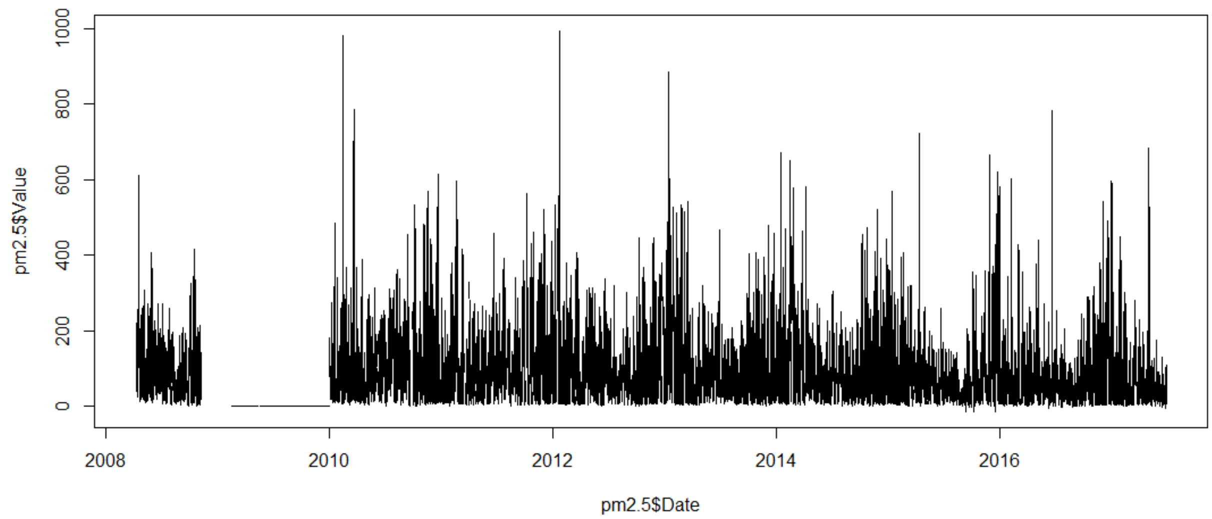
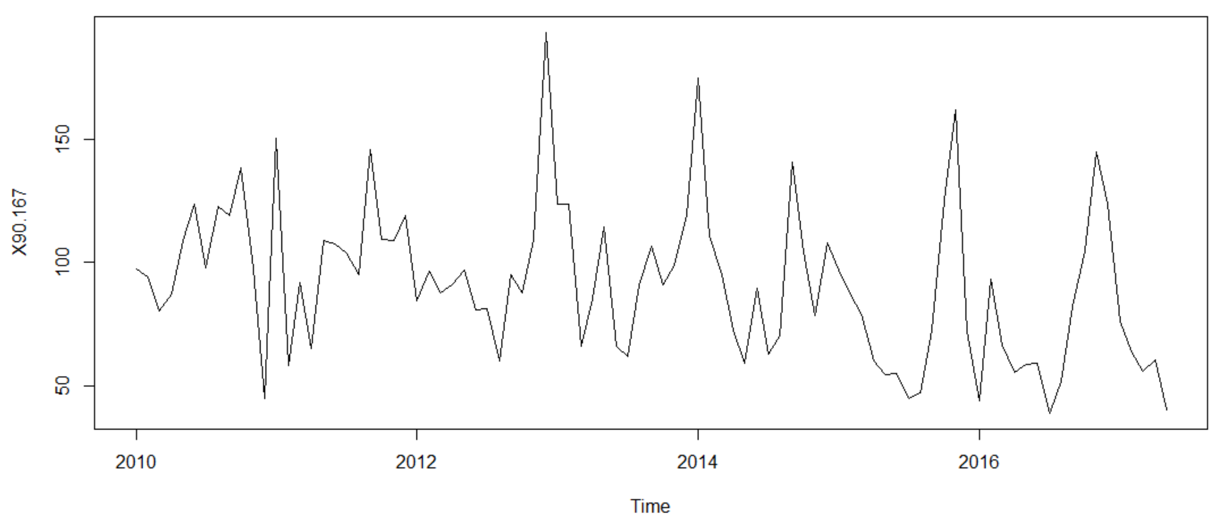
Firstly, I make a time plot.



