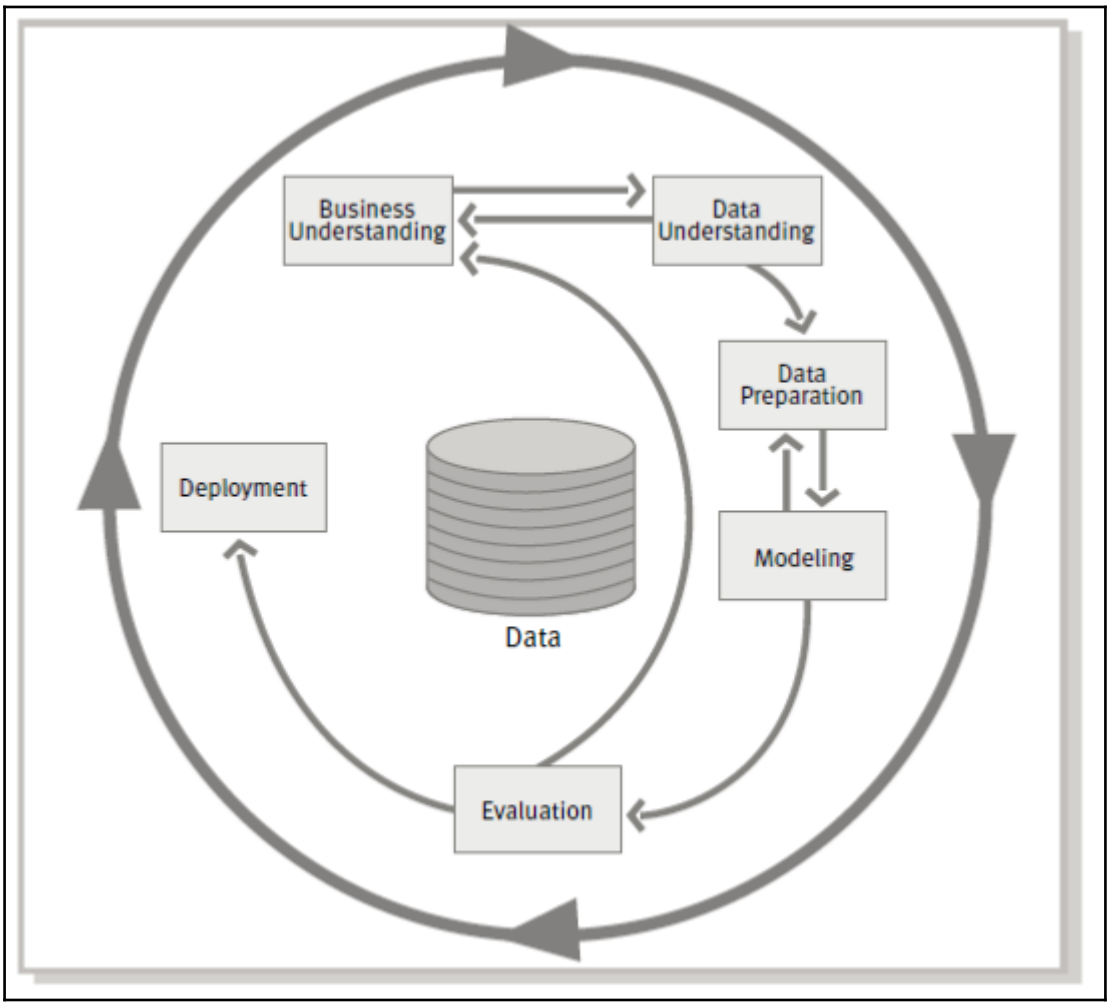
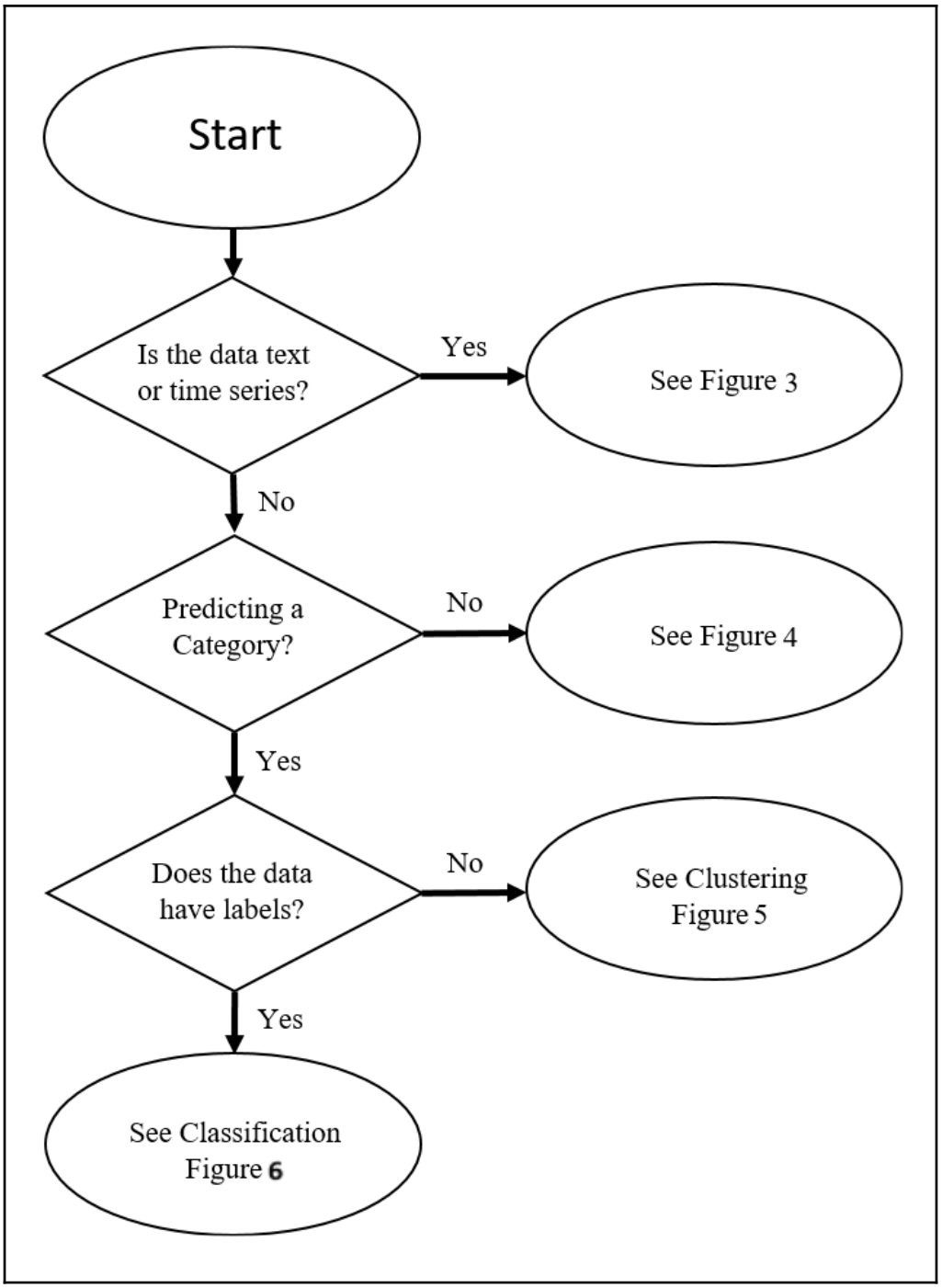
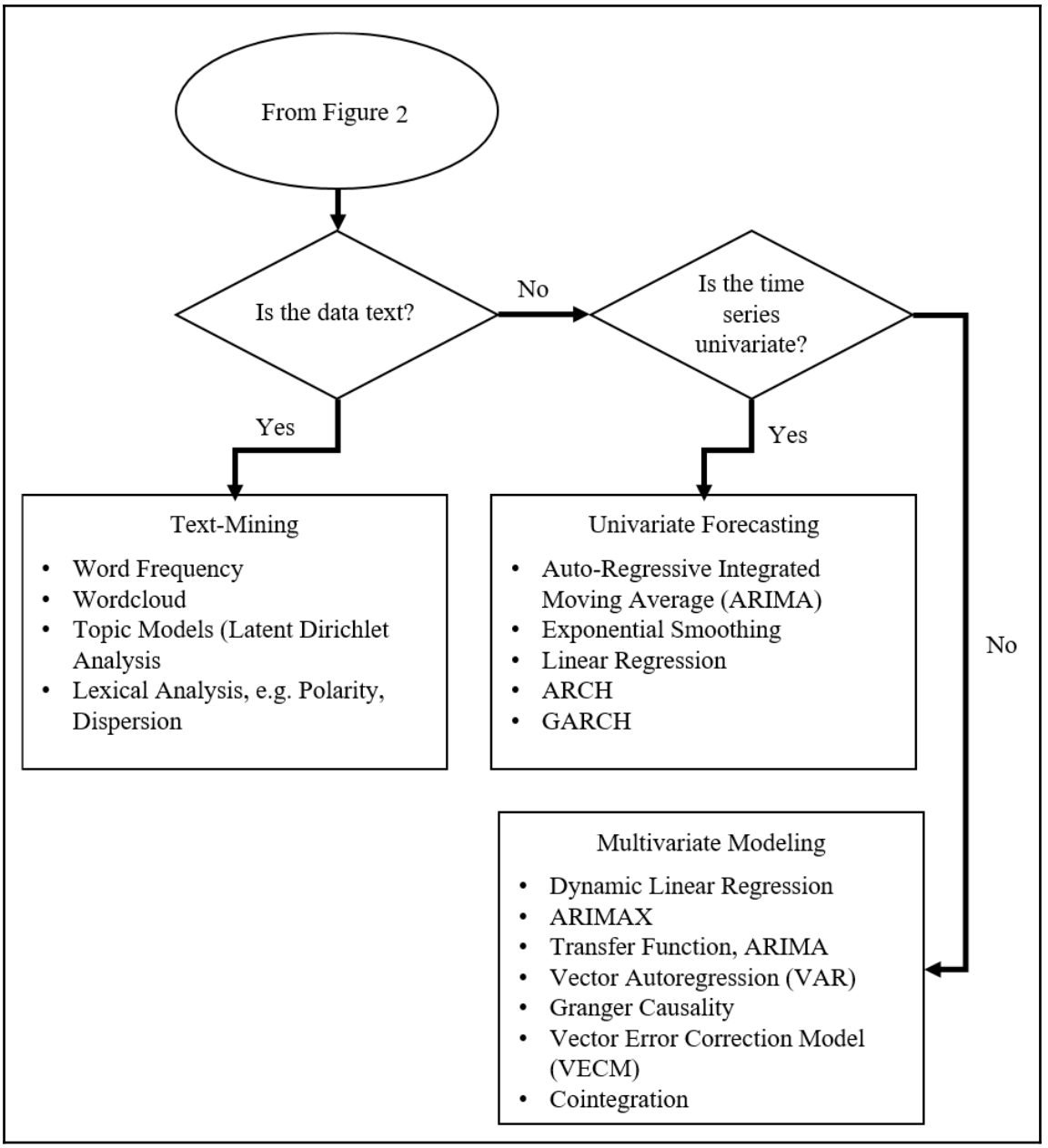
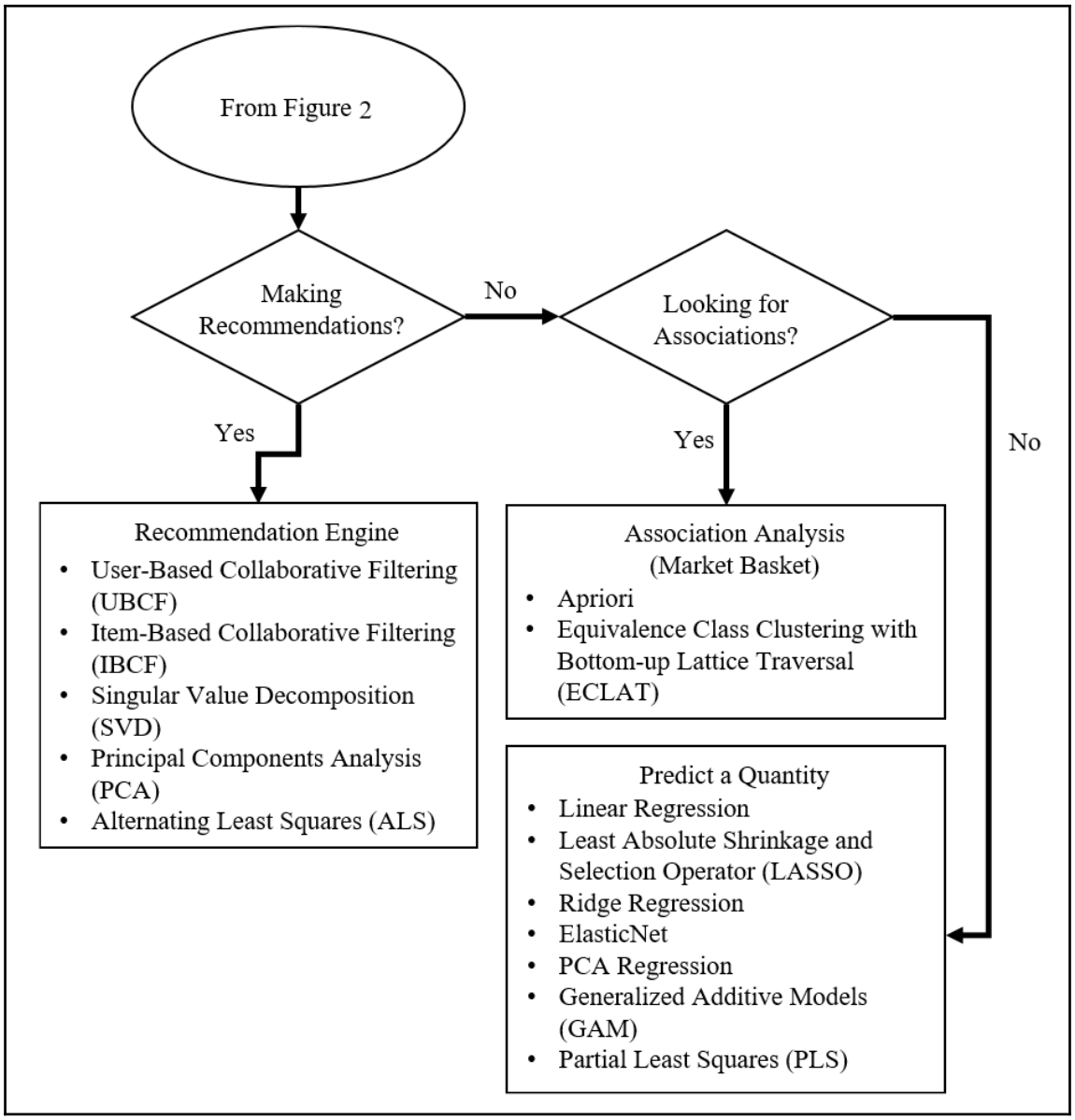
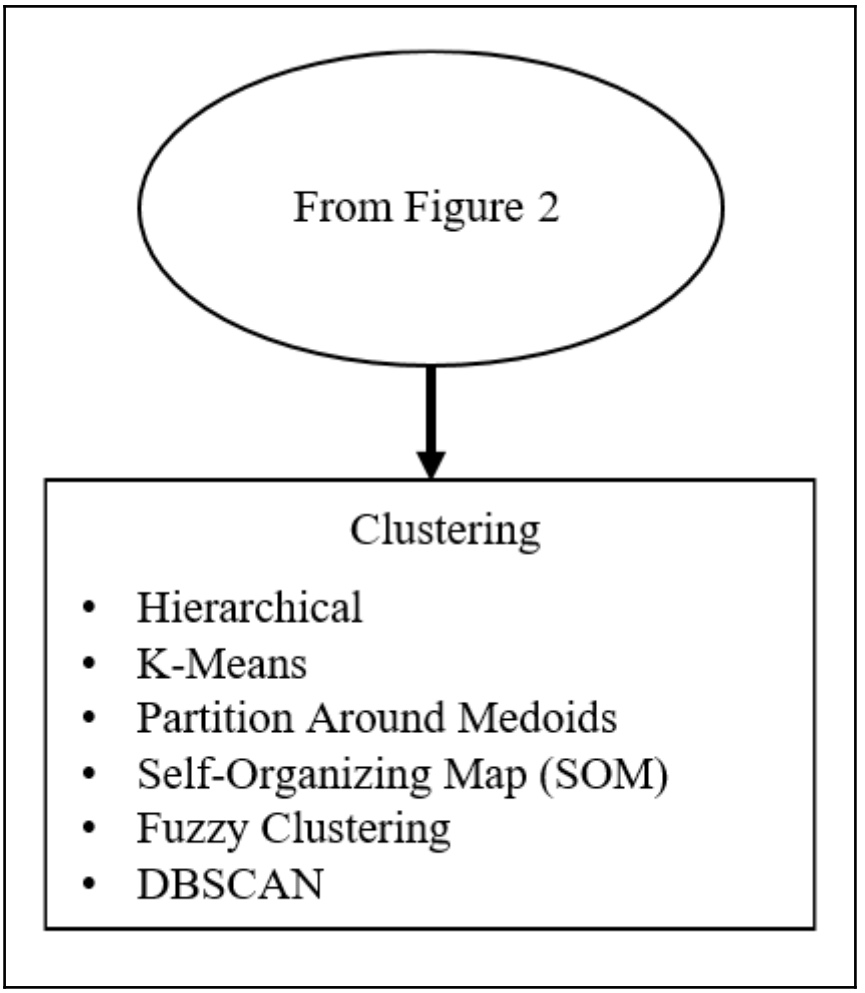
Chapter 1: A Process for Success

