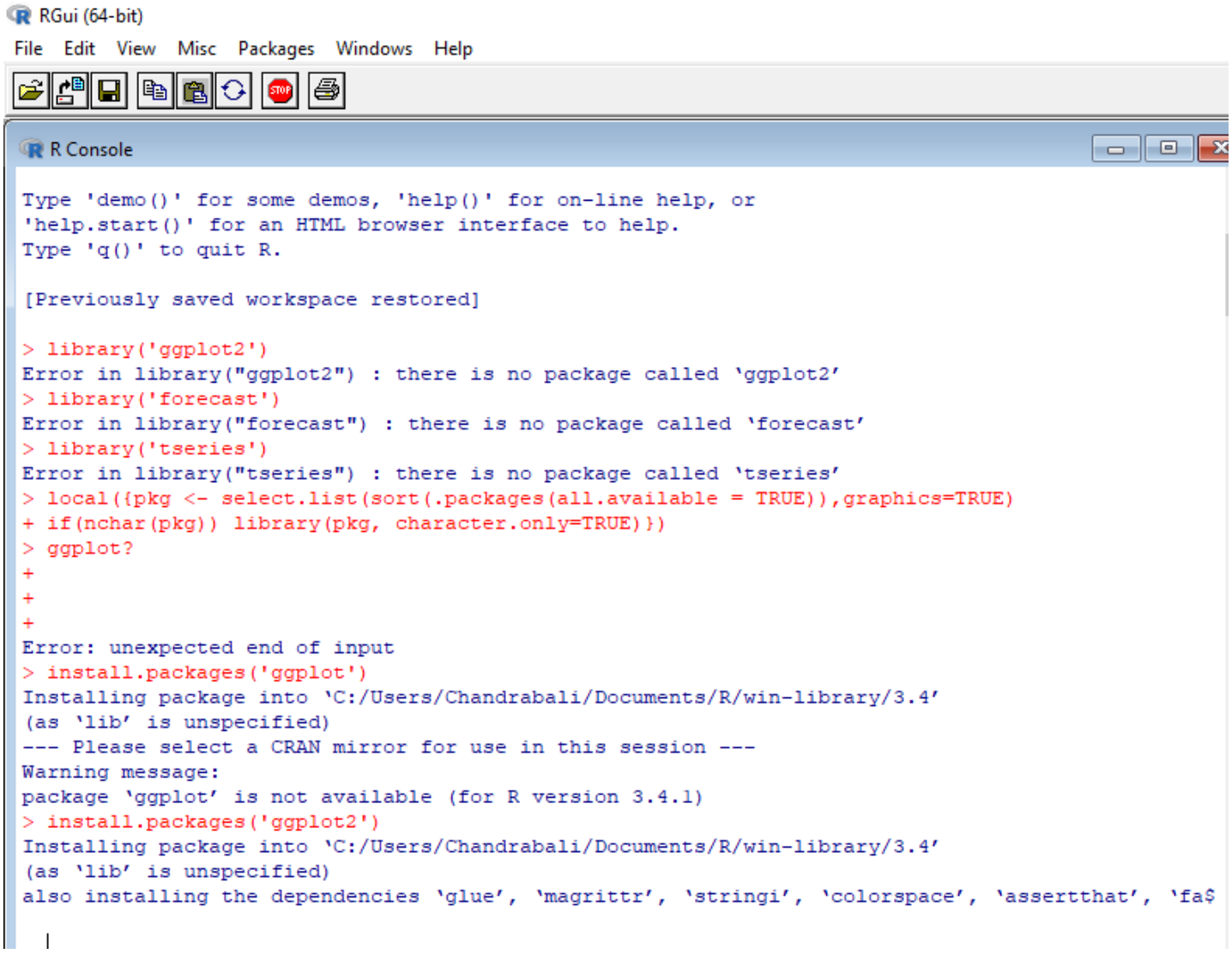
Time series experiment in R with India's annual mean temperature data