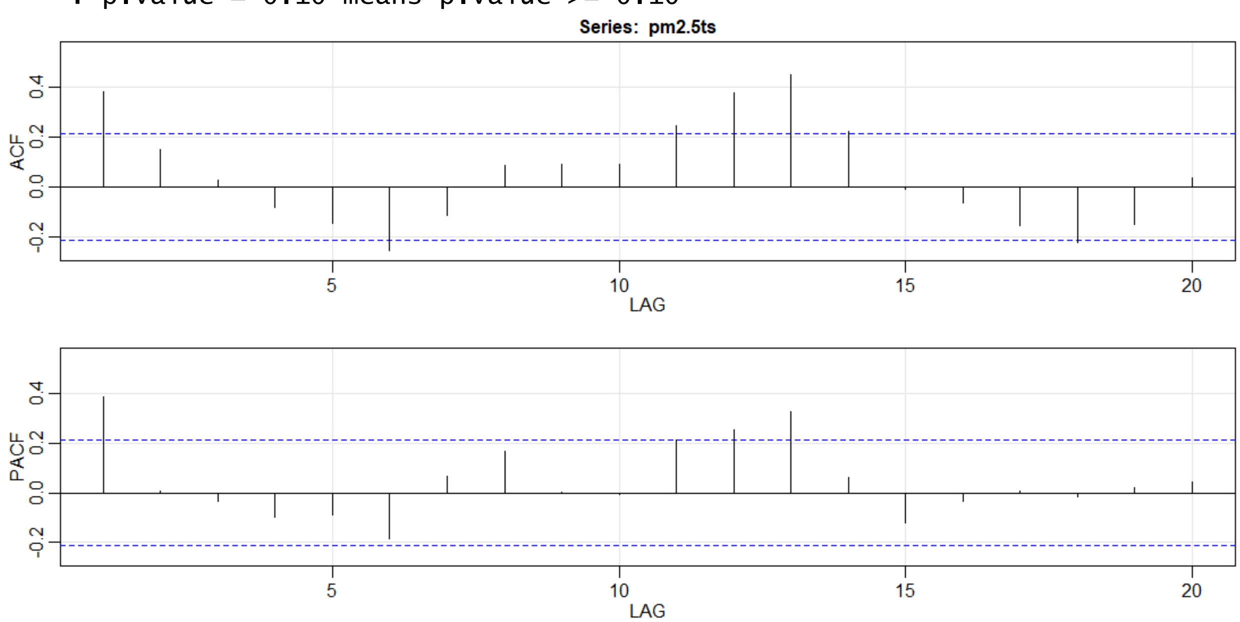
But there are too many missing values, and it's hard to analyze. So I delete all the missing values and calculate the average monthly pm2.5. Details are included in the dataset "monthly2.txt".



