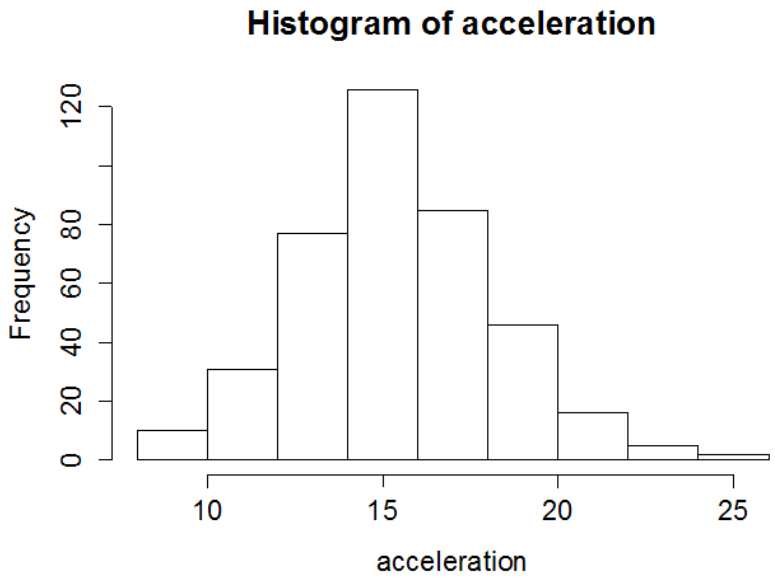
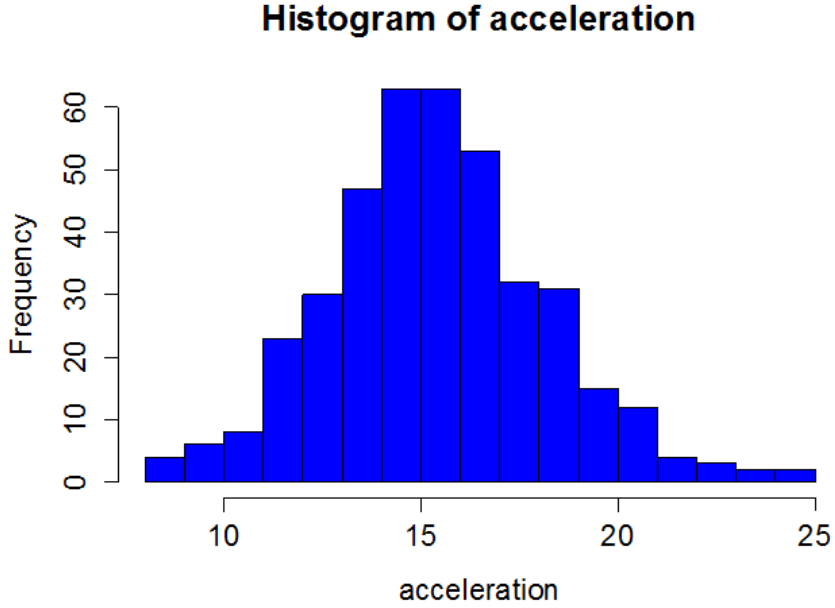
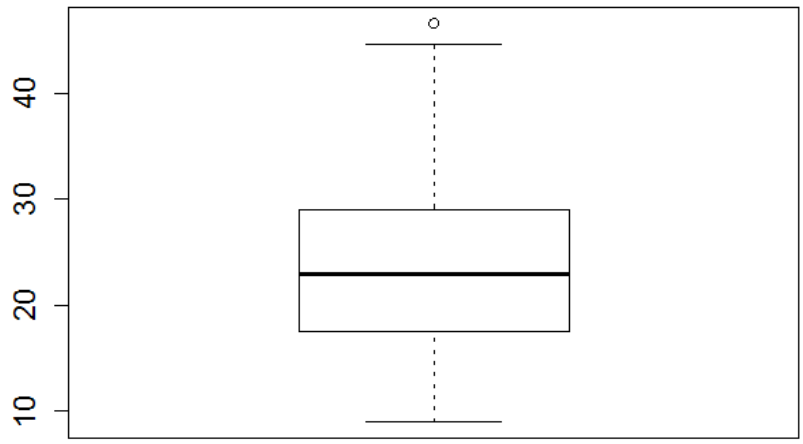
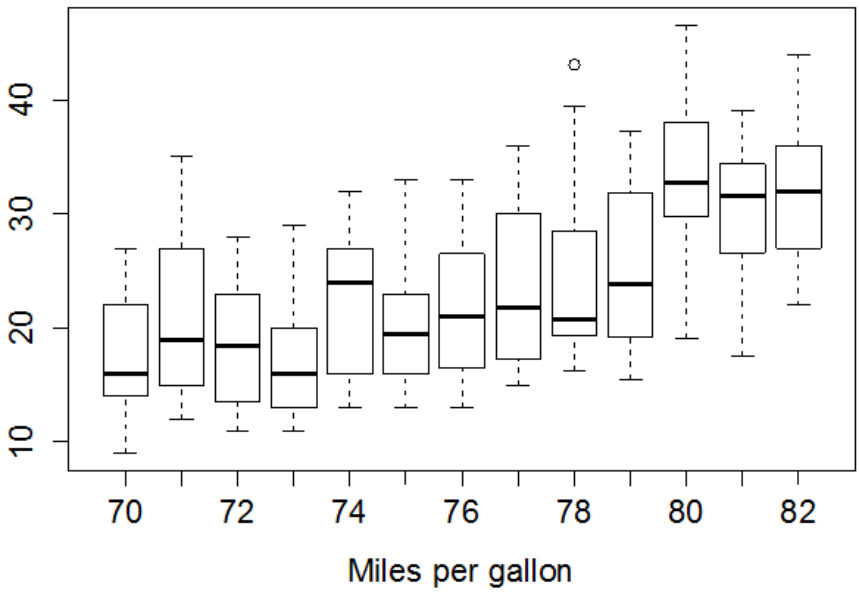
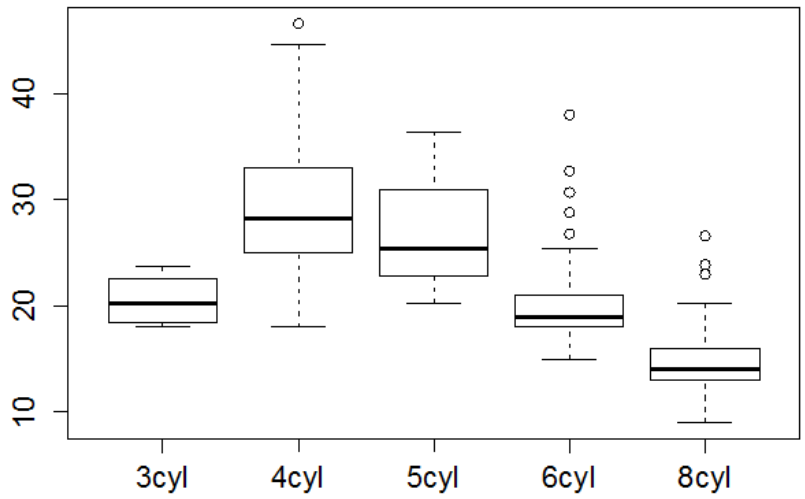
Chapter 2

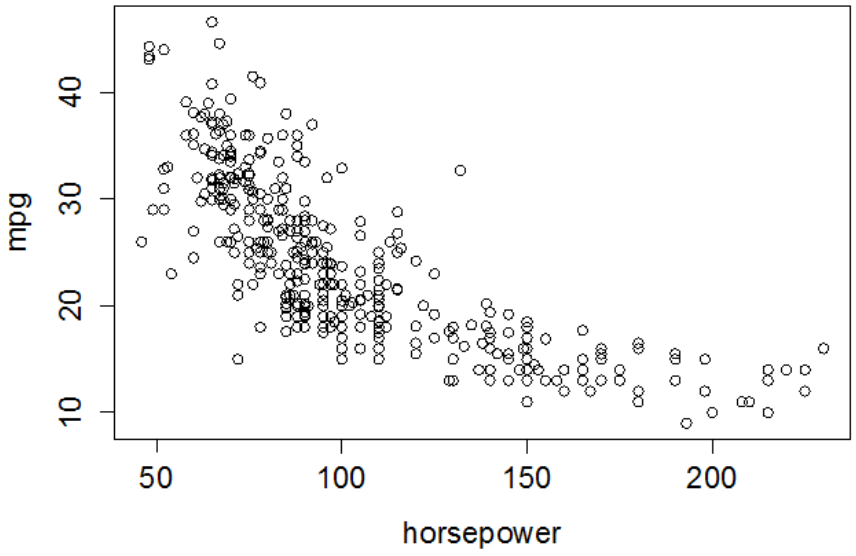

Histogram of acceleration
Frequency
0
20
40
60
80
120
10
15
20
25
acceleration

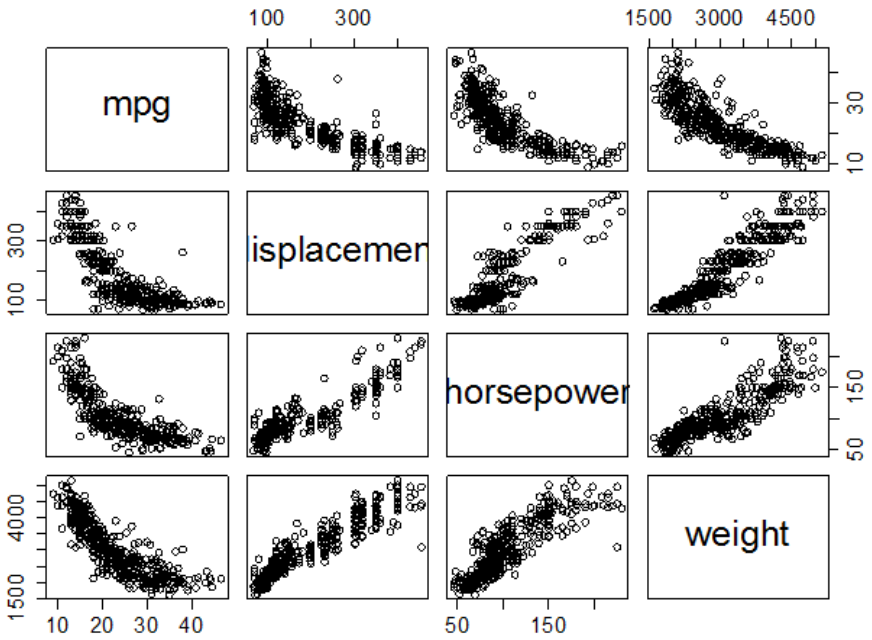

Histogram of acceleration
Frequency
0
10
20
30
40
50
60
10
15
20
25
acceleration

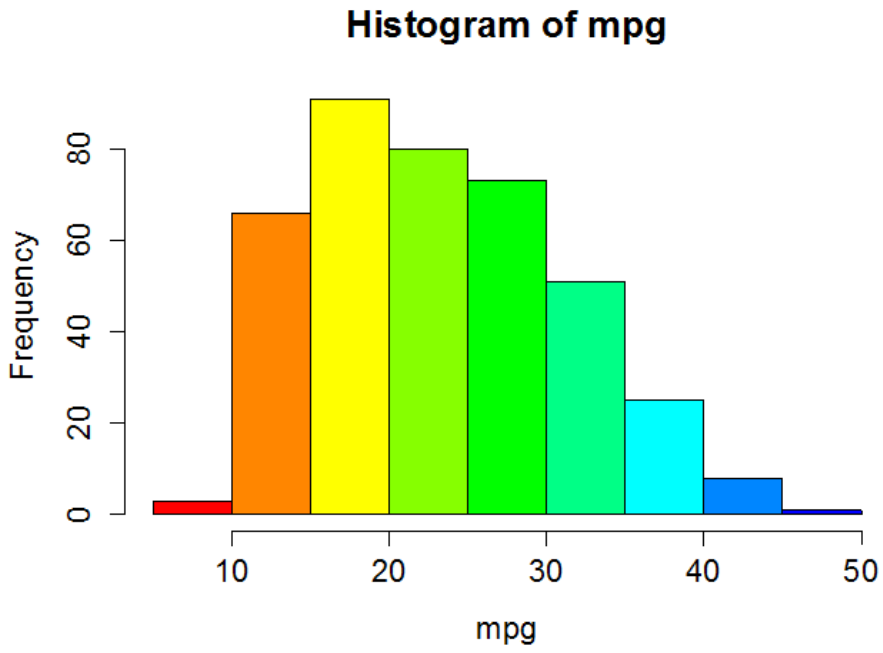

10
20
30
40
Miles per gallon

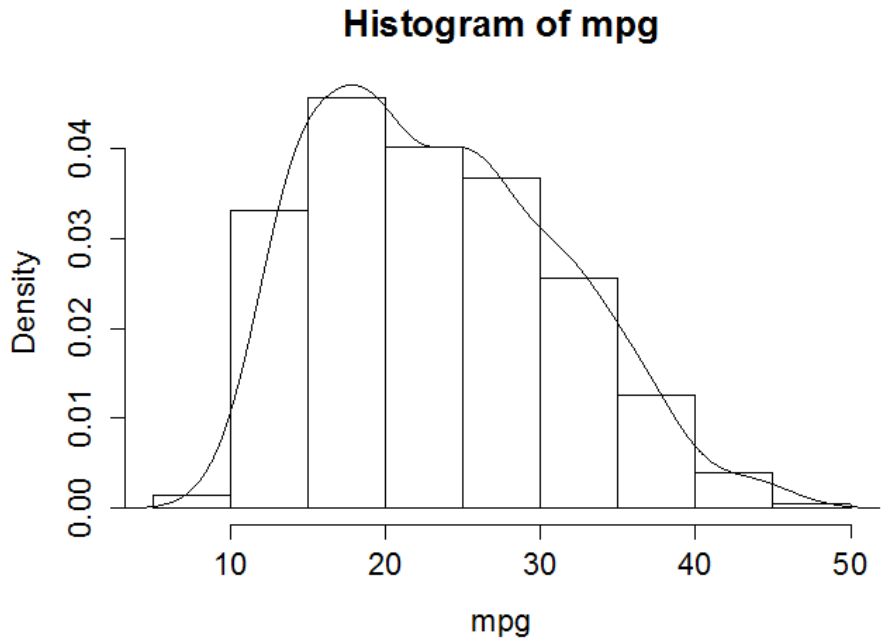

10
20
30
40
70
72
74
76
78
80
82
Miles per gallon

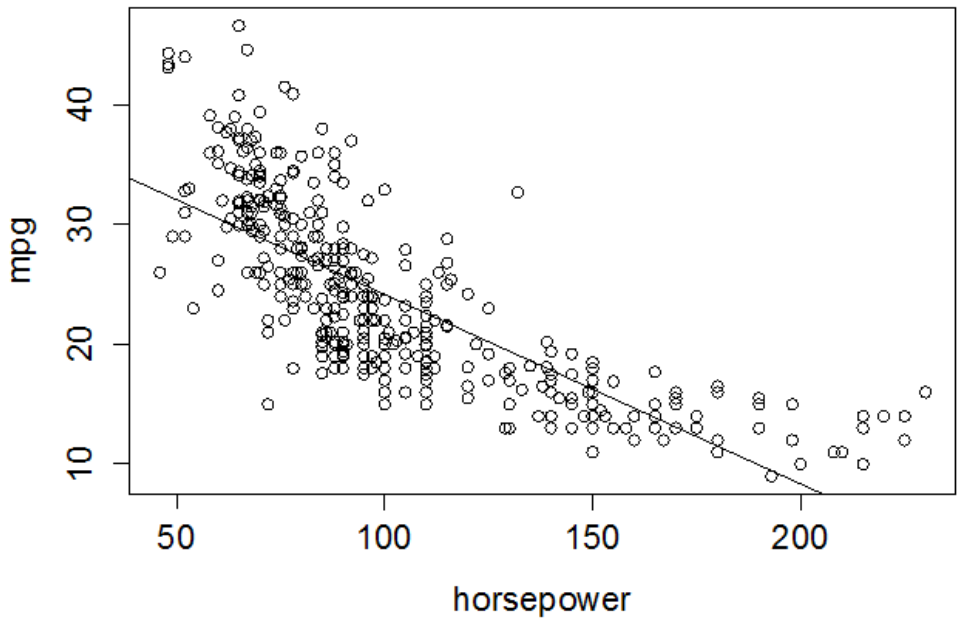

10
20
30
40
3cyl
4cyl
5cyl
6cyl
8cyl

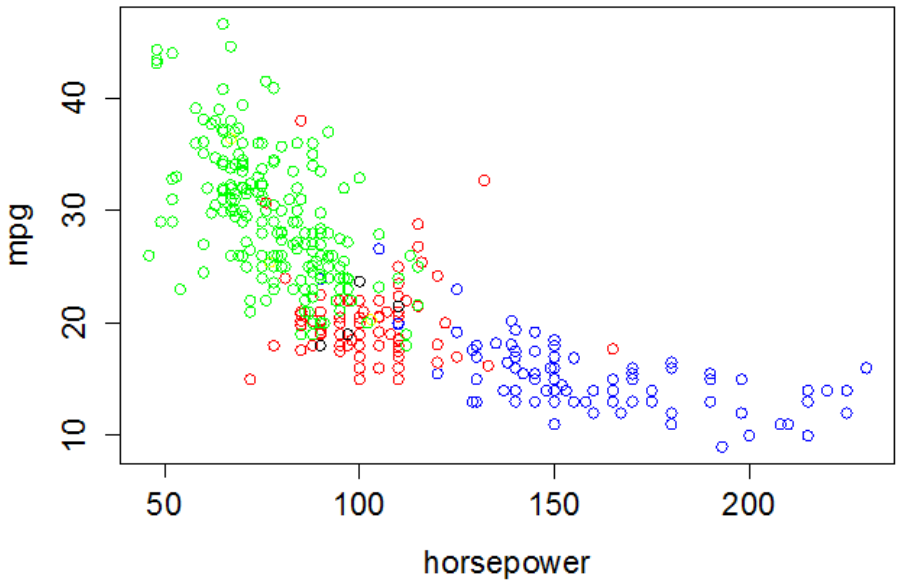

mpg
10
20
30
40
50
100
150
200
horsepower

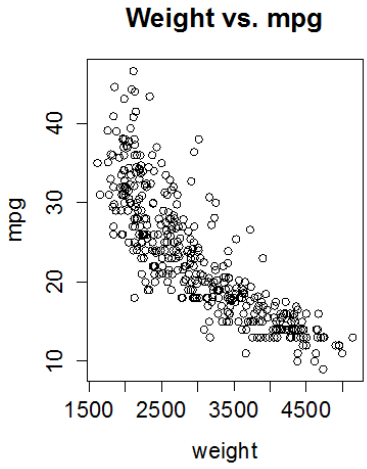

mpg
isplacemer
horsepower
weight
100
300
1500
3000
4500
10
30
100
300
50
150
1500
4000
10
20
30
40
50
150