At first, I load the necessary packages. I don't have them so I have to install them before I can actually load them.



```
> library(ggplot2)
Warning message:
package 'ggplot2' was built under R version 3.4.4
> install.packages('forecast')
Installing package into 'C:/Users/Chandrabali/Documents/R/win-library/3.4'
(as 'lib' is unspecified)
also installing the dependencies 'xts', 'TTR', 'curl', 'quadprog', 'quantmod', 'fracdiff', 'lr

trying URL 'https://ftp.iitm.ac.in/cran/bin/windows/contrib/3.4/xts_0.11-2.zip'
Content type 'application/zip' length 757167 bytes (739 KB)
downloaded 739 KB

trying URL 'https://ftp.iitm.ac.in/cran/bin/windows/contrib/3.4/TTR_0.23-4.zip'
Content type 'application/zip' length 443214 bytes (432 KB)
```

```
> library('forecast')
Warning message:
package 'forecast' was built under R version 3.4.4
> install.packages('tseries')
Installing package into 'C:/Users/Chandrabali/Documents/R/win-library/3.4'
(as 'lib' is unspecified)
trying URL 'https://ftp.iitm.ac.in/cran/bin/windows/contrib/3.4/tseries_0.10-46.zip'
Content type 'application/zip' length 327628 bytes (319 KB)
downloaded 319 KB

package 'tseries' successfully unpacked and MD5 sums checked
Warning: cannot remove prior installation of package 'tseries'

The downloaded binary packages are in
        C:\Users\Chandrabali\AppData\Local\Temp\RtmpkJgRaM\downloaded_packages
>
```

Load the data.

```
> data=read.csv(file.choose())
>
```

Convert the 'year' column to 'date' datatype. I had to mention the format is it in.

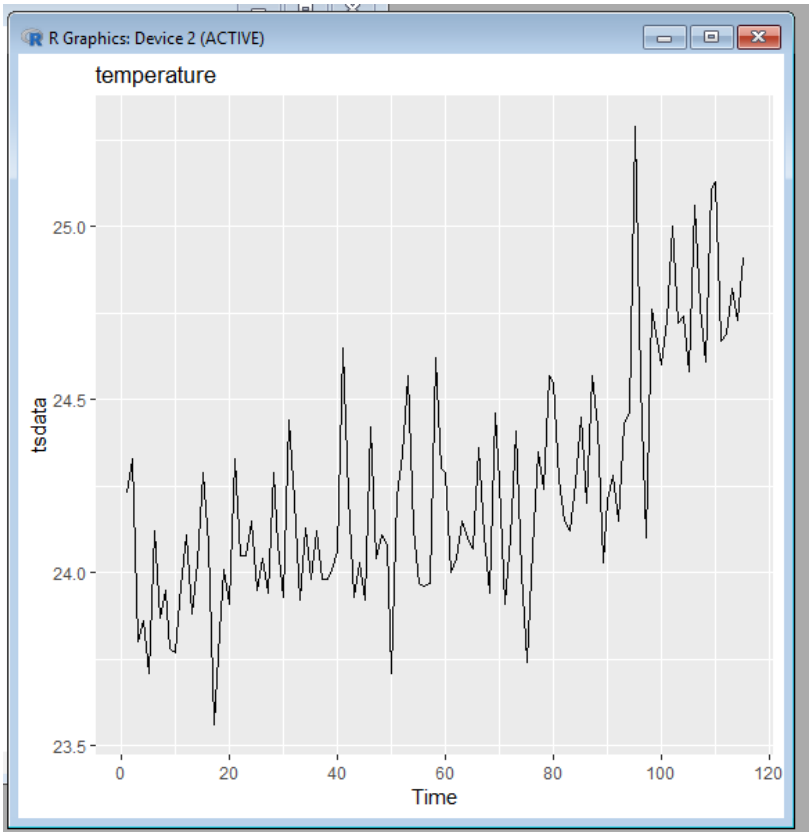
```
> data$year = as.Date(data$year)
Error in charToDate(x) :
  character string is not in a standard unambiguous format
> data$year = as.Date(data$year,format = "%Y-%m-%d")
>
```

Convert the data to a "ts" object with as.ts() function , just for visualization. Save it as "tsdata"

```
> tsdata=as.ts(data)
> tsdata
Time Series:
Start = 1
End = 115
Frequency = 1
     year temperature
  1 1901       24.23
  2 1902       24.33
  3 1903       23.80
  4 1904       23.86
  5 1905       23.71
  6 1906       24.12
  7 1907       23.87
  8 1908       23.95
  9 1909       23.78
```