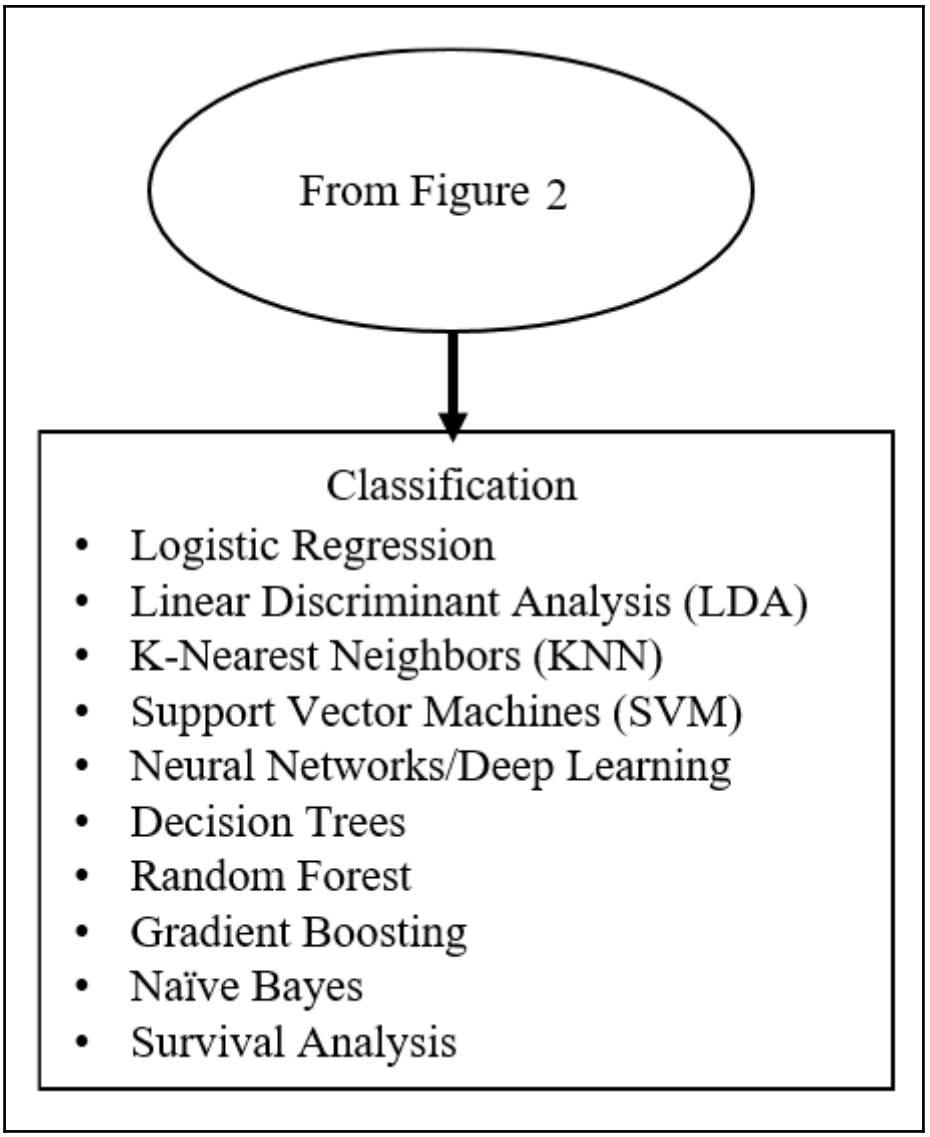



Start
Is the data text or time series?
Yes
See Figure 3
No
Predicting a Category?
No
See Figure 4
Yes
Does the data have labels?
No
See Clustering Figure 5
Yes
See Classification Figure 6