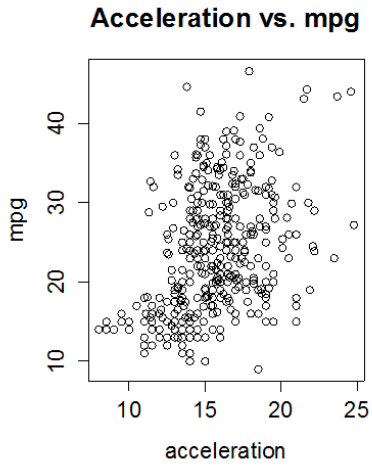

Histogram of mpg
Frequency
0
20
40
60
80
10
20
30
40
50
mpg


Histogram of mpg
Density
0.00
0.01
0.02
0.03
0.04
10
20
30
40
50
mpg

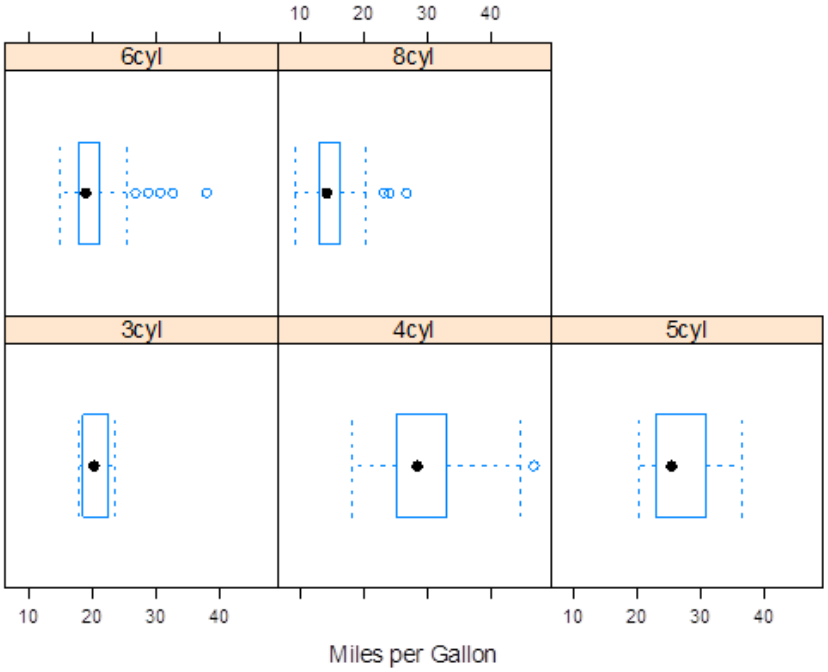

mpg
10
20
30
40
50
100
150
200
horsepower


mpg
10
20
30
40
50
100
150
200
horsepower

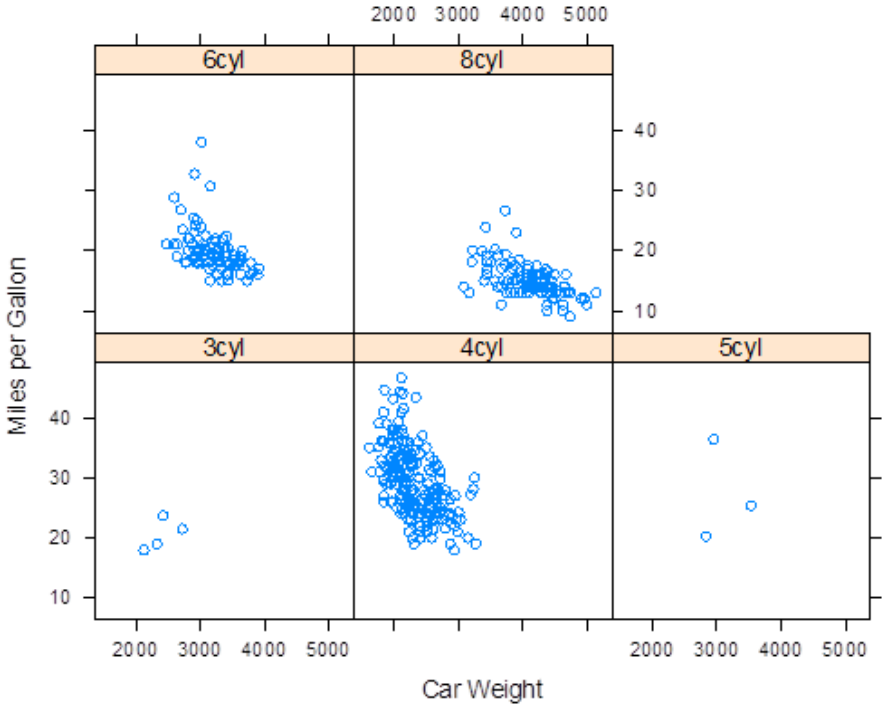

Weight vs. mpg
mpg
10
20
30
40
1500
2500
3500
4500
weight

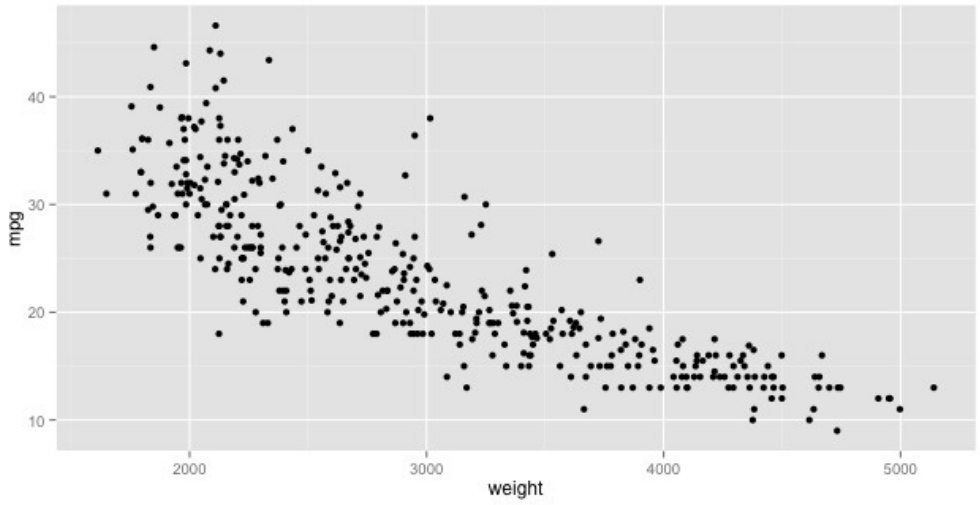

Acceleration vs. mpg
mpg
10
20
30
40
10
15
20
25
acceleration

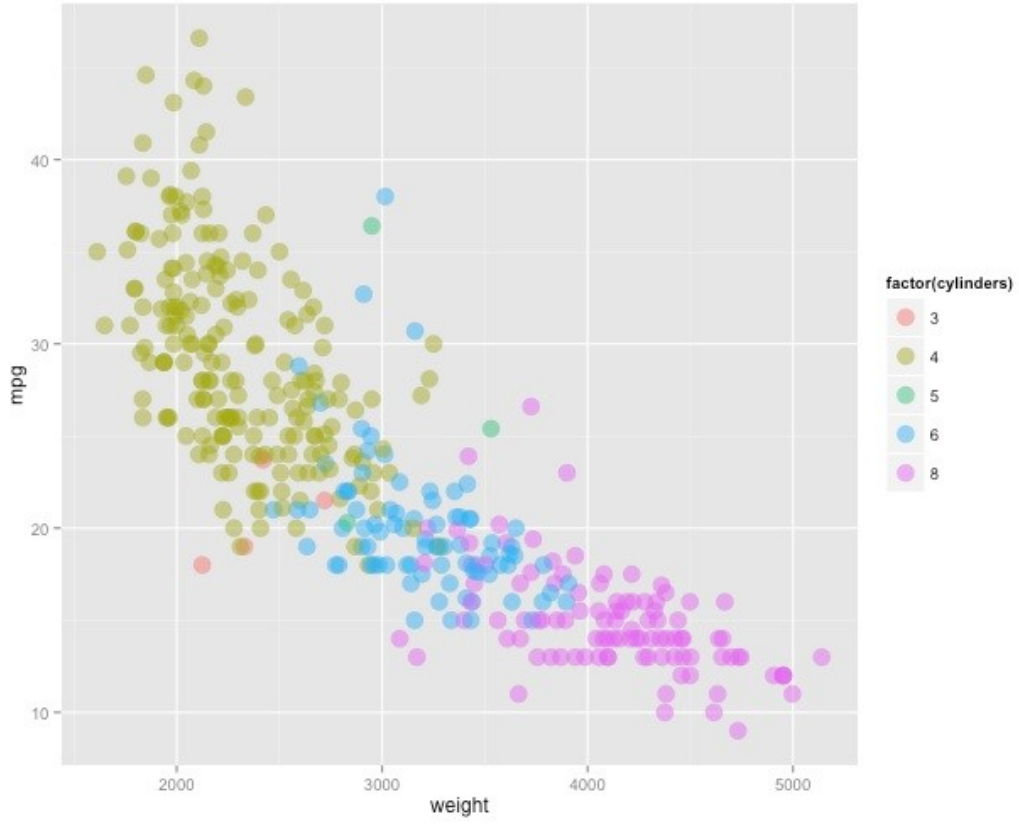
MPG by Number Of Cylinders

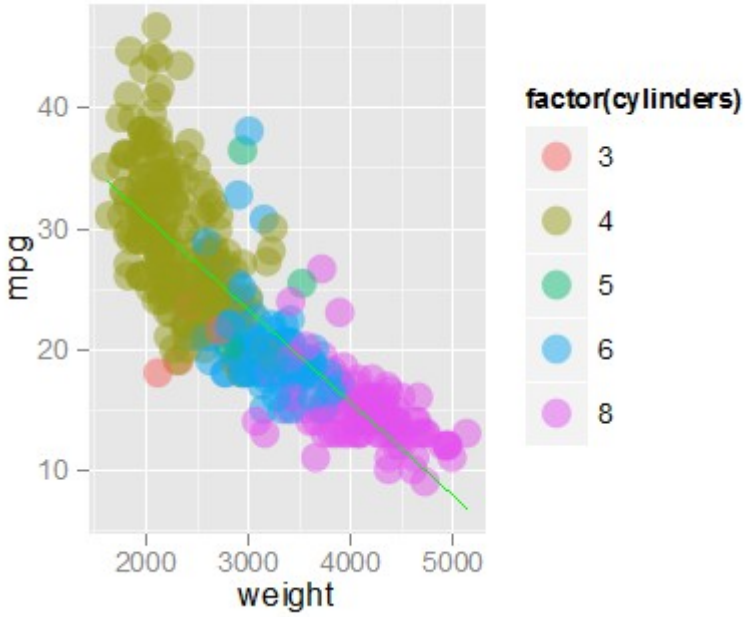


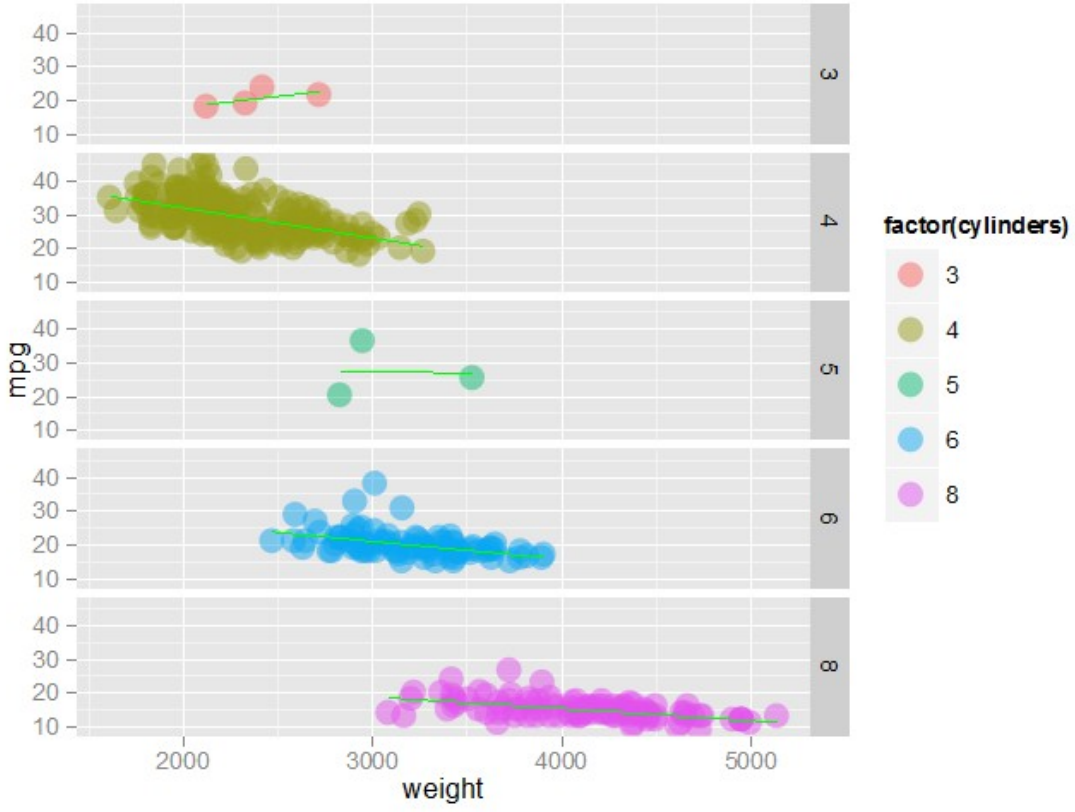
Weight Vs MPG by Number of Cylinders

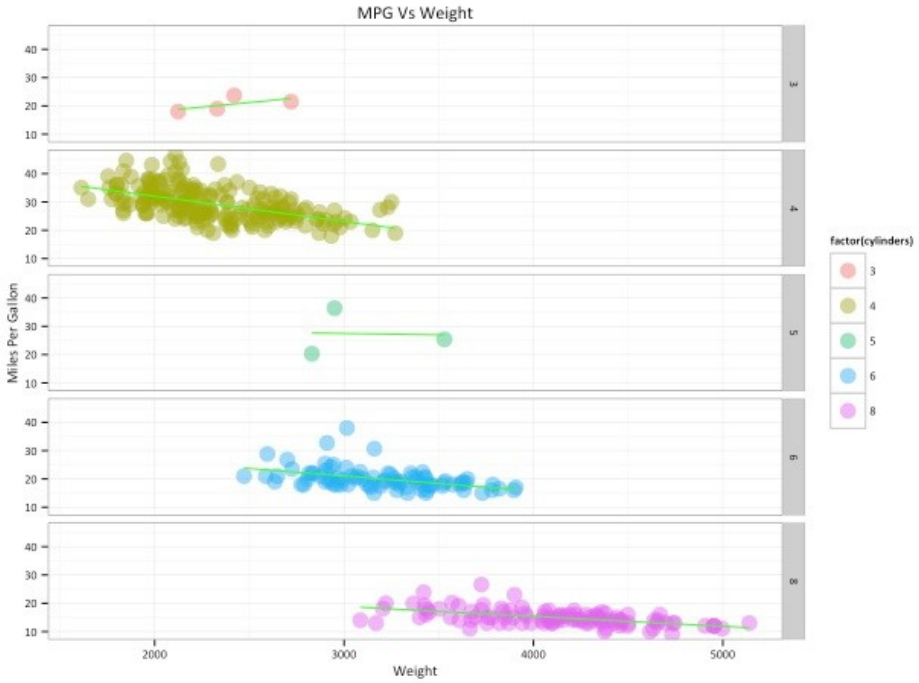


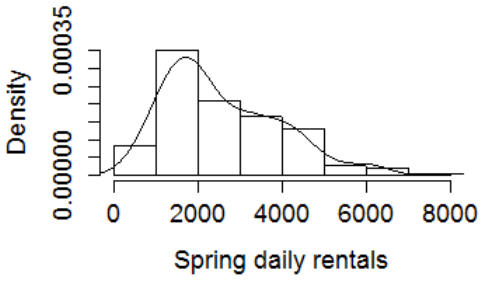

mpg
40
30
20
10
2000
3000
4000
5000
weight

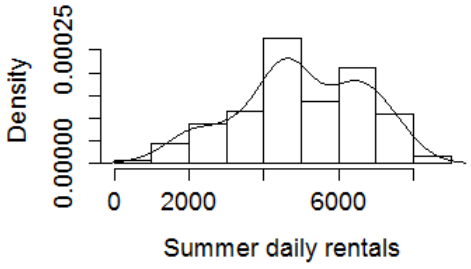

mpg
40
30
20
10
2000
3000
4000
5000
weight
factor(cylinders)
3
4
5
6
8

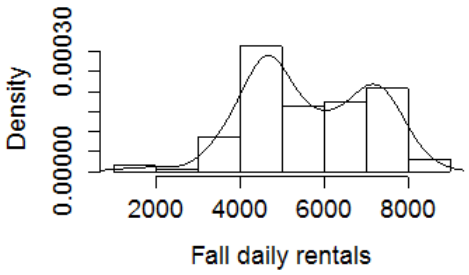

mpg
40
30
20
10
2000
3000
4000
5000
weight
factor(cylinders)
3
4
5
6
8

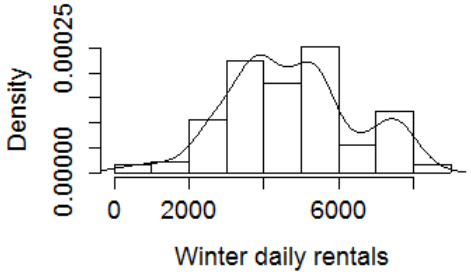

mpg
40
30
20
10
3
4
5
6
8
2000
3000
4000
5000
weight
factor(cylinders)
3
4
5
6
8

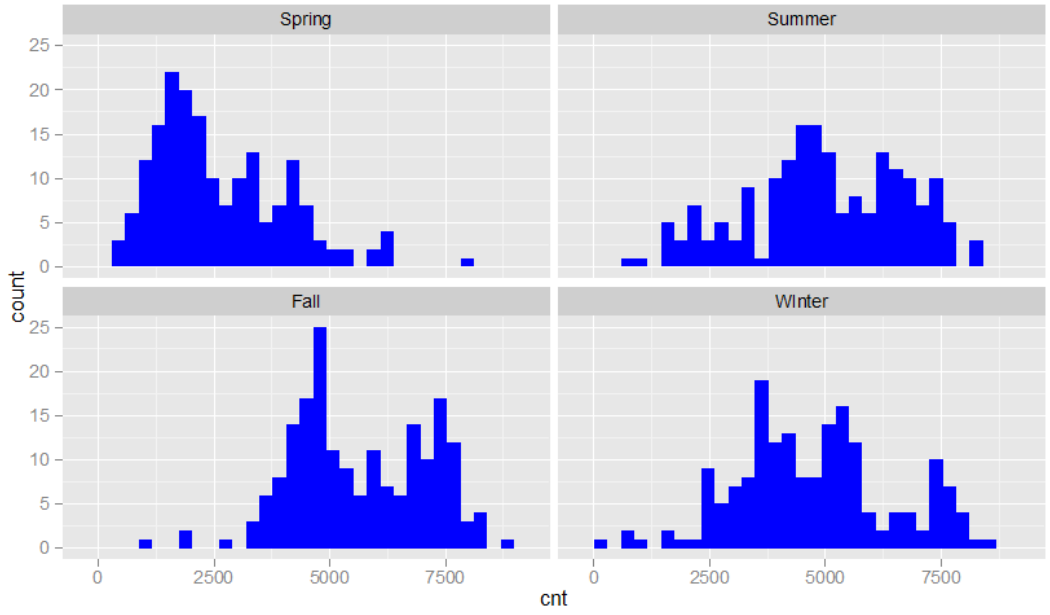

MPG Vs Weight
Miles Per Gallon
Weight
factor(cylinders)
3
4
5
6
8

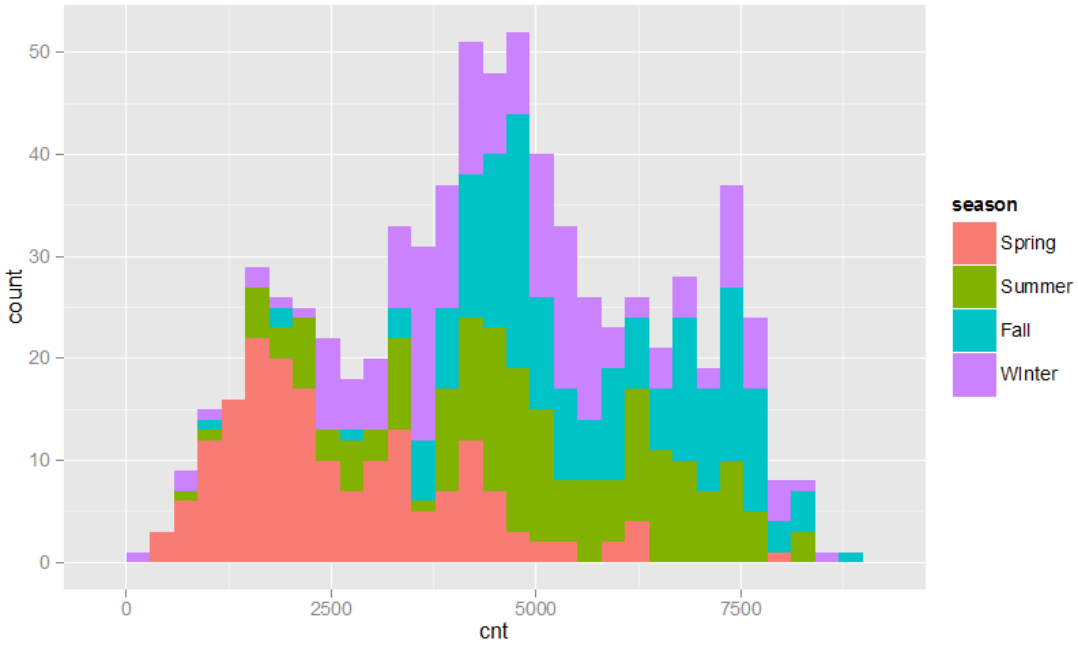

Density
0.00000
0.00035
0
2000
4000
6000
8000
Spring daily rentals

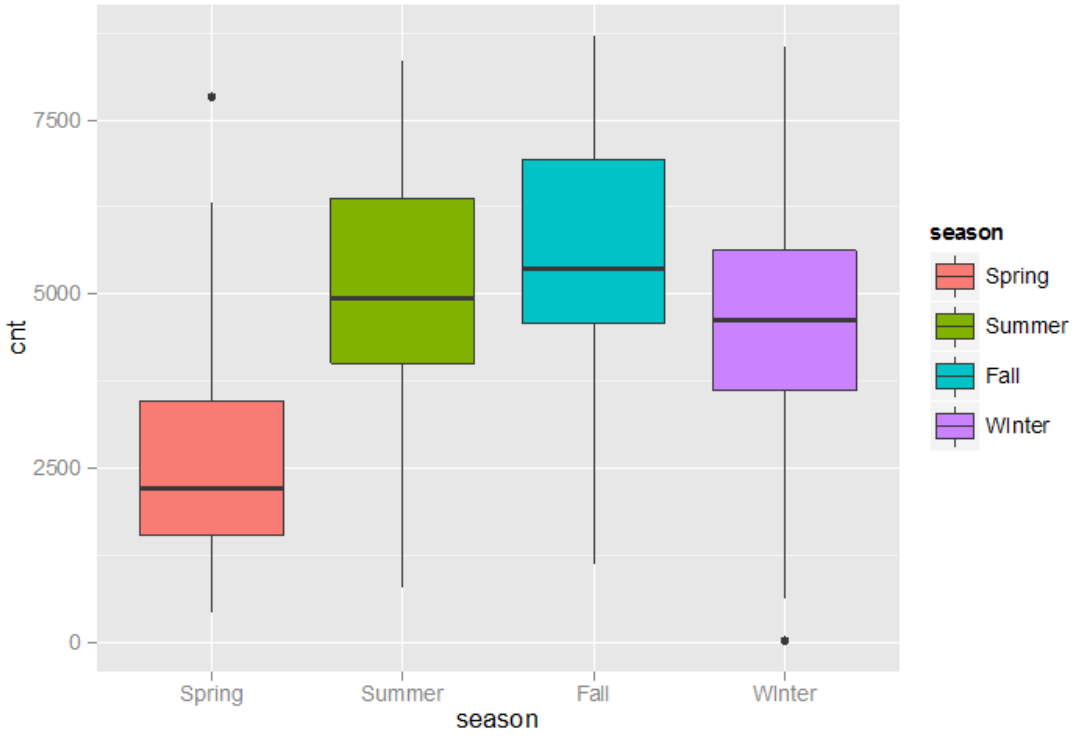

Density
0.00000
0.00025
0
2000
6000
Summer daily rentals

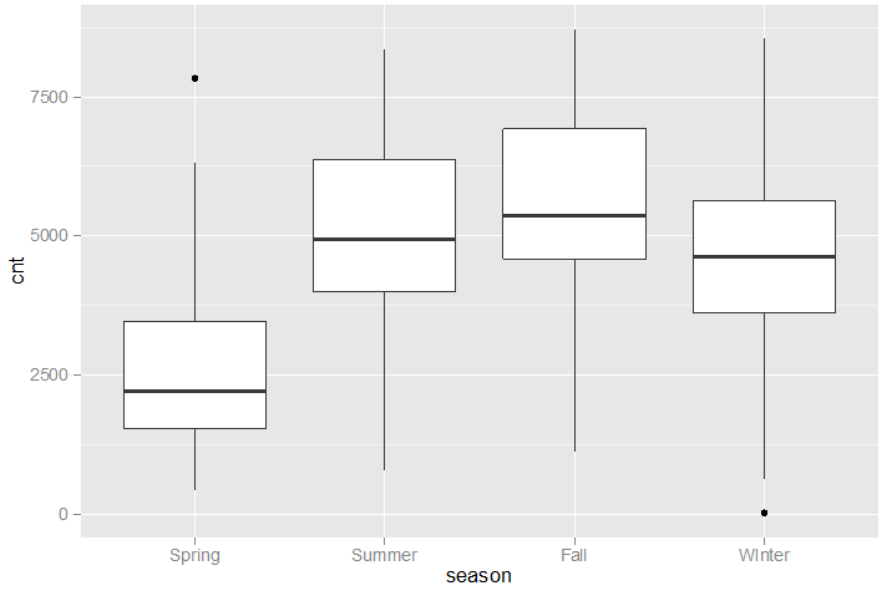

Density
0.00000
0.00030
2000
4000
6000
8000
Fall daily rentals

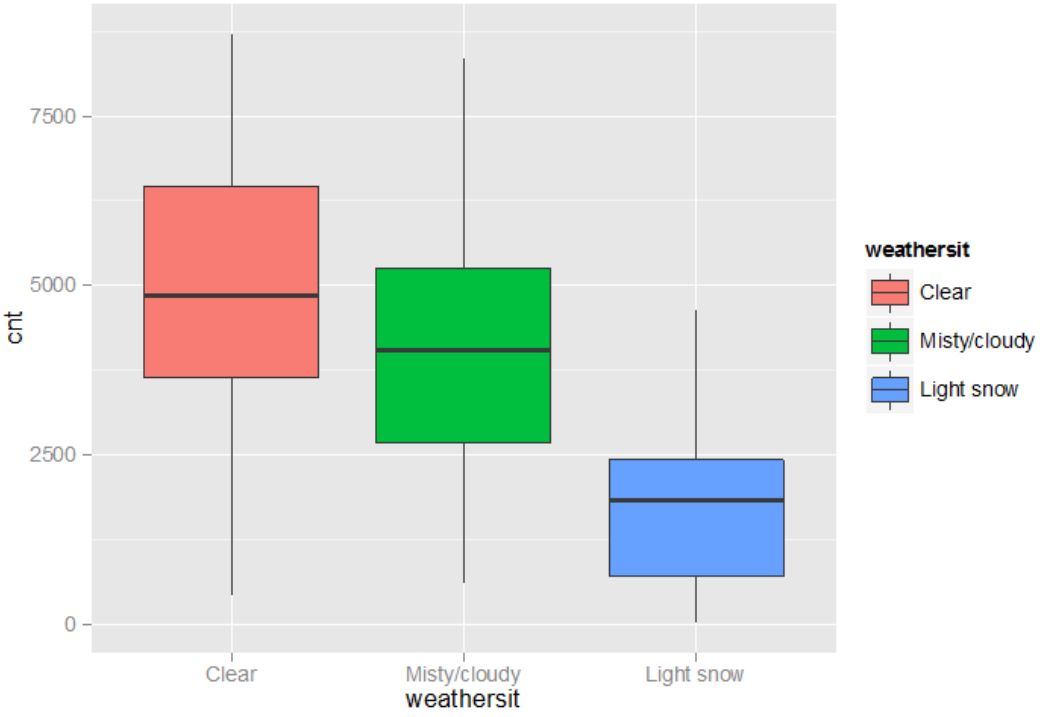

Density
0.00000
0.00025
0
2000
6000
Winter daily rentals

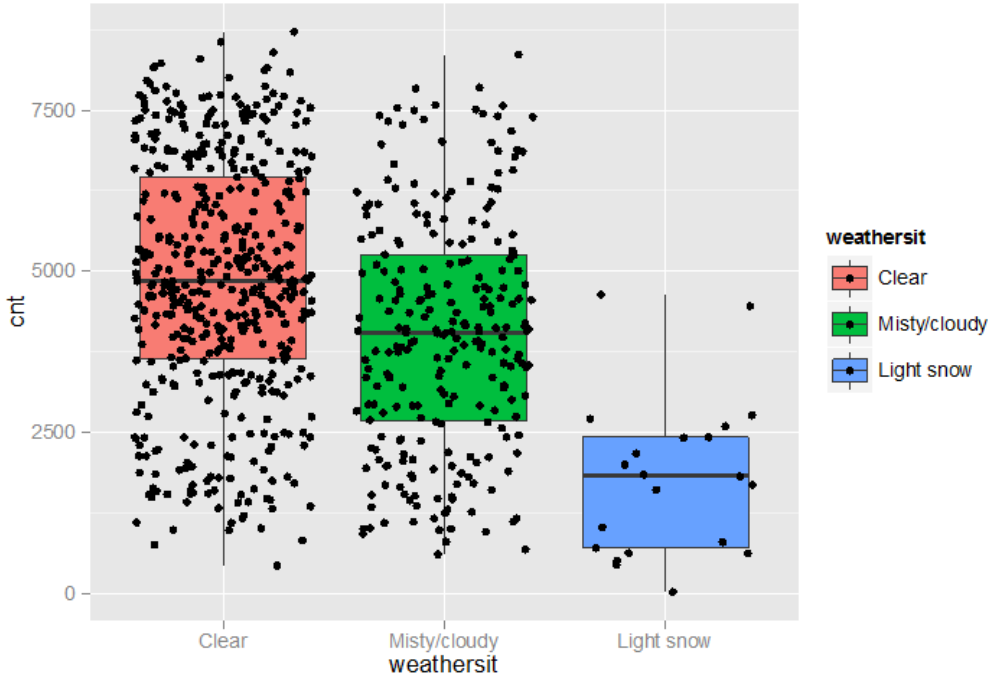

Spring
Summer
Fall
WInter
count
cnt
0
5
10
15
20
25
0
2500
5000
7500

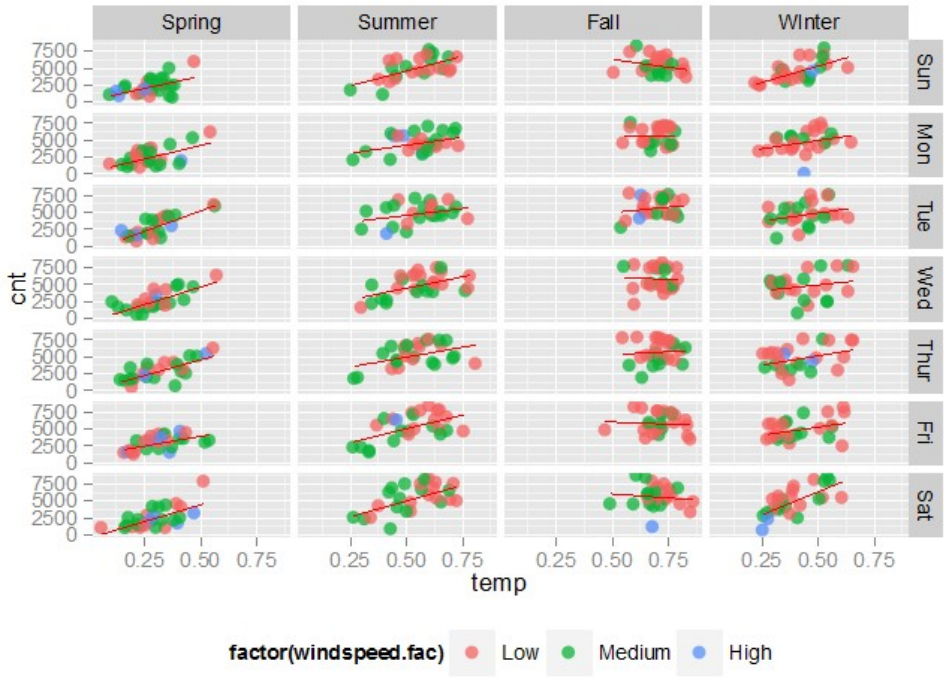

count
