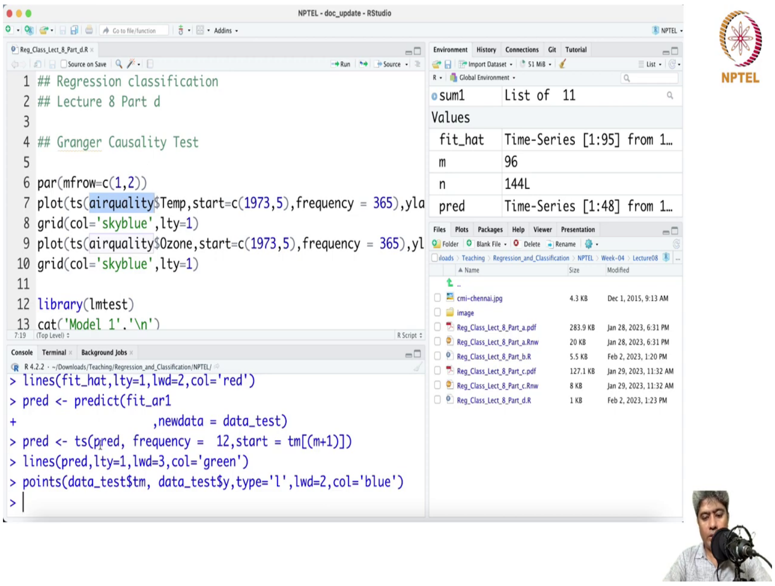
Predictive Analytics - Regression and Classification
Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 29
Hands on with R Part - 7

Hello all. Welcome back to the last part of video lecture series of 8. Now, we are going to do the Hands-on for Granger causality test.

(Refer Slide Time: 00:29)



```r
## Regression classification
## Lecture 8 Part d

## Granger Causality Test

par(mfrow=c(1,2))
plot(ts(airquality$Temp,start=c(1973,5),frequency = 365),yla
grid(col='skyblue',lty=1)
plot(ts(airquality$Ozone,start=c(1973,5),frequency = 365),yl
grid(col='skyblue',lty=1)

library(lmtest)
cat('Model 1','\n')
```

```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
> lines(fit_hat,lty=1,lwd=2,col='red')
> pred <- predict(fit_ar1
+                        ,newdata = data_test)
> pred <- ts(pred, frequency =  12,start = tm[(m+1)])
> lines(pred,lty=1,lwd=3,col='green')
> points(data_test$tm, data_test$y,type='l',lwd=2,col='blue')
> |
```

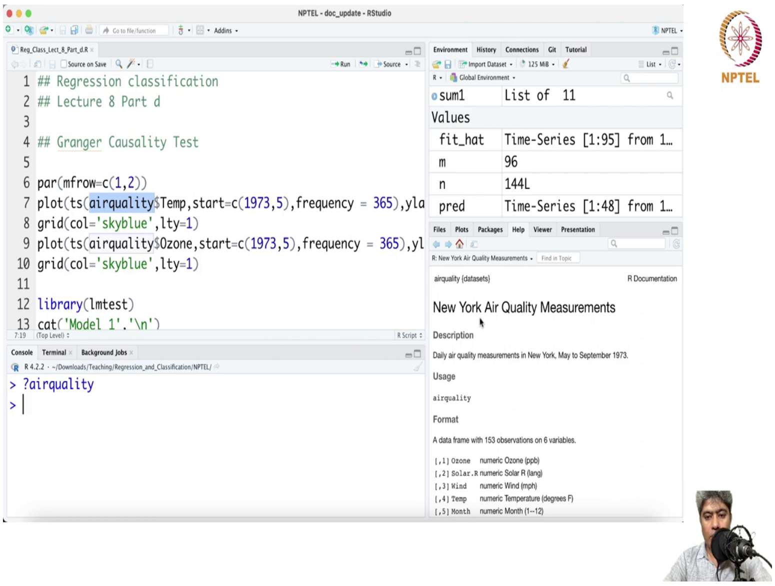
sum1	List of 11
Values	
fit_hat	Time-Series [1:95] from 1…
m	96
n	144L
pred	Time-Series [1:48] from 1…

Files
Plots
Packages
Help
Viewer
Presentation
Folder
Blank File
Delete
Rename
loads › Teaching › Regression_and_Classification › NPTEL › Week-04 › Lecture08

Name	Size	Modified
..		
cmi-chennai.jpg	4.3 KB	Dec 1, 2015, 9:13 AM
image		
Reg_Class_Lect_8_Part_a.pdf	283.9 KB	Jan 28, 2023, 6:31 PM
Reg_Class_Lect_8_Part_a.Rnw	20 KB	Jan 28, 2023, 6:31 PM
Reg_Class_Lect_8_Part_b.R	5.5 KB	Feb 2, 2023, 1:20 PM
Reg_Class_Lect_8_Part_c.pdf	127.1 KB	Jan 29, 2023, 11:32 AM
Reg_Class_Lect_8_Part_c.Rnw	8 KB	Jan 29, 2023, 11:32 AM
Reg_Class_Lect_8_Part_d.R	1 KB	Feb 2, 2023, 1:23 PM

So, first we will consider this airquality dataset.

(Refer Slide Time: 00:45)



```
## Regression classification
## Lecture 8 Part d

## Granger Causality Test

par(mfrow=c(1,2))
plot(ts(airquality$Temp,start=c(1973,5),frequency = 365),yla
grid(col='skyblue',lty=1)
plot(ts(airquality$Ozone,start=c(1973,5),frequency = 365),yl
grid(col='skyblue',lty=1)

library(lmtest)
cat('Model 1','\n')
```

```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
> ?airquality
> |
```

sum1	List of 11
Values	
fit_hat	Time-Series [1:95] from 1…
m	96
n	144L
pred	Time-Series [1:48] from 1…

airquality {datasets} R Documentation

New York Air Quality Measurements

Description

Daily air quality measurements in New York, May to September 1973.

Usage

```
airquality
```

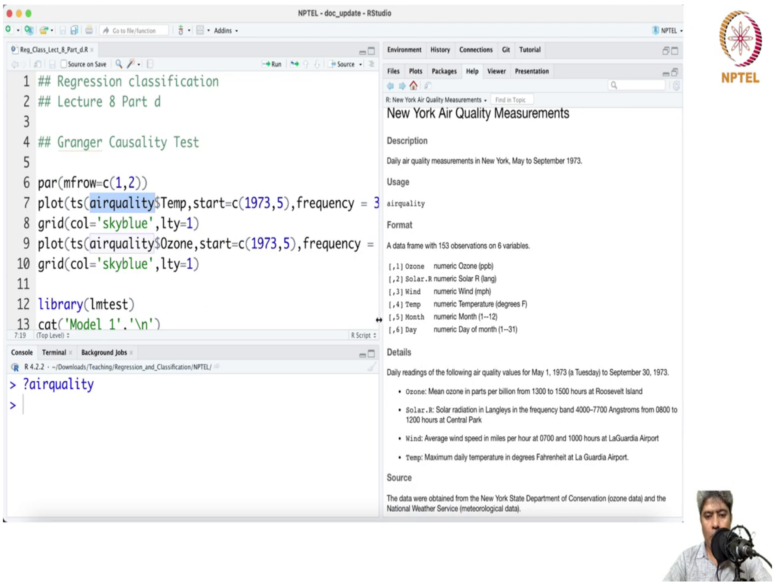
Format

A data frame with 153 observations on 6 variables.

```
[,1] Ozone   numeric Ozone (ppb)
[,2] Solar.R numeric Solar R (lang)
[,3] Wind    numeric Wind (mph)
[,4] Temp    numeric Temperature (degrees F)
[,5] Month   numeric Month (1--12)
```

Say this test comes with the basic datasets package.

(Refer Slide Time: 00:51)



NPTEL - doc_update - RStudio

Reg_Class_Lect_8_Part_d.R

```r
## Regression classification
## Lecture 8 Part d

## Granger Causality Test

par(mfrow=c(1,2))
plot(ts(airquality$Temp,start=c(1973,5),frequency = 3
grid(col='skyblue',lty=1)
plot(ts(airquality$Ozone,start=c(1973,5),frequency =
grid(col='skyblue',lty=1)

library(lmtest)
cat('Model 1','\n')
```

```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
> ?airquality
>
```

R: New York Air Quality Measurements

New York Air Quality Measurements

Description

Daily air quality measurements in New York, May to September 1973.

Usage

`airquality`

Format

A data frame with 153 observations on 6 variables.

```
[,1] Ozone    numeric Ozone (ppb)
[,2] Solar.R  numeric Solar R (lang)
[,3] Wind     numeric Wind (mph)
[,4] Temp     numeric Temperature (degrees F)
[,5] Month    numeric Month (1--12)
[,6] Day      numeric Day of month (1--31)
```

Details

Daily readings of the following air quality values for May 1, 1973 (a Tuesday) to September 30, 1973.

- Ozone: Mean ozone in parts per billion from 1300 to 1500 hours at Roosevelt Island
- Solar.R: Solar radiation in Langleys in the frequency band 4000–7700 Angstroms from 0800 to 1200 hours at Central Park
- Wind: Average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport
- Temp: Maximum daily temperature in degrees Fahrenheit at La Guardia Airport.

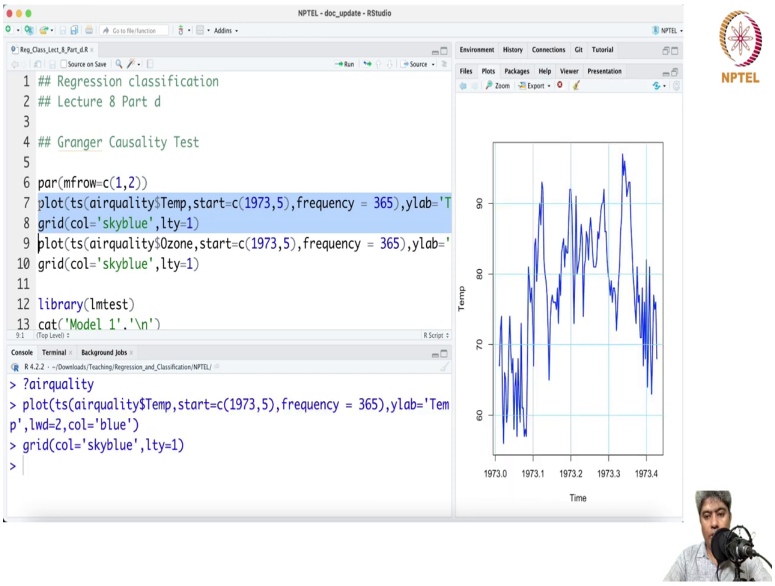
Source

The data were obtained from the New York State Department of Conservation (ozone data) and the National Weather Service (meteorological data).

And in the air, it is if you look into the description of the dataset its New York Air Quality Measurement. So, Daily Air Quality Measurements in New York between May and September of 1973. So, data its a data frame it has 153 observation on 6 variables ozone, solar, wind temperature, month and day.

Now, we are only focusing on temperature and ozone, but you can try with the other variable as well. So, the first thing we are going to plot is temperature dataset.