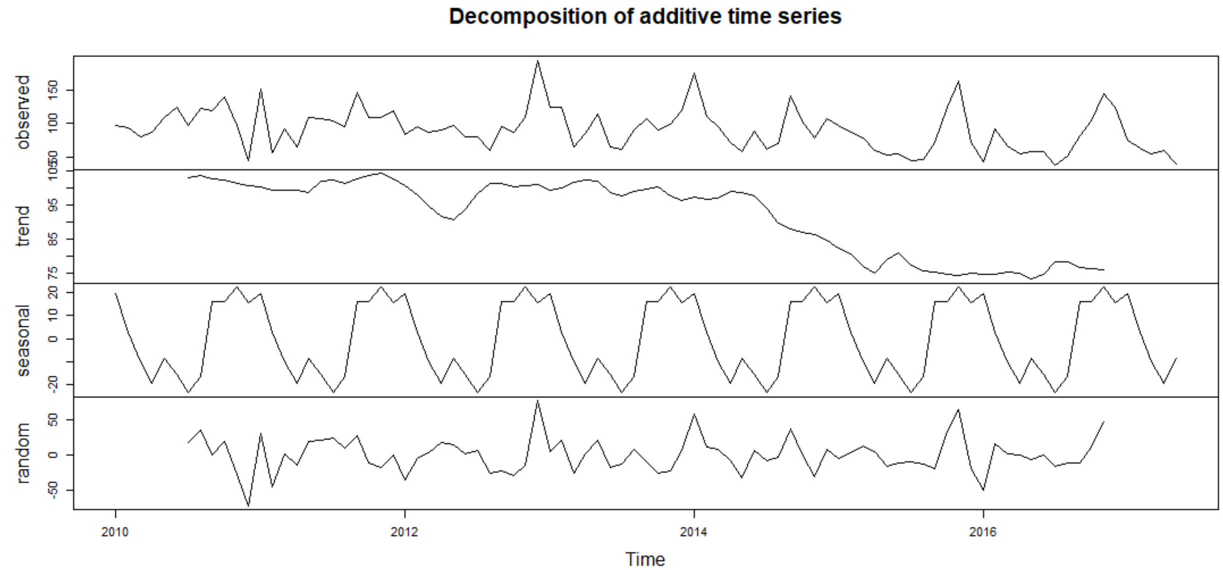
Then stationary tests are performed and we can see the ACF and PACF.

```
> adf.test(pm2.5ts)

        Augmented Dickey-Fuller Test

data:  pm2.5ts
Dickey-Fuller = -5.2312, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

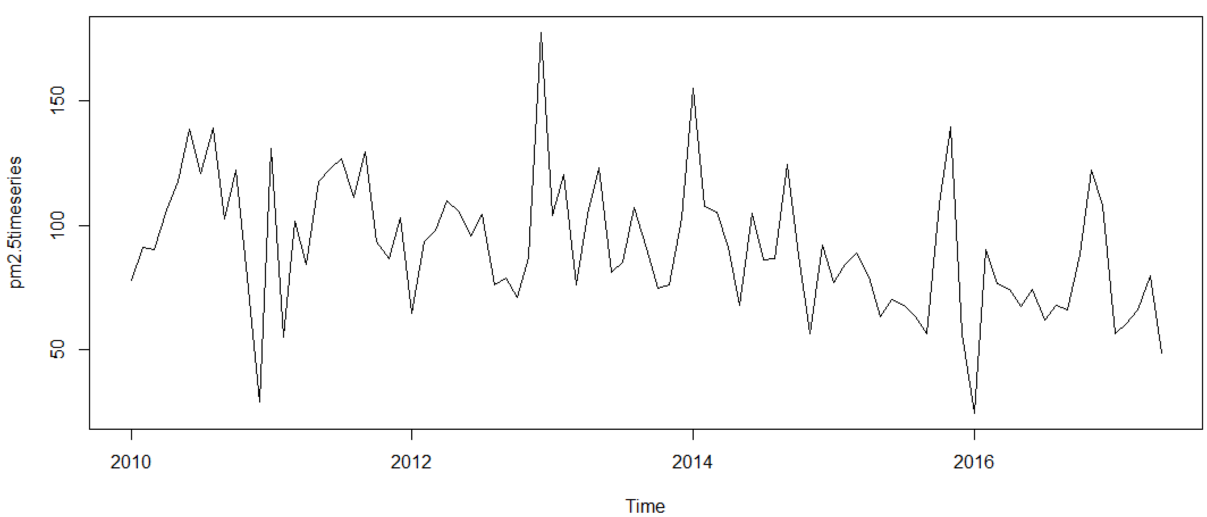
```
> stationary.test(pm2.5ts, method = "pp") # same as pp.test(x)
Phillips-Perron Unit Root Test
alternative: stationary

Type 1: no drift no trend
 lag Z_rho p.value
   3 -3.47   0.278
-----
 Type 2: with drift no trend
 lag Z_rho p.value
   3   -54    0.01
-----
 Type 3: with drift and trend
 lag Z_rho p.value
   3 -58.8    0.01
---------------
Note: p-value = 0.01 means p.value <= 0.01
> stationary.test(pm2.5ts, method = "kpss") # same as kpss.test(x)
KPSS Unit Root Test
alternative: nonstationary