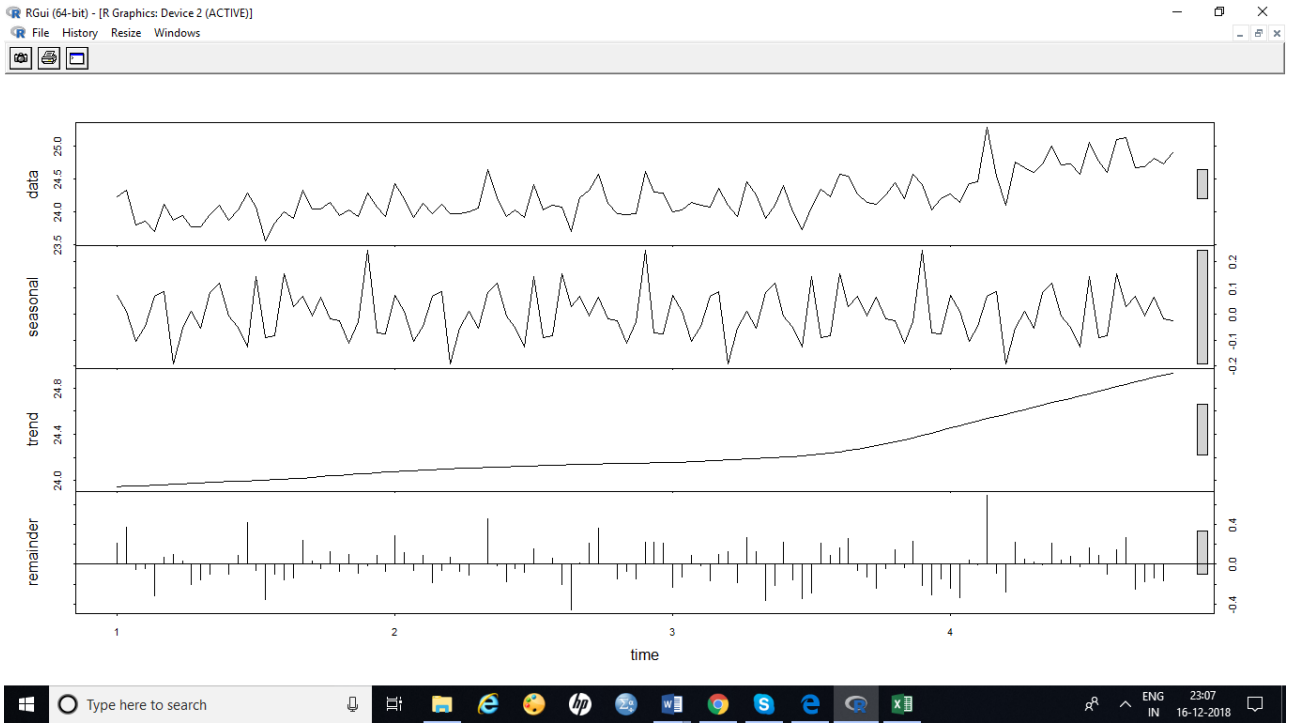
Plot the time series object with ggfortify's autoplot method.

```
> tsdata=as.ts(data)
> autoplot(tsdata) + labs(title="temperature")
```



First, we calculate seasonal component of the data using stl().

```
> temp=ts(na.omit(data$temperature), frequency=30)
> decomp = stl(temp,s.window="periodic")
> deseasonal_temp=seasadj(decomp)
> deseasonal_temp
Time Series:
Start = c(1, 1)
End = c(4, 25)
Frequency = 30
  [1] 24.15766 24.31967 23.90419 23.90871 23.64072 24.03274 24.06158 24.00542 23.76926 23.82310
 [11] 23.87944 23.99090 23.88985 24.08130 24.41525 23.93671 23.65222 23.91274 23.85575 23.88127
 [21] 24.26179 24.05710 23.98491 24.16772 23.97553 24.15236 23.96896 24.04556 24.15216 24.00543
 [31] 24.36766 24.19967 24.02419 24.17871 23.91072 24.03274 24.17158 24.03542 23.99926 24.11310
 [41] 24.56944 24.10090 23.93985 24.08130 24.04525 24.27671 24.13222 24.19274 23.92575 23.68127
 [51] 24.15179 24.34710 24.50491 24.14772 23.99553 24.07236 23.99896 24.37556 24.37216 24.36543
 [61] 23.92766 24.02967 24.25419 24.14871 24.00072 24.27274 24.30158 23.99542 24.44926 24.31310
 [71] 23.82944 23.98090 24.41985 24.05130 23.86525 23.92671 24.44222 24.32274 24.41575 24.52127
 [81] 24.20179 24.15710 24.05491 24.27772 24.47553 24.31236 24.59896 24.17556 24.10216 24.28543
 [91] 24.20766 24.13967 24.53419 24.50871 25.22072 24.46274 24.29158 24.81542 24.65926 24.65310
[101] 24.64944 24.88090 24.72985 24.79130 24.70525 24.91671 24.86222 24.69274 24.95575 25.10127
[111] 24.60179 24.69710 24.75491 24.74772 24.93553
> plot(decomp)
>
```