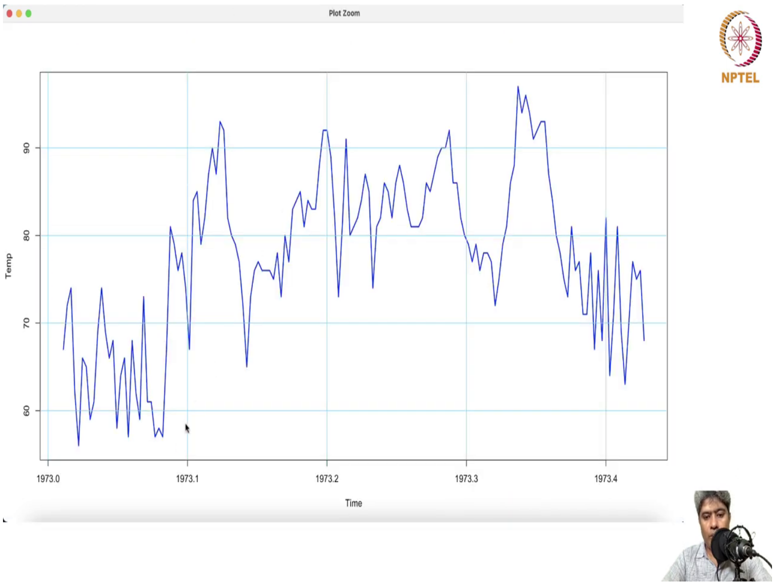
(Refer Slide Time: 01:42)



```
## Regression classification
## Lecture 8 Part d

## Granger Causality Test

par(mfrow=c(1,2))
plot(ts(airquality$Temp,start=c(1973,5),frequency = 365),ylab='T
grid(col='skyblue',lty=1)
plot(ts(airquality$Ozone,start=c(1973,5),frequency = 365),ylab='
grid(col='skyblue',lty=1)

library(lmtest)
cat('Model 1','\n')
```

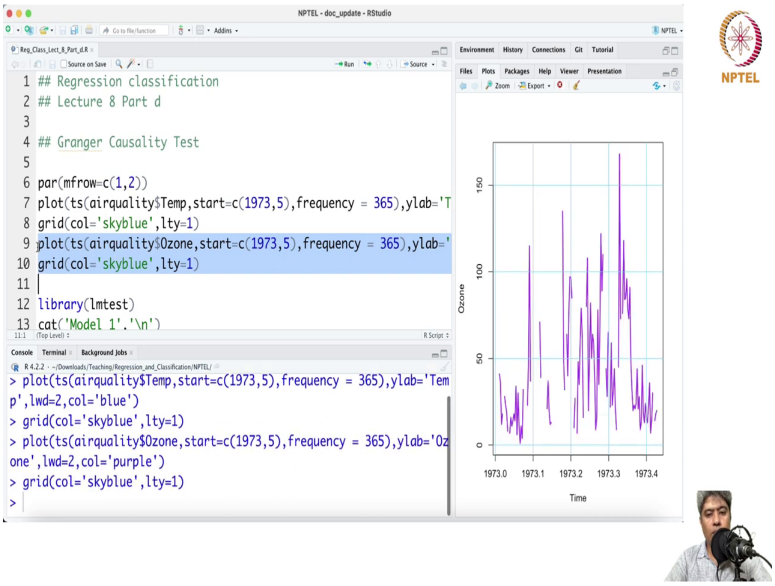
```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
> ?airquality
> plot(ts(airquality$Temp,start=c(1973,5),frequency = 365),ylab='Tem
p',lwd=2,col='blue')
> grid(col='skyblue',lty=1)
>
```

(Refer Slide Time: 01:46)



That is how like temperature dataset has.

(Refer Slide Time: 01:52)

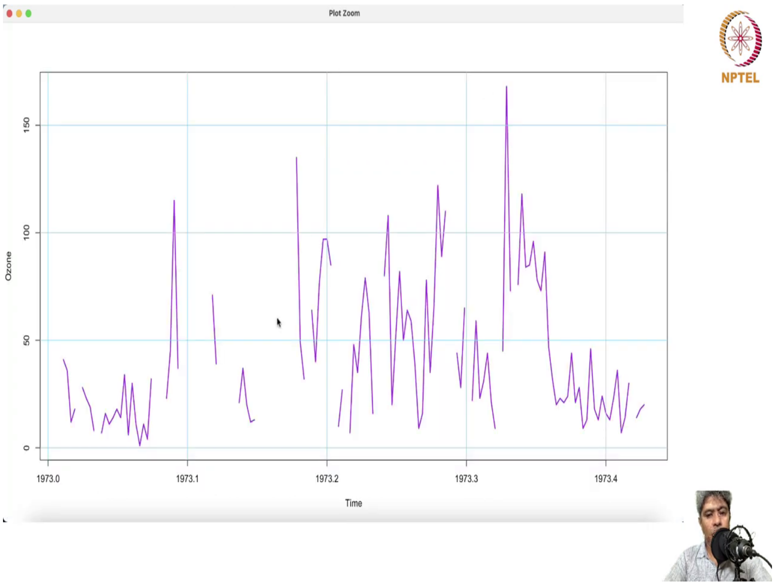


```
## Regression classification
## Lecture 8 Part d

## Granger Causality Test

par(mfrow=c(1,2))
plot(ts(airquality$Temp,start=c(1973,5),frequency = 365),ylab='T
grid(col='skyblue',lty=1)
plot(ts(airquality$Ozone,start=c(1973,5),frequency = 365),ylab='
grid(col='skyblue',lty=1)

library(lmtest)
cat('Model 1','\n')
```

```
> plot(ts(airquality$Temp,start=c(1973,5),frequency = 365),ylab='Tem
p',lwd=2,col='blue')
> grid(col='skyblue',lty=1)
> plot(ts(airquality$Ozone,start=c(1973,5),frequency = 365),ylab='Oz
one',lwd=2,col='purple')
> grid(col='skyblue',lty=1)
>
```

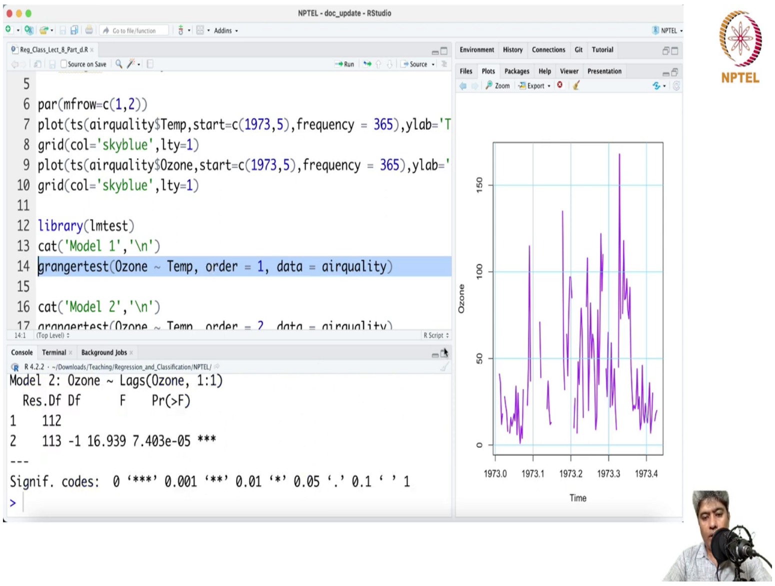
This is the ozone dataset as we have seen.

(Refer Slide Time: 01:55)



Looks like there are some missing values are also there in the ozone time series.

(Refer Slide Time: 02:00)

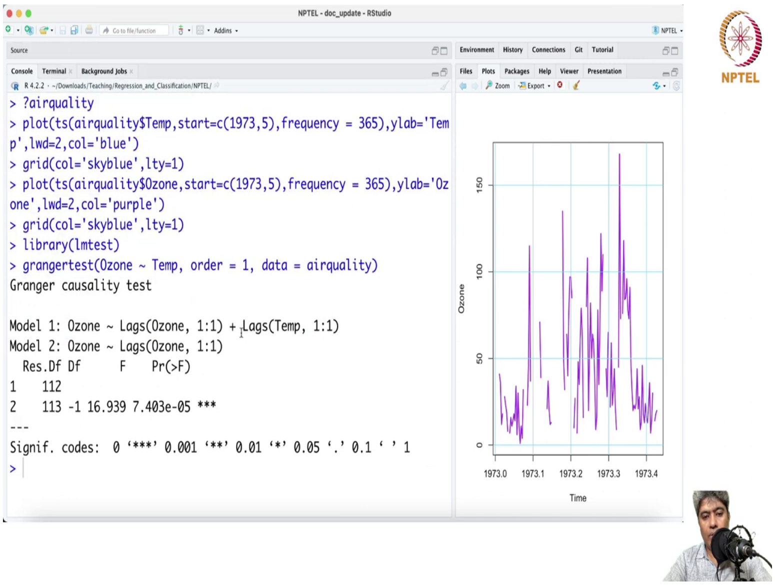


```r

par(mfrow=c(1,2))
plot(ts(airquality$Temp,start=c(1973,5),frequency = 365),ylab='T
grid(col='skyblue',lty=1)
plot(ts(airquality$Ozone,start=c(1973,5),frequency = 365),ylab='
grid(col='skyblue',lty=1)

library(lmtest)
cat('Model 1','\n')
grangertest(Ozone ~ Temp, order = 1, data = airquality)

cat('Model 2','\n')
grangertest(Ozone ~ Temp, order = 2, data = airquality)
```

```
Model 2: Ozone ~ Lags(Ozone, 1:1)
  Res.Df Df      F    Pr(>F)
1    112
2    113 -1 16.939 7.403e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

So, the first thing we are running here is in from the LM test. The first model that we are trying to fit is the grangertest where Ozone is a function of Temperature and with only lag 1 and if we run this.

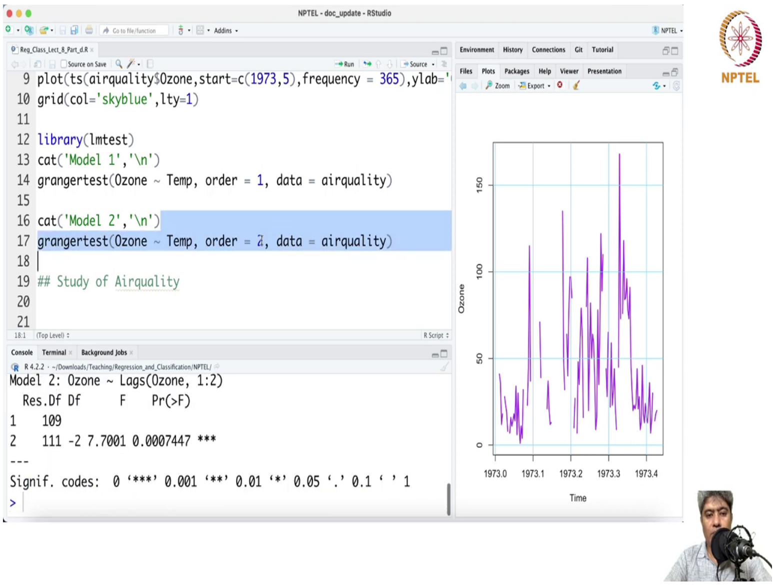
(Refer Slide Time: 02:25)



```
> ?airquality
> plot(ts(airquality$Temp,start=c(1973,5),frequency = 365),ylab='Tem
p',lwd=2,col='blue')
> grid(col='skyblue',lty=1)
> plot(ts(airquality$Ozone,start=c(1973,5),frequency = 365),ylab='Oz
one',lwd=2,col='purple')
> grid(col='skyblue',lty=1)
> library(lmtest)
> grangertest(Ozone ~ Temp, order = 1, data = airquality)
Granger causality test

Model 1: Ozone ~ Lags(Ozone, 1:1) + Lags(Temp, 1:1)
Model 2: Ozone ~ Lags(Ozone, 1:1)
  Res.Df Df      F    Pr(>F)
1    112
2    113 -1 16.939 7.403e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

And data is over air quality we run this the P value this is the F test run. So, the Model 1 is Ozone with Lag 1 Ozone and Lag 1 Temperature and then the second model is only the null model and then it did a test whether the lag model has effect or not and then it says that ok the null model is not right.