Type 1: no drift no trend
 lag stat p.value
   2 1.72  0.0464
-----
 Type 2: with drift no trend
 lag  stat p.value
   2 0.495  0.0427
-----
 Type 1: with drift and trend
 lag   stat p.value
   2 0.0396     0.1
-----------
Note: p.value = 0.01 means p.value <= 0.01
    : p.value = 0.10 means p.value >= 0.10
```



Though we can conclude if the time series is stationary or nor, from the figure above it's seasonal. Thus, decomposition of the time series is performed.



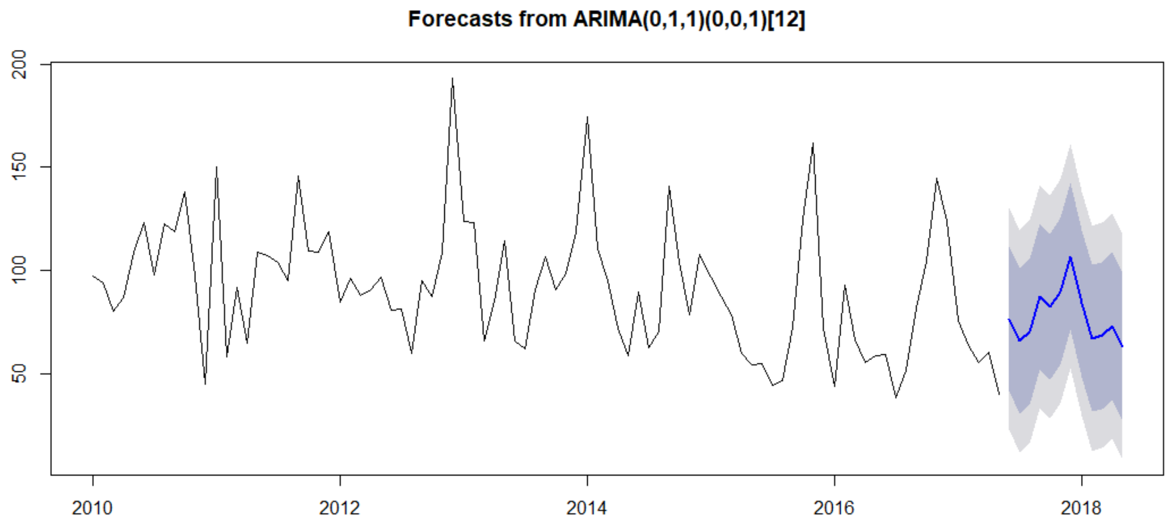
From the figure above, obviously there is a decreasing trend of the time series, which can prove that pm2.5 value of Beijing is basically reducing.

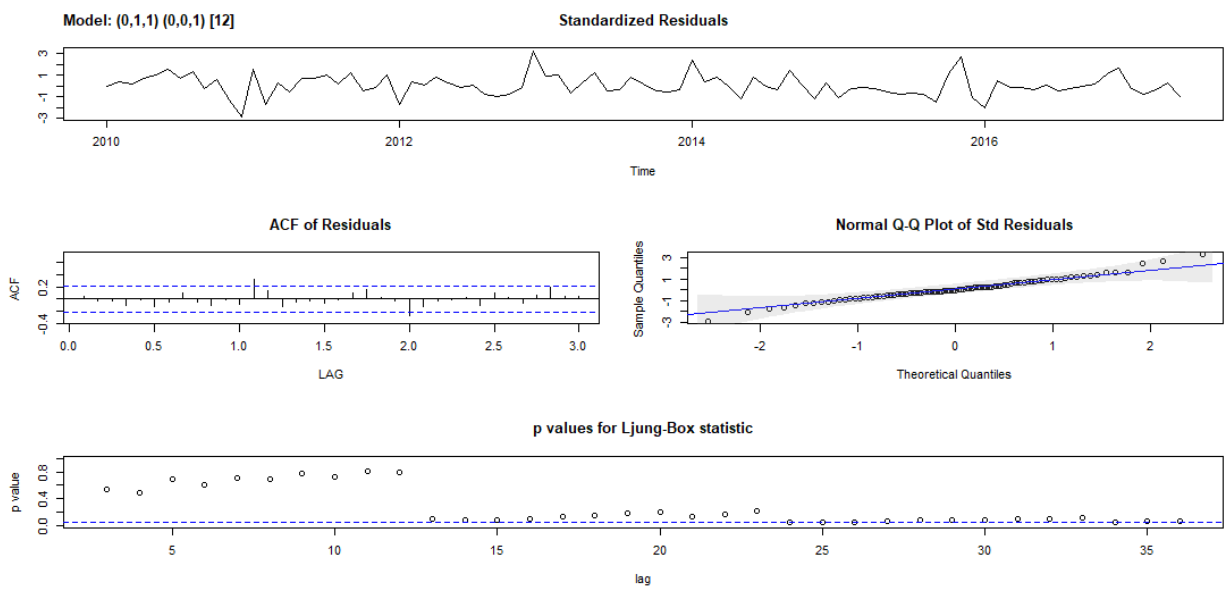