cnt
season
Spring
Summer
Fall
WInter
0
10
20
30
40
50
0
2500
5000
7500


7500
5000
2500
0
cnt
Spring
Summer
Fall
Winter
season
season
Spring
Summer
Fall
Winter

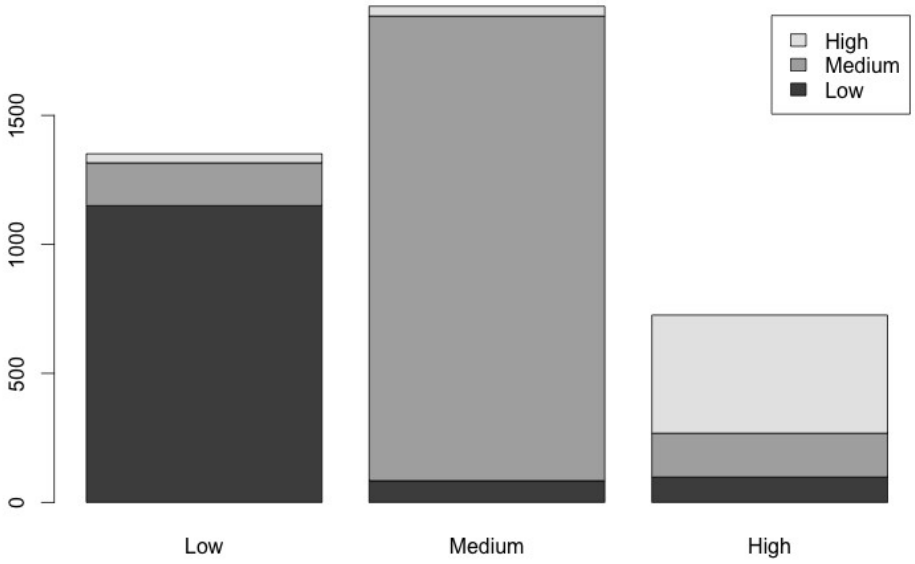

7500
5000
2500
0
cnt
Spring
Summer
Fall
Winter
season

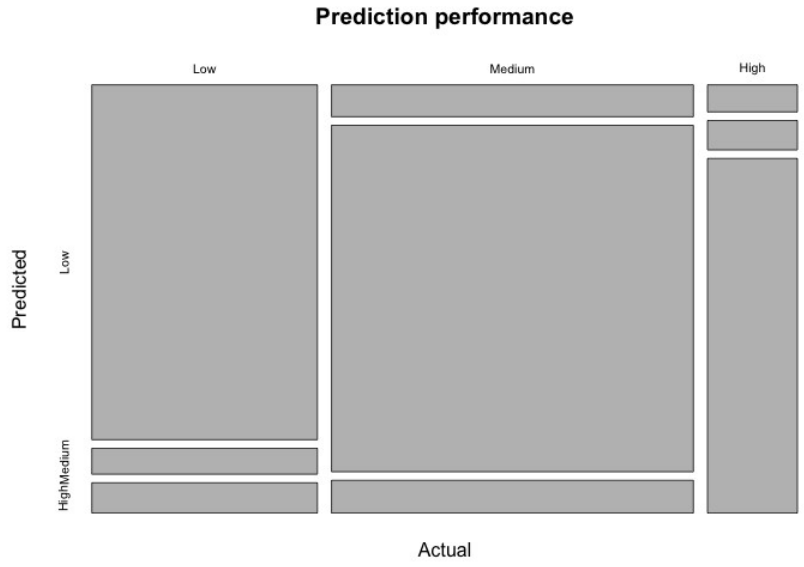

7500
5000
2500
0
cnt
Clear
Misty/cloudy
Light snow
weathersit
weathersit
Clear
Misty/cloudy
Light snow

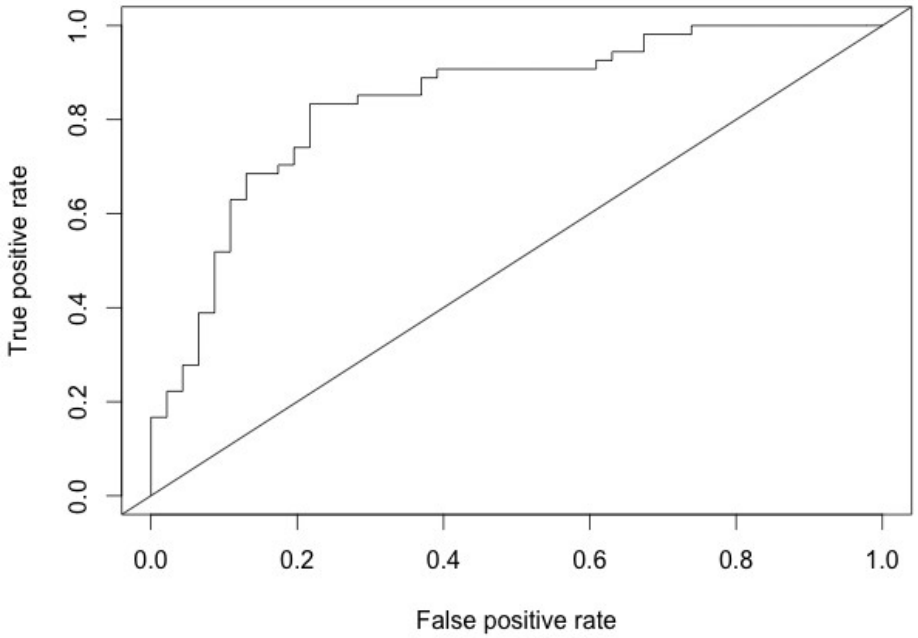

7500
5000
2500
0
cnt
Clear
Misty/cloudy
Light snow
weathersit
weathersit
Clear
Misty/cloudy
Light snow

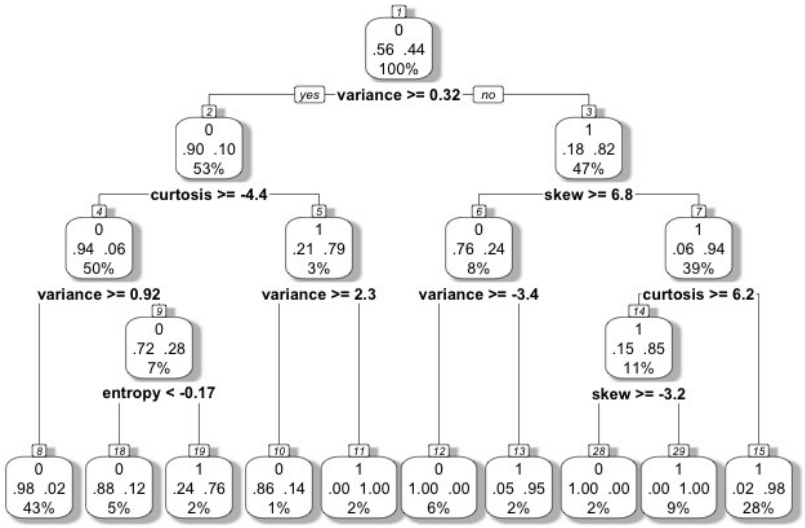

Spring
Summer
Fall
WInter
Sun
Mon
Tue
Wed
Thur
Fri
Sat
cnt
7500
5000
2500
0
0.25
0.50
0.75
temp
factor(windspeed.fac)
Low
Medium
High

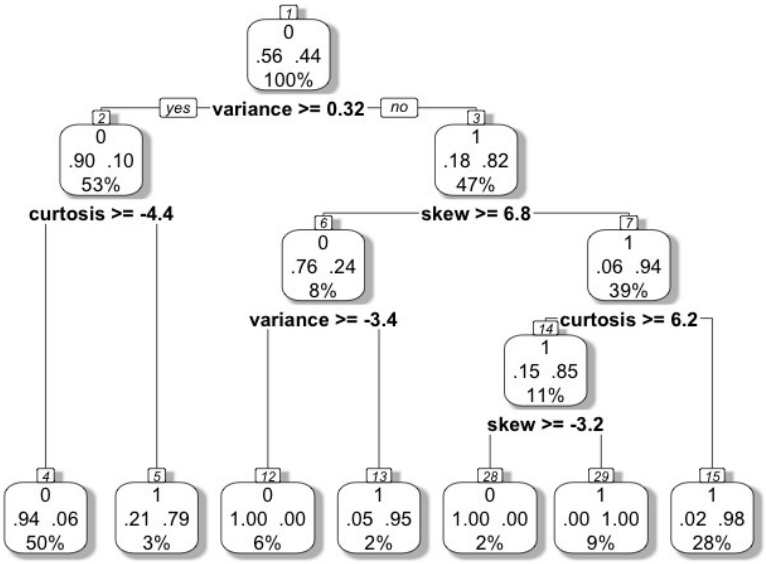
Chapter 3


High
Medium
Low
1500
1000
500
0
Low
Medium
High


Prediction performance
Low
Medium
High
Predicted
Low
High
Medium
Actual


True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


1
0
.56 .44
100%
yes
variance >= 0.32
no
2
0
.90 .10
53%
3
1
.18 .82
47%
curtosis >= -4.4
skew >= 6.8
4
0
.94 .06
50%
5
1
.21 .79
3%
6
0
.76 .24
8%
7
1
.06 .94
39%
variance >= 0.92
variance >= 2.3
variance >= -3.4
curtosis >= 6.2
9
0
.72 .28
7%
14
1
.15 .85
11%
entropy < -0.17
skew >= -3.2
8
0
.98 .02
43%
18
0
.88 .12
5%
19
1
.24 .76
2%
10
0
.86 .14
1%
11
1
.00 1.00
2%
12
0
1.00 .00
6%
13
1
.05 .95
2%
28
0
1.00 .00
2%
29
1
.00 1.00
9%
15
1
.02 .98
28%



Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes. default(x=X, y=Y, laplace= laplace)

A-priori probabilities:
Y
No Yes
0.5 0.5

A-priori probabilities of each class

Conditional probabilities:
Education

Y	A	B	C
No	0.5000000	0.3333333	0.1666667
Yes	0.1666667	0.3333333	0.5000000

P(Education=B|Purchase=Yes)

Gender

Y	F	M
No	0.5000000	0.5000000
Yes	0.3333333	0.6666667

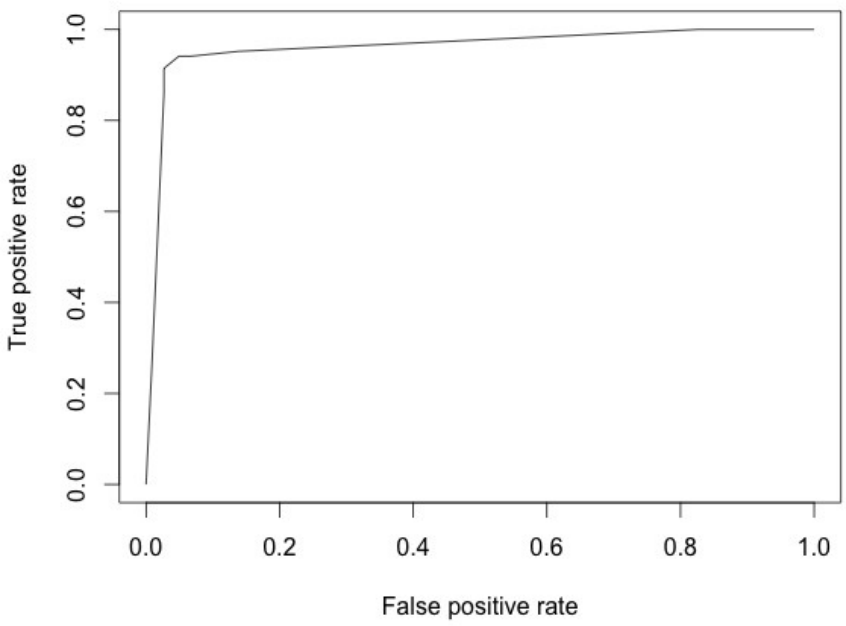
P(Gender=M|Purchase=No)