(Refer Slide Time: 02:52)

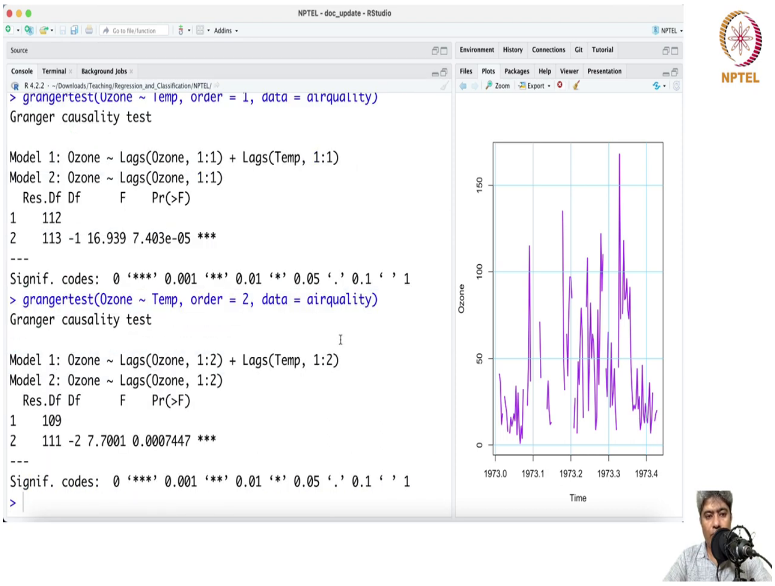


```
plot(ts(airquality$Ozone,start=c(1973,5),frequency = 365),ylab='
grid(col='skyblue',lty=1)

library(lmtest)
cat('Model 1','\n')
grangertest(Ozone ~ Temp, order = 1, data = airquality)

cat('Model 2','\n')
grangertest(Ozone ~ Temp, order = 2, data = airquality)

## Study of Airquality
```

```
Model 2: Ozone ~ Lags(Ozone, 1:2)
  Res.Df Df      F    Pr(>F)
1    109
2    111 -2 7.7001 0.0007447 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

So, that means, temperature does have a effect on the Ozone. Similarly, we run the second model and in this ozone again as a function of Temperature, but now we are going up to the second order of the granger causal test model.

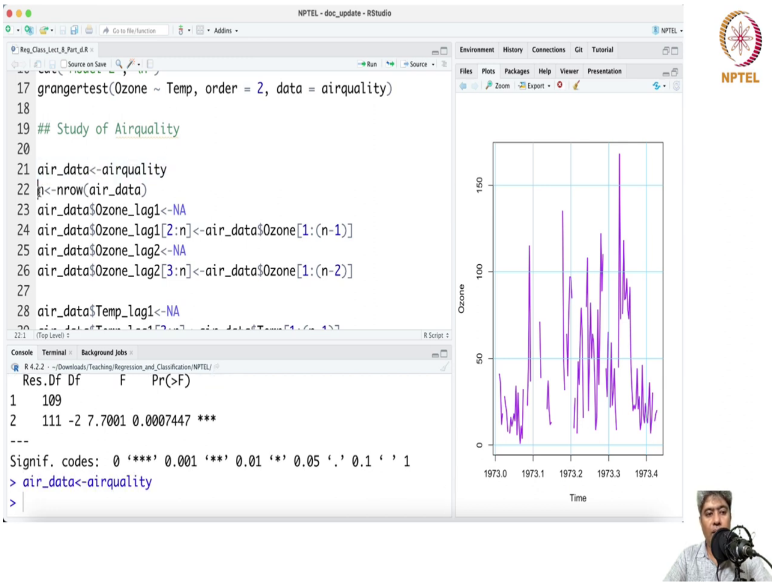
(Refer Slide Time: 03:15)



```
> grangertest(Ozone ~ Temp, order = 1, data = airquality)
Granger causality test

Model 1: Ozone ~ Lags(Ozone, 1:1) + Lags(Temp, 1:1)
Model 2: Ozone ~ Lags(Ozone, 1:1)
  Res.Df Df      F    Pr(>F)
1    112
2    113 -1 16.939 7.403e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> grangertest(Ozone ~ Temp, order = 2, data = airquality)
Granger causality test

Model 1: Ozone ~ Lags(Ozone, 1:2) + Lags(Temp, 1:2)
Model 2: Ozone ~ Lags(Ozone, 1:2)
  Res.Df Df      F    Pr(>F)
1    109
2    111 -2 7.7001 0.0007447 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Its essentially autoregressive model essentially its autoregressive model and P value is still small. So, we can say that temperature does have effect on ozone.

(Refer Slide Time: 03:31)

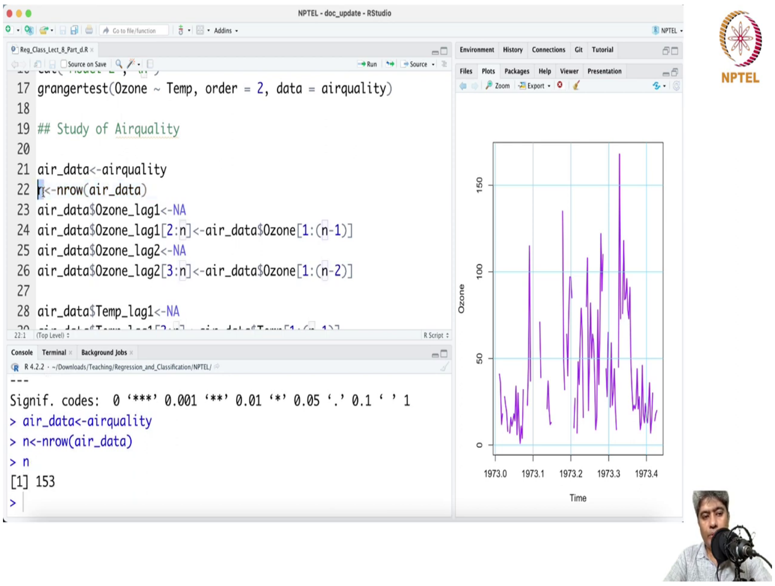


```
grangertest(Ozone ~ Temp, order = 2, data = airquality)

## Study of Airquality

air_data<-airquality
n<-nrow(air_data)
air_data$Ozone_lag1<-NA
air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
air_data$Ozone_lag2<-NA
air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]

air_data$Temp_lag1<-NA
```

```
  Res.Df Df      F    Pr(>F)
1    109
2    111 -2 7.7001 0.0007447 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> air_data<-airquality
> 
```

So, in order to understand a little bit more how this whole thing is working. So, we take the air quality data.

(Refer Slide Time: 03:43)

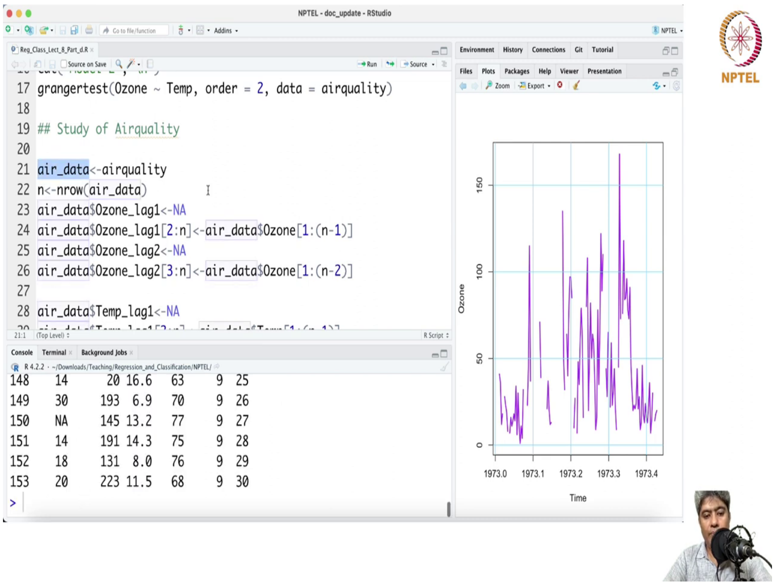


```r
grangertest(Ozone ~ Temp, order = 2, data = airquality)

## Study of Airquality

air_data<-airquality
n<-nrow(air_data)
air_data$Ozone_lag1<-NA
air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
air_data$Ozone_lag2<-NA
air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]

air_data$Temp_lag1<-NA
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> air_data<-airquality
> n<-nrow(air_data)
> n
[1] 153
>
```

n is the number of samples. First what we and now if you look into the let us look into the air_data.

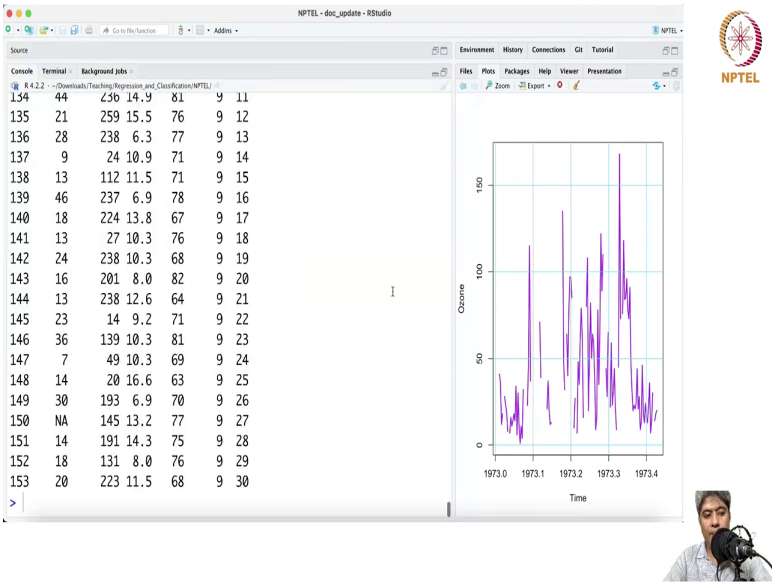
(Refer Slide Time: 03:53)



```r
grangertest(Ozone ~ Temp, order = 2, data = airquality)

## Study of Airquality

air_data<-airquality
n<-nrow(air_data)
air_data$Ozone_lag1<-NA
air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
air_data$Ozone_lag2<-NA
air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]

air_data$Temp_lag1<-NA
```

```
148    14      20 16.6   63     9  25
149    30     193  6.9   70     9  26
150    NA     145 13.2   77     9  27
151    14     191 14.3   75     9  28
152    18     131  8.0   76     9  29
153    20     223 11.5   68     9  30
>
```

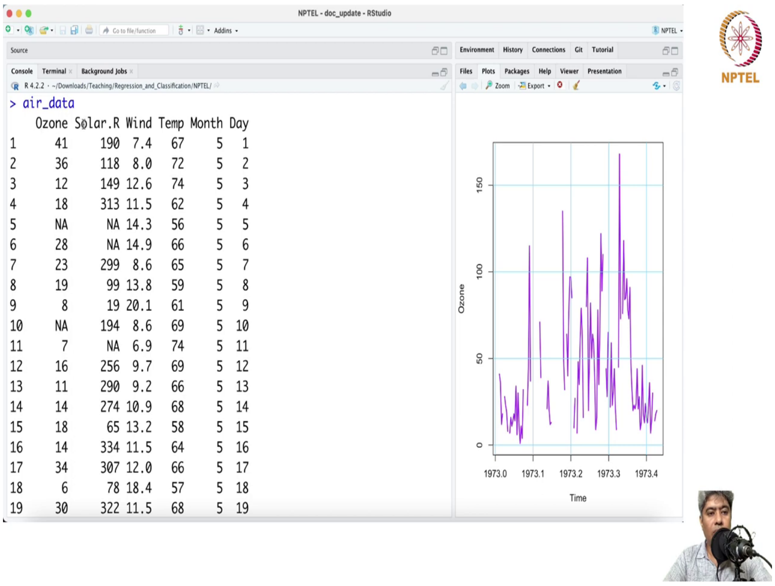
(Refer Slide Time: 03:56)



```
134    44       236 14.9   81      9  11
135    21       259 15.5   76      9  12
136    28       238  6.3   77      9  13
137     9        24 10.9   71      9  14
138    13       112 11.5   71      9  15
139    46       237  6.9   78      9  16
140    18       224 13.8   67      9  17
141    13        27 10.3   76      9  18
142    24       238 10.3   68      9  19
143    16       201  8.0   82      9  20
144    13       238 12.6   64      9  21
145    23        14  9.2   71      9  22
146    36       139 10.3   81      9  23
147     7        49 10.3   69      9  24
148    14        20 16.6   63      9  25
149    30       193  6.9   70      9  26
150    NA       145 13.2   77      9  27
151    14       191 14.3   75      9  28
152    18       131  8.0   76      9  29
153    20       223 11.5   68      9  30
> 
```

So, that is how the data looks like, ok.