```
> adf.test(pm2.5noseason)

        Augmented Dickey-Fuller Test

data:  pm2.5noseason
Dickey-Fuller = -4.7766, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
> stationary.test(pm2.5noseason, method = "pp") # same as pp.test(x)
```

```
Phillips-Perron Unit Root Test
alternative: stationary

Type 1: no drift no trend
 lag Z_rho p.value
   3 -2.41   0.368
-----
 Type 2: with drift no trend
 lag Z_rho p.value
   3 -65.4    0.01
-----
 Type 3: with drift and trend
 lag Z_rho p.value
   3 -71.5    0.01
---------------
Note: p-value = 0.01 means p.value <= 0.01
> stationary.test(pm2.5noseason, method = "kpss") # same as kpss.test(x)
KPSS Unit Root Test
alternative: nonstationary

Type 1: no drift no trend
 lag stat p.value
   2 2.01  0.0306
-----
 Type 2: with drift no trend
 lag  stat p.value
   2 0.944    0.01
-----
 Type 1: with drift and trend
 lag   stat p.value
   2 0.0535     0.1
-----------
Note: p.value = 0.01 means p.value <= 0.01
    : p.value = 0.10 means p.value >= 0.10
> fit1 <- auto.arima(pm2.5timeseries,seasonal = TRUE, approximation = FALSE,
stepwise = FALSE)
> fit1
Series: pm2.5timeseries
ARIMA(0,1,1)(0,0,1)[12]

Coefficients:
          ma1    sma1
      -0.9471  0.5040
s.e.   0.0332  0.1146

sigma^2 estimated as 754.5:  log likelihood=-418.06
AIC=842.12   AICc=842.41   BIC=849.55
> sarima(pm2.5noseason,0,1,1,0,0,1,12)
initial  value 3.461508
iter   2 value 3.343969
iter   3 value 3.272602
iter   4 value 3.267929
iter   5 value 3.264526
iter   6 value 3.254229
iter   7 value 3.253207
iter   8 value 3.248024
iter   9 value 3.247992
```

```
iter  10 value 3.247943
iter  11 value 3.247868
iter  12 value 3.247852
iter  12 value 3.247852
iter  12 value 3.247852
final  value 3.247852
converged
initial  value 3.229920
iter   2 value 3.228418
iter   3 value 3.214790
iter   4 value 3.210124
iter   5 value 3.209778
iter   6 value 3.209678
iter   7 value 3.209653
iter   8 value 3.209649
iter   8 value 3.209649
final  value 3.209649
converged
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
    Q), period = S), xreg = constant, optim.control = list(trace = trc, REPOR
T = 1,
    reltol = tol))

Coefficients:
          ma1    sma1  constant
      -1.0000  0.2717   -0.3904
s.e.   0.0332  0.1615    0.1170

sigma^2 estimated as 579.8:  log likelihood = -407.32,  aic = 822.63

$degrees_of_freedom
[1] 85

$ttable
         Estimate     SE  t.value p.value
ma1       -1.0000 0.0332 -30.1159  0.0000
sma1       0.2717 0.1615   1.6828  0.0961
constant  -0.3904 0.1170  -3.3379  0.0013

$AIC
[1] 7.430084

$AICc
[1] 7.457906

$BIC
[1] 6.513971
```


Forecasts from ARIMA(0,1,1)(0,0,1)[12]
200
150
100
50
2010
2012
2014
2016
2018


Model: (0,1,1) (0,0,1) [12]
Standardized Residuals
Time
ACF of Residuals
ACF
LAG
Normal Q-Q Plot of Std Residuals
Sample Quantiles
Theoretical Quantiles
p values for Ljung-Box statistic
p value
lag