Now we go to stationary checking. Fitting an ARIMA model requires the series to be stationary. A series is said to be stationary when its mean, variance, and autocovariance are time invariant.

From the above graph, the temperature data does not seem to be stationary.

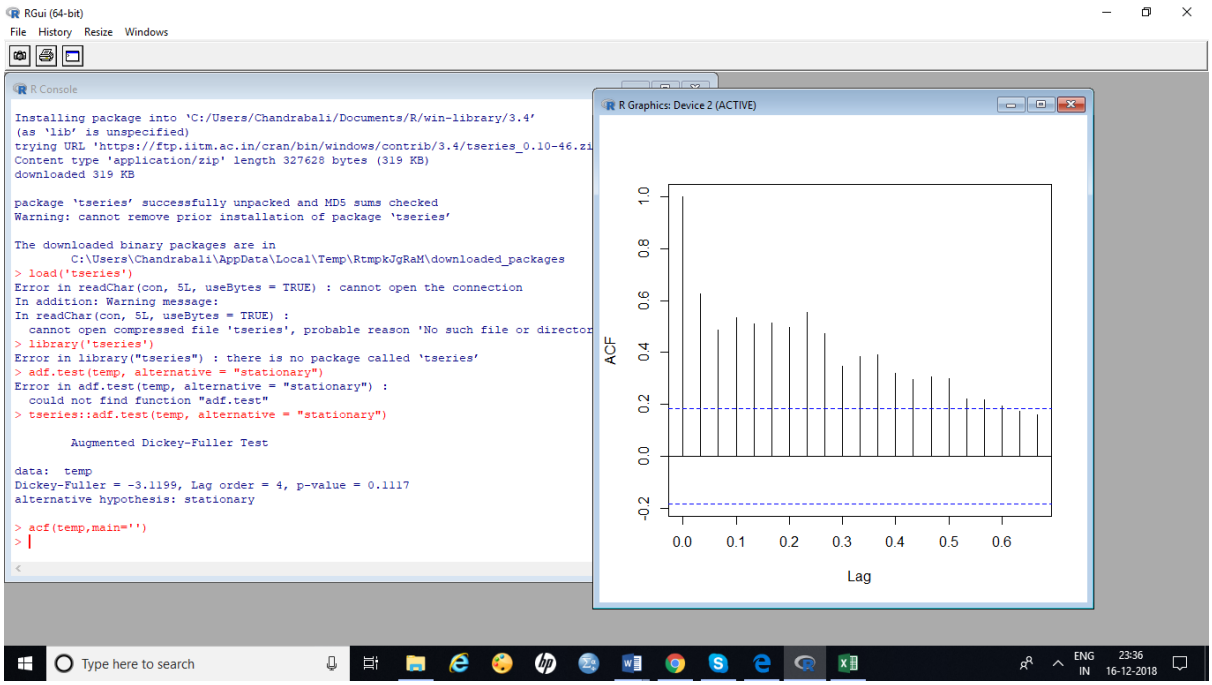
Let’s check with the augmented Dicky Fuller test.