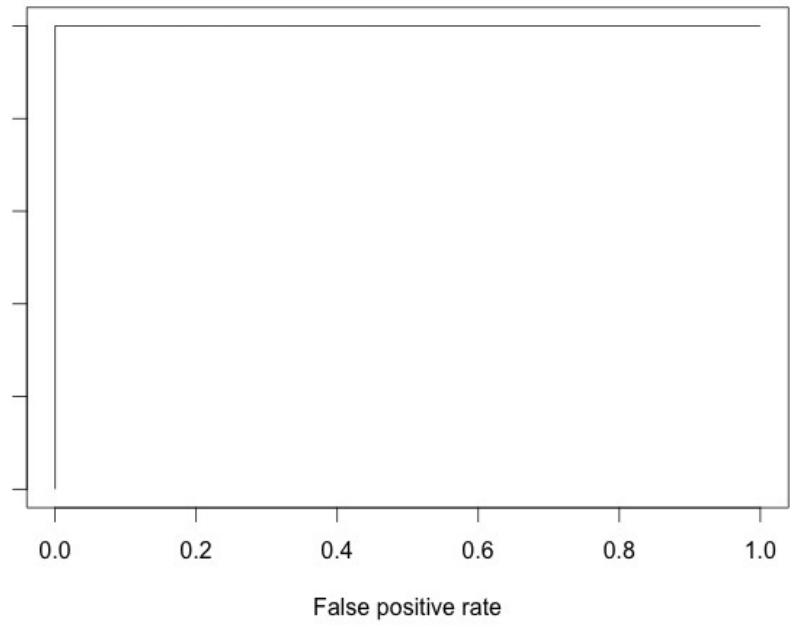
Smart_ph

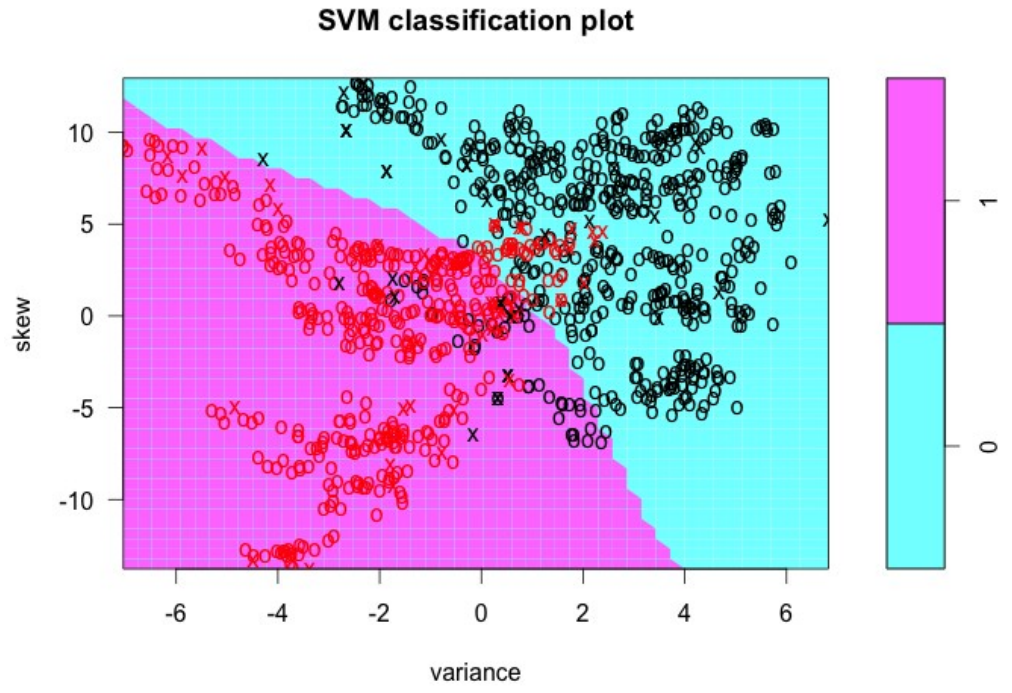
Y	N	Y
No	0.5	0.5
Yes	0.5	0.5

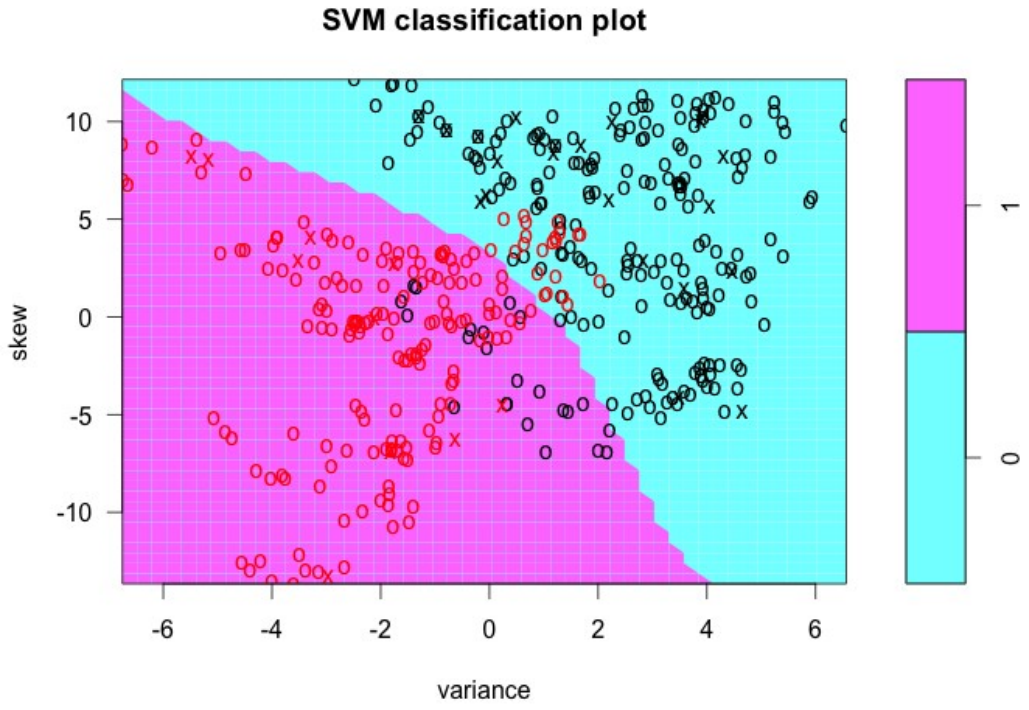
Tablet

Y	N	Y
No	0.1666667	0.3333333
Yes	0.1666667	0.8333333

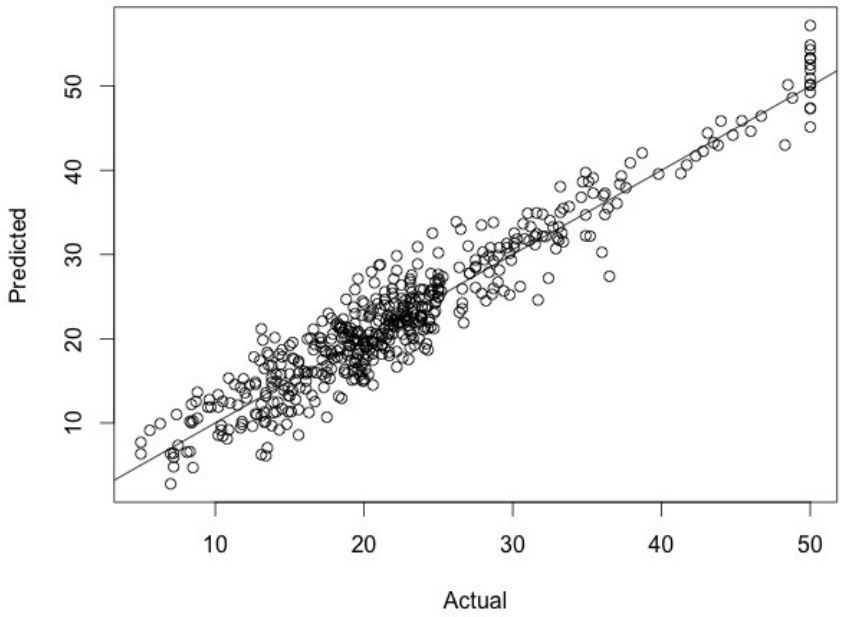

True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate

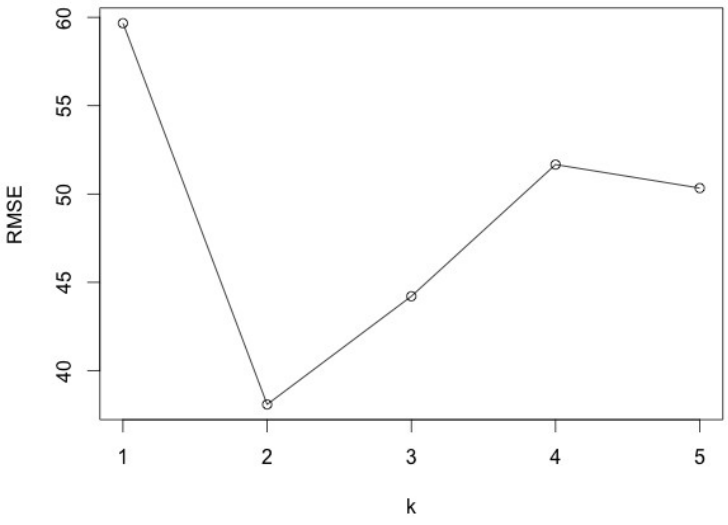

0.0
0.2
0.4
0.6
0.8
1.0
False positive rate

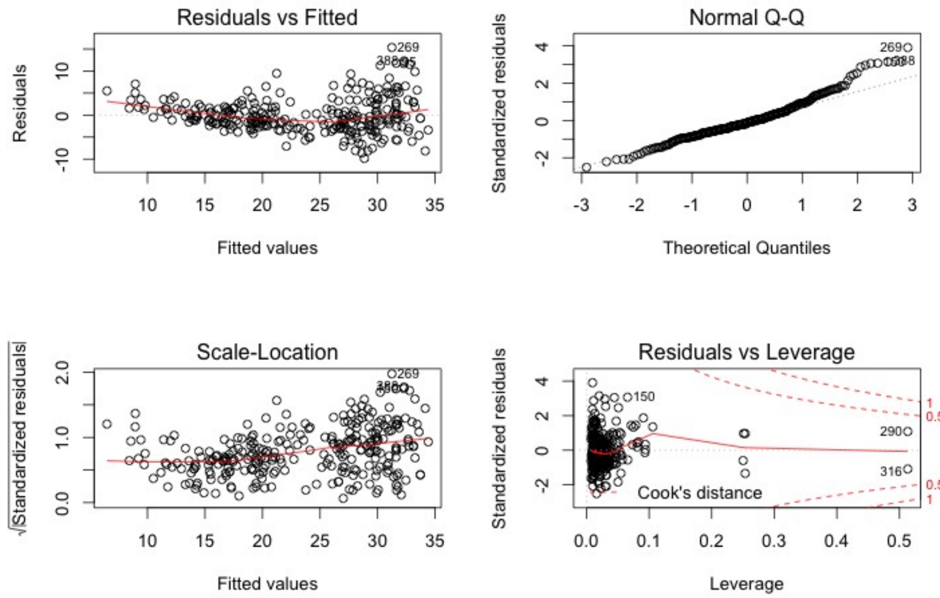

SVM classification plot
skew
variance
10
5
0
-5
-10
-6
-4
-2
0
2
4
6
1
0


SVM classification plot
skew
variance
10
5
0
-5
-10
-6
-4
-2
0
2
4
6
1
0

Chapter 4


Predicted
10
20
30
40
50
Actual
10
20
30
40
50


RMSE
40
45
50
55
60
k
1
2
3
4
5




```
Residuals:
    Min      1Q  Median      3Q     Max
-9.8488 -2.4015 -0.5022  1.8422 15.3597

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   39.4504219  3.3806186  11.670  < 2e-16 ***
cylinders4cyl  6.4665111  2.1248876   3.043  0.00257 **
cylinders5cyl  4.7697941  3.5603033   1.340  0.18146
cylinders6cyl  1.96741[illegible]  0.42803
cylinders8cyl  6.29193[illegible]5  0.03451 *
displacement   0.00478[illegible]9  0.66100
horsepower    -0.0816418  0.0200237  -4.077 5.99e-05 ***
weight        -0.0046663  0.0009857  -4.734 3.55e-06 ***
acceleration   0.0035761  0.1426022   0.025  0.98001
---
Signif. codes:  0 '***[illegible]' 0.05 '.' 0.1
' ' 1

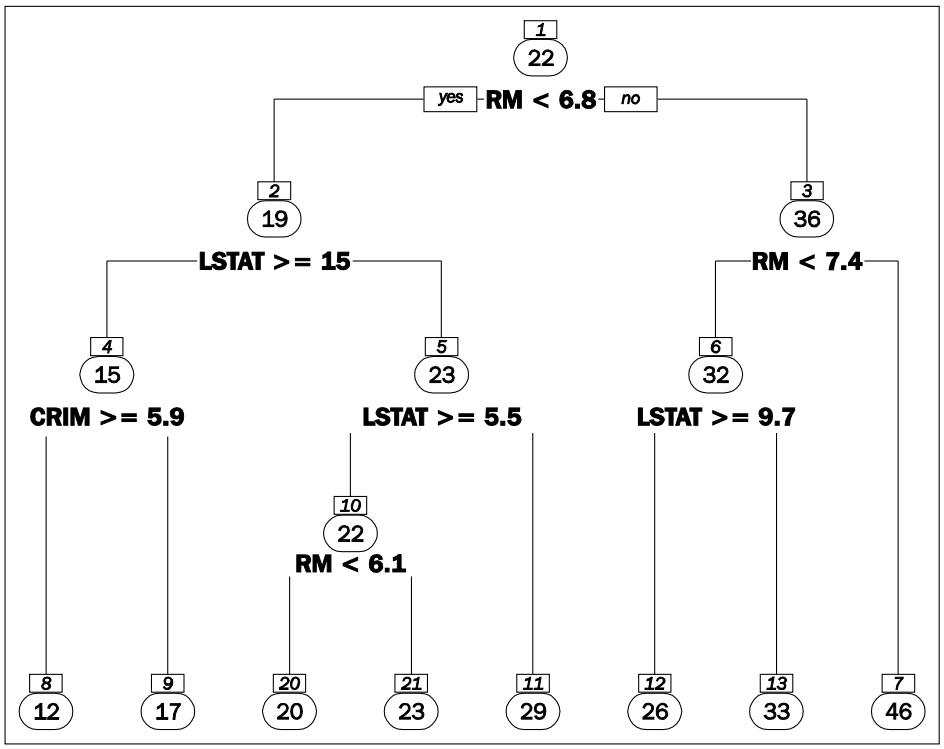
Residual standard error[illegible]rees of freedom
Multiple R-squared:  0[illegible] R-squared:
0.7488
F-statistic:   105 on 8 and 271 DF,  p-value: < 2.2e-16
```

Residual quartiles

Information about regression coefficients

Legend for significance codes used above

Model performance summary



$$Y \sim P$$

$$\hat{Y} \sim \beta_0 + \beta_1 P$$

$$Y \sim P + Q$$

$$\hat{Y} \sim \beta_0 + \beta_1 P + \beta_2 Q$$

$$Y \sim -1 + P$$

$$\hat{Y} \sim \beta_1 P$$

$$Y \sim P : Q$$

$$\hat{Y} \sim \beta_0 + \beta_1 PQ$$

$$Y \sim P * Q$$

$$\hat{Y} \sim \beta_0 + \beta_1 P + \beta_2 Q + \beta_3 PQ$$

$$Y \sim P + Q + P : Q$$

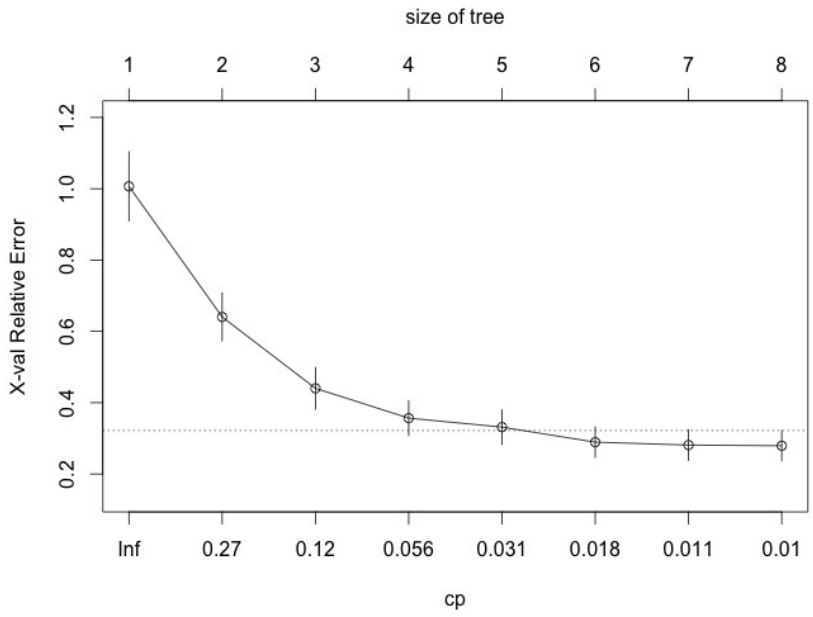
$$Y \sim P + I\left(\log\left(Q\right)\right)$$

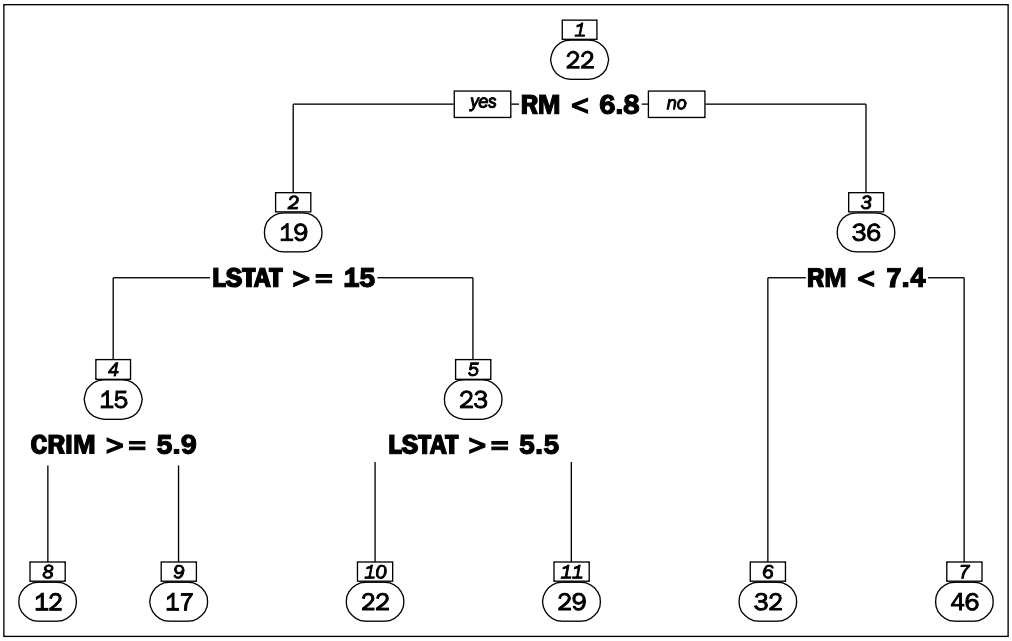
$$Y \sim \left(P + Q + R\right) \wedge 2$$

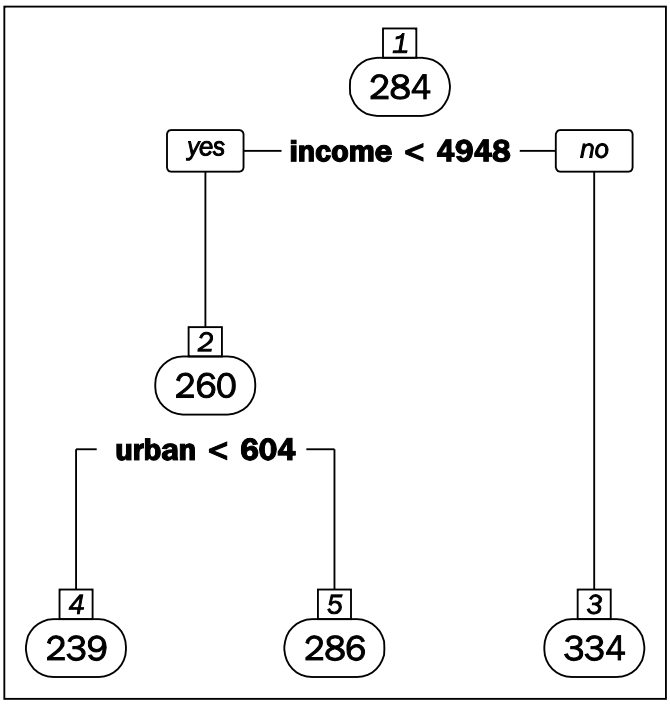
$$\hat{Y} \sim \beta_0 + \beta_1 P + \beta_2 Q + \beta_3 R + \beta_4 PQ + \beta_5 QR + \beta_6 PR$$

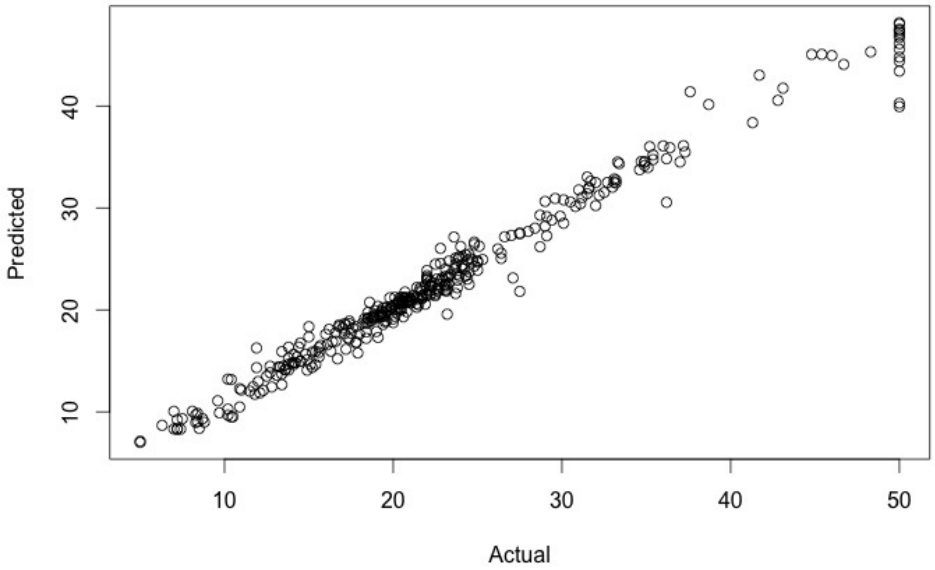
$$Y \sim P * Q * R - P : Q : R$$

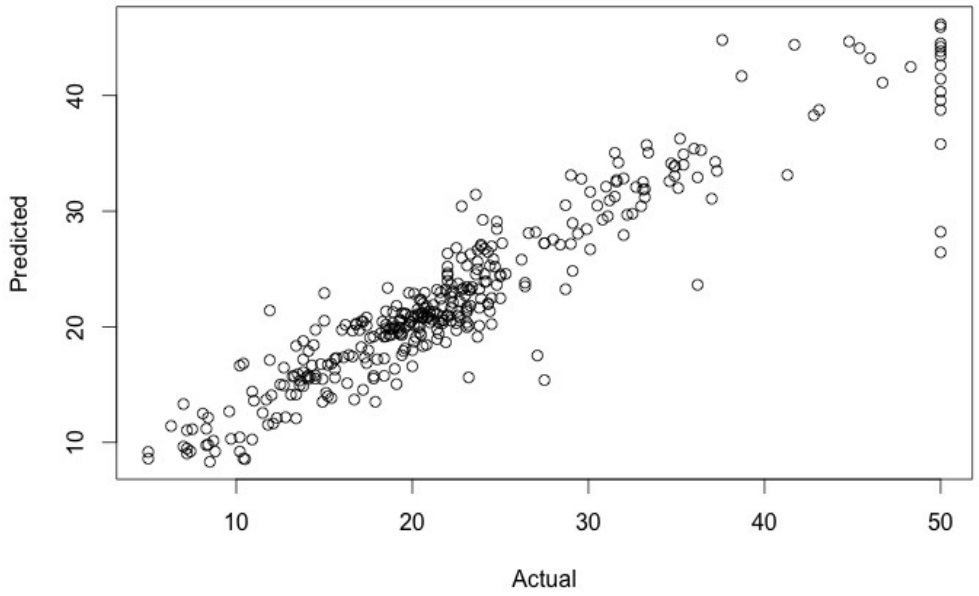
$$\hat{Y} \sim \beta_0 + \beta_1 P + \beta_2 \log\left(Q\right)$$

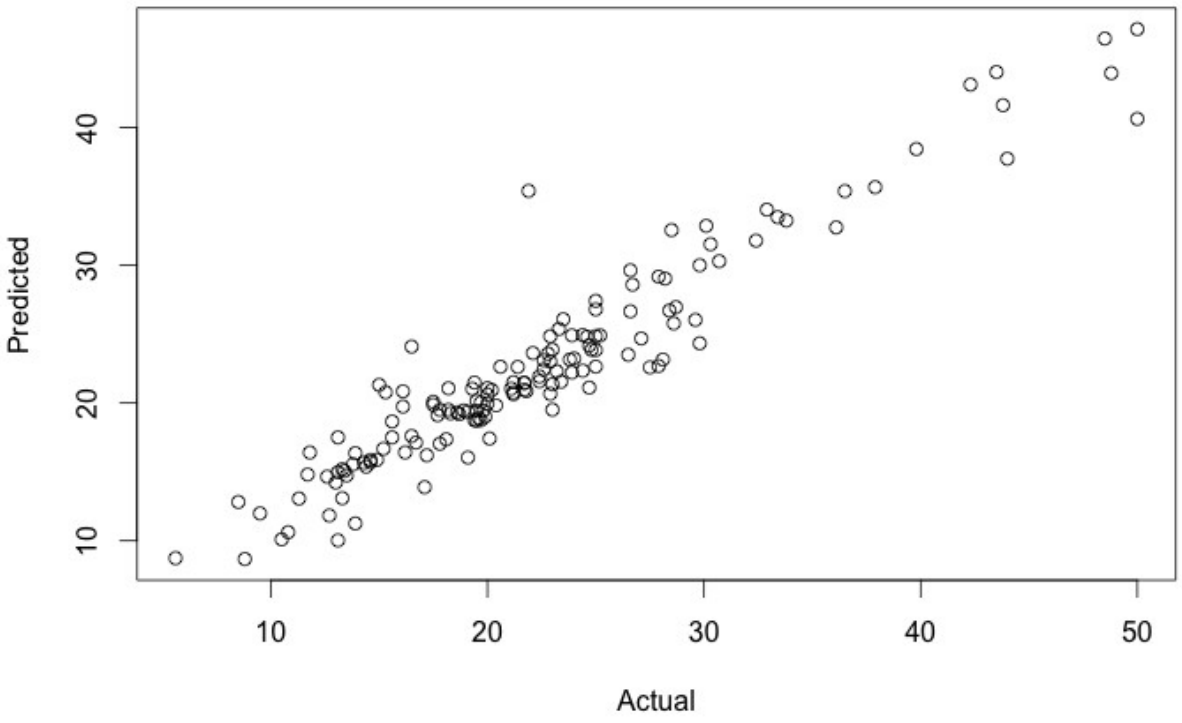

size of tree
1
2
3
4
5
6
7
8
X-val Relative Error
0.2
0.4
0.6
0.8
1.0
1.2
Inf
0.27
0.12
0.056
0.031
0.018
0.011
0.01
cp