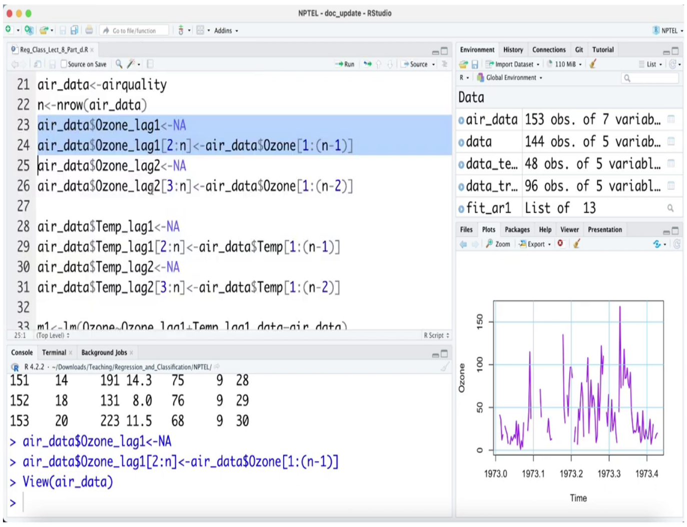
(Refer Slide Time: 04:01)



```
> air_data
   Ozone Solar.R Wind Temp Month Day
1     41     190  7.4   67     5   1
2     36     118  8.0   72     5   2
3     12     149 12.6   74     5   3
4     18     313 11.5   62     5   4
5     NA      NA 14.3   56     5   5
6     28      NA 14.9   66     5   6
7     23     299  8.6   65     5   7
8     19      99 13.8   59     5   8
9      8      19 20.1   61     5   9
10    NA     194  8.6   69     5  10
11     7      NA  6.9   74     5  11
12    16     256  9.7   69     5  12
13    11     290  9.2   66     5  13
14    14     274 10.9   68     5  14
15    18      65 13.2   58     5  15
16    14     334 11.5   64     5  16
17    34     307 12.0   66     5  17
18     6      78 18.4   57     5  18
19    30     322 11.5   68     5  19
```

That is how the Ozone Solar Solar.R radiation Wind Solar.R stands for radiation Wind Temperature Month and Day. Now, there are some NA observations are available.

(Refer Slide Time: 04:25)



```r
air_data<-airquality
n<-nrow(air_data)
air_data$Ozone_lag1<-NA
air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
air_data$Ozone_lag2<-NA
air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]

air_data$Temp_lag1<-NA
air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
air_data$Temp_lag2<-NA
air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]

m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
```

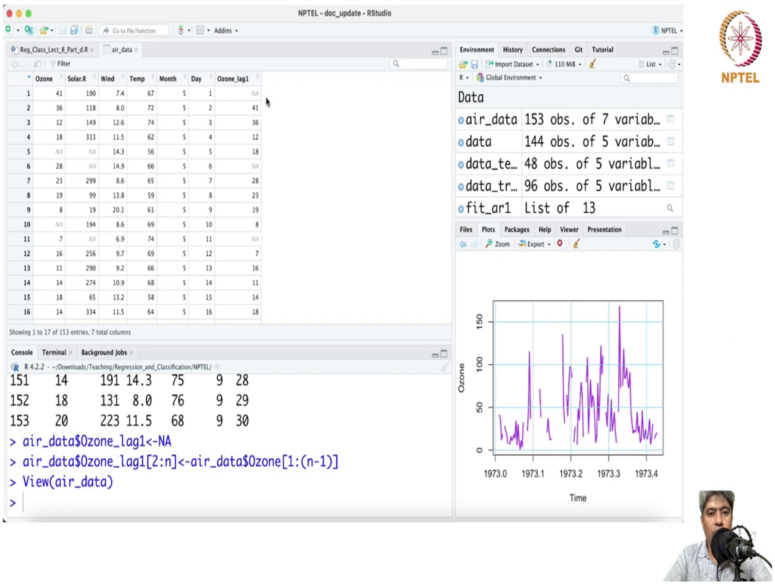
Data

air_data	153 obs. of 7 variab…
data	144 obs. of 5 variab…
data_te…	48 obs. of 5 variabl…
data_tr…	96 obs. of 5 variabl…
fit_ar1	List of 13

```
151     14      191 14.3    75      9  28
152     18      131  8.0    76      9  29
153     20      223 11.5    68      9  30
> air_data$Ozone_lag1<-NA
> air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
> View(air_data)
> 
```

So, it is always bit difficult how you do this you know imputation, but for the time being we are not handling the missing data. We are only going to use the data which is fully available to us, but for the time being. So, we suppose we want to fit the first model we create the lag data set ok.

(Refer Slide Time: 04:47)



	Ozone	Solar.R	Wind	Temp	Month	Day	Ozone_lag1
1	41	190	7.4	67	5	1	NA
2	36	118	8.0	72	5	2	41
3	12	149	12.6	74	5	3	36
4	18	313	11.5	62	5	4	12
5	NA	NA	14.3	56	5	5	18
6	28	NA	14.9	66	5	6	NA
7	23	299	8.6	65	5	7	28
8	19	99	13.8	59	5	8	23
9	8	19	20.1	61	5	9	19
10	NA	194	8.6	69	5	10	8
11	7	NA	6.9	74	5	11	NA
12	16	256	9.7	69	5	12	7
13	11	290	9.2	66	5	13	16
14	14	274	10.9	68	5	14	11
15	18	65	13.2	58	5	15	14
16	14	334	11.5	64	5	16	18

Showing 1 to 17 of 153 entries, 7 total columns

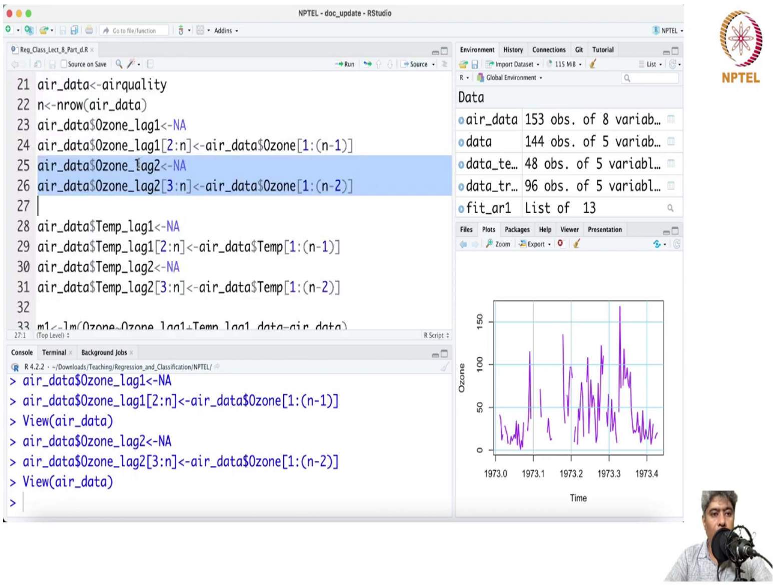
```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
151     14      191 14.3    75      9  28
152     18      131  8.0    76      9  29
153     20      223 11.5    68      9  30
> air_data$Ozone_lag1<-NA
> air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
> View(air_data)
>
```

Data

air_data	153 obs. of 7 variab…
data	144 obs. of 5 variab…
data_te…	48 obs. of 5 variabl…
data_tr…	96 obs. of 5 variabl…
fit_ar1	List of 13

So, here we have created the lag. So, 41 was here we just created lag 36 just brought it down by 1.

(Refer Slide Time: 04:58)



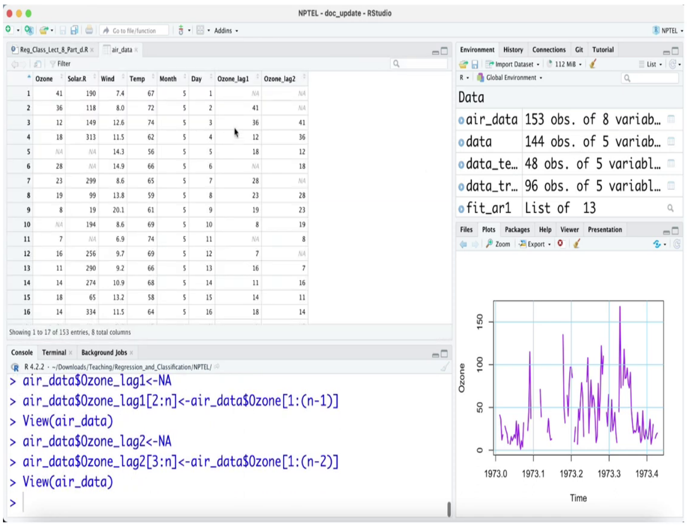
```r
air_data<-airquality
n<-nrow(air_data)
air_data$Ozone_lag1<-NA
air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
air_data$Ozone_lag2<-NA
air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]

air_data$Temp_lag1<-NA
air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
air_data$Temp_lag2<-NA
air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]

m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
```

```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
> air_data$Ozone_lag1<-NA
> air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
> View(air_data)
> air_data$Ozone_lag2<-NA
> air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]
> View(air_data)
>
```

```
Data
air_data  153 obs. of 8 variab…
data      144 obs. of 5 variab…
data_te…  48 obs. of 5 variabl…
data_tr…  96 obs. of 5 variabl…
fit_ar1   List of 13
```

Then we created the lag2 variable.

(Refer Slide Time: 05:00)



	Ozone	Solar.R	Wind	Temp	Month	Day	Ozone_lag1	Ozone_lag2
1	41	190	7.4	67	5	1	NA	NA
2	36	118	8.0	72	5	2	41	NA
3	12	149	12.6	74	5	3	36	41
4	18	313	11.5	62	5	4	12	36
5	NA	NA	14.3	56	5	5	18	12
6	28	NA	14.9	66	5	6	NA	18
7	23	299	8.6	65	5	7	28	NA
8	19	99	13.8	59	5	8	23	28
9	8	19	20.1	61	5	9	19	23
10	NA	194	8.6	69	5	10	8	19
11	7	NA	6.9	74	5	11	NA	8
12	16	256	9.7	69	5	12	7	NA
13	11	290	9.2	66	5	13	16	7
14	14	274	10.9	68	5	14	11	16
15	18	65	13.2	58	5	15	14	11
16	14	334	11.5	64	5	16	18	14

Showing 1 to 17 of 153 entries, 8 total columns

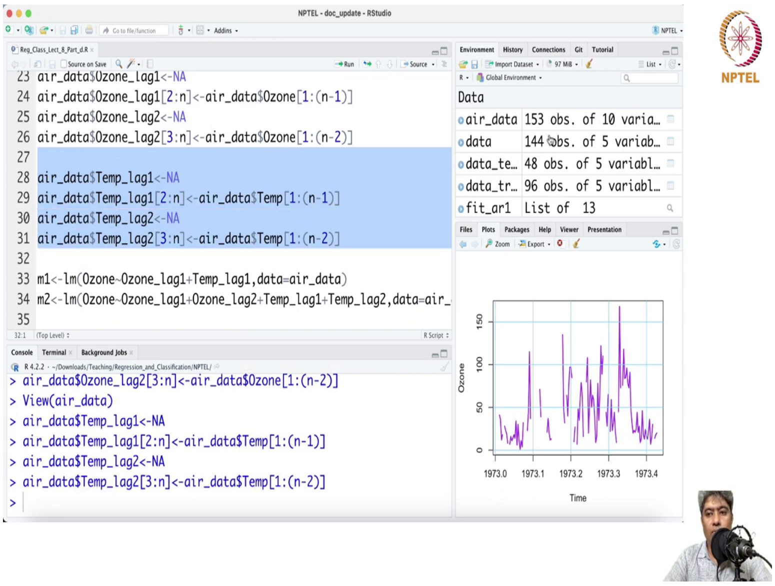
Data

air_data	153 obs. of 8 variab…
data	144 obs. of 5 variab…
data_te…	48 obs. of 5 variabl…
data_tr…	96 obs. of 5 variabl…
fit_ar1	List of 13

```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
> air_data$Ozone_lag1<-NA
> air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
> View(air_data)
> air_data$Ozone_lag2<-NA
> air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]
> View(air_data)
>
```

So, now you have see the lag2 variables have been created ok.