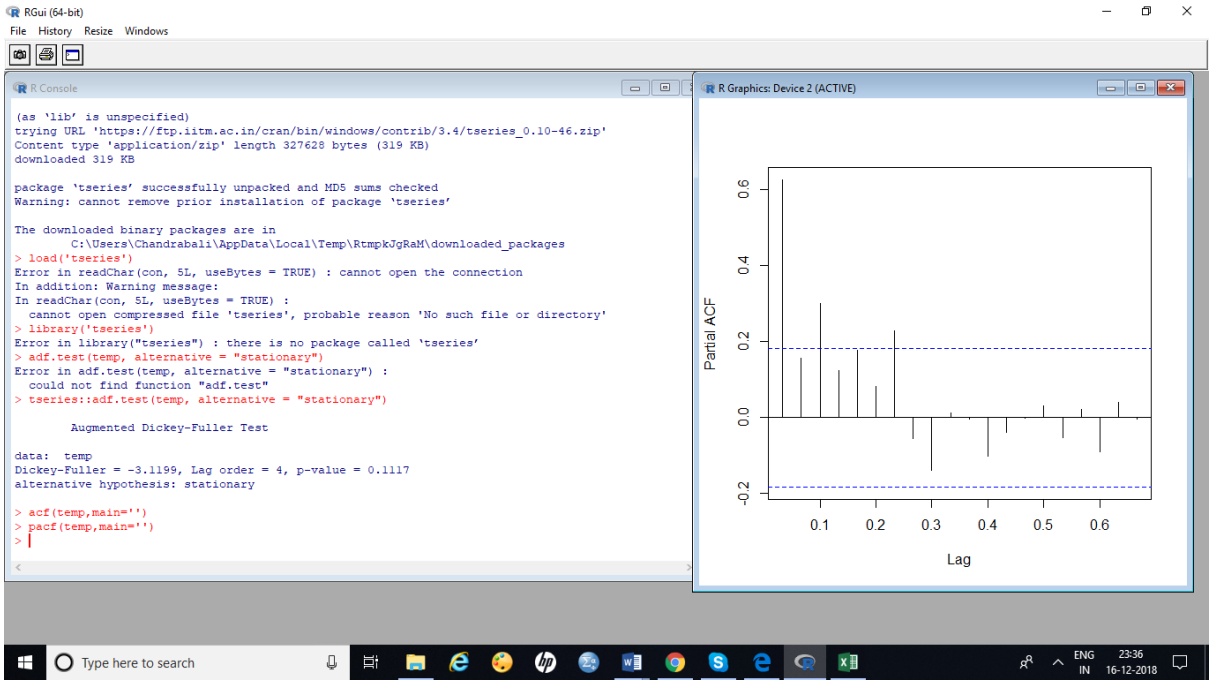
```
> tseries::adf.test(temp, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  temp
Dickey-Fuller = -3.1199, Lag order = 4, p-value = 0.1117
alternative hypothesis: stationary
```

We cannot reject H0, the null hypothesis with 5% significance level so we conclude that the data is non-stationary.

So, we’ve got to stationarize the data.

But to cross-check, we use ACF and PACF plots.



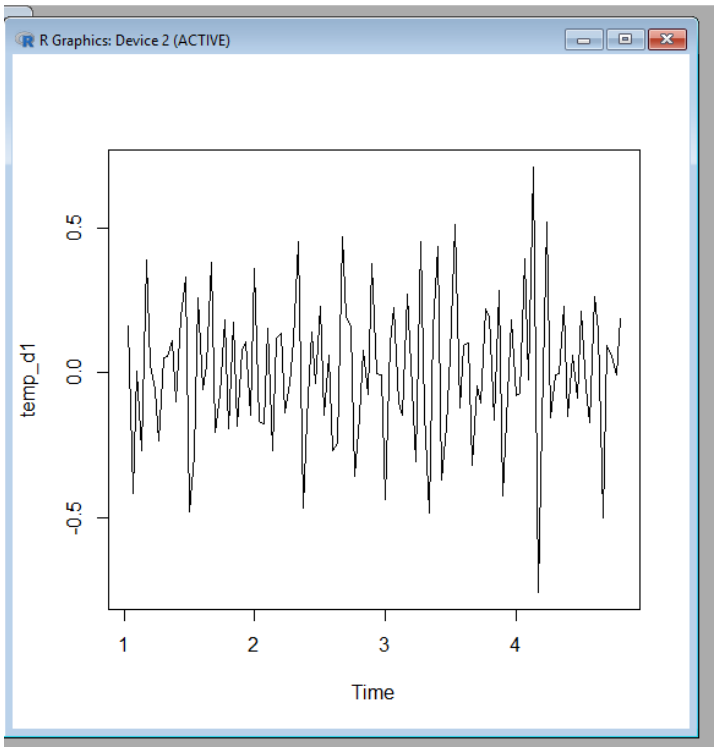


Both the plots agree with the result of adf test.

So, we need to stationarize the series.