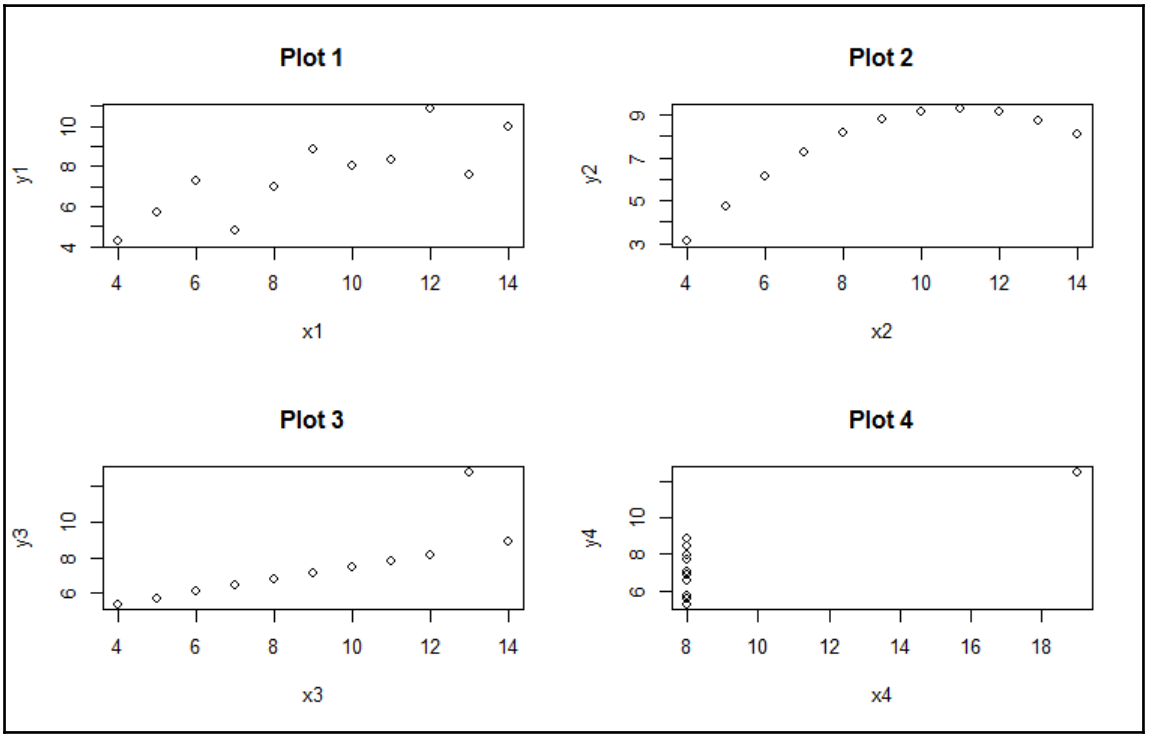

From Figure 2
Is the data text?
No
Is the time series univariate?
Yes
Yes
No
Text-Mining
• Word Frequency
• Wordcloud
• Topic Models (Latent Dirichlet Analysis
• Lexical Analysis, e.g. Polarity, Dispersion
Univariate Forecasting
• Auto-Regressive Integrated Moving Average (ARIMA)
• Exponential Smoothing
• Linear Regression
• ARCH
• GARCH
Multivariate Modeling
• Dynamic Linear Regression
• ARIMAX
• Transfer Function, ARIMA
• Vector Autoregression (VAR)
• Granger Causality
• Vector Error Correction Model (VECM)
• Cointegration

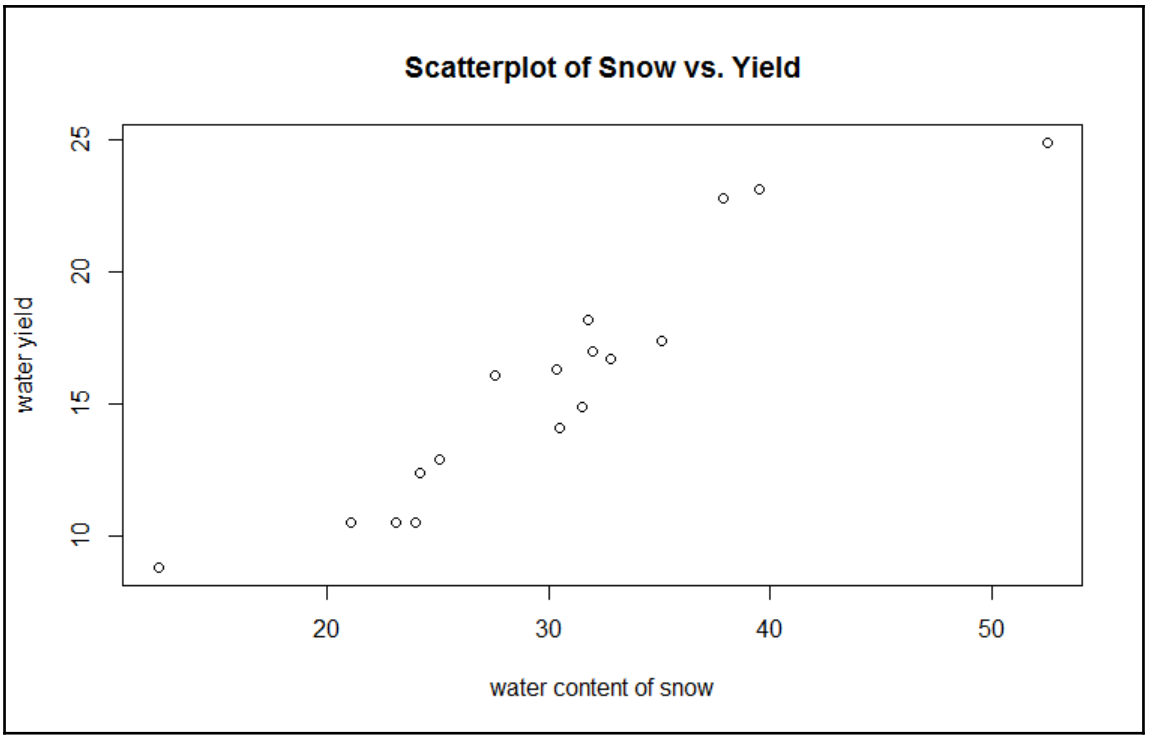

From Figure 2
Making Recommendations?
No
Looking for Associations?
Yes
Yes
No
Recommendation Engine
• User-Based Collaborative Filtering (UBCF)
• Item-Based Collaborative Filtering (IBCF)
• Singular Value Decomposition (SVD)
• Principal Components Analysis (PCA)
• Alternating Least Squares (ALS)
Association Analysis (Market Basket)
• Apriori
• Equivalence Class Clustering with Bottom-up Lattice Traversal (ECLAT)
Predict a Quantity
• Linear Regression
• Least Absolute Shrinkage and Selection Operator (LASSO)
• Ridge Regression
• ElasticNet
• PCA Regression
• Generalized Additive Models (GAM)
• Partial Least Squares (PLS)

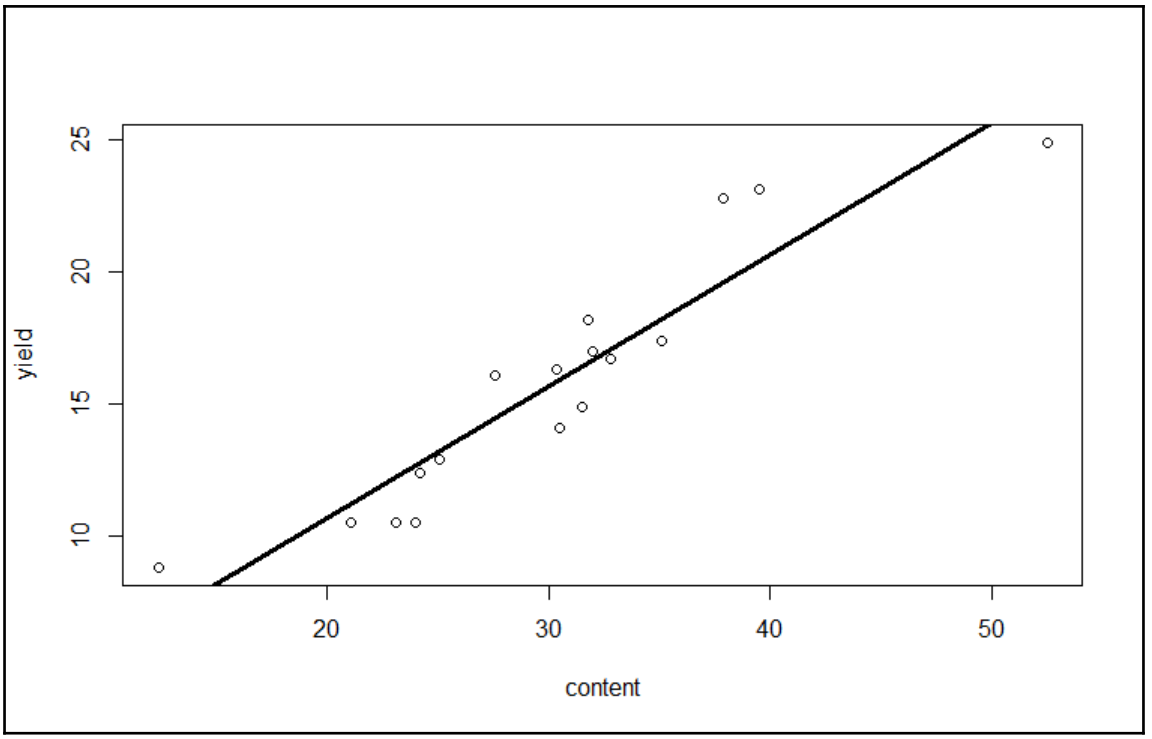

From Figure 2
Clustering
• Hierarchical
• K-Means
• Partition Around Medoids
• Self-Organizing Map (SOM)
• Fuzzy Clustering
• DBSCAN

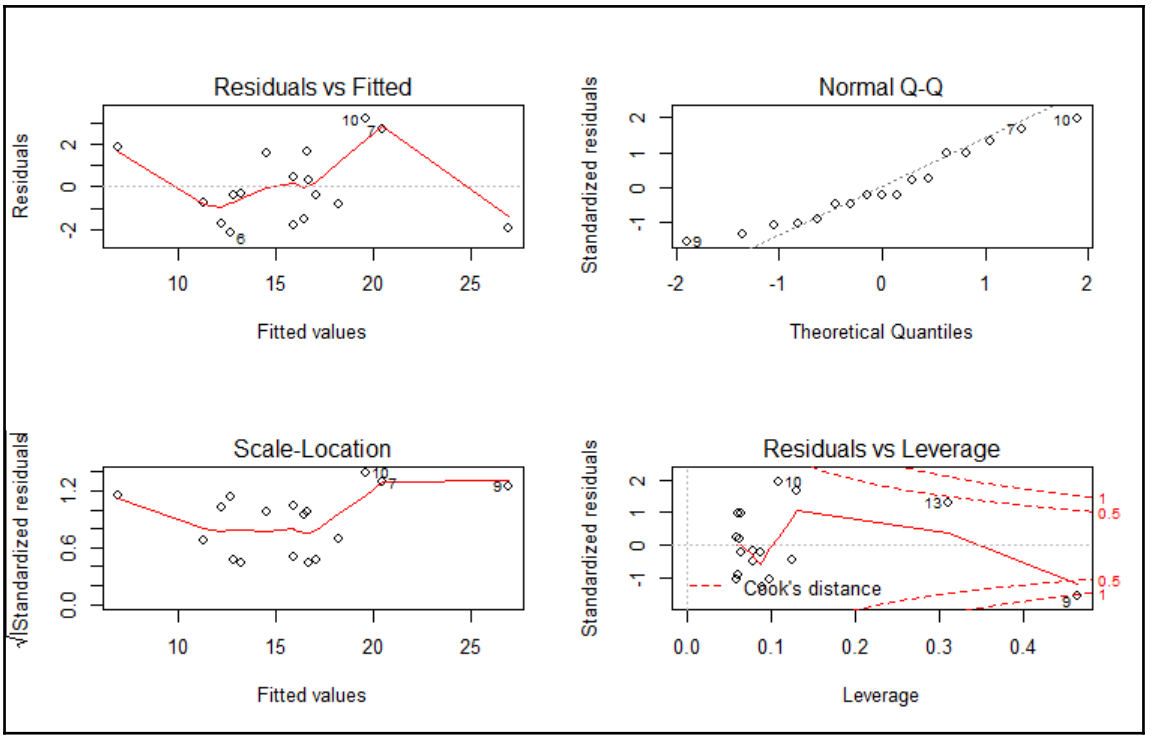

From Figure 2
Classification
• Logistic Regression
• Linear Discriminant Analysis (LDA)
• K-Nearest Neighbors (KNN)
• Support Vector Machines (SVM)
• Neural Networks/Deep Learning
• Decision Trees
• Random Forest
• Gradient Boosting
• Naïve Bayes
• Survival Analysis