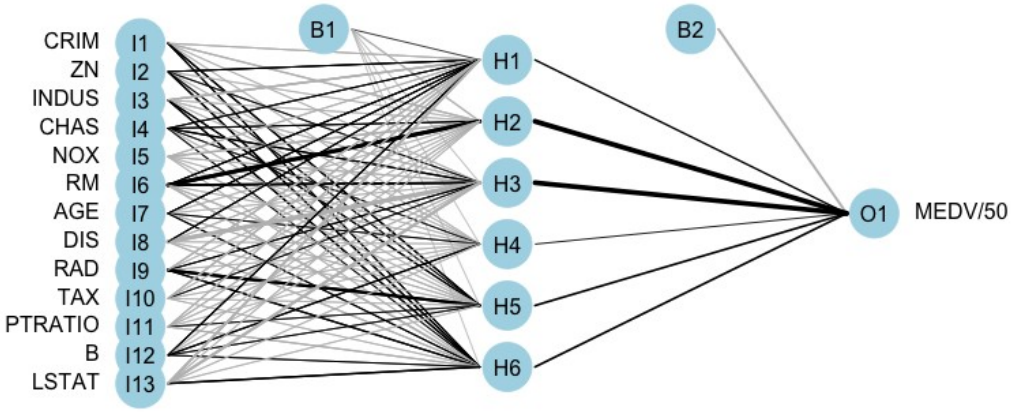

1
22
yes
RM < 6.8
no
2
19
3
36
LSTAT >= 15
RM < 7.4
4
15
5
23
CRIM >= 5.9
LSTAT >= 5.5
8
12
9
17
10
22
11
29
6
32
7
46


1
284
yes
income < 4948
no
2
260
urban < 604
4
239
5
286
3
334

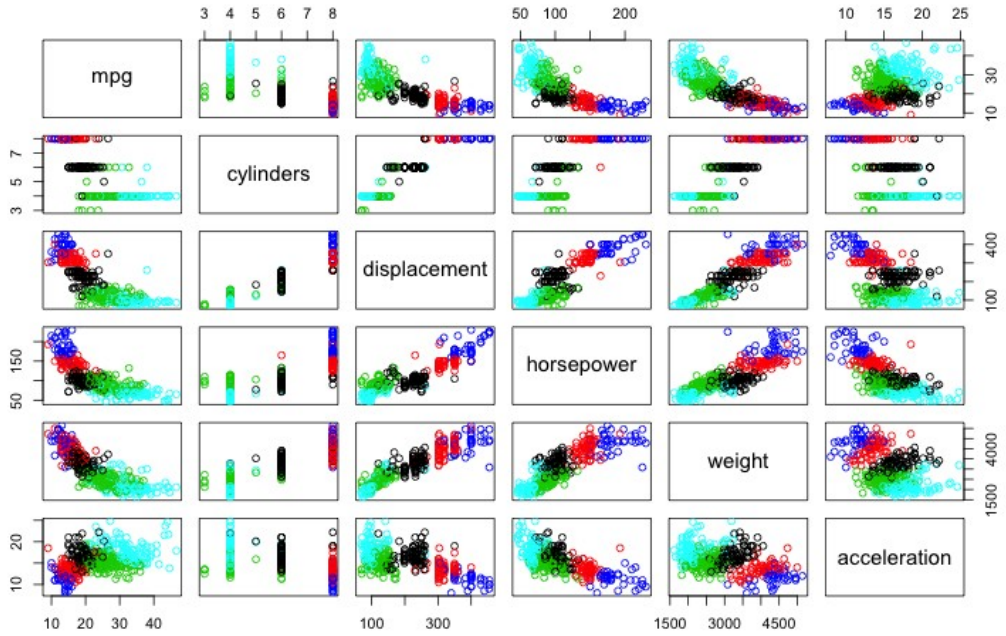

Predicted
10
20
30
40
Actual
10
20
30
40
50

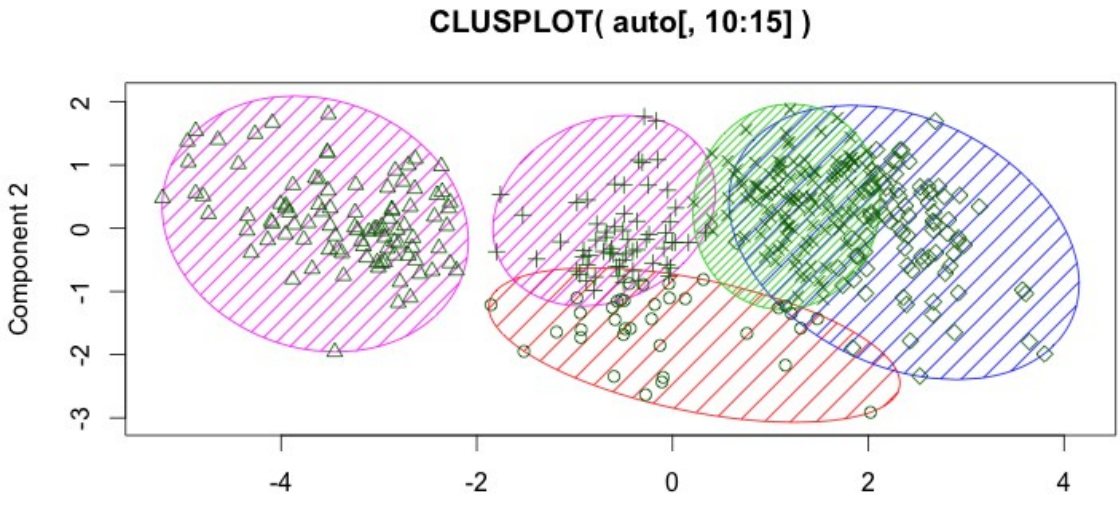

Predicted
10
20
30
40
10
20
30
40
50
Actual

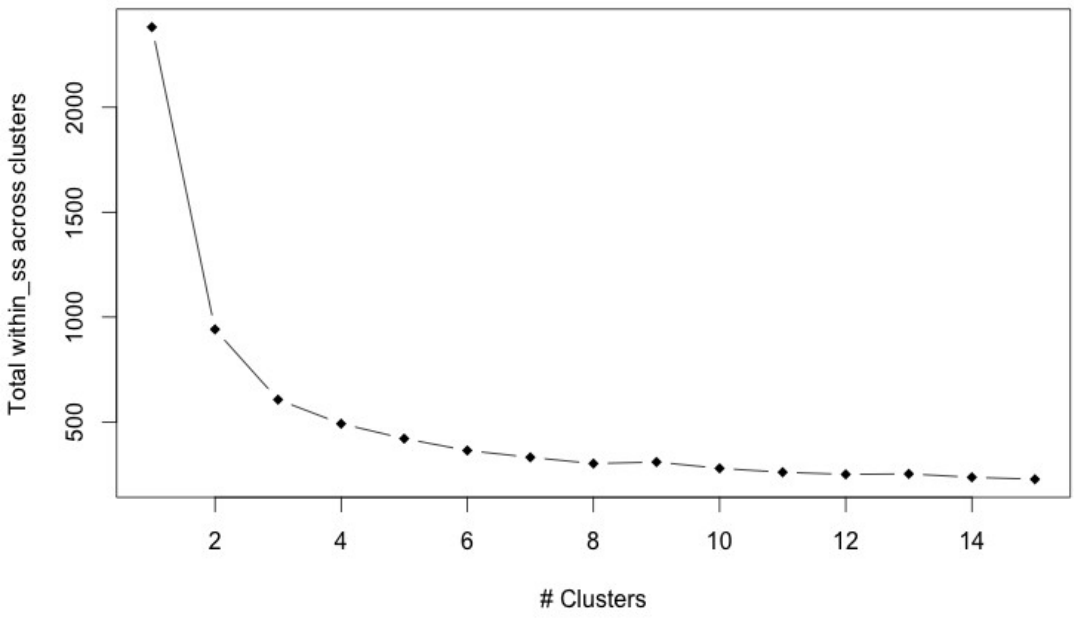

Predicted
10
20
30
40
10
20
30
40
50
Actual

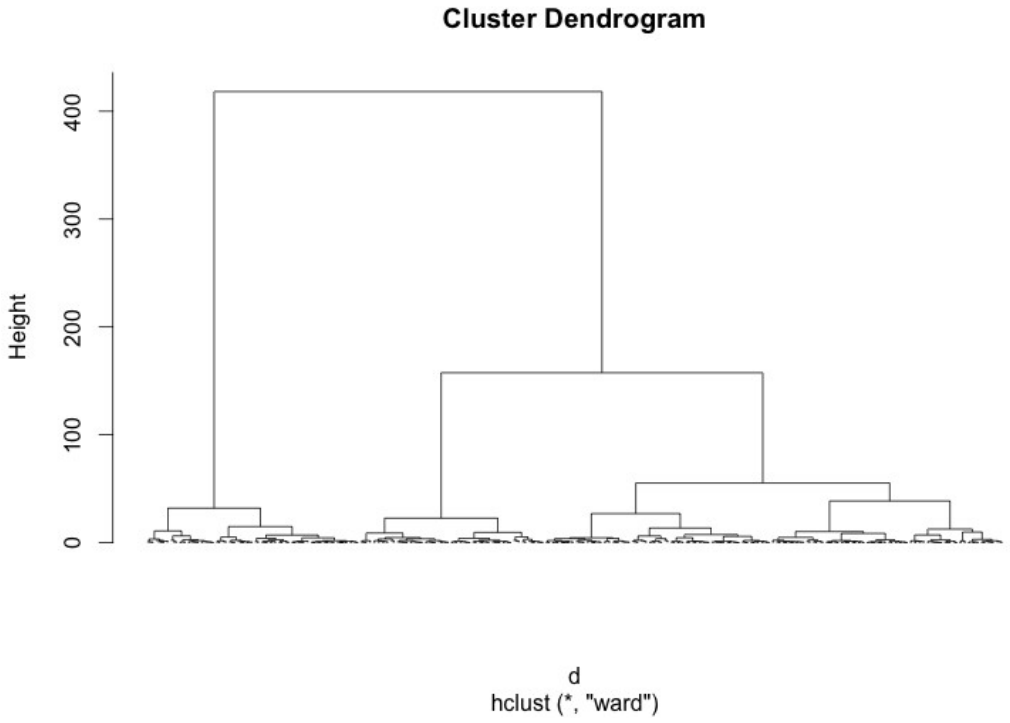

CRIM
ZN
INDUS
CHAS
NOX
RM
AGE
DIS
RAD
TAX
PTRATIO
B
LSTAT
I1
I2
I3
I4
I5
I6
I7
I8
I9
I10
I11
I12
I13
B1
H1
H2
H3
H4
H5
H6
B2
O1
MEDV/50

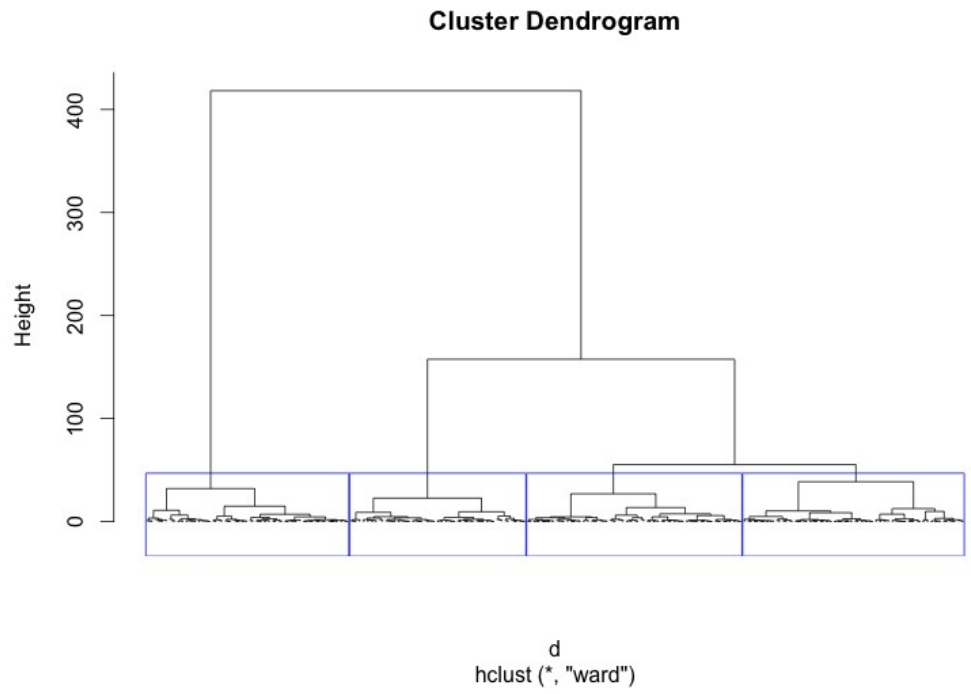
Chapter 5

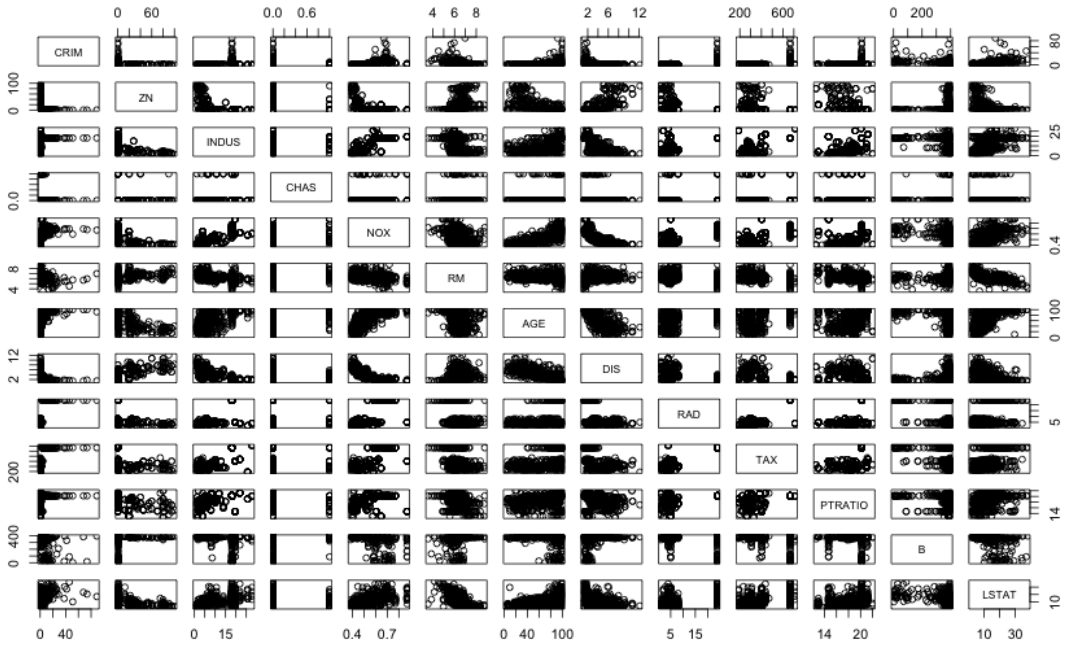

mpg
cylinders
displacement
horsepower
weight
acceleration

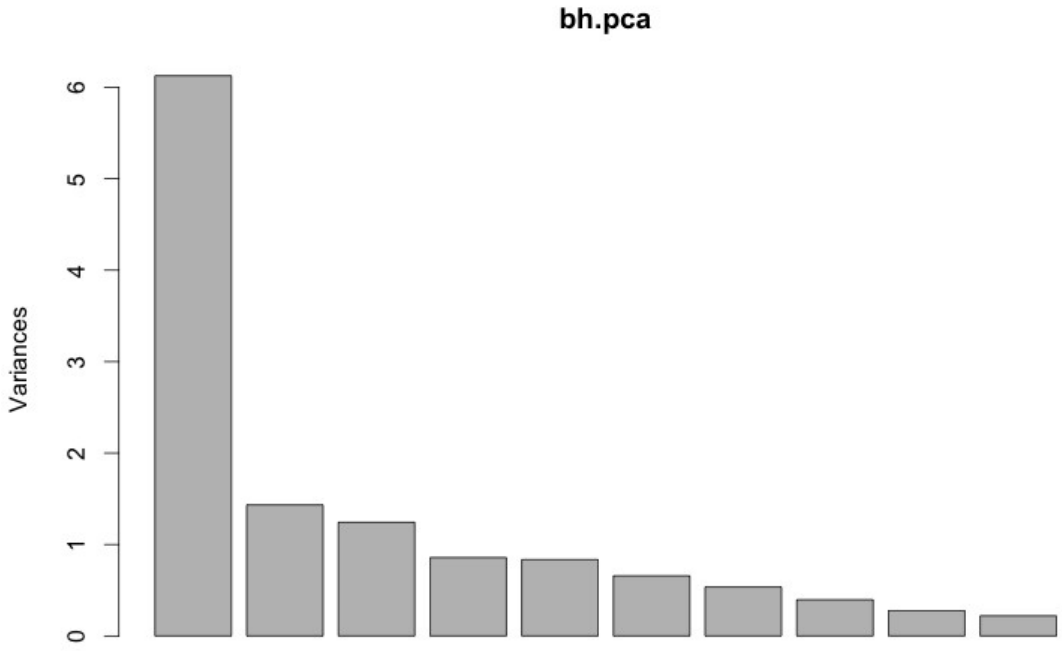

CLUSPLOT(auto[, 10:15])
Component 2
Component 1
These two components explain 91.9 % of the point variability.

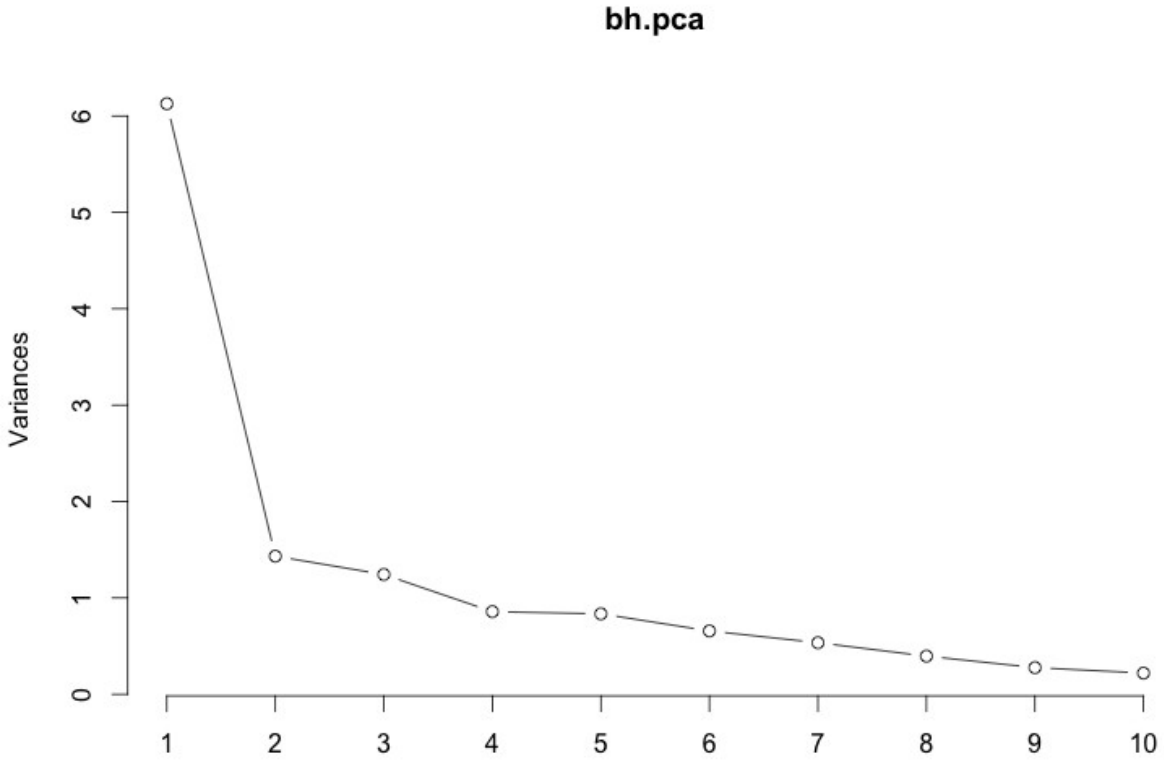

Total within_ss across clusters
500
1000
1500
2000
2
4
6
8
10
12
14
Clusters


Cluster Dendrogram
Height
0
100
200
300
400
d
hclust (*, "ward")