First we try with differentiation of first order.

```
> temp_dl = diff(deseasonal_temp, differences = 1)
> plot(temp_dl)
>
```



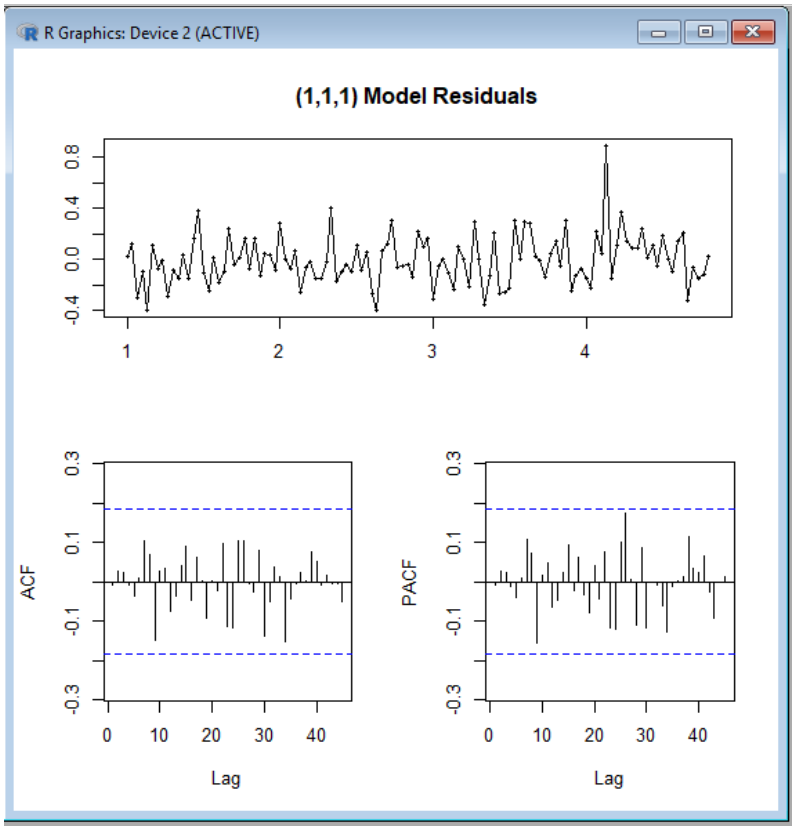

```
  could not find function "adf.test"
> tseries::adf.test(temp_d1, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  temp_d1
Dickey-Fuller = -8.4299, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In tseries::adf.test(temp_d1, alternative = "stationary") :
  p-value smaller than printed p-value
>
```

So, we fit the ARIMA model

```
> auto.arima(deseasonal_temp, seasonal=FALSE)
Series: deseasonal_temp
ARIMA(3,1,3) with drift

Coefficients:
         ar1     ar2      ar3      ma1      ma2     ma3   drift
      0.0984  0.3304  -0.2016  -0.7859  -0.6717  0.6140  0.0075
s.e.  0.3166  0.2566   0.1765   0.3017   0.3545  0.1844  0.0038

sigma^2 estimated as 0.04074:  log likelihood=23.41
AIC=-30.81   AICc=-29.44   BIC=-8.92
>
```

Now we fitted a model. But to trust it, it should be evaluated.

```
> fit<-auto.arima(deseasonal_temp, seasonal=FALSE)
> tsdisplay(residuals(fit), lag.max=45, main='(1,1,1) Model Residuals')
>
```



The residuals look almost random.so we keep this model.