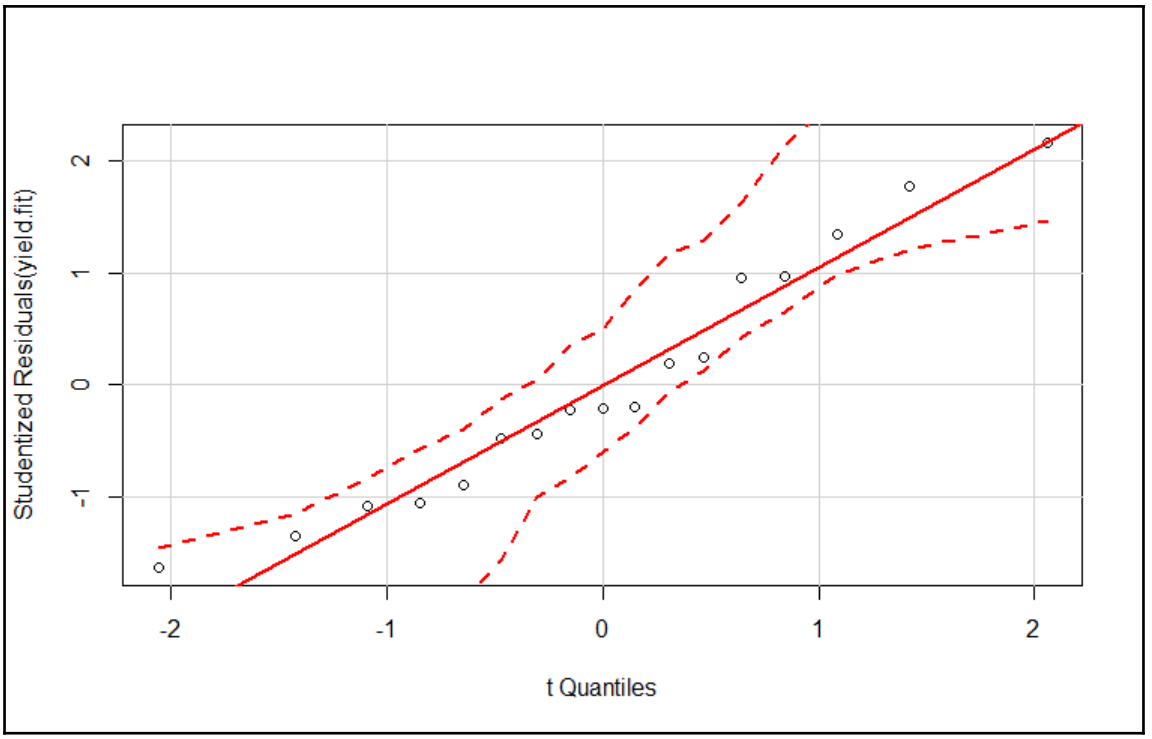
Chapter 2: Linear Regression - The Blocking and Tackling of Machine Learning

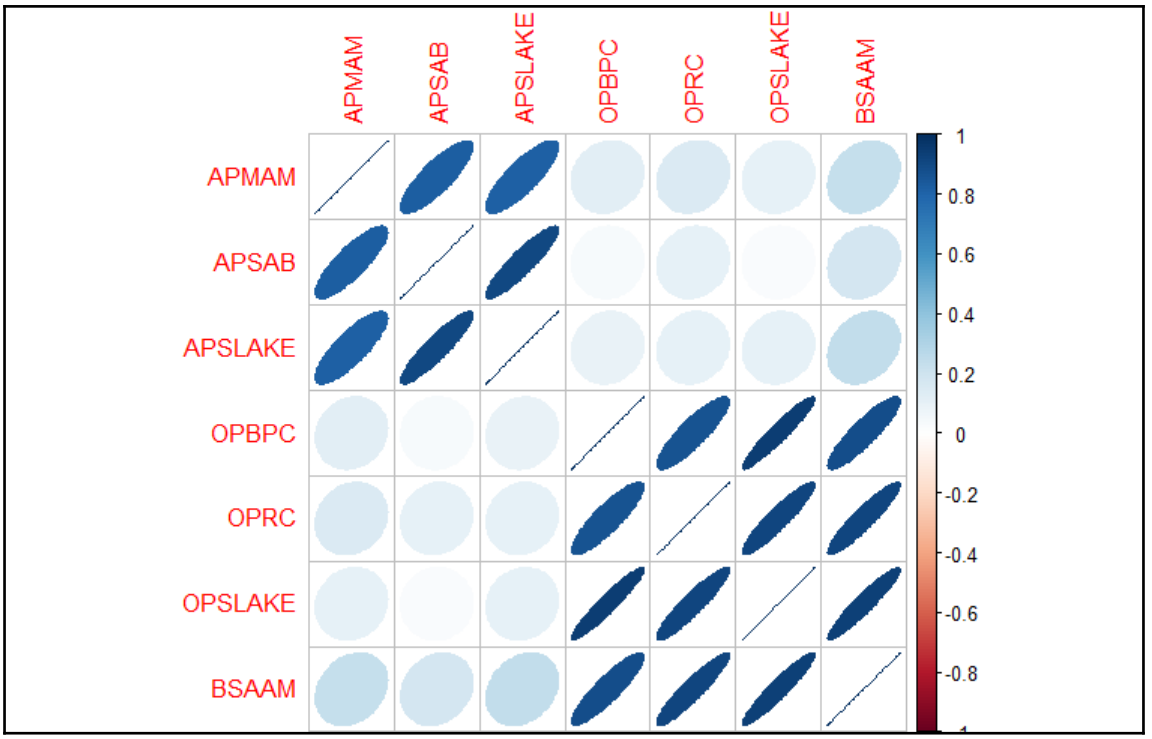


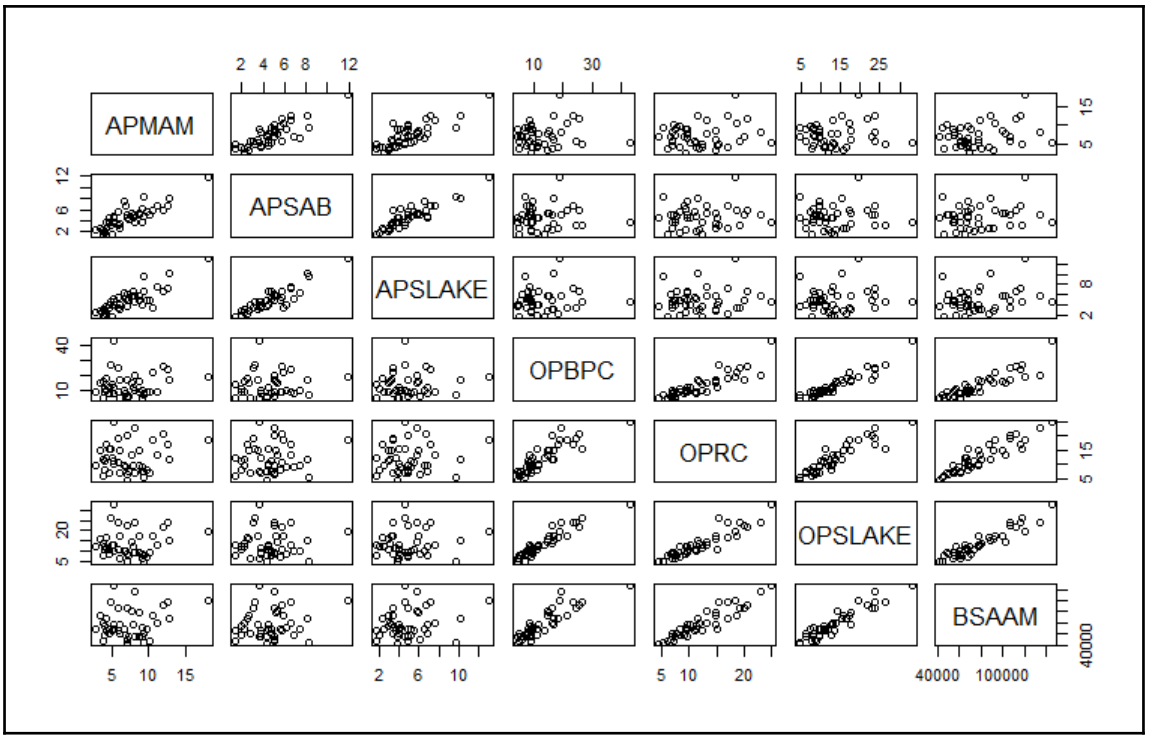

Scatterplot of Snow vs. Yield
water yield
water content of snow
10
15
20
25
20
30
40
50