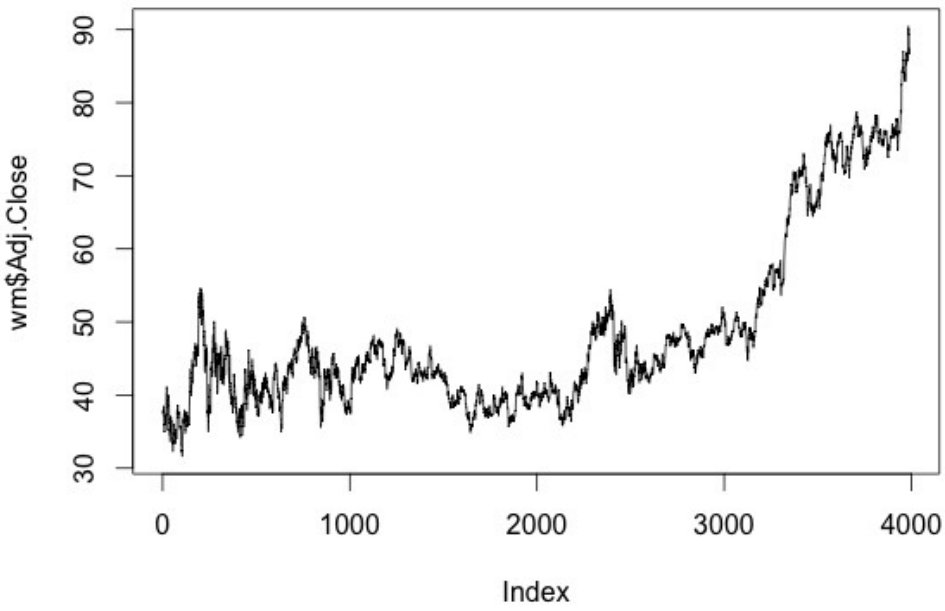

Cluster Dendrogram
Height
0
100
200
300
400
d
hclust (*, "ward")

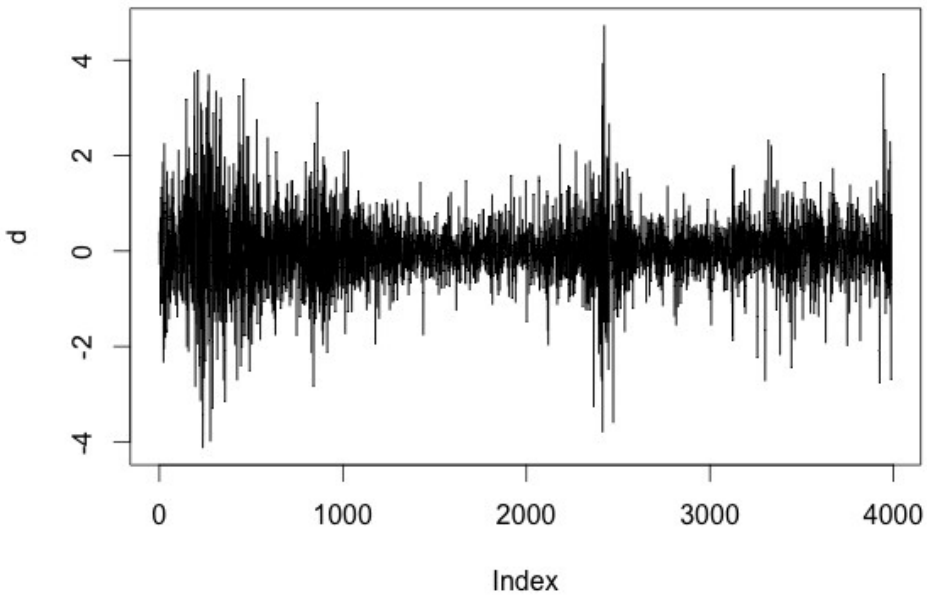

CRIM
ZN
INDUS
CHAS
NOX
RM
AGE
DIS
RAD
TAX
PTRATIO
B
LSTAT

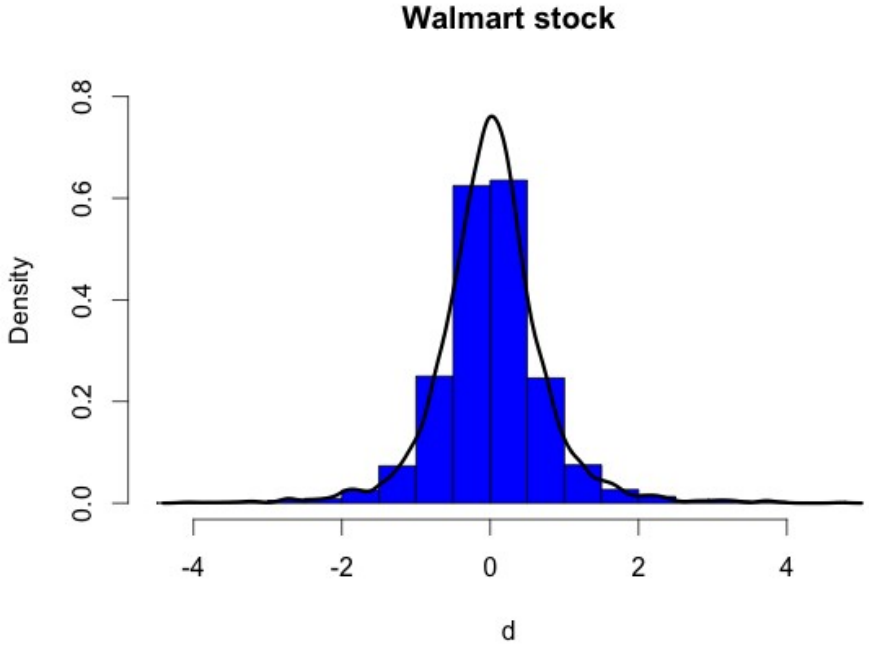

bh.pca
Variances
0
1
2
3
4
5
6

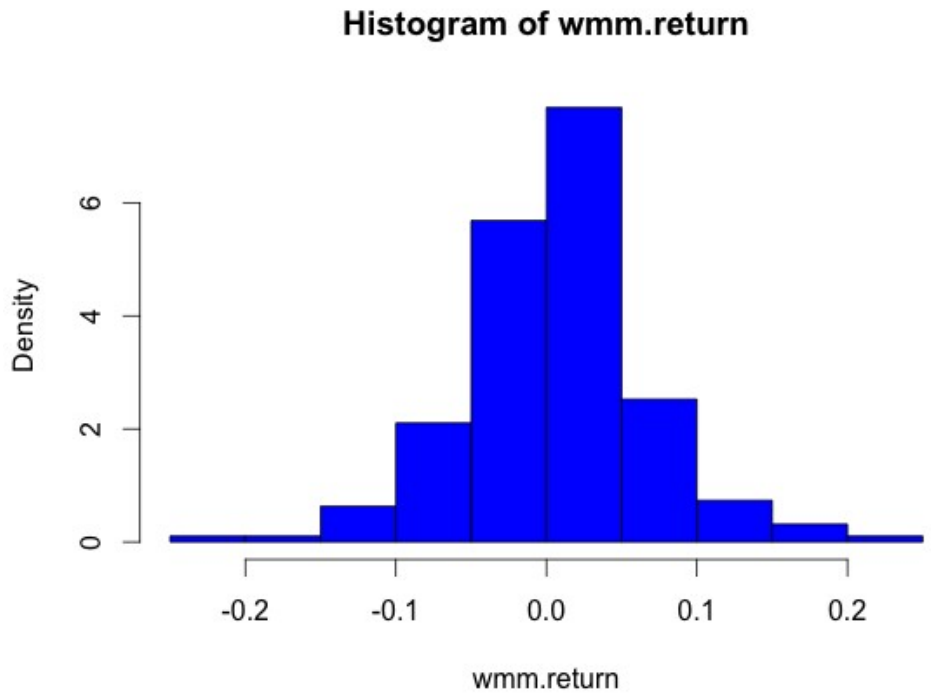

bh.pca
Variances
0
1
2
3
4
5
6
1
2
3
4
5
6
7
8
9
10

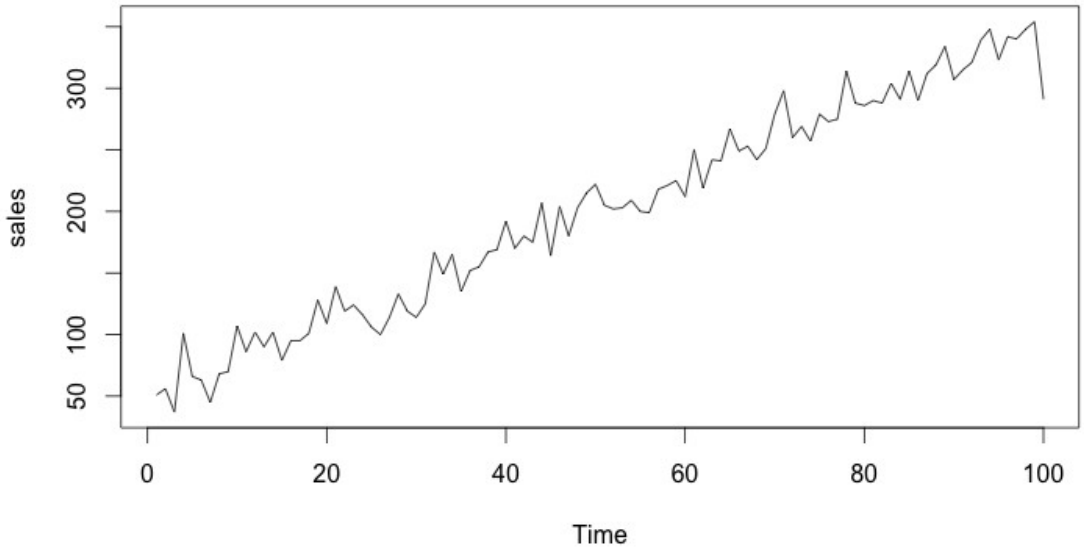
Chapter 6

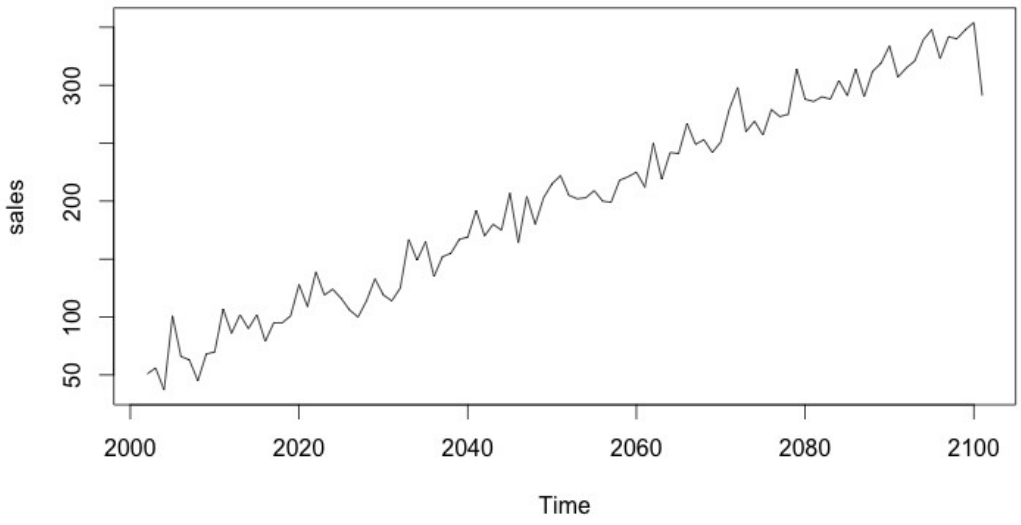


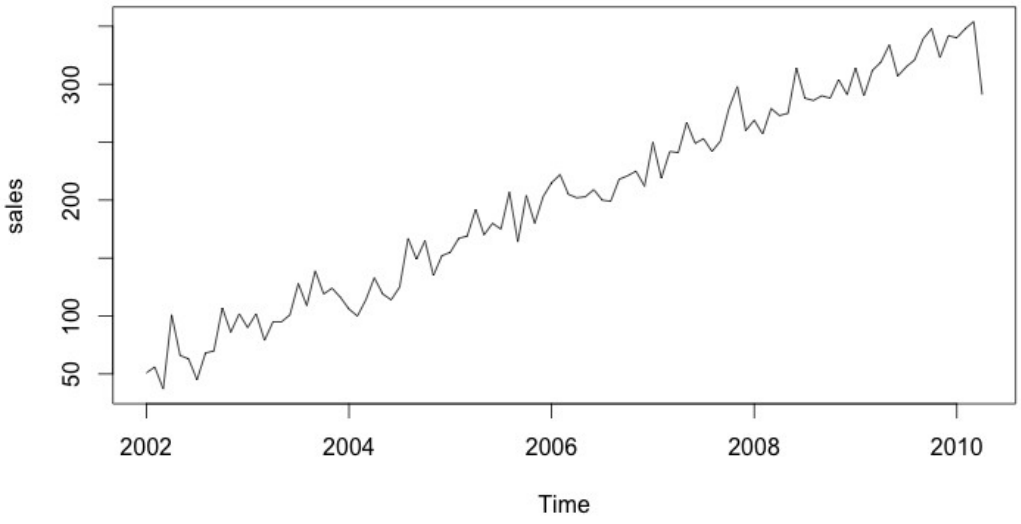

d
4
2
0
-2
-4
0
1000
2000
3000
4000
Index

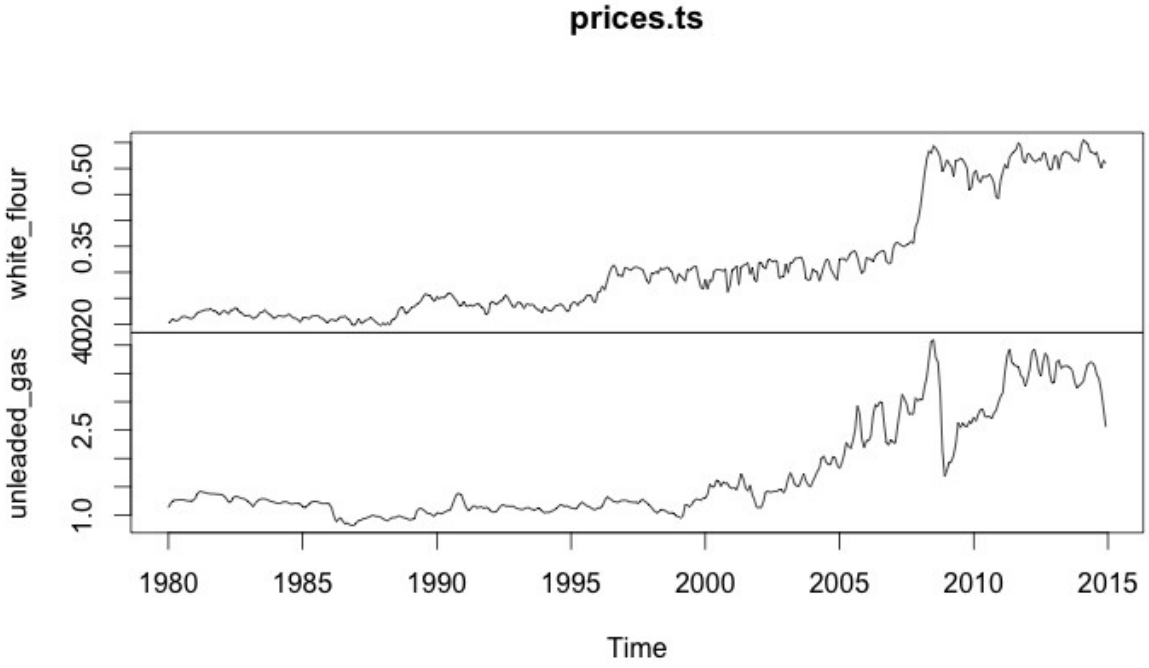

Walmart stock
Density
0.8
0.6
0.4
0.2
0.0
-4
-2
0
2
4
d

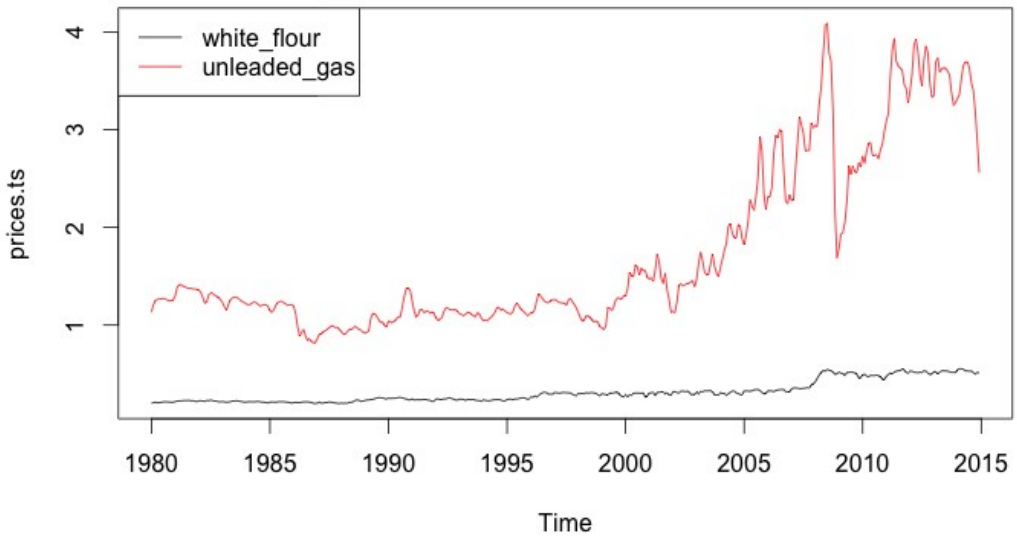

Histogram of wmm.return
Density
0
2
4
6
-0.2
-0.1
0.0
0.1
0.2
wmm.return

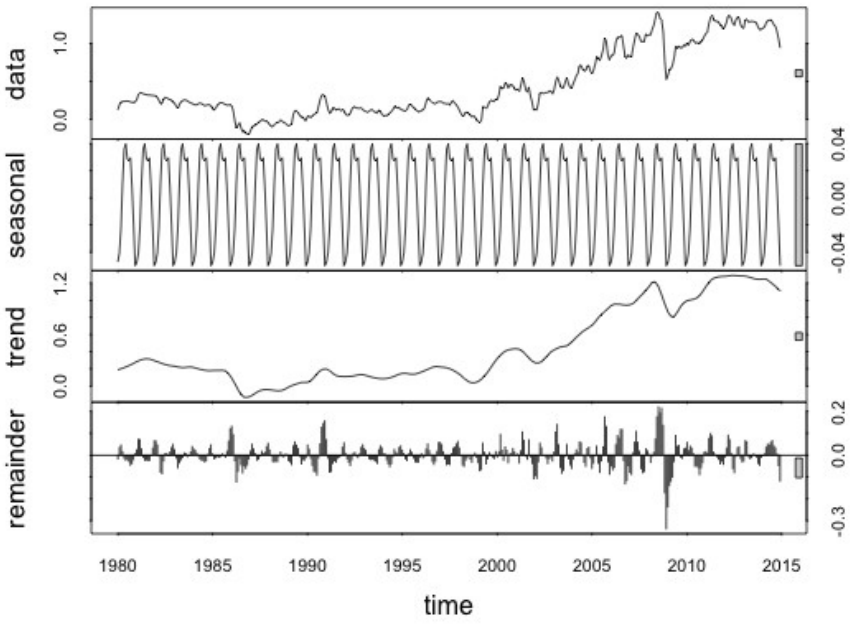

sales
50
100
200
300
0
20
40
60
80
100
Time

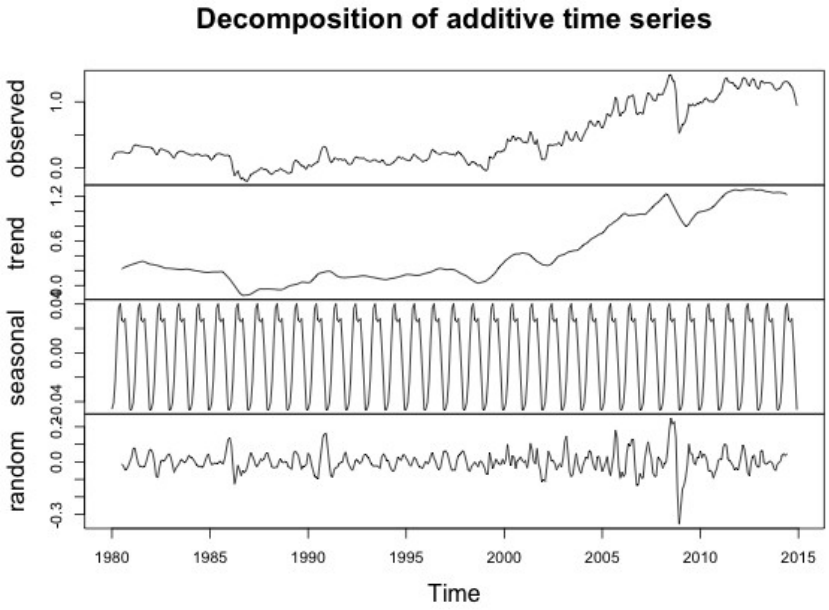

sales
50
100
200
300
2000
2020
2040
2060
2080
2100
Time

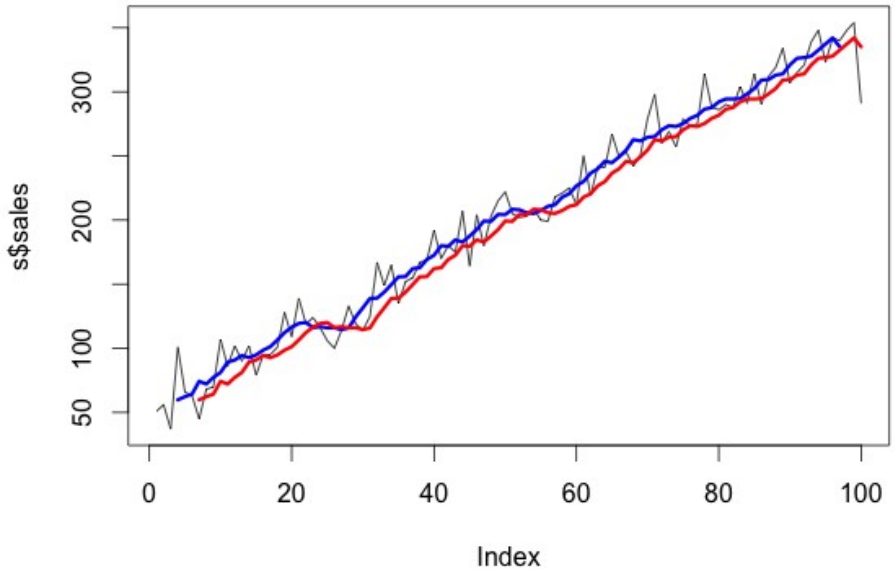

sales
50
100
200
300
2002
2004
2006
2008
2010
Time

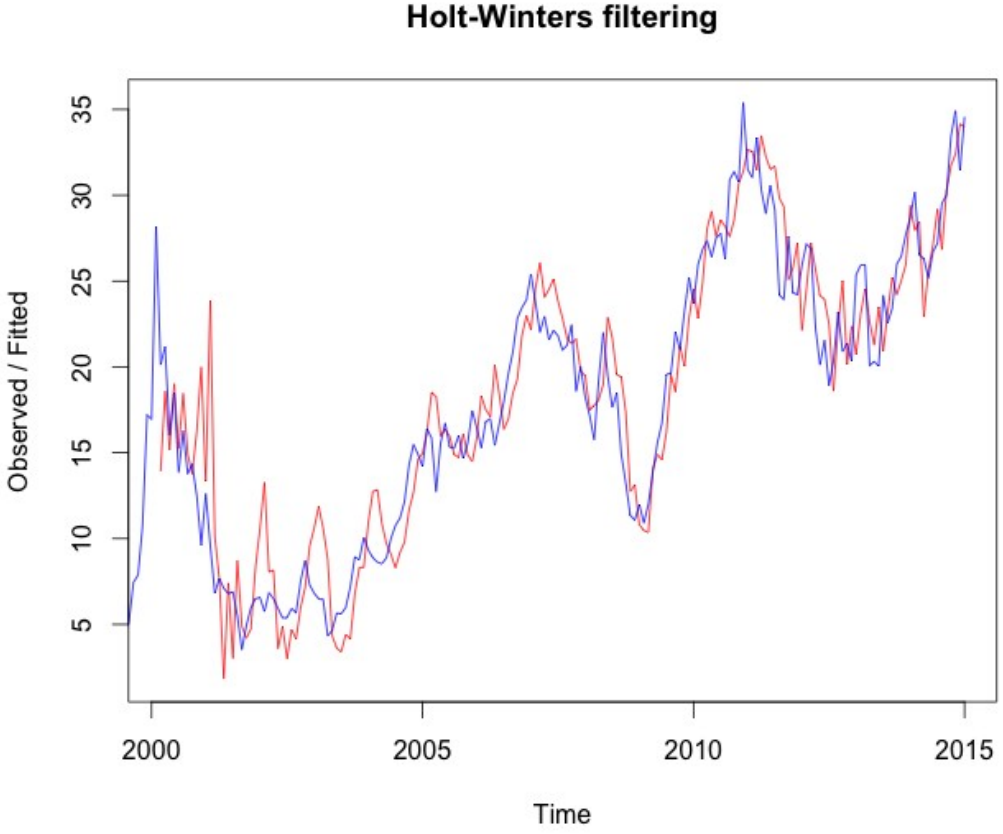

prices.ts
white_flour
unleaded_gas
0.50
0.35
4
3
2
1.0
2.5
1980
1985
1990
1995
2000
2005
2010
2015
Time