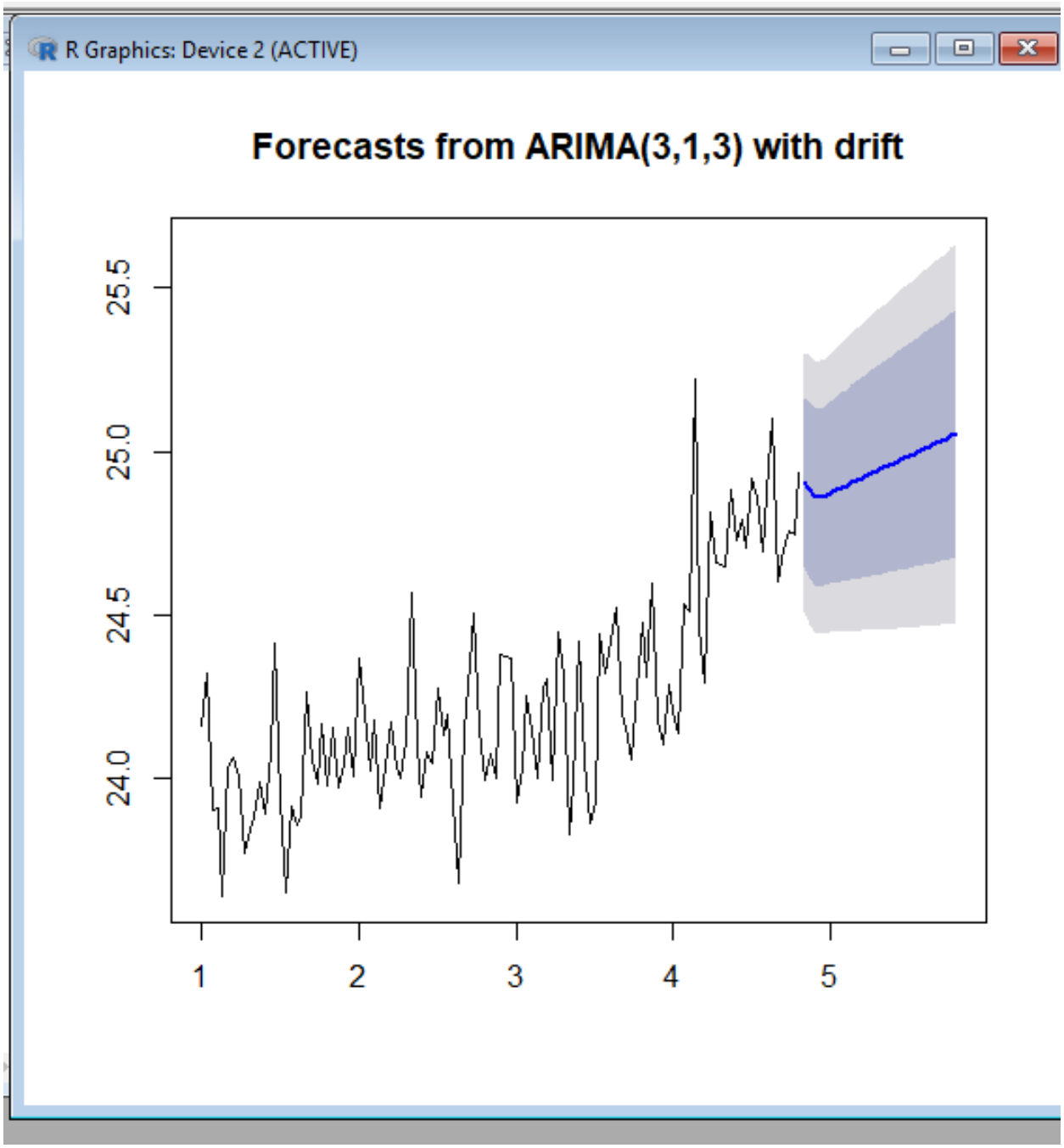
Predict from the model:

RGui (64-bit)

File History Resize Windows

R Console

```
Error in adf.test(temp_dl, alternative = "stationary") :
  could not find function "adf.test"
> tseries::adf.test(temp_dl, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  temp_dl
Dickey-Fuller = -8.4299, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In tseries::adf.test(temp_dl, alternative = "stationary") :
  p-value smaller than printed p-value
> auto.arima(deseasonal_temp, seasonal=FALSE)
Series: deseasonal_temp
ARIMA(3,1,3) with drift

Coefficients:
         ar1     ar2      ar3      ma1      ma2     ma3   drift
      0.0984  0.3304  -0.2016  -0.7859  -0.6717  0.6140  0.0075
s.e.  0.3166  0.2566   0.1765   0.3017   0.3545  0.1844  0.0038

sigma^2 estimated as 0.04074:  log likelihood=23.41
AIC=-30.81   AICc=-29.44   BIC=-8.92
> fit<-auto.arima(deseasonal_temp, seasonal=FALSE)
> tsdisplay(residuals(fit), lag.max=45, main='(1,1,1) Model Residuals')
> fcast <- forecast(fit2, h=30)
Error in forecast(fit2, h = 30) : object 'fit2' not found
> fcast <- forecast(fit, h=30)
> plot(fcast)
>
```


R Graphics: Device 2 (ACTIVE)
Forecasts from ARIMA(3,1,3) with drift
25.5
25.0
24.5
24.0
1
2
3
4
5