(Refer Slide Time: 05:07)



```
air_data$Ozone_lag1<-NA
air_data$Ozone_lag1[2:n]<-air_data$Ozone[1:(n-1)]
air_data$Ozone_lag2<-NA
air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]

air_data$Temp_lag1<-NA
air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
air_data$Temp_lag2<-NA
air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]

m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_
```

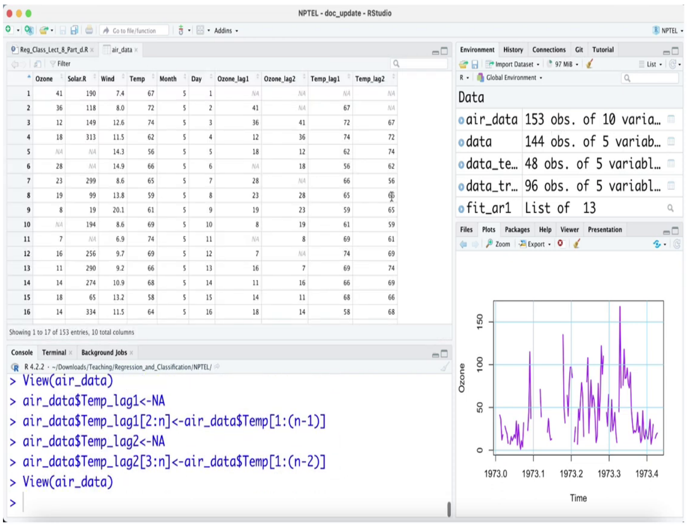
Data

air_data	153 obs. of 10 varia…
data	144 obs. of 5 variab…
data_te…	48 obs. of 5 variabl…
data_tr…	96 obs. of 5 variabl…
fit_ar1	List of 13

```
> air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]
> View(air_data)
> air_data$Temp_lag1<-NA
> air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
> air_data$Temp_lag2<-NA
> air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]
>
```

Then similarly we create the Temperature_lag1 and Temperature_lag2 variables.

(Refer Slide Time: 05:12)



NPTEL - doc_update - RStudio

Reg_Class_Lect_8_Part_d.R × | air_data ×

	Ozone	Solar.R	Wind	Temp	Month	Day	Ozone_lag1	Ozone_lag2	Temp_lag1	Temp_lag2
1	41	190	7.4	67	5	1	NA	NA	NA	NA
2	36	118	8.0	72	5	2	41	NA	67	NA
3	12	149	12.6	74	5	3	36	41	72	67
4	18	313	11.5	62	5	4	12	36	74	72
5	NA	NA	14.3	56	5	5	18	12	62	74
6	28	NA	14.9	66	5	6	NA	18	56	62
7	23	299	8.6	65	5	7	28	NA	66	56
8	19	99	13.8	59	5	8	23	28	65	66
9	8	19	20.1	61	5	9	19	23	59	65
10	NA	194	8.6	69	5	10	8	19	61	59
11	7	NA	6.9	74	5	11	NA	8	69	61
12	16	256	9.7	69	5	12	7	NA	74	69
13	11	290	9.2	66	5	13	16	7	69	74
14	14	274	10.9	68	5	14	11	16	66	69
15	18	65	13.2	58	5	15	14	11	68	66
16	14	334	11.5	64	5	16	18	14	58	68

Showing 1 to 17 of 153 entries, 10 total columns

R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```
> View(air_data)
> air_data$Temp_lag1<-NA
> air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
> air_data$Temp_lag2<-NA
> air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]
> View(air_data)
>
```

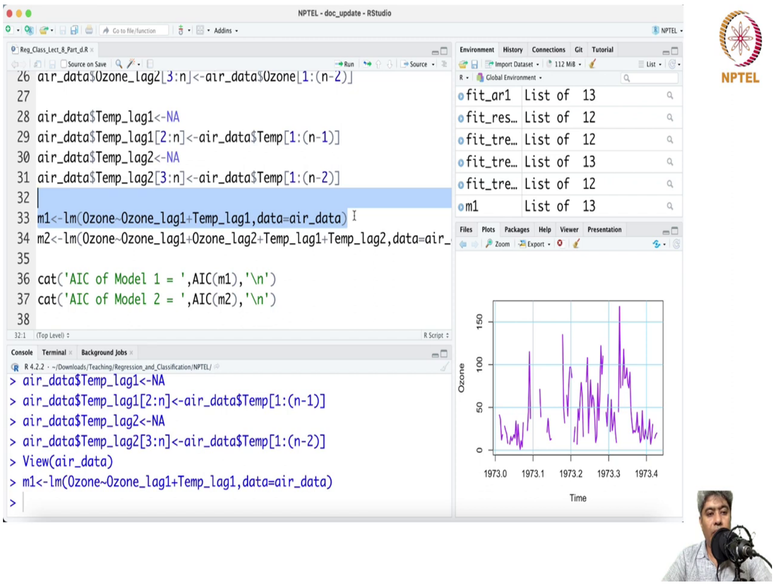
Global Environment

Data

air_data	153 obs. of 10 varia…
data	144 obs. of 5 variab…
data_te…	48 obs. of 5 variabl…
data_tr…	96 obs. of 5 variabl…
fit_ar1	List of 13

So, Temperature_lag1, Temperature_lag2 variable have been created.

(Refer Slide Time: 05:19)



NPTEL - doc_update - RStudio

Reg_Class_Lect_8_Part_d.R

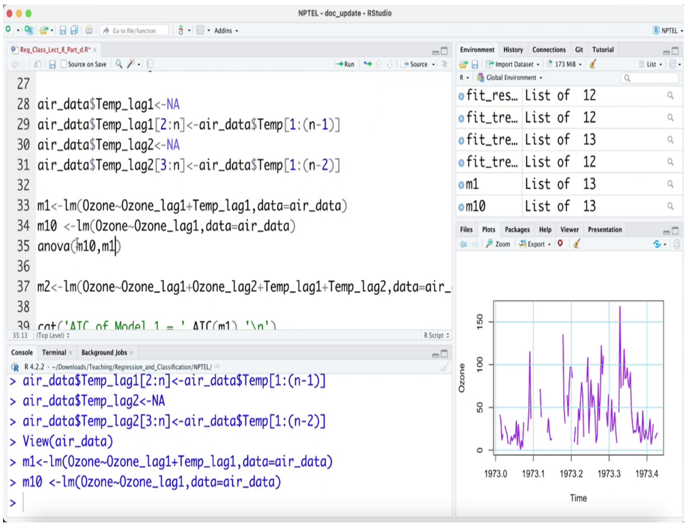
```
air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]

air_data$Temp_lag1<-NA
air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
air_data$Temp_lag2<-NA
air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]

m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_

cat('AIC of Model 1 = ',AIC(m1),'\n')
cat('AIC of Model 2 = ',AIC(m2),'\n')
```

```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
> air_data$Temp_lag1<-NA
> air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
> air_data$Temp_lag2<-NA
> air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]
> View(air_data)
> m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
> 
```

fit_ar1	List of 13
fit_res…	List of 12
fit_tre…	List of 12
fit_tre…	List of 13
fit_tre…	List of 12
m1	List of 13

Now, if you fit the model with Ozone_lag1 and Temperature _ag1 ok.

(Refer Slide Time: 05:36)

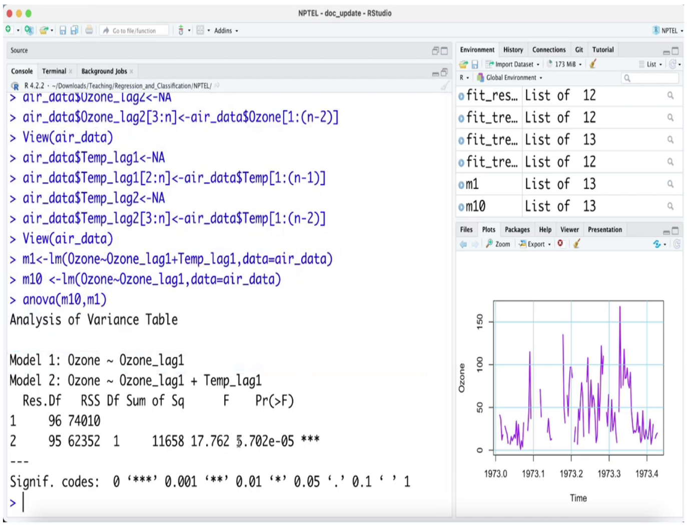


```r

air_data$Temp_lag1<-NA
air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
air_data$Temp_lag2<-NA
air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]

m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
m10 <-lm(Ozone~Ozone_lag1,data=air_data)
anova(m10,m1)

m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_

cat('AIC of Model 1 = ' AIC(m1) '\n')
```

```
> air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
> air_data$Temp_lag2<-NA
> air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]
> View(air_data)
> m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
> m10 <-lm(Ozone~Ozone_lag1,data=air_data)
> 
```

fit_res…	List of	12
fit_tre…	List of	12
fit_tre…	List of	13
fit_tre…	List of	12
m1	List of	13
m10	List of	13

(Refer Slide Time: 05:12)



NPTEL - doc_update - RStudio

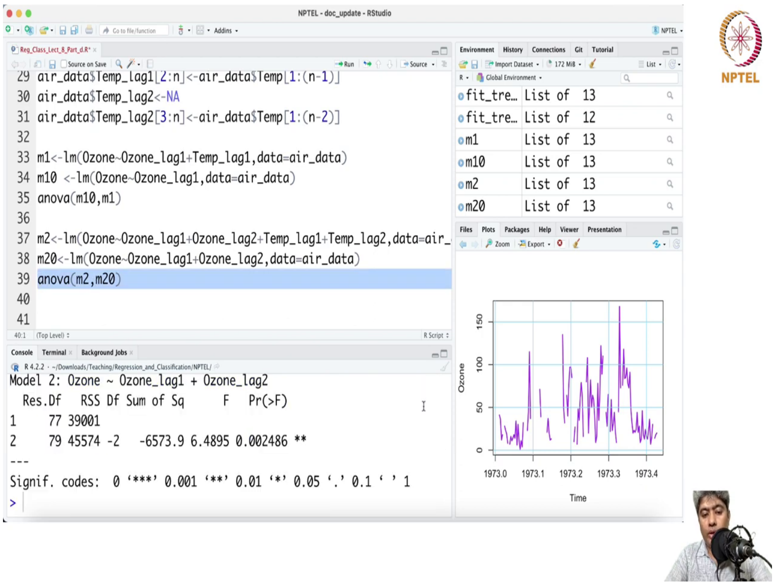
```
> air_data$Ozone_lag2<-NA
> air_data$Ozone_lag2[3:n]<-air_data$Ozone[1:(n-2)]
> View(air_data)
> air_data$Temp_lag1<-NA
> air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
> air_data$Temp_lag2<-NA
> air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]
> View(air_data)
> m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
> m10 <-lm(Ozone~Ozone_lag1,data=air_data)
> anova(m10,m1)
Analysis of Variance Table

Model 1: Ozone ~ Ozone_lag1
Model 2: Ozone ~ Ozone_lag1 + Temp_lag1
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1     96 74010
2     95 62352  1     11658 17.762 5.702e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 
```