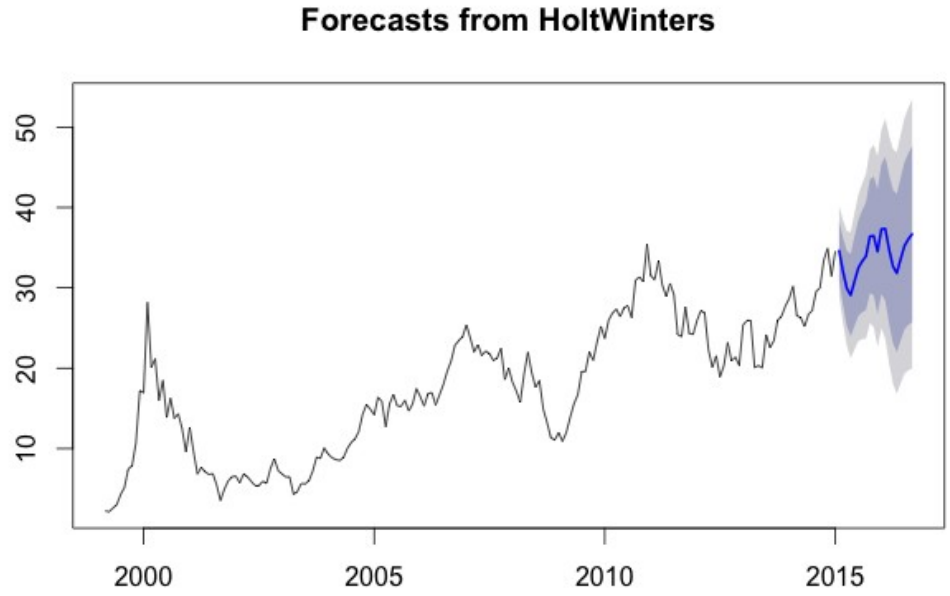

white_flour
unleaded_gas
prices.ts
4
3
2
1
1980
1985
1990
1995
2000
2005
2010
2015
Time

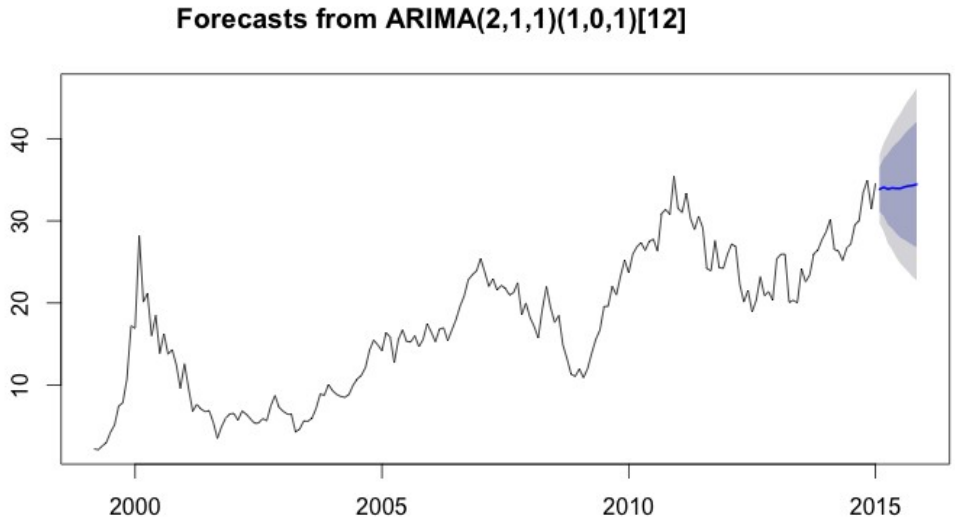

data
seasonal
trend
remainder
1.0
0.0
0.04
0.00
-0.04
1.2
0.6
0.0
0.2
0.0
-0.3
1980
1985
1990
1995
2000
2005
2010
2015
time


Decomposition of additive time series
observed
trend
seasonal
random
1.0
0.0
1.2
0.6
0.0
0.04
0.00
0.04
0.2
0.0
-0.3
1980
1985
1990
1995
2000
2005
2010
2015
Time


s$sales
50
100
200
300
0
20
40
60
80
100
Index


Holt-Winters filtering
Observed / Fitted
5
10
15
20
25
30
35
2000
2005
2010
2015
Time


Forecasts from HoltWinters
10
20
30
40
50
2000
2005
2010
2015


Forecasts from ARIMA(2,1,1)(1,0,1)[12]
10
20
30
40
2000
2005
2010
2015

Chapter 7

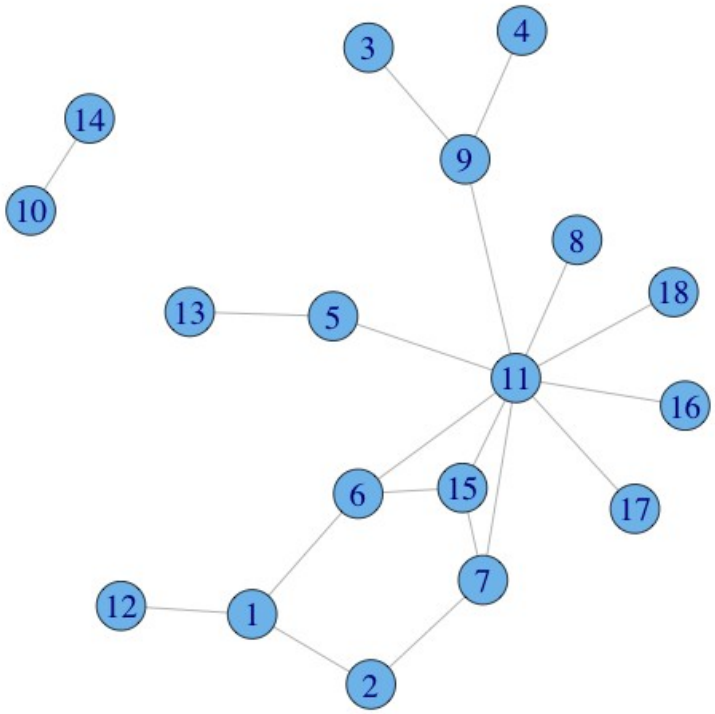
Users

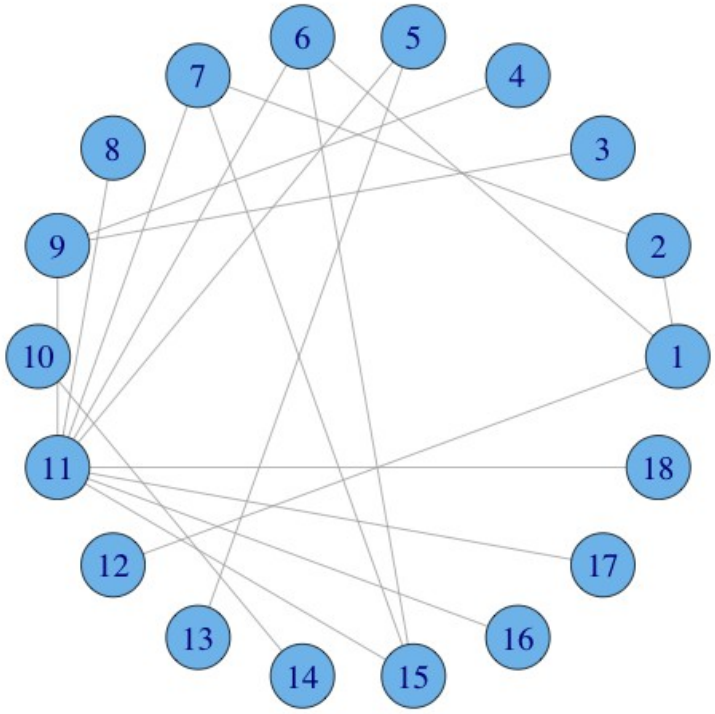
Groups	1	2	3	4	5	6	7	8	9
1	1	.	.	1	.	.	1	.	.
2	1	.	1	1	.	1	.	.	.
3	1	1	.	1	.	1	1	.	1
4	.	1	1	.	.	1	.	.	1

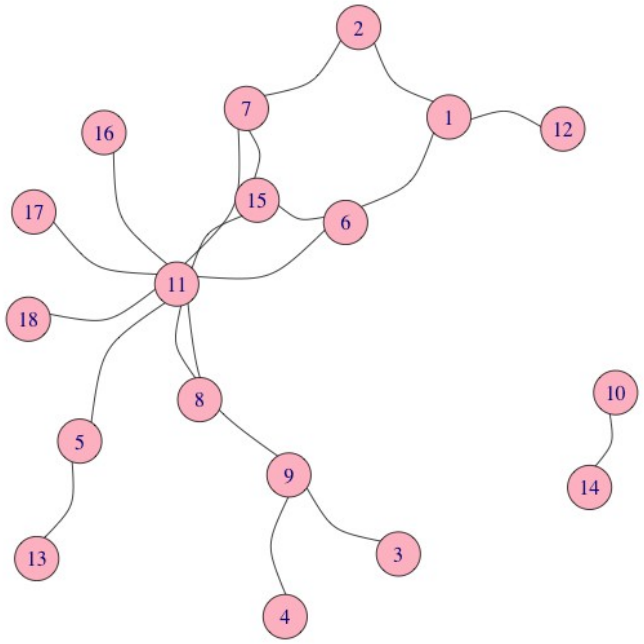
Users

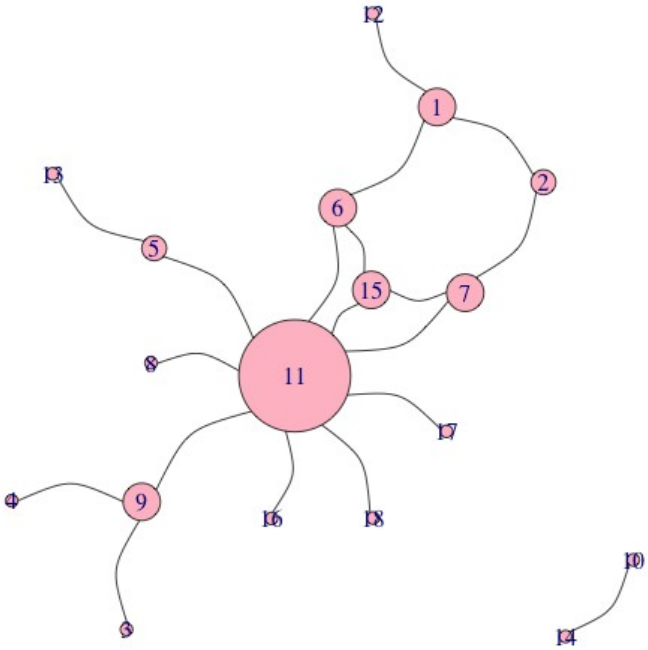
Users	1	2	3	4	5	6	7	8	9
1	3	1	1	3	0	2	2	0	1
2	1	2	1	1	0	2	1	0	2
3	1	1	2	1	0	2	0	0	1
4	3	1	1	3	0	2	2	0	1
5	0	0	0	0	0	0	0	0	0
6	2	2	2	2	0	3	1	0	2
7	2	1	0	2	0	1	2	0	1
8	0	0	0	0	0	0	0	0	0
9	1	2	1	1	0	2	1	0	2

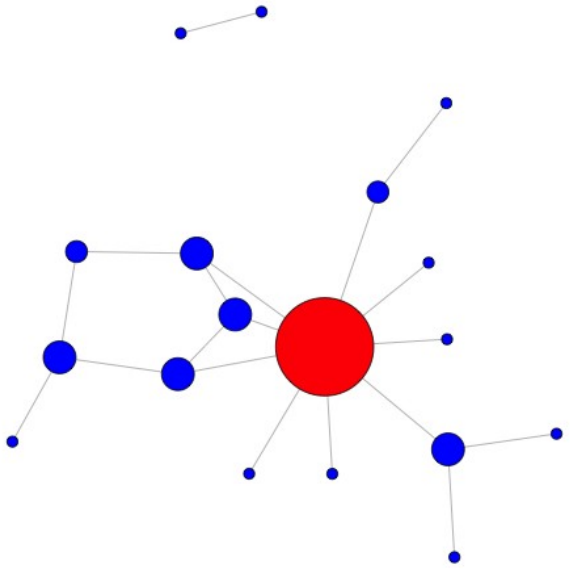
i	j	x
1	1	3
2	1	1
3	1	1
4	1	3
6	1	2
7	1	2
9	1	1
1	2	1
2	2	2
3	2	1
...	...	...

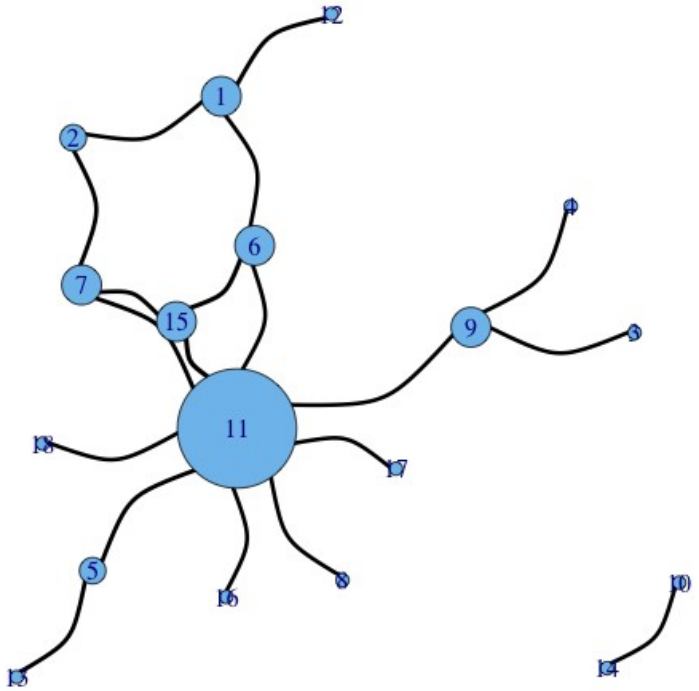

3
4
14
9
10
8
18
13
5
11
16
6
15
17
7
12
1
2

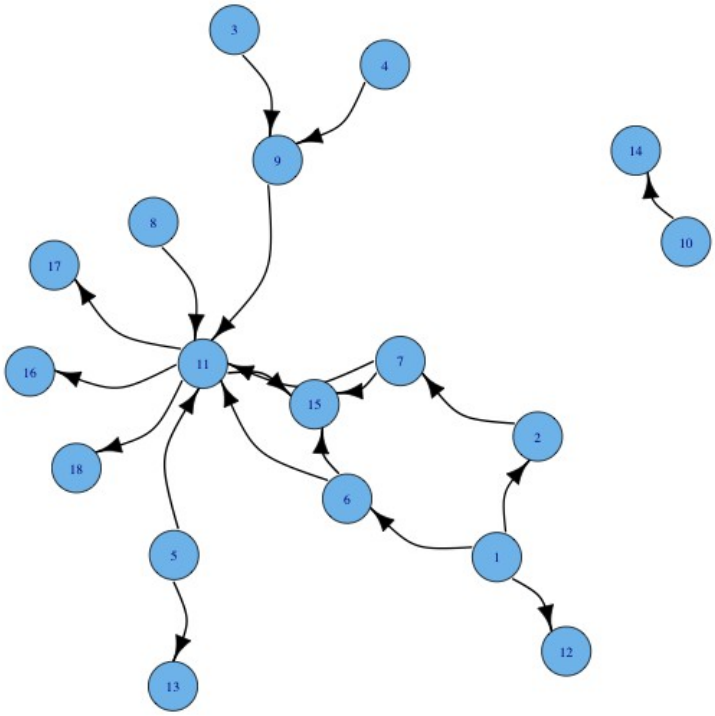

6
5
7
4
8
3
9
2
10
1
11
18
12
17
13
16
14
15

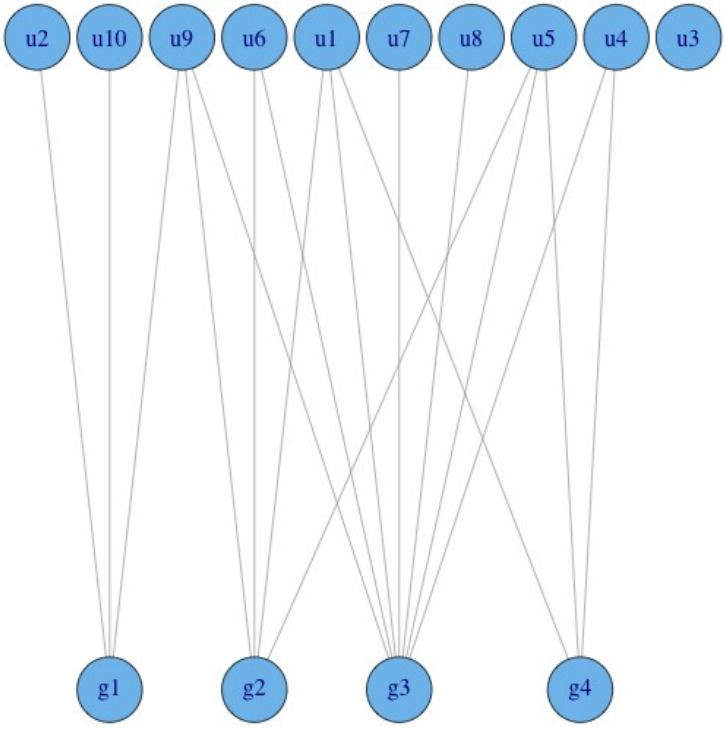

2
7
1
12
16
15
6
17
11
18
8
10
5
9
14
13
3
4

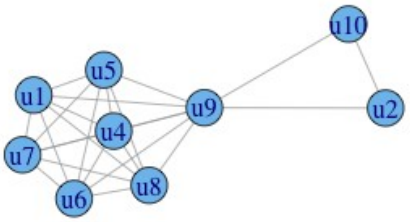

12
1
13
2
6
5
15
7
8
11
17
4
9
16
18
3
10
14



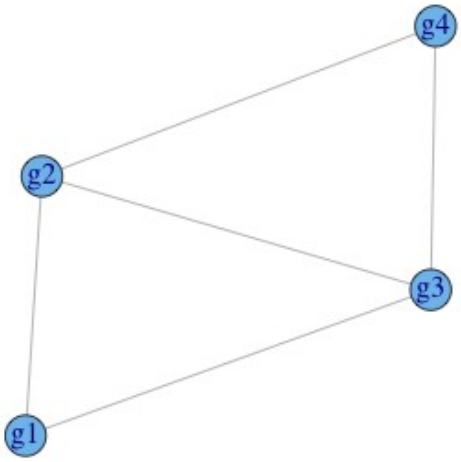

12
1
2
6
7
15
9
4
3
11
18
17
5
16
8
13
10
14


3
4
9
14
10
8
17
11
7
16
15
2
18
6
5
1
12
13


u2
u10
u9
u6
u1
u7
u8
u5
u4
u3
g1
g2
g3
g4


u10
u5
u1
u9
u2
u4
u7
u8
u6


u3


g4
g2
g3
g1

Chapter 8

```
---
title: "Introduction"
author: "Shanthi Viswanathan"
date: "March 21, 2015"
output: html_document
---

This is an R Markdown document. Markdown is a simple formatting
syntax for authoring HTML, PDF, and MS Word documents. For more
details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that
includes both content as well as the output of any embedded R code
chunks within the document. You can embed an R code chunk like this:

```{r}
summary(cars)
```

You can also embed plots, for example:

```{r, echo=FALSE}
plot(cars)
```

Note that the `echo = FALSE` parameter was added to the code chunk
to prevent printing of the R code that generated the plot.
```

Markdown Document

Shanthi Viswanathan

December 8, 2014

Introduction

This is an *R Markdown document*. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

HTML Content

This is a new paragraph written with the HTML tag

Pros	Cons
Easy to use	Need to Plan ahead

Embed Code

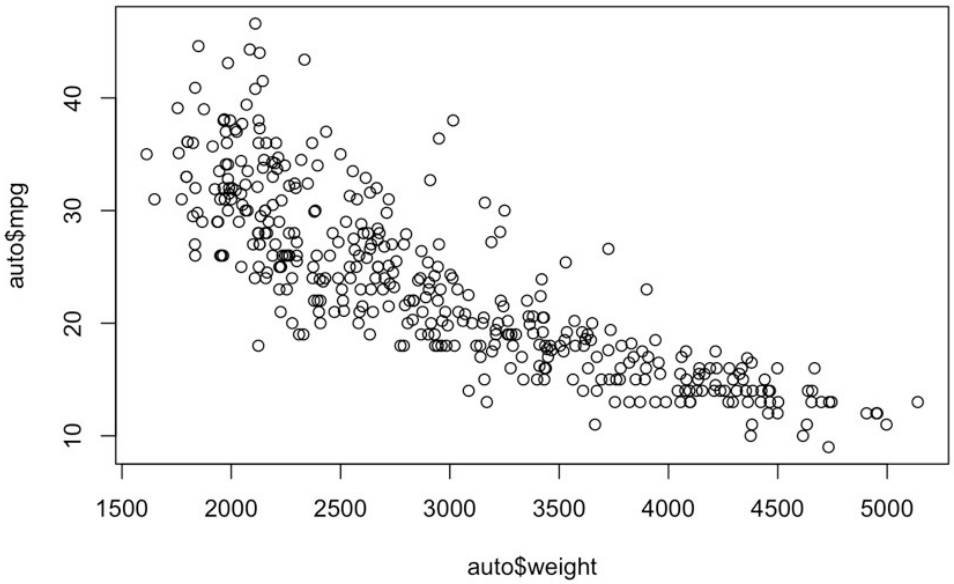
Set Directory

You can embed any R code chunk within 3 ticks. If you add echo=FALSE the code chunk is not displayed in the document. We can set knitr options either globally or within a code segment. The options set globally are used throughout the document.

