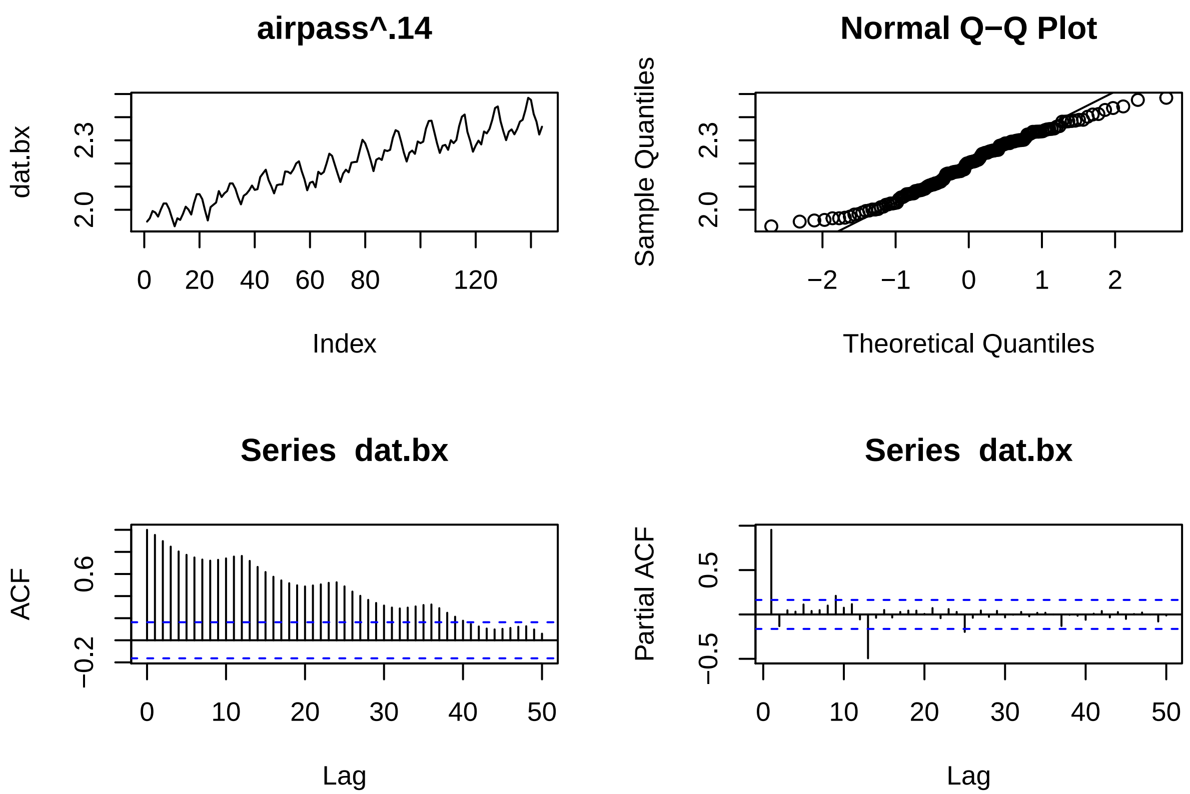
$\lambda < 0$ indicates the decreasing variance

```
lambda = fit$x[which.max(fit$y)]
lambda
```

```
## [1] 0.1414141
```

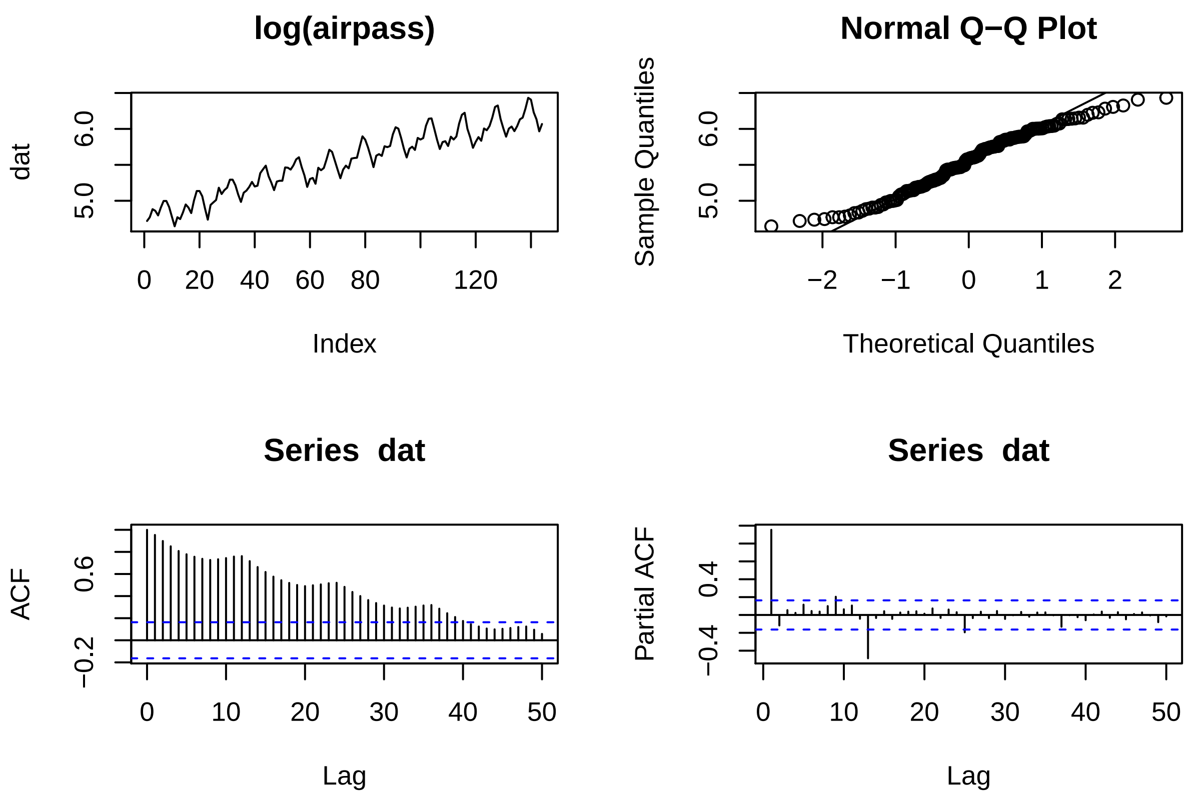
It suggests to take $data^{.14}$ transformation. Then, it gives the following results.

```
dat.bx = data^lambda;
layout(matrix(c(1,2,3,4), 2, 2, byrow = TRUE))
plot(dat.bx, type="l")
title("airpass^.14")
qqnorm(dat.bx); qqline(dat.bx);
acf(dat.bx, lag=50);
pacf(dat.bx, lag=50);
```



Box-Cox transforamtion makes the variance constant over time. However, in practice, taking the power transformation makes the model interpretation hard. In my own experiences, we prefer to take either $\log(y)$ or $\sqrt{y}$ transformation if needed.

```
dat = log(data);
layout(matrix(c(1,2,3,4), 2, 2, byrow = TRUE))
plot(dat, type="l")
title("log(airpass)")
qqnorm(dat); qqline(dat);
acf(dat, lag=50);
pacf(dat, lag=50);
```



Practice: See HW problem with wind data.