```
fit_res… List of 12
fit_tre… List of 12
fit_tre… List of 13
fit_tre… List of 12
m1       List of 13
m10      List of 13
```

Ozone_lag1 and Temperature_lag1 and similarly m1 say for null I will use 0, but just copy this entire thing and but instead of Temperature_lag1 I am dropping the Temperature_lag1. So, I am just saying the Ozone is only function of its own ok. Now, you see anova if you just run anova between m10 versus m1 ok.

You can see that it is the F test reject the null hypothesis and it says that the lag1 temperature does have effect on Ozone. So, you can do the Granger causal test in this way as well.

(Refer Slide Time: 06:30)



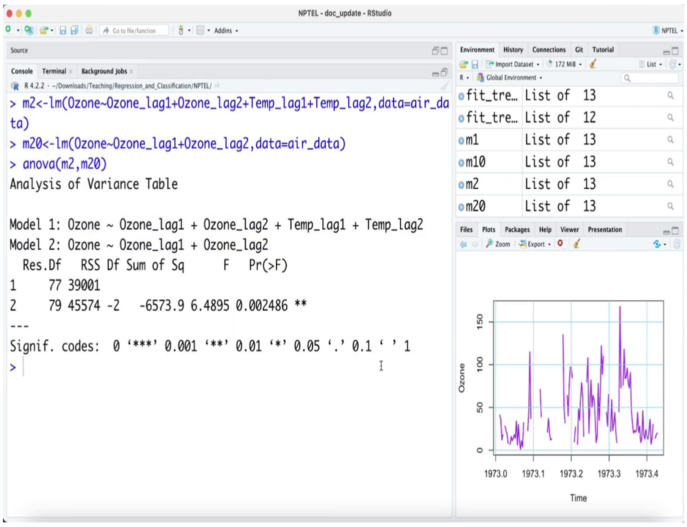
```
air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
air_data$Temp_lag2<-NA
air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]

m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
m10 <-lm(Ozone~Ozone_lag1,data=air_data)
anova(m10,m1)

m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_
m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
anova(m2,m20)
```

```
fit_tre… List of 13
fit_tre… List of 12
m1       List of 13
m10      List of 13
m2       List of 13
m20      List of 13
```

```
Model 2: Ozone ~ Ozone_lag1 + Ozone_lag2
  Res.Df   RSS Df Sum of Sq      F   Pr(>F)
1     77 39001
2     79 45574 -2   -6573.9 6.4895 0.002486 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Similarly, this is m2 this test with the 2 lag and then I can define the model with null model here, but only with the lag Ozone not with the temperature I am dropping the temperature and then we run the anova m test or F test 2 to comma m20 ok.

(Refer Slide Time: 07:11)

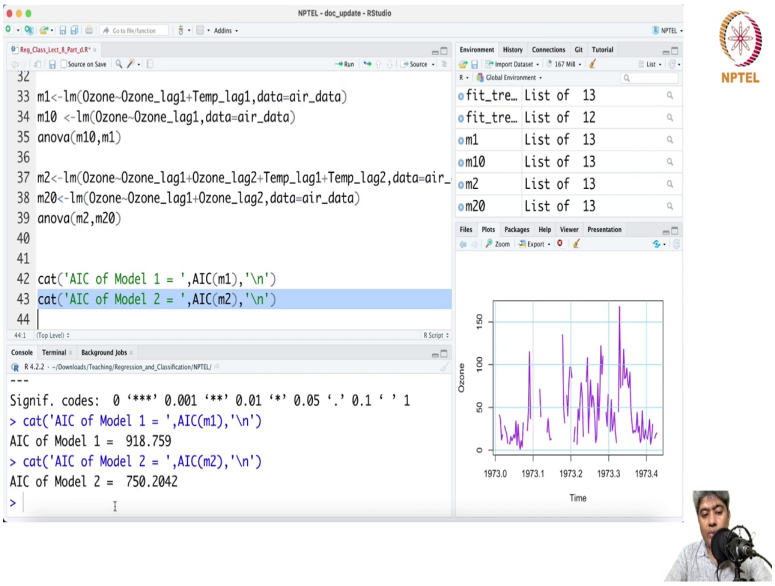


```
> m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_da
ta)
> m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
> anova(m2,m20)
Analysis of Variance Table

Model 1: Ozone ~ Ozone_lag1 + Ozone_lag2 + Temp_lag1 + Temp_lag2
Model 2: Ozone ~ Ozone_lag1 + Ozone_lag2
  Res.Df   RSS Df Sum of Sq      F   Pr(>F)
1     77 39001
2     79 45574 -2   -6573.9 6.4895 0.002486 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

fit_tre…	List of	13
fit_tre…	List of	12
m1	List of	13
m10	List of	13
m2	List of	13
m20	List of	13

And you can see you can reject the F test. So, lag2 have effect on the Ozone.

(Refer Slide Time: 07:21)



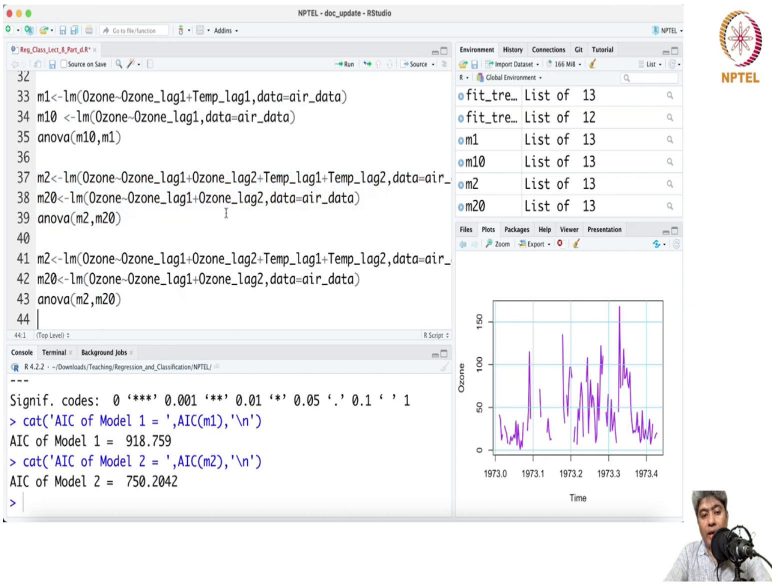
```r
m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
m10 <-lm(Ozone~Ozone_lag1,data=air_data)
anova(m10,m1)

m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_
m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
anova(m2,m20)


cat('AIC of Model 1 = ',AIC(m1),'\n')
cat('AIC of Model 2 = ',AIC(m2),'\n')
```

fit_tre…	List of	13
fit_tre…	List of	12
m1	List of	13
m10	List of	13
m2	List of	13
m20	List of	13

```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> cat('AIC of Model 1 = ',AIC(m1),'\n')
AIC of Model 1 =  918.759
> cat('AIC of Model 2 = ',AIC(m2),'\n')
AIC of Model 2 =  750.2042
>
```

Now, if I compute the AIC of Model 1 and Model 2 clearly the Model 1 has a more AIC than the Model 2 the Model 2 has a lower mic AIC. So, we can we would prefer Model 2 over the Model 1.

(Refer Slide Time: 07:51)



```
m1<-lm(Ozone~Ozone_lag1+Temp_lag1,data=air_data)
m10 <-lm(Ozone~Ozone_lag1,data=air_data)
anova(m10,m1)

m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_
m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
anova(m2,m20)

m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_
m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
anova(m2,m20)
```

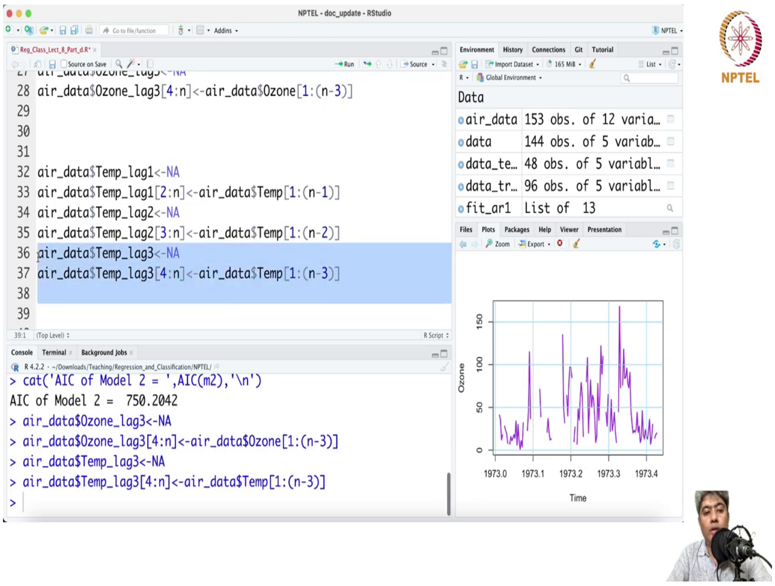
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> cat('AIC of Model 1 = ',AIC(m1),'\n')
AIC of Model 1 =  918.759
> cat('AIC of Model 2 = ',AIC(m2),'\n')
AIC of Model 2 =  750.2042
>
```

fit_tre…	List of	13
fit_tre…	List of	12
m1	List of	13
m10	List of	13
m2	List of	13
m20	List of	13

Similarly, you can go for the third variable third lag we have to create a lag here. So, we can always create these lines.

(Refer Slide Time: 08:00)



```
air_data$Ozone_lag3<-NA
air_data$Ozone_lag3[4:n]<-air_data$Ozone[1:(n-3)]



air_data$Temp_lag1<-NA
air_data$Temp_lag1[2:n]<-air_data$Temp[1:(n-1)]
air_data$Temp_lag2<-NA
air_data$Temp_lag2[3:n]<-air_data$Temp[1:(n-2)]
air_data$Temp_lag3<-NA
air_data$Temp_lag3[4:n]<-air_data$Temp[1:(n-3)]
```

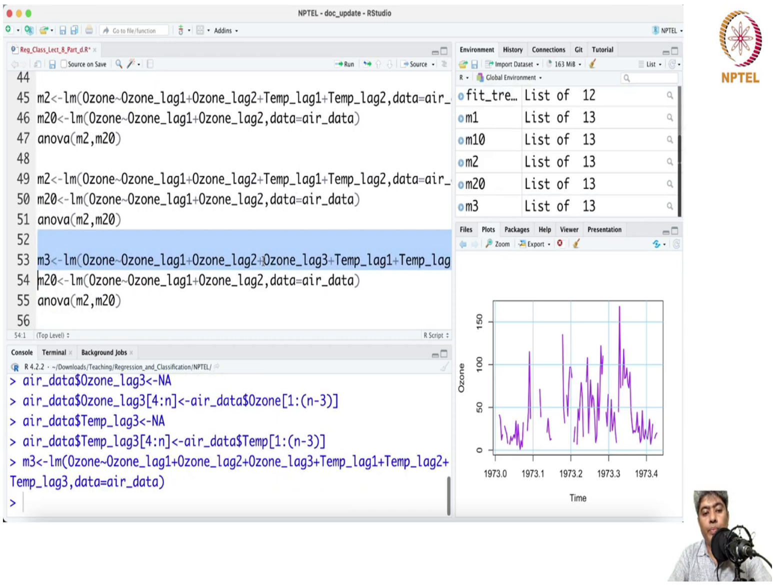
```
> cat('AIC of Model 2 = ',AIC(m2),'\n')
AIC of Model 2 =  750.2042
> air_data$Ozone_lag3<-NA
> air_data$Ozone_lag3[4:n]<-air_data$Ozone[1:(n-3)]
> air_data$Temp_lag3<-NA
> air_data$Temp_lag3[4:n]<-air_data$Temp[1:(n-3)]
>
```

Data

air_data	153 obs. of 12 varia…
data	144 obs. of 5 variab…
data_te…	48 obs. of 5 variabl…
data_tr…	96 obs. of 5 variabl…
fit_ar1	List of 13

First, we have to create these lines. So, maybe lag3, lag3, 2, 4, 3.