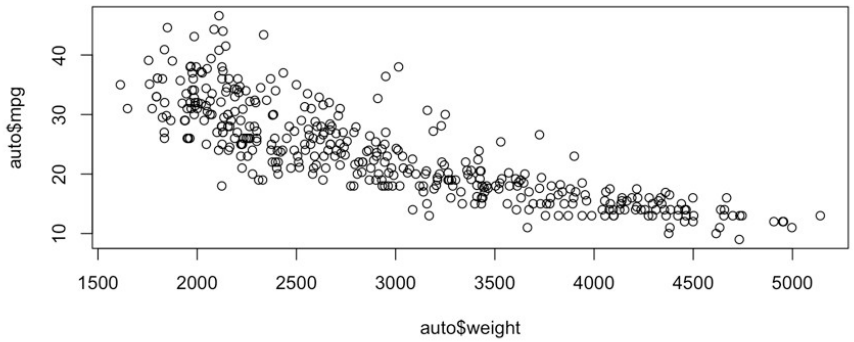
We set the root.dir before loading any files. By enabling cache=TRUE, a code chunk is executed only when there is a change from the prior execution. This enhances knitr performance.

Plot Data

```
plot(auto$mpg~auto$weight)
```



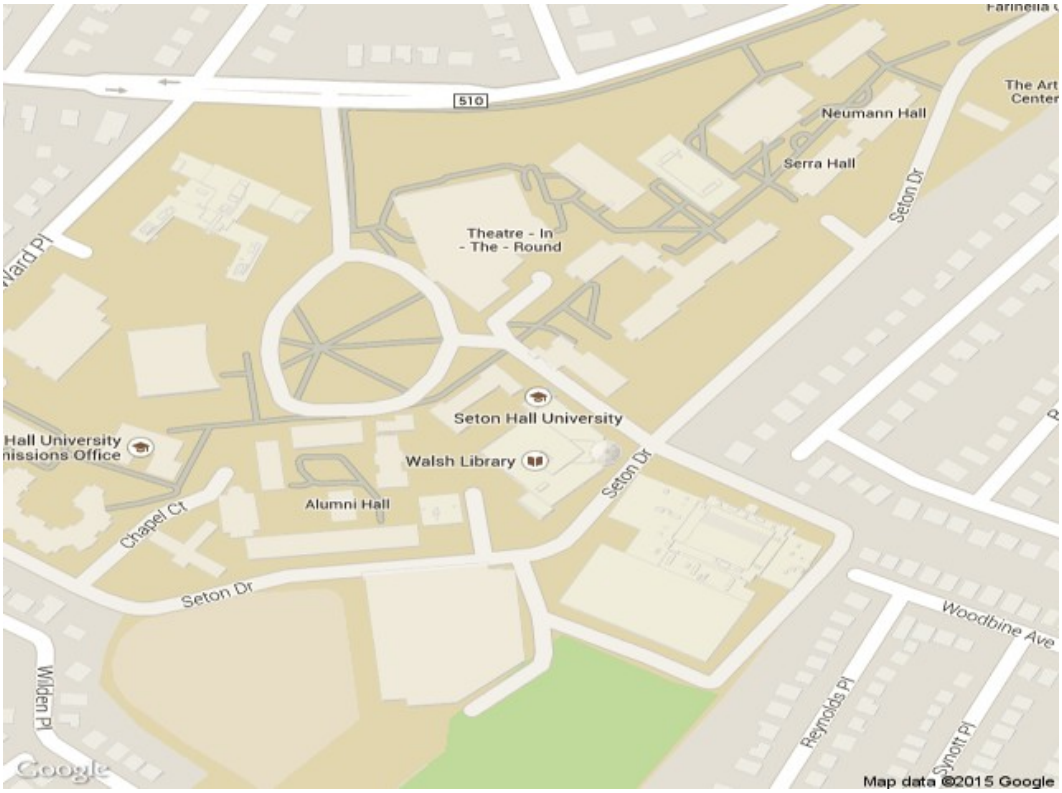
Plot with format options



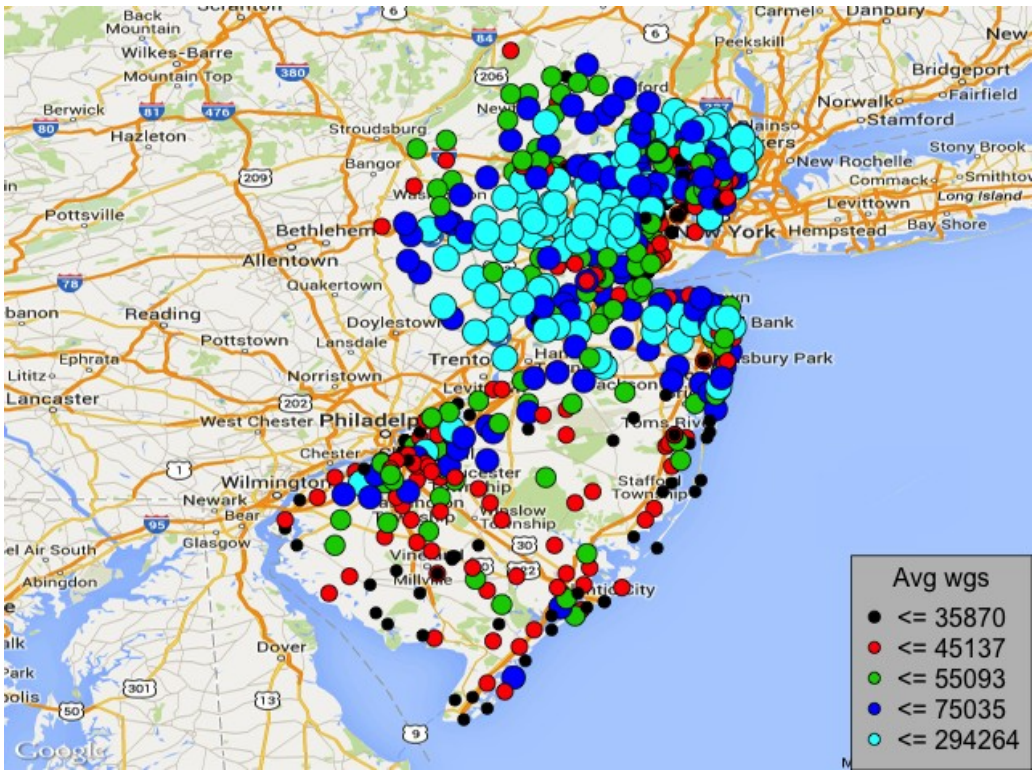

```
## 'data.frame':    398 obs. of  9 variables:
##  $ No           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ mpg          : num  28 19 36 28 21 23 15.5 32.9
16 13 ...
##  $ cylinders    : int  4 3 4 4 6 4 8 4 6 8 ...
##  $ displacement: num  140 70 107 97 199 115 304 1
19 250 318 ...
##  $ horsepower   : int  90 97 75 92 90 95 120 100 1
05 150 ...
##  $ weight       : int  2264 2330 2205 2288 2648 26
94 3962 2615 3897 3755 ...
##  $ acceleration: num  15.5 13.5 14.5 17 15 15 13.
9 14.8 18.5 14 ...
##  $ model_year   : int  71 72 82 72 70 75 76 81 75
76 ...
##  $ car_name     : Factor w/ 305 levels "amc ambass
ador brougham",..: 66 184 165 86 8 18 11 79 42 112 .
..
```

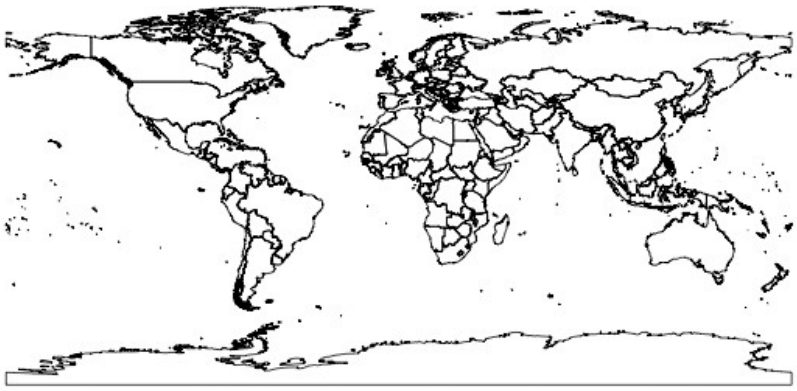
There are 398 cars in the auto data set.

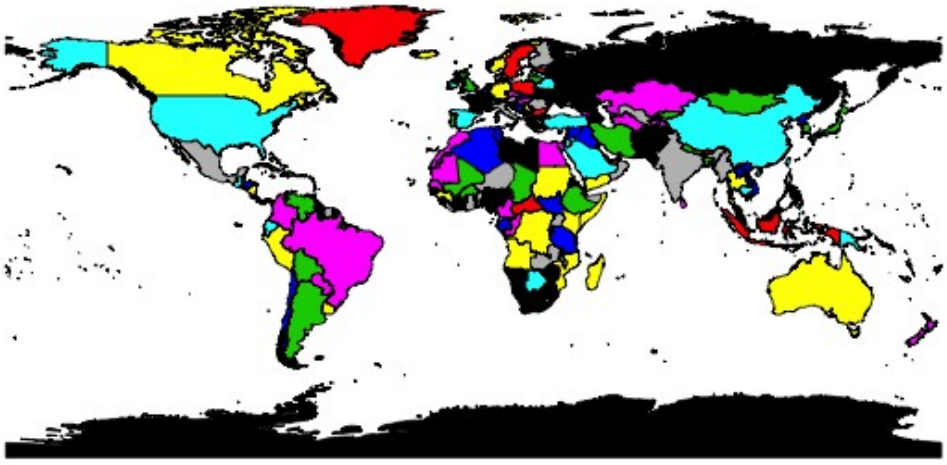
Chapter 10

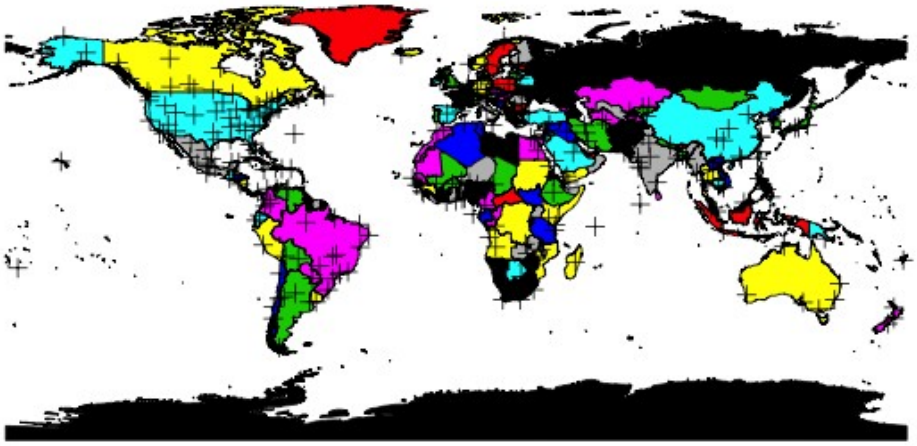

510
Neumann Hall
Serra Hall
Theatre - In - The - Round
Seton Dr
Seton Hall University
Walsh Library
Alumni Hall
Chapel Ct
Seton Dr
Seton Dr
Woodbine Ave
Reynolds Pl
Symott Pl
Widen Pl
Google
Map data ©2015 Google

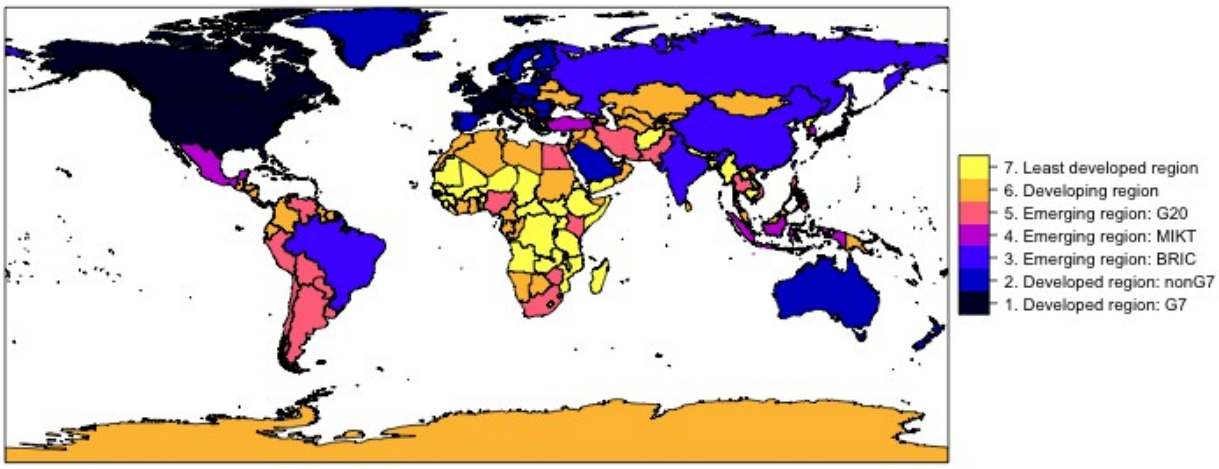

Google
Imagery ©2015 , Bluesky, DigitalGlobe, USDA Farm Service Agency

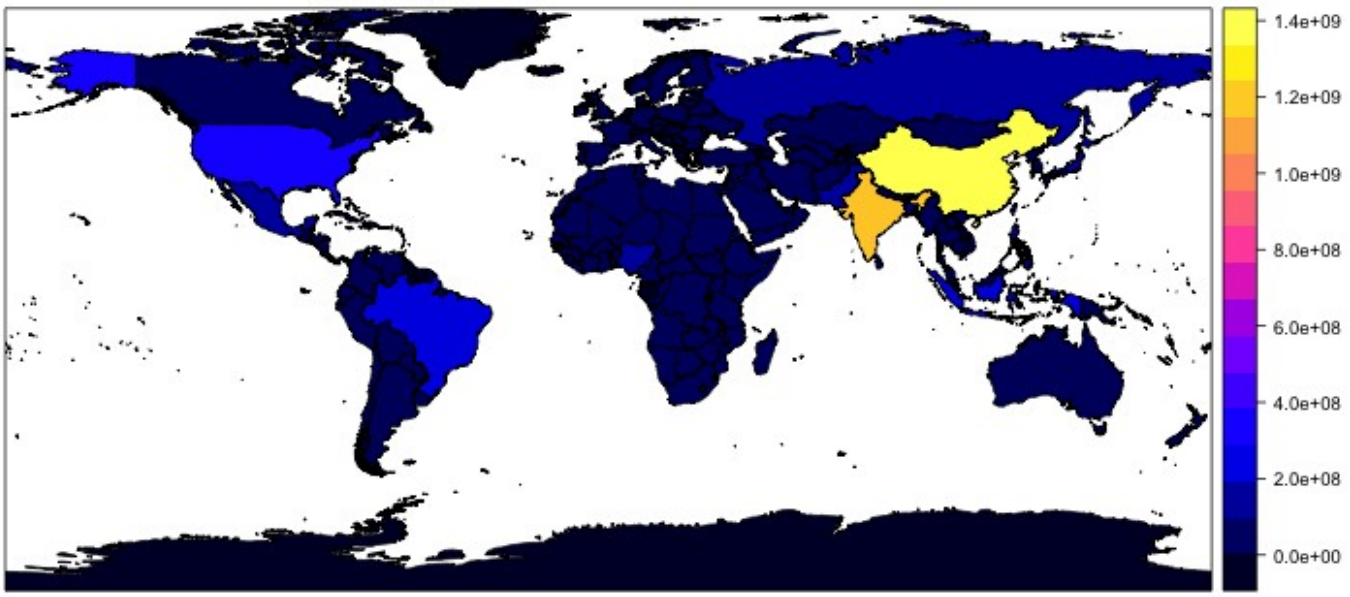

Avg wgs
<= 35870
<= 45137
<= 55093
<= 75035
<= 294264
Google

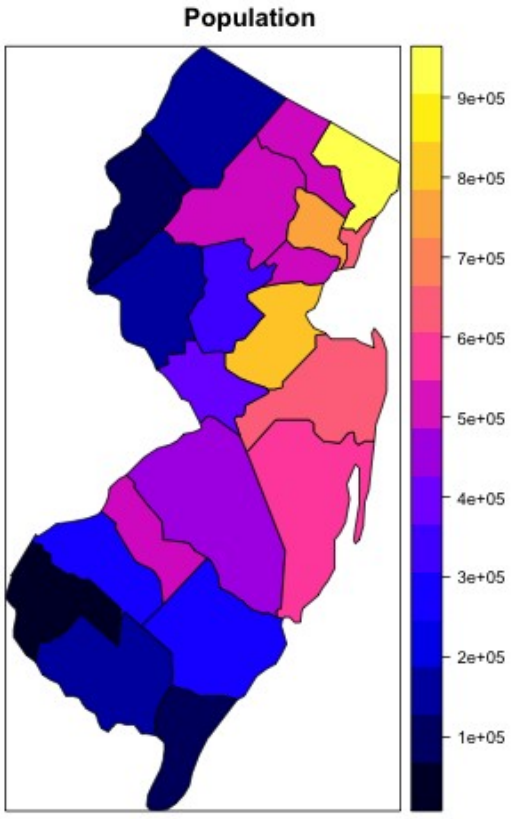


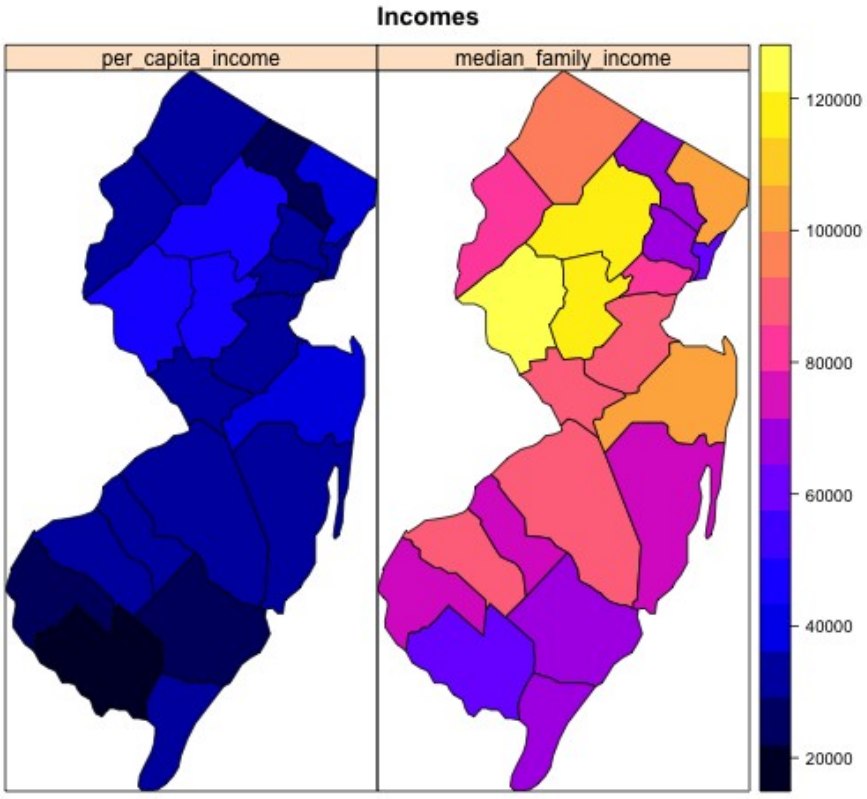


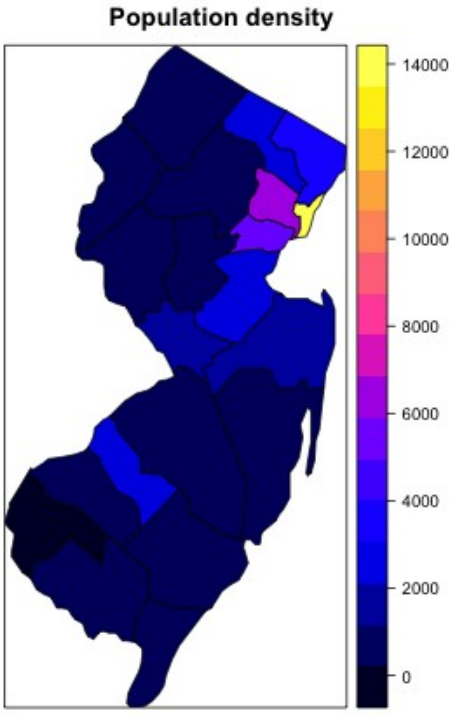



7. Least developed region
6. Developing region
5. Emerging region: G20
4. Emerging region: MIKT
3. Emerging region: BRIC
2. Developed region: nonG7
1. Developed region: G7


1.4e+09
1.2e+09
1.0e+09
8.0e+08
6.0e+08
4.0e+08
2.0e+08
0.0e+00


Population
9e+05
8e+05
7e+05
6e+05
5e+05
4e+05
3e+05
2e+05
1e+05


Incomes
per_capita_income
median_family_income
120000
100000
80000
60000
40000
20000


Population density
14000
12000
10000
8000
6000
4000
2000
0

Chapter 11