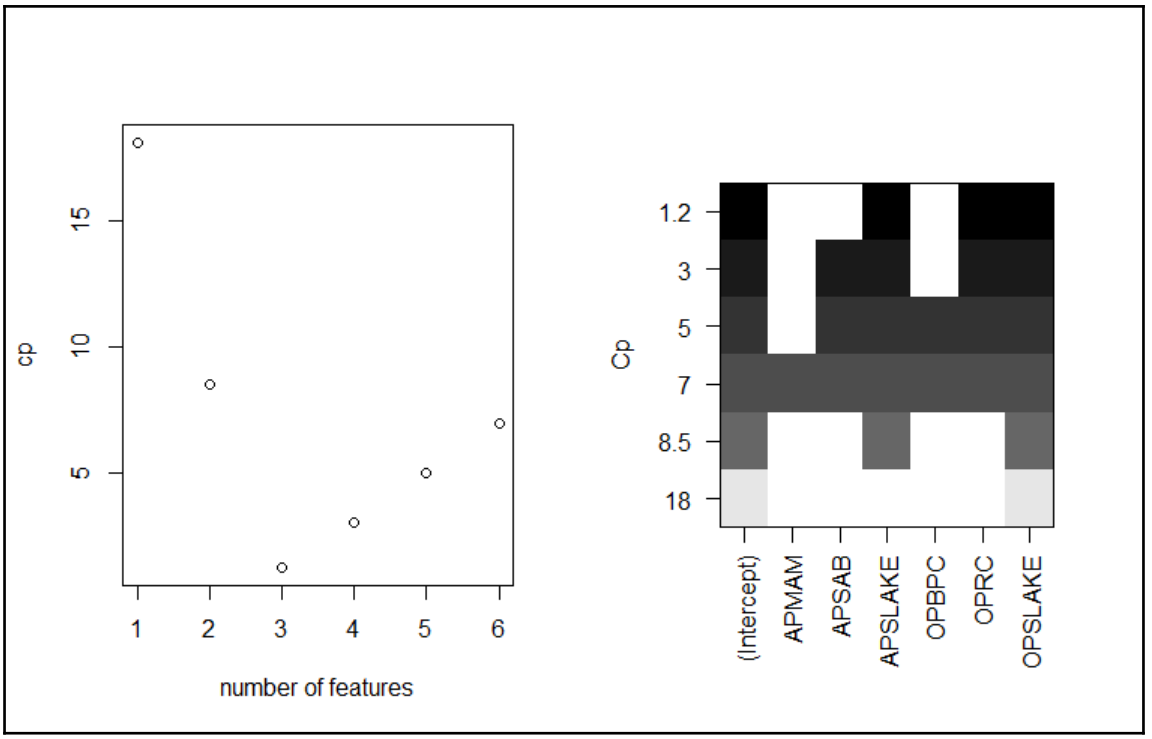

yield
content
10
15
20
25
20
30
40
50

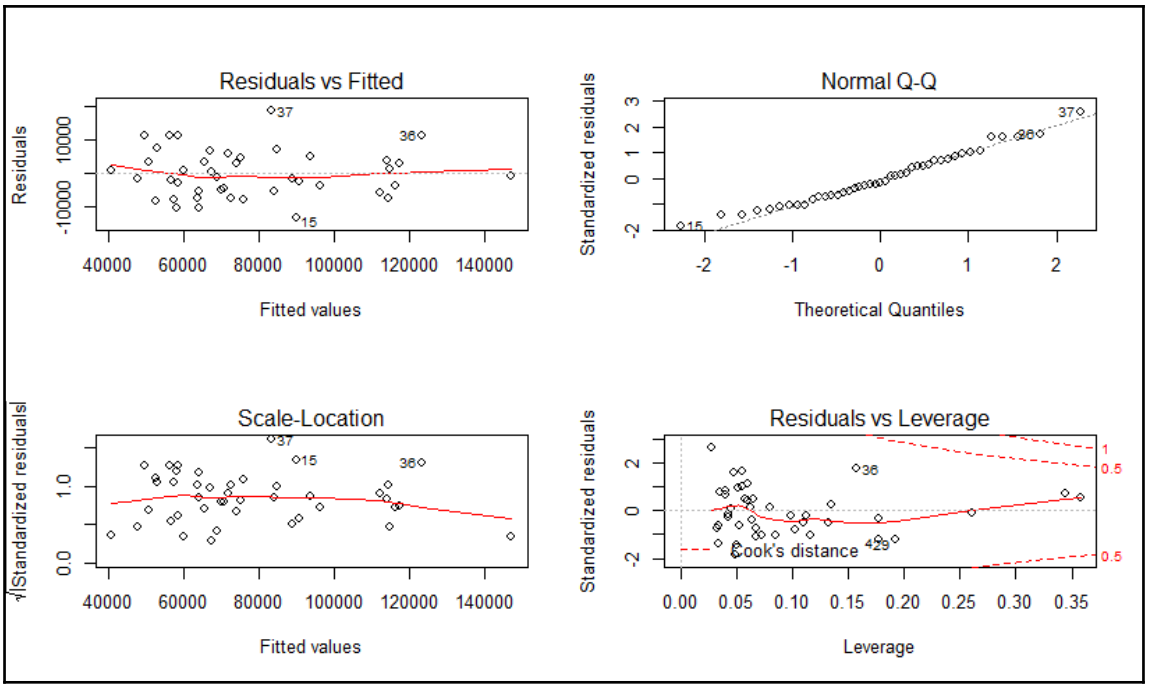

Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

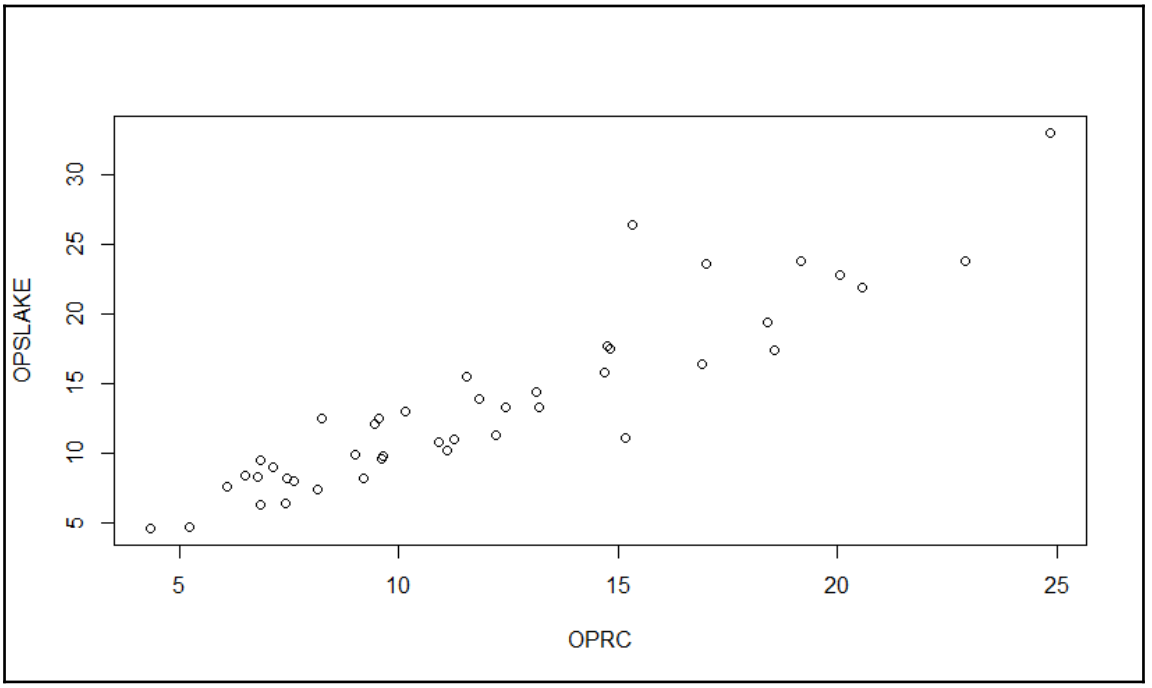

Studentized Residuals(yield.fit)
t Quantiles
-2
-1
0
1
2
-1
0
1
2

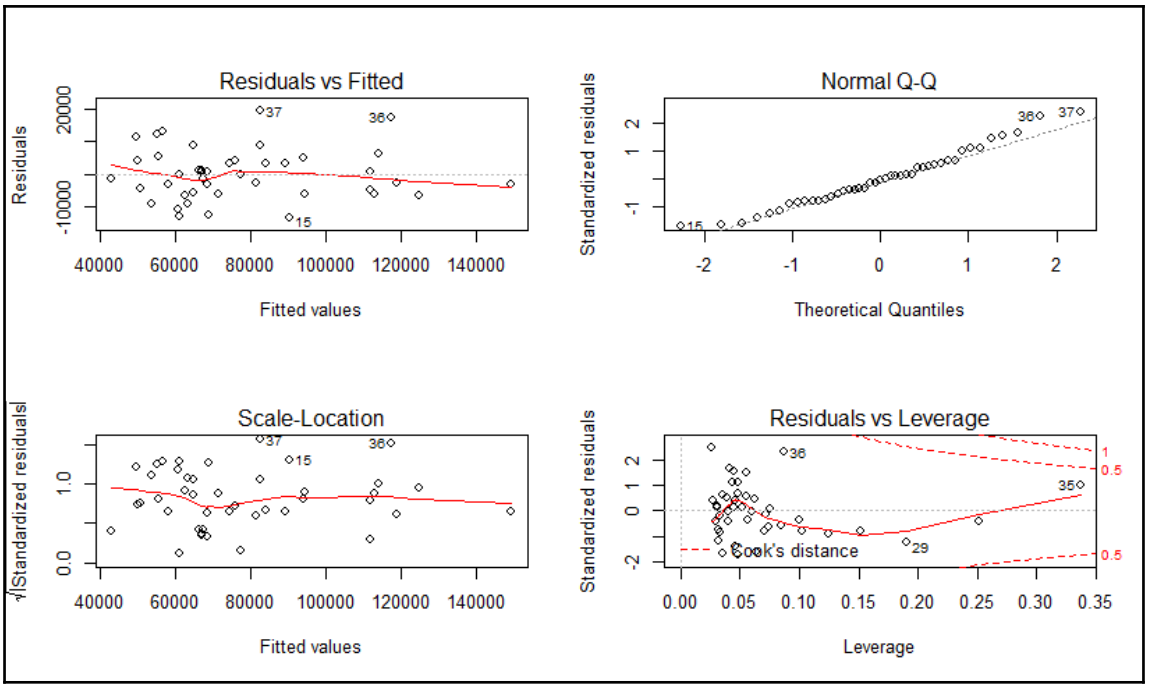

APMAM
APSAB
APSLAKE
OPBPC
OPRC
OPSLAKE
BSAAM
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8

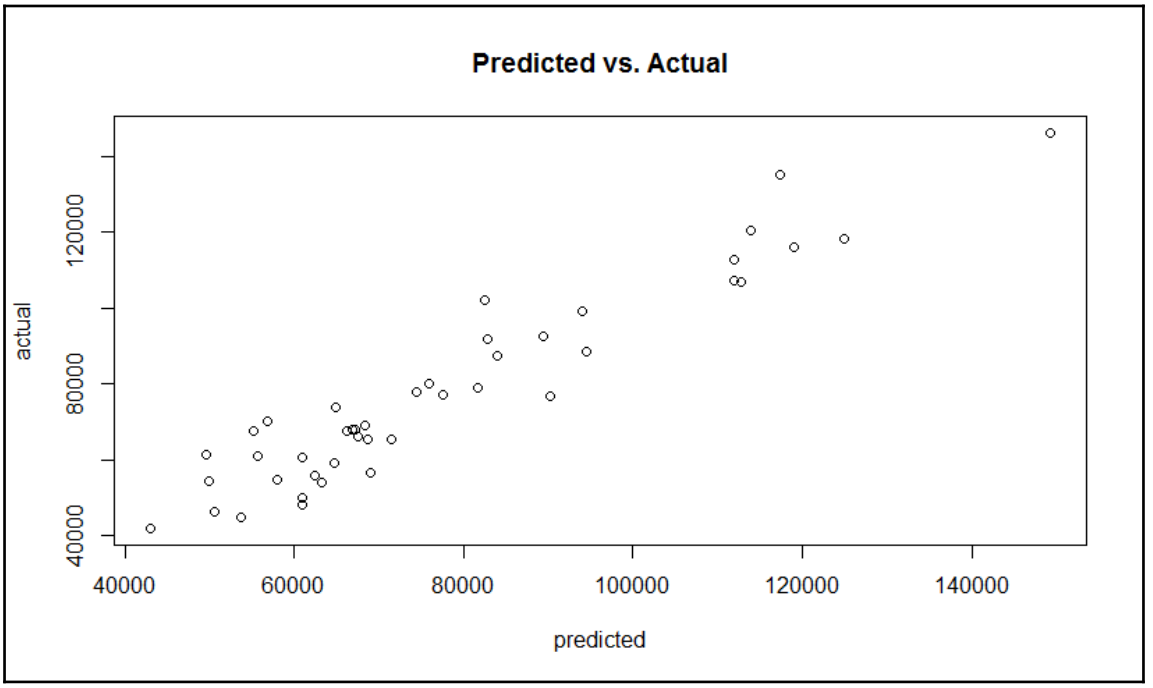

APMAM
APSAB
APSLAKE
OPBPC
OPRC
OPSLAKE
BSAAM
2 4 6 8 12
10 30
5 15 25
5 15
2 6 12
2 8
10 40
5 15
5 20
40000
5 10 15
2 6 10
5 10 20
40000 100000