(Refer Slide Time: 08:26)



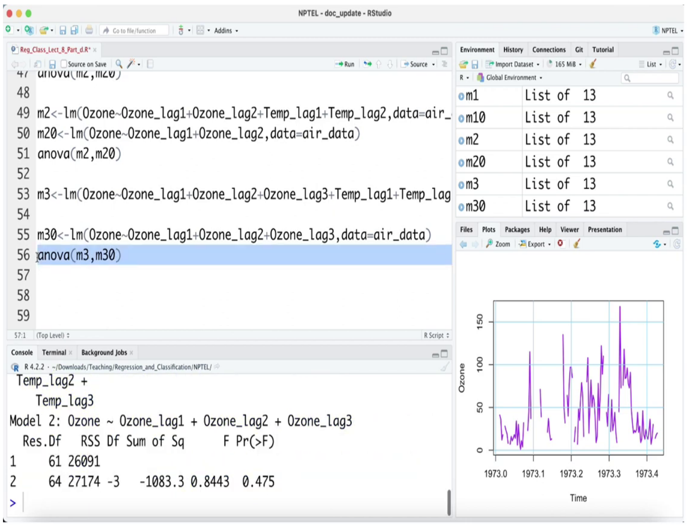
```

m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_
m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
anova(m2,m20)

m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_
m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
anova(m2,m20)

m3<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3+Temp_lag1+Temp_lag
m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
anova(m2,m20)

```

```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
> air_data$Ozone_lag3<-NA
> air_data$Ozone_lag3[4:n]<-air_data$Ozone[1:(n-3)]
> air_data$Temp_lag3<-NA
> air_data$Temp_lag3[4:n]<-air_data$Temp[1:(n-3)]
> m3<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3+Temp_lag1+Temp_lag2+
Temp_lag3,data=air_data)
>
```

fit_tre…	List of	12
m1	List of	13
m10	List of	13
m2	List of	13
m20	List of	13
m3	List of	13

So, we have to create these lags for ozone and then now we have to create these lags for temperature third lag plus 3. And then what we need is model for third lag and Ozone plus 3 lag plus 3 and ok.

(Refer Slide Time: 09:10)



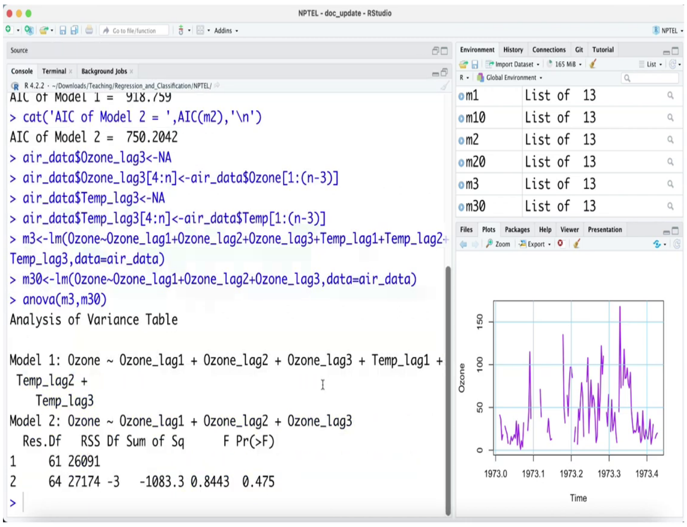
```
anova(m2,m20)

m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_
m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
anova(m2,m20)

m3<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3+Temp_lag1+Temp_lag

m30<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3,data=air_data)
anova(m3,m30)
```

```
 Temp_lag2 +
    Temp_lag3
Model 2: Ozone ~ Ozone_lag1 + Ozone_lag2 + Ozone_lag3
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1     61 26091
2     64 27174 -3   -1083.3 0.8443  0.475
>
```

m1	List of 13
m10	List of 13
m2	List of 13
m20	List of 13
m3	List of 13
m30	List of 13

And the 3 naught null we just Ozone it will be only function of Ozone and then we will see if it is still effective.

(Refer Slide Time: 09:34)

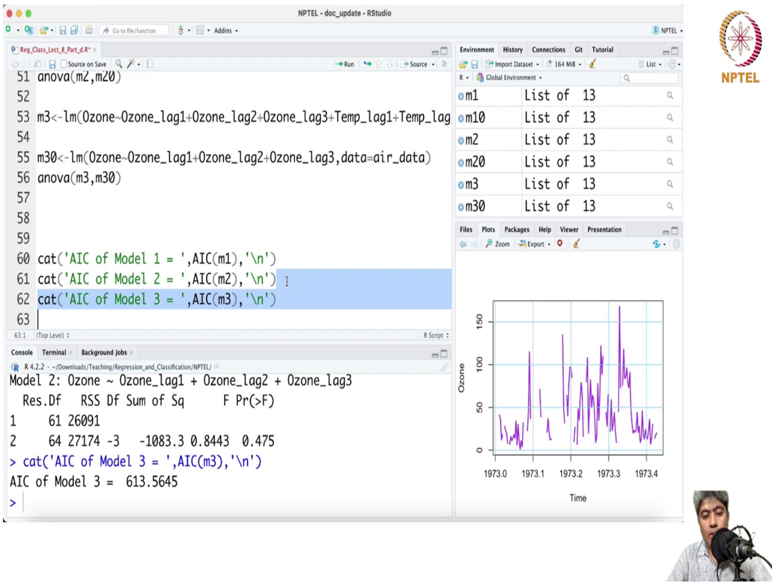


```
AIC of Model 1 =  918.759
> cat('AIC of Model 2 = ',AIC(m2),'\n')
AIC of Model 2 =  750.2042
> air_data$Ozone_lag3<-NA
> air_data$Ozone_lag3[4:n]<-air_data$Ozone[1:(n-3)]
> air_data$Temp_lag3<-NA
> air_data$Temp_lag3[4:n]<-air_data$Temp[1:(n-3)]
> m3<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3+Temp_lag1+Temp_lag2+
Temp_lag3,data=air_data)
> m30<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3,data=air_data)
> anova(m3,m30)
Analysis of Variance Table

Model 1: Ozone ~ Ozone_lag1 + Ozone_lag2 + Ozone_lag3 + Temp_lag1 +
 Temp_lag2 +
    Temp_lag3
Model 2: Ozone ~ Ozone_lag1 + Ozone_lag2 + Ozone_lag3
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1     61 26091
2     64 27174 -3   -1083.3 0.8443  0.475
>
```

m1	List of 13
m10	List of 13
m2	List of 13
m20	List of 13
m3	List of 13
m30	List of 13

And now when we are doing third lag we are adding third lag its not effective anymore.

(Refer Slide Time: 09:38)



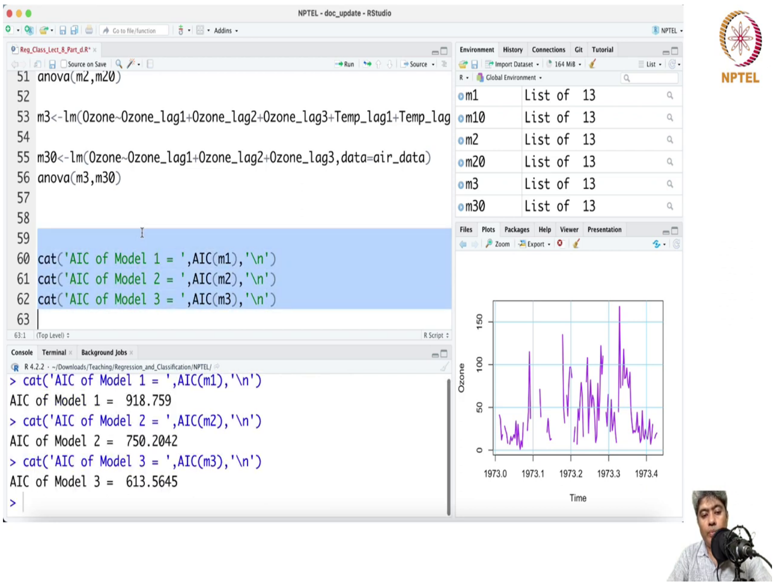
```
anova(m2,m20)

m3<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3+Temp_lag1+Temp_lag

m30<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3,data=air_data)
anova(m3,m30)



cat('AIC of Model 1 = ',AIC(m1),'\n')
cat('AIC of Model 2 = ',AIC(m2),'\n')
cat('AIC of Model 3 = ',AIC(m3),'\n')

```

```
Model 2: Ozone ~ Ozone_lag1 + Ozone_lag2 + Ozone_lag3
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1     61 26091
2     64 27174 -3   -1083.3 0.8443  0.475
> cat('AIC of Model 3 = ',AIC(m3),'\n')
AIC of Model 3 =  613.5645
>
```

m1	List of 13
m10	List of 13
m2	List of 13
m20	List of 13
m3	List of 13
m30	List of 13

But can we choose m3 what is the AIC of model third model.

(Refer Slide Time: 09:56)



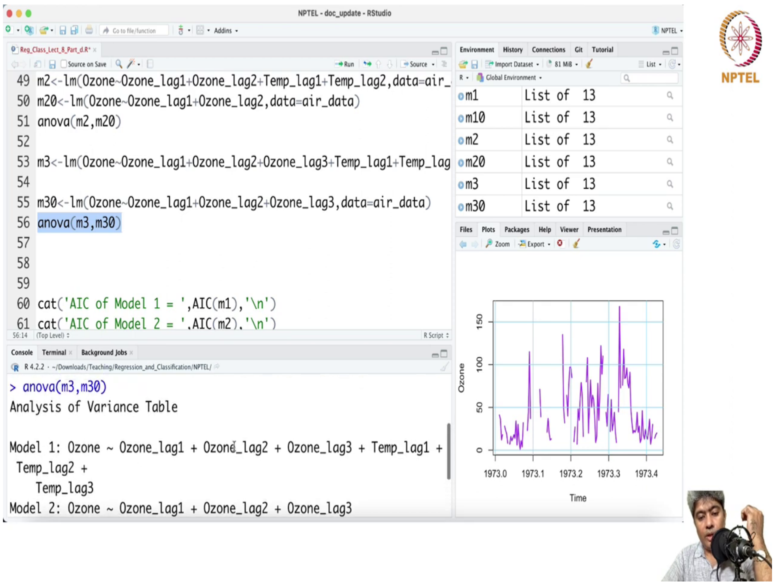
```
anova(m2,m20)

m3<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3+Temp_lag1+Temp_lag

m30<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3,data=air_data)
anova(m3,m30)



cat('AIC of Model 1 = ',AIC(m1),'\n')
cat('AIC of Model 2 = ',AIC(m2),'\n')
cat('AIC of Model 3 = ',AIC(m3),'\n')
```

```
R 4.2.2 · ~/Downloads/Teaching/Regression_and_Classification/NPTEL/
> cat('AIC of Model 1 = ',AIC(m1),'\n')
AIC of Model 1 =  918.759
> cat('AIC of Model 2 = ',AIC(m2),'\n')
AIC of Model 2 =  750.2042
> cat('AIC of Model 3 = ',AIC(m3),'\n')
AIC of Model 3 =  613.5645
>
```

```
m1   List of  13
m10  List of  13
m2   List of  13
m20  List of  13
m3   List of  13
m30  List of  13
```

Now, if you interestingly if you see what is happening. You see what is happening here the AIC is constantly dropping whereas, model; that means, third model is probably better than the first and second model.

(Refer Slide Time: 10:21)



```r
m2<-lm(Ozone~Ozone_lag1+Ozone_lag2+Temp_lag1+Temp_lag2,data=air_
m20<-lm(Ozone~Ozone_lag1+Ozone_lag2,data=air_data)
anova(m2,m20)

m3<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3+Temp_lag1+Temp_lag

m30<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3,data=air_data)
anova(m3,m30)



cat('AIC of Model 1 = ',AIC(m1),'\n')
cat('AIC of Model 2 = ',AIC(m2),'\n')
```

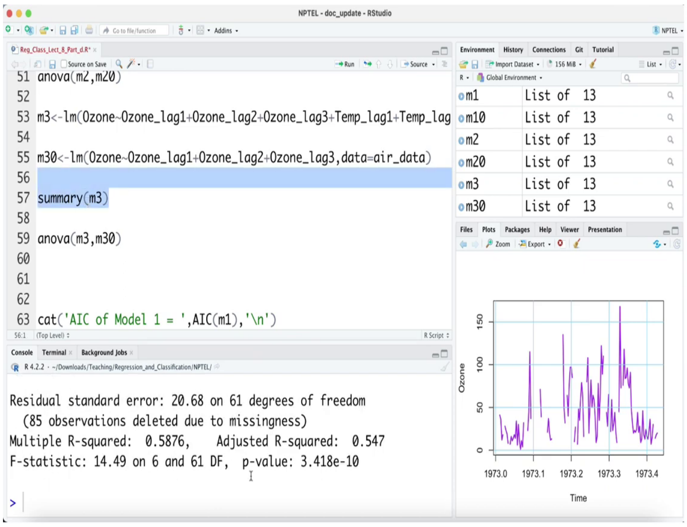
```
m1    List of 13
m10   List of 13
m2    List of 13
m20   List of 13
m3    List of 13
m30   List of 13
```

```
> anova(m3,m30)
Analysis of Variance Table

Model 1: Ozone ~ Ozone_lag1 + Ozone_lag2 + Ozone_lag3 + Temp_lag1 +
 Temp_lag2 +
    Temp_lag3
Model 2: Ozone ~ Ozone_lag1 + Ozone_lag2 + Ozone_lag3
```

However, what we are seeing that third model the lag does not have any effect. So, that means, we cannot really use we cannot say that third lag has any effect on the 3 lag model of a granger causal model does any effect any of temperature has no effect on the Ozone.