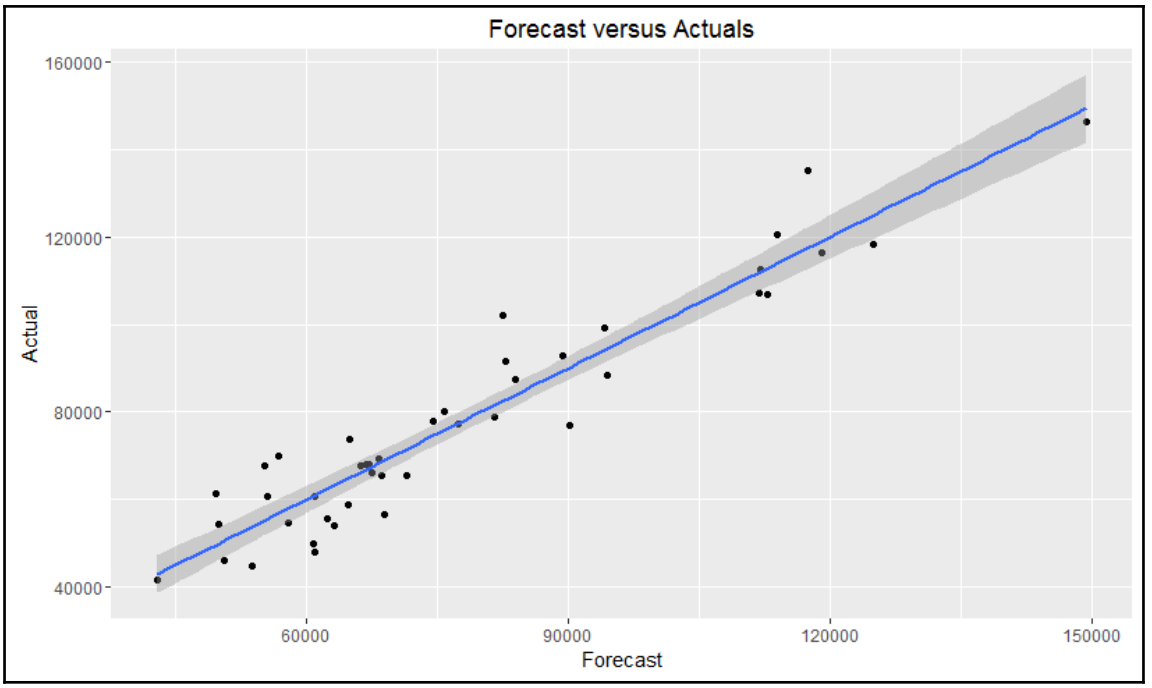

cp
5
10
15
1
2
3
4
5
6
number of features
Cp
1.2
3
5
7
8.5
18
(Intercept)
APMAM
APSAB
APSLAKE
OPBPC
OPRC
OPSLAKE


Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

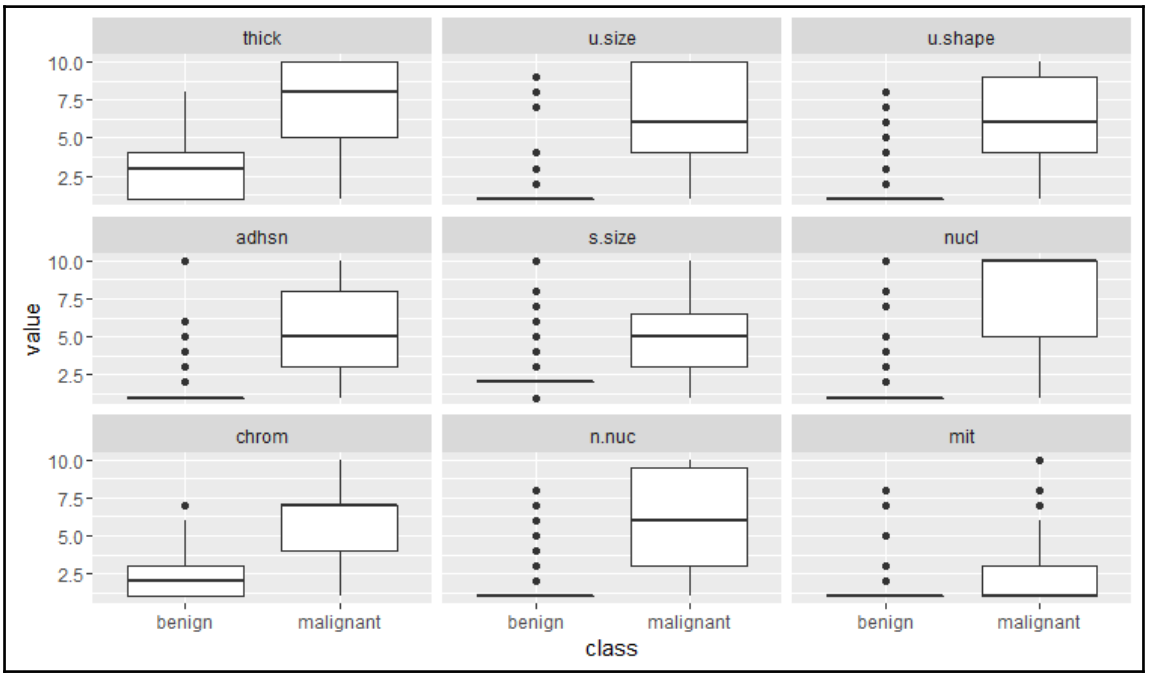

OPSLAKE
OPRC

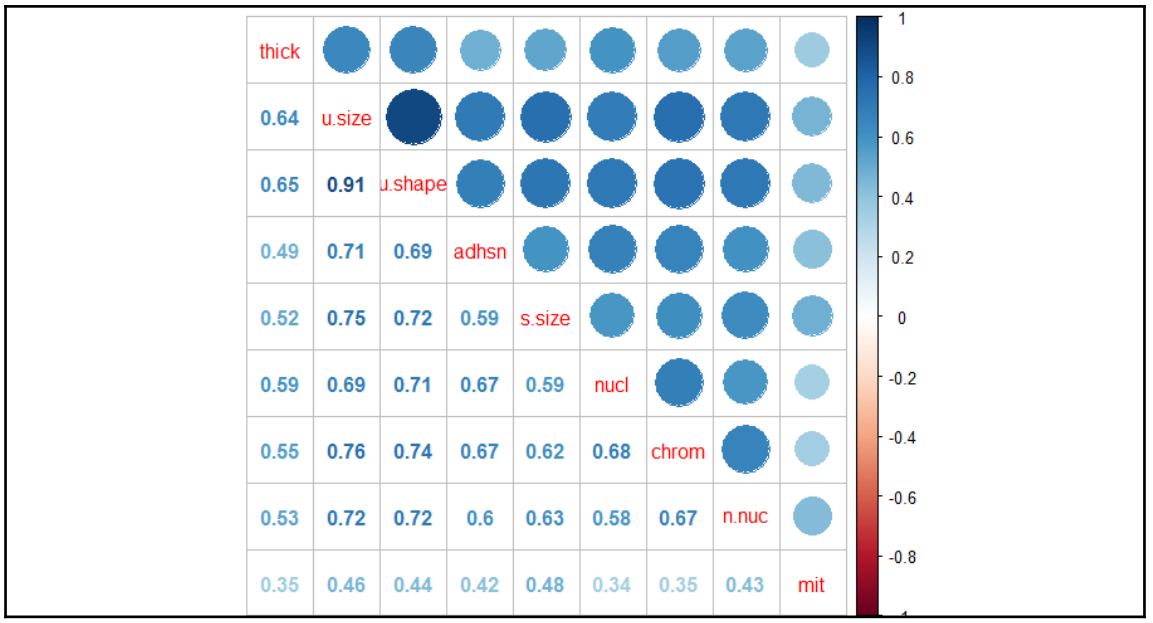

Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

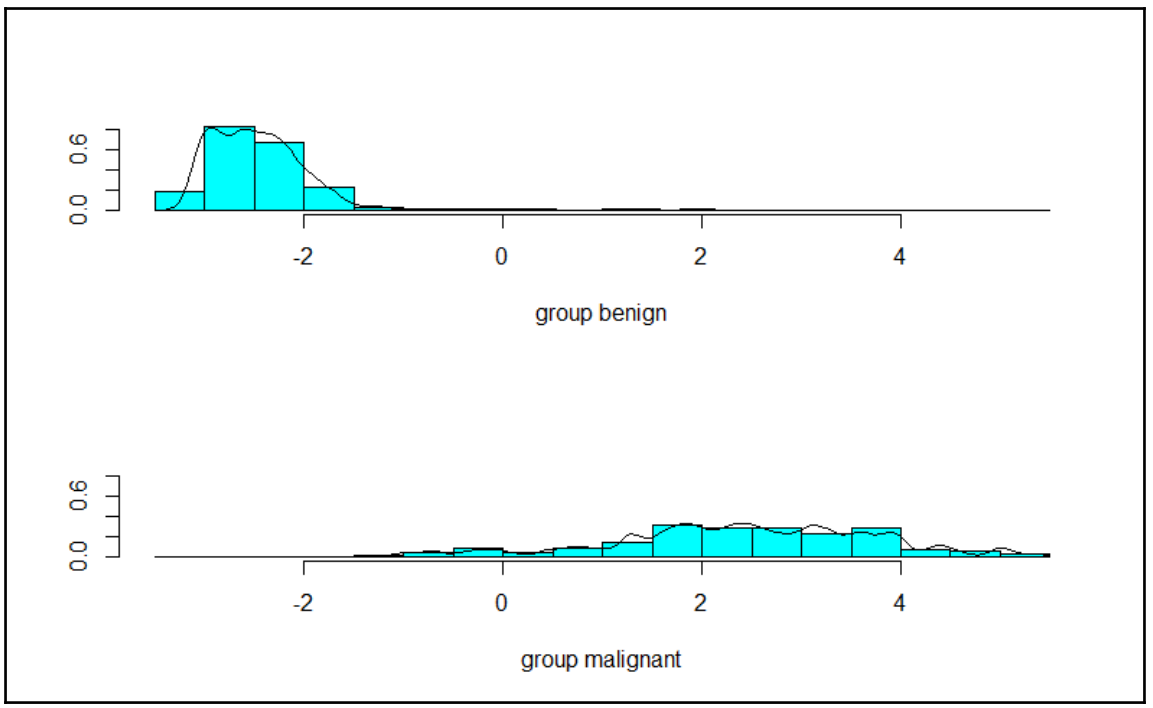

Predicted vs. Actual
actual
predicted

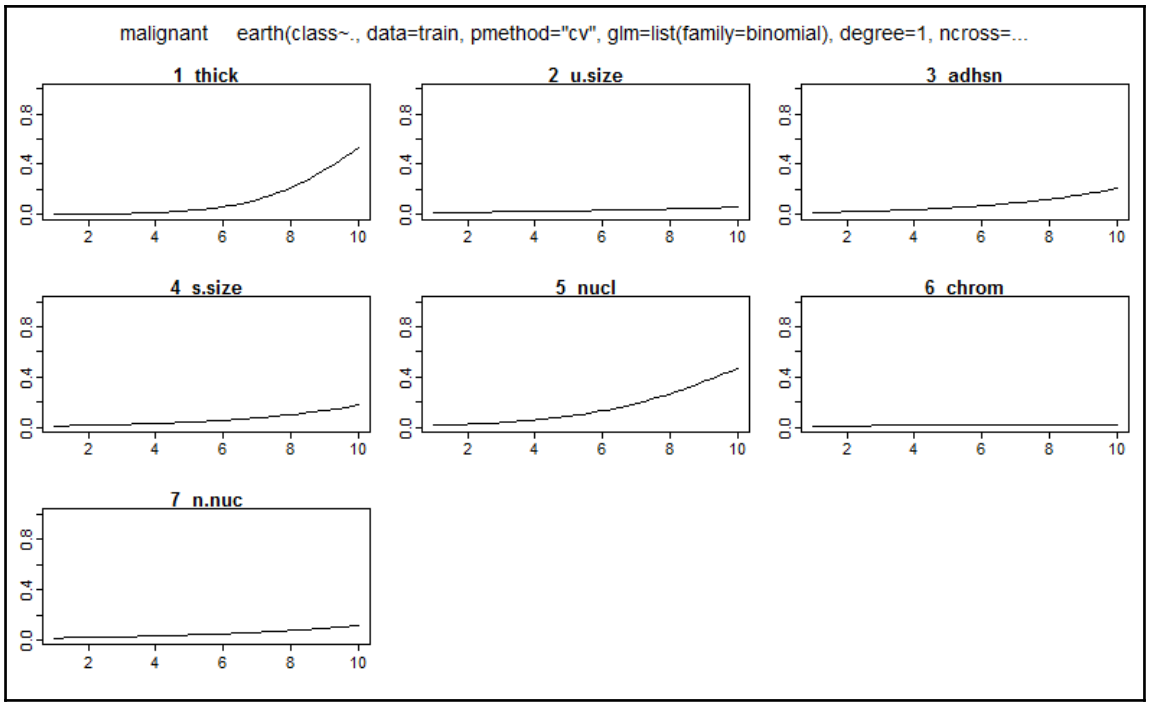

Forecast versus Actuals
160000
120000
80000
40000
Actual
60000
90000
120000
150000
Forecast

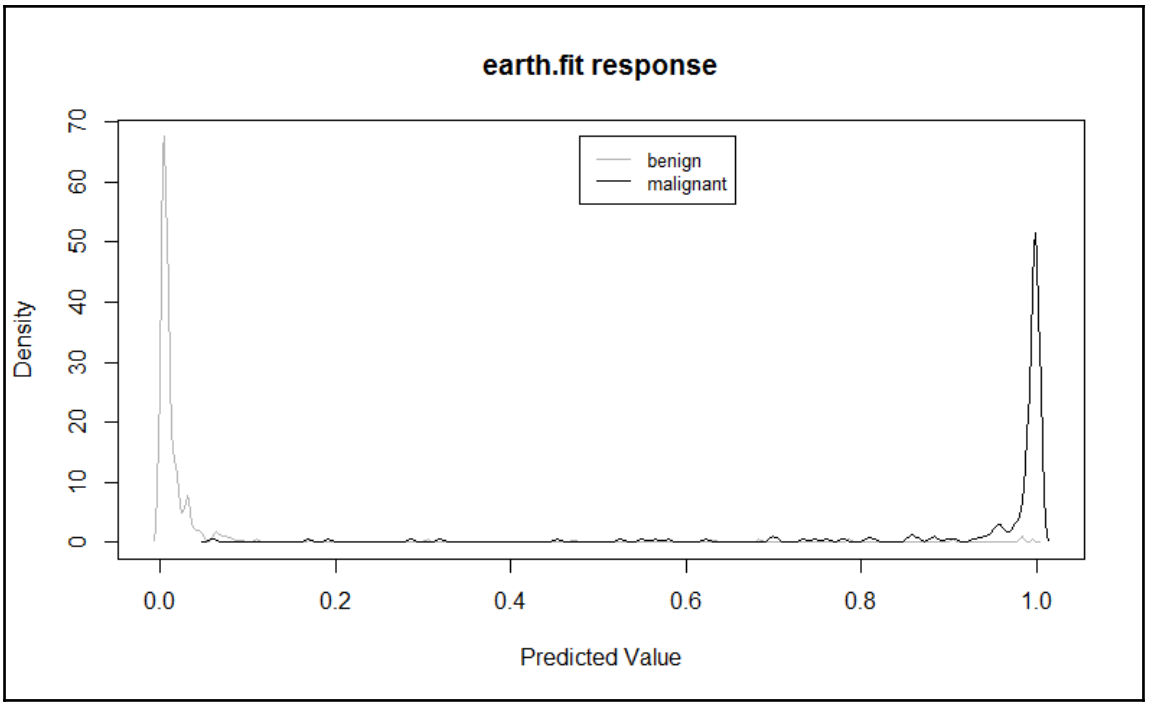
Chapter 3: Logistic Regression and Discriminant Analysis

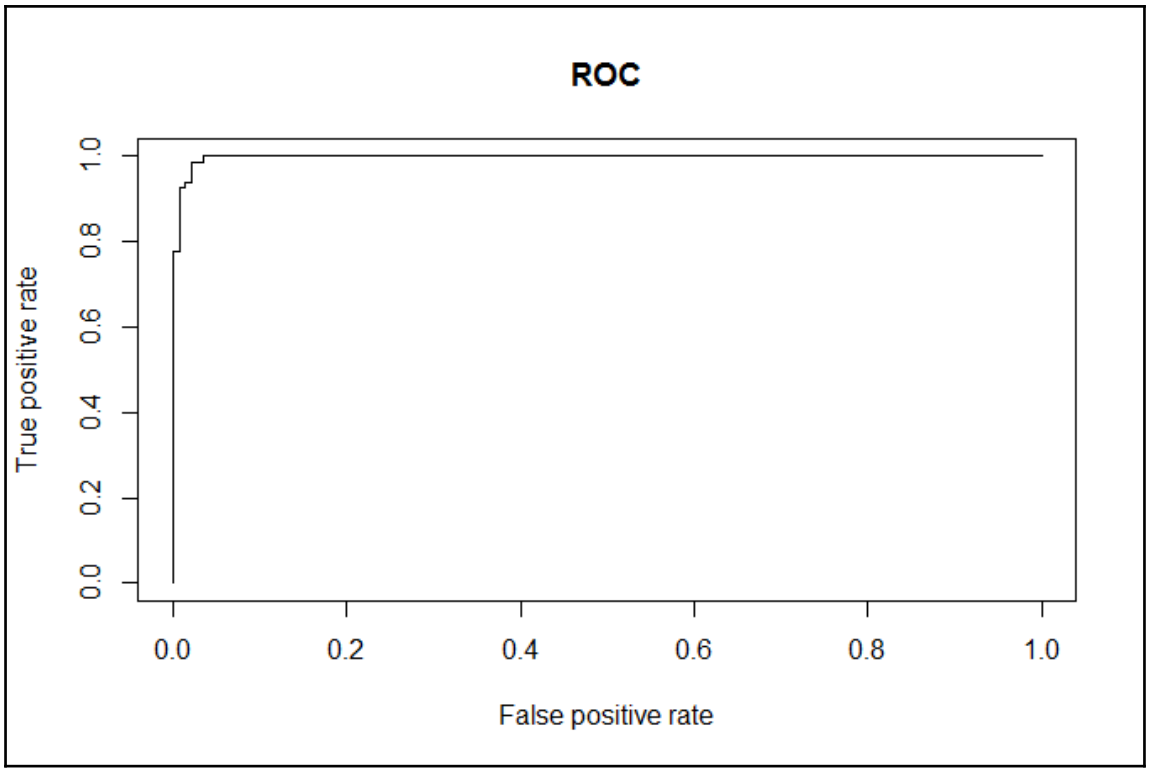


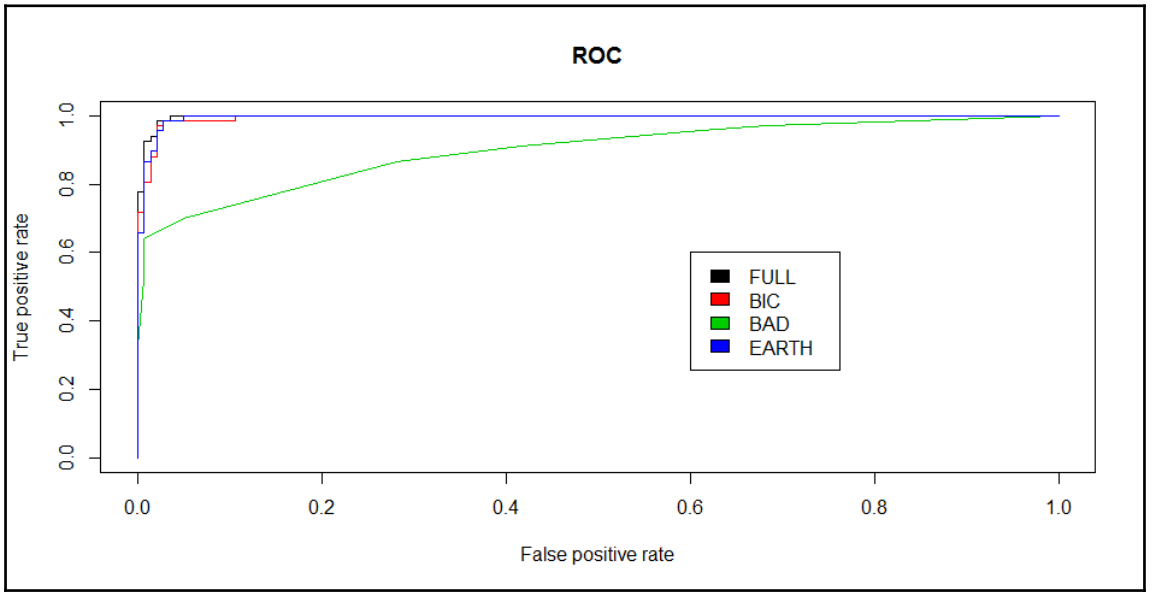

thick
u.size
u.shape
adhsn
s.size
nucl
chrom
n.nuc
mit
0.64
0.65 0.91
0.49 0.71 0.69
0.52 0.75 0.72 0.59
0.59 0.69 0.71 0.67 0.59
0.55 0.76 0.74 0.67 0.62 0.68
0.53 0.72 0.72 0.6 0.63 0.58 0.67
0.35 0.46 0.44 0.42 0.48 0.34 0.35 0.43
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8