So, that means, as we are putting as we are you know put more and more lags. So, naturally what is happening the model complexity increases and its going to higher and higher dimension as model complexity increases it is doing some overfitting because my AIC is constantly going down. But we know, but most likely it is effectively losing its interpretability and that is why probably its not running its not giving any effect.

(Refer Slide Time: 11:21)

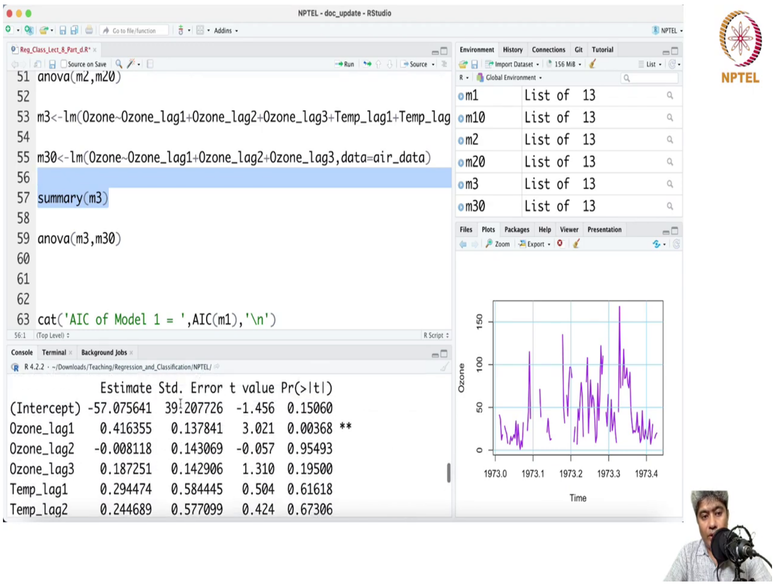


```r
anova(m2,m20)

m3<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3+Temp_lag1+Temp_lag

m30<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3,data=air_data)


summary(m3)

anova(m3,m30)




cat('AIC of Model 1 = ',AIC(m1),'\n')
```

m1	List of 13
m10	List of 13
m2	List of 13
m20	List of 13
m3	List of 13
m30	List of 13

```
Residual standard error: 20.68 on 61 degrees of freedom
  (85 observations deleted due to missingness)
Multiple R-squared:  0.5876,    Adjusted R-squared:  0.547
F-statistic: 14.49 on 6 and 61 DF,  p-value: 3.418e-10

>
```

(Refer Slide Time: 11:32)



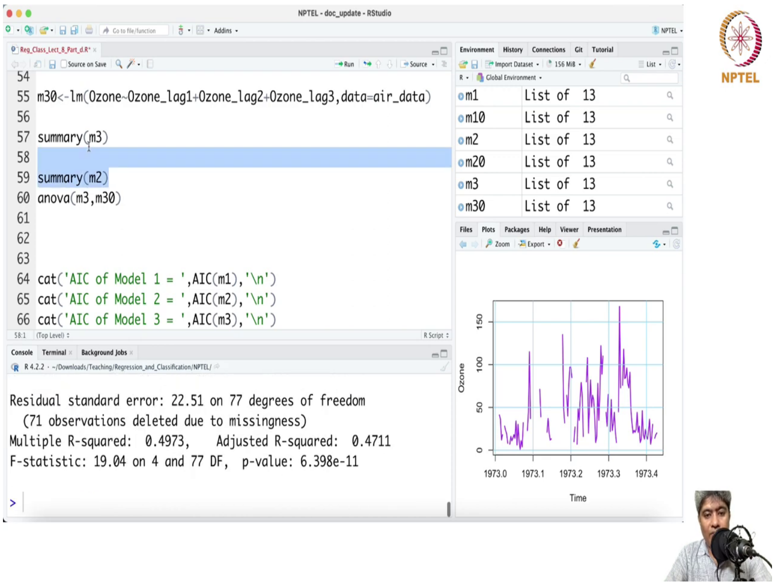
```
anova(m2,m20)

m3<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3+Temp_lag1+Temp_lag

m30<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3,data=air_data)

summary(m3)

anova(m3,m30)



cat('AIC of Model 1 = ',AIC(m1),'\n')
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -57.075641  39.207726  -1.456  0.15060
Ozone_lag1    0.416355   0.137841   3.021  0.00368 **
Ozone_lag2   -0.008118   0.143069  -0.057  0.95493
Ozone_lag3    0.187251   0.142906   1.310  0.19500
Temp_lag1     0.294474   0.584445   0.504  0.61618
Temp_lag2     0.244689   0.577099   0.424  0.67306
```

m1	List of 13
m10	List of 13
m2	List of 13
m20	List of 13
m3	List of 13
m30	List of 13

But we though we know that you know lag1 and lag2 does have a effect. So, we can see that you know you know very high M3 in case of third model except the lag1 model does not have any effect.

(Refer Slide Time: 11:49)

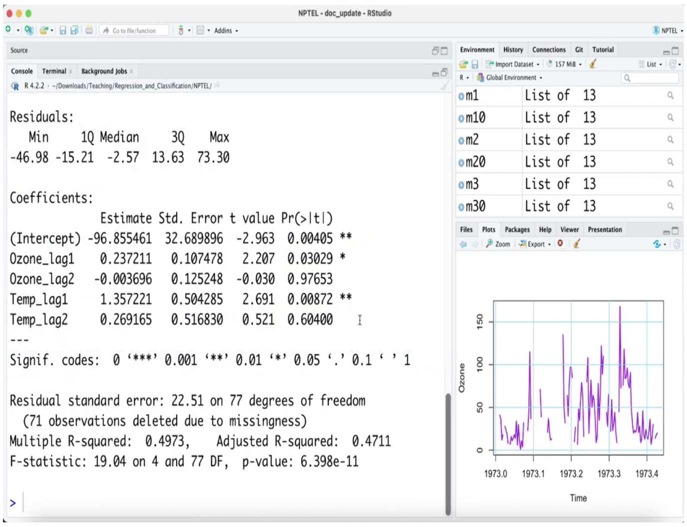


```
m30<-lm(Ozone~Ozone_lag1+Ozone_lag2+Ozone_lag3,data=air_data)

summary(m3)

summary(m2)
anova(m3,m30)



cat('AIC of Model 1 = ',AIC(m1),'\n')
cat('AIC of Model 2 = ',AIC(m2),'\n')
cat('AIC of Model 3 = ',AIC(m3),'\n')
```

```
Residual standard error: 22.51 on 77 degrees of freedom
  (71 observations deleted due to missingness)
Multiple R-squared:  0.4973,    Adjusted R-squared:  0.4711
F-statistic: 19.04 on 4 and 77 DF,  p-value: 6.398e-11

>
```

m1	List of 13
m10	List of 13
m2	List of 13
m20	List of 13
m3	List of 13
m30	List of 13

(Refer Slide Time: 11:53)

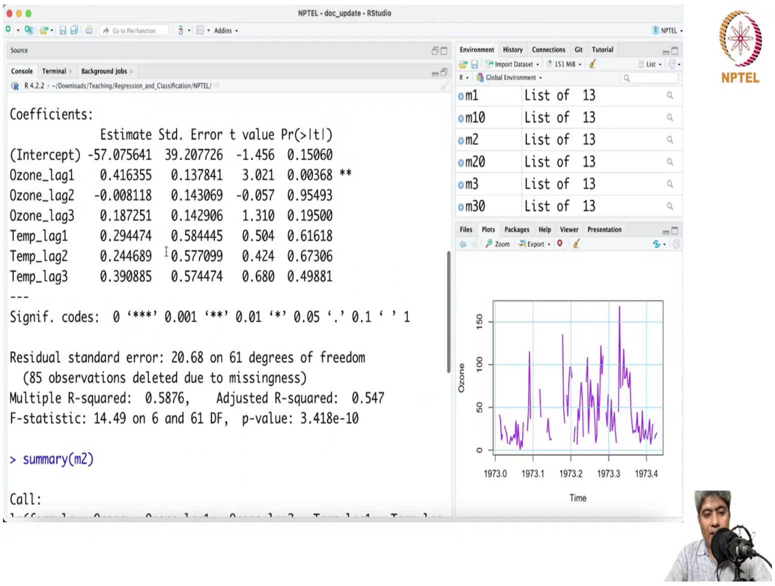


```
Residuals:
   Min     1Q Median     3Q    Max 
-46.98 -15.21  -2.57  13.63  73.30 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept) -96.855461  32.689896  -2.963  0.00405 **
Ozone_lag1    0.237211   0.107478   2.207  0.03029 * 
Ozone_lag2   -0.003696   0.125248  -0.030  0.97653   
Temp_lag1     1.357221   0.504285   2.691  0.00872 **
Temp_lag2     0.269165   0.516830   0.521  0.60400   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 22.51 on 77 degrees of freedom
  (71 observations deleted due to missingness)
Multiple R-squared:  0.4973,	Adjusted R-squared:  0.4711 
F-statistic: 19.04 on 4 and 77 DF,  p-value: 6.398e-11

> 
```

m1	List of 13
m10	List of 13
m2	List of 13
m20	List of 13
m3	List of 13
m30	List of 13

Whereas, m2 if you see m2 lag1 temperature does have a effect we see it does have a effect whereas, the lag2 does not have a effect.

(Refer Slide Time: 12:00)



```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -57.075641  39.207726  -1.456  0.15060
Ozone_lag1    0.416355   0.137841   3.021  0.00368 **
Ozone_lag2   -0.008118   0.143069  -0.057  0.95493
Ozone_lag3    0.187251   0.142906   1.310  0.19500
Temp_lag1     0.294474   0.584445   0.504  0.61618
Temp_lag2     0.244689   0.577099   0.424  0.67306
Temp_lag3     0.390885   0.574474   0.680  0.49881
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.68 on 61 degrees of freedom
  (85 observations deleted due to missingness)
Multiple R-squared:  0.5876,    Adjusted R-squared:  0.547
F-statistic: 14.49 on 6 and 61 DF,  p-value: 3.418e-10

> summary(m2)

Call:
```

```
m1    List of  13
m10   List of  13
m2    List of  13
m20   List of  13
m3    List of  13
m30   List of  13
```

Also, if you look into the standard error standard error in third model is 0.577 whereas, in the second model for lag 1 is 0.50. So, that means, standard error is increasing. So, there is a high possibility that a multicollinearity also creeping in because of the multicollinearity these lags are not any more effective because their standard error is going up because of the multicollinearity.

So, you have to be very careful about how you do the interpretation of these statistical inference and these you have to be very careful about these statistical inferences when particularly when you are adding more and more features just because every time you are increasing lag.

That means, your model is getting complex your complexity of your model is increasing it will have a higher tendency to overfit because end of the day your training datasets is finite most of the time your data training dataset is not increasing. So, as a result your model will overfit and as it overfits you do not want your model to overfit because your bias will when it happens your bias will be small.

But you will have a very high variance in out of the sample it will not do very well. So, you have to be very careful about model fitting when for adding more complex lag variable and it is better to obtain a parsimonious small model and just stop there, ok. So, far this week we will this is that is how this much we will be discussing. Next week we will see you with a new video with a new chapter.

Thank you very much. Take care. Bye.