0.0
0.6
-2
0
2
4
group benign
0.0
0.6
-2
0
2
4
group malignant

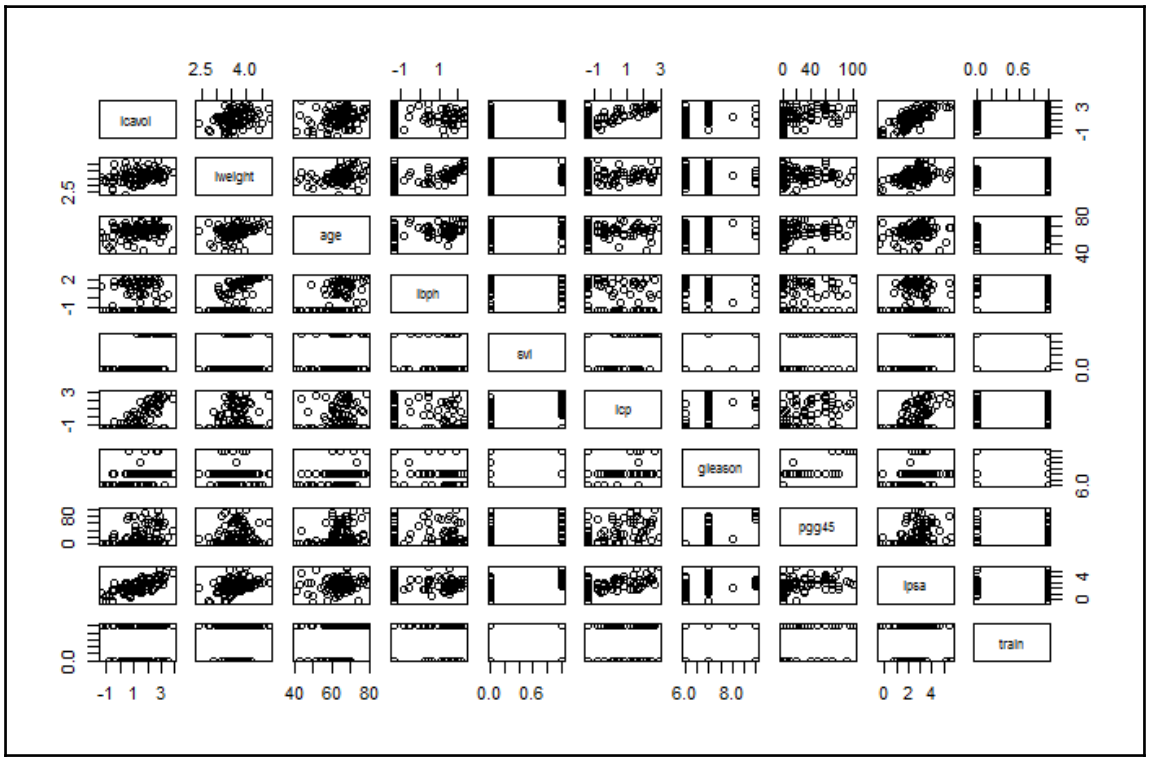

malignant earth(class~., data=train, pmethod="cv", glm=list(family=binomial), degree=1, ncross=...
1 thick
2 u.size
3 adhsn
4 s.size
5 nucl
6 chrom
7 n.nuc
0.0
0.4
0.8
2
4
6
8
10

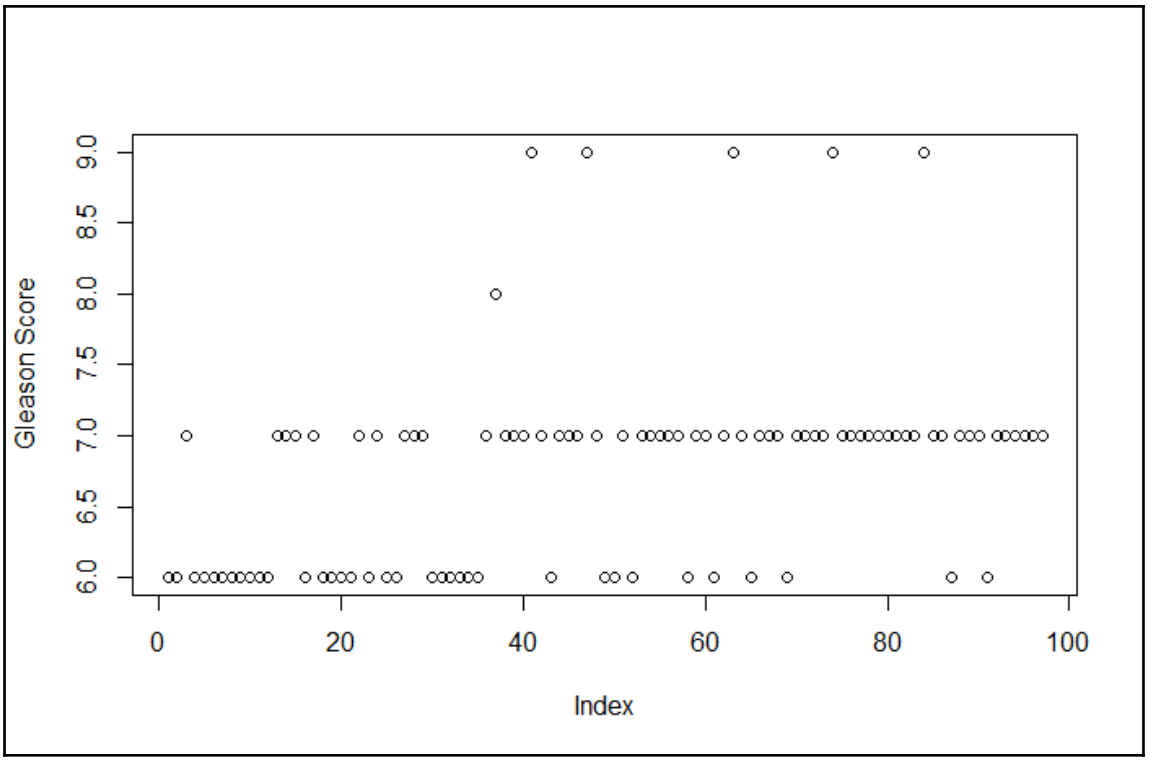

earth.fit response
benign
malignant
Density
0
10
20
30
40
50
60
70
0.0
0.2
0.4
0.6
0.8
1.0
Predicted Value

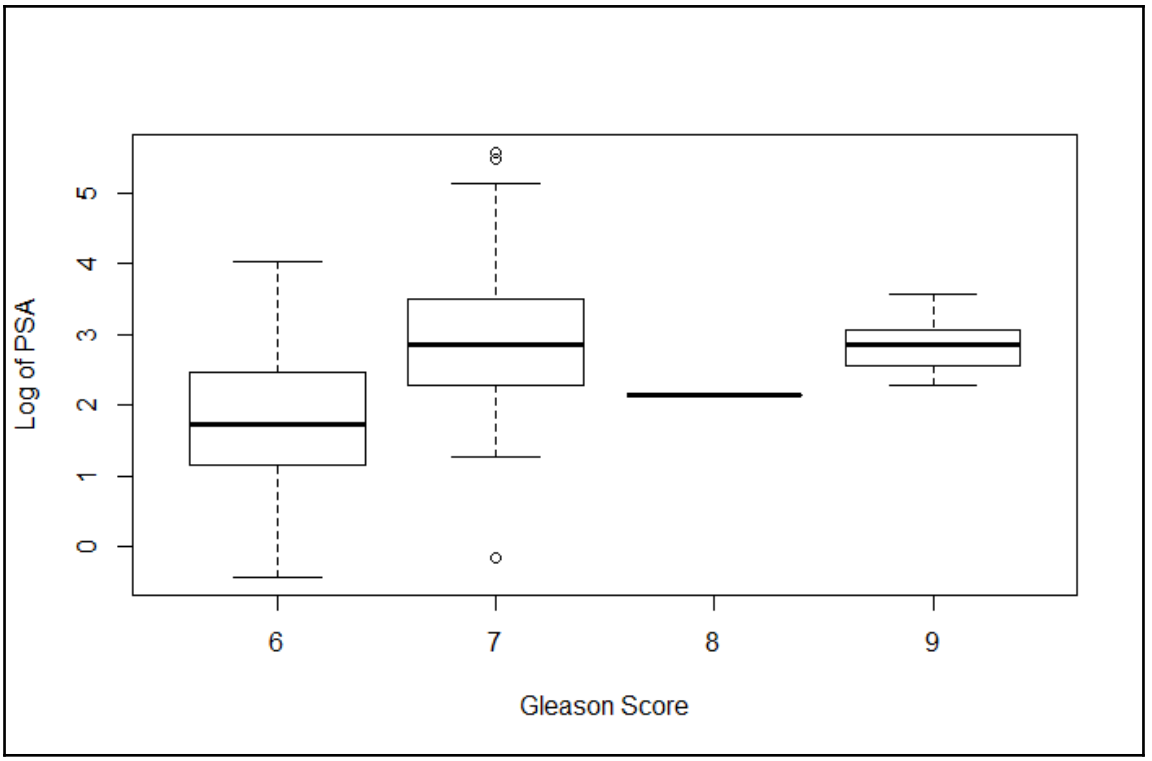

ROC
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate

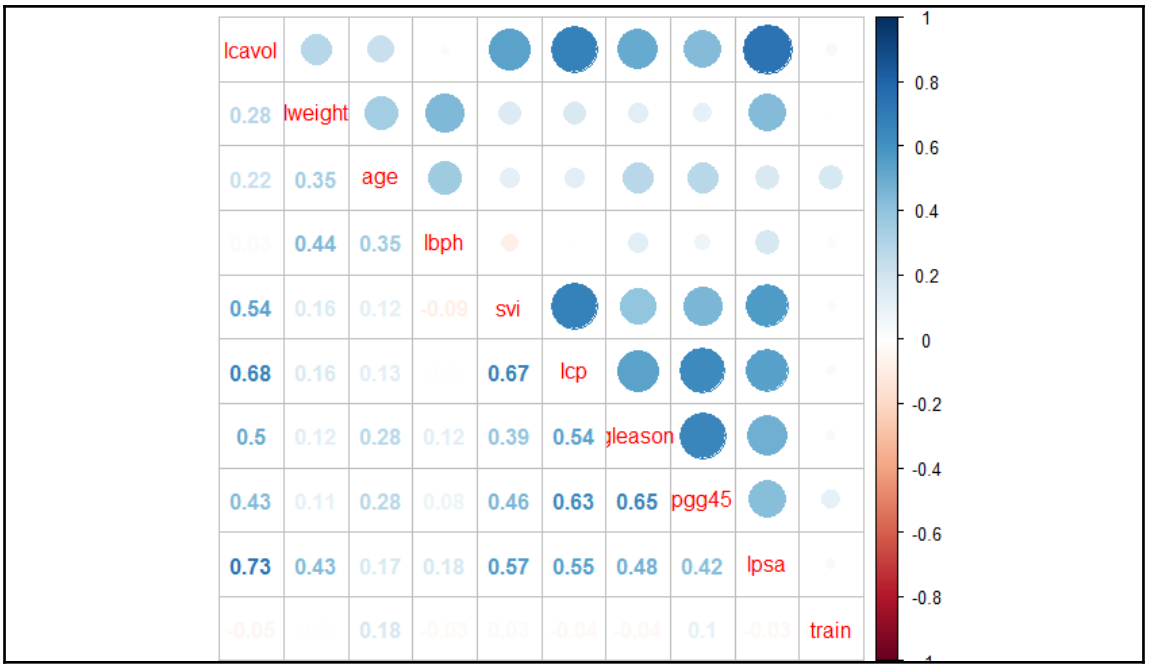

ROC
True positive rate
False positive rate
FULL
BIC
BAD
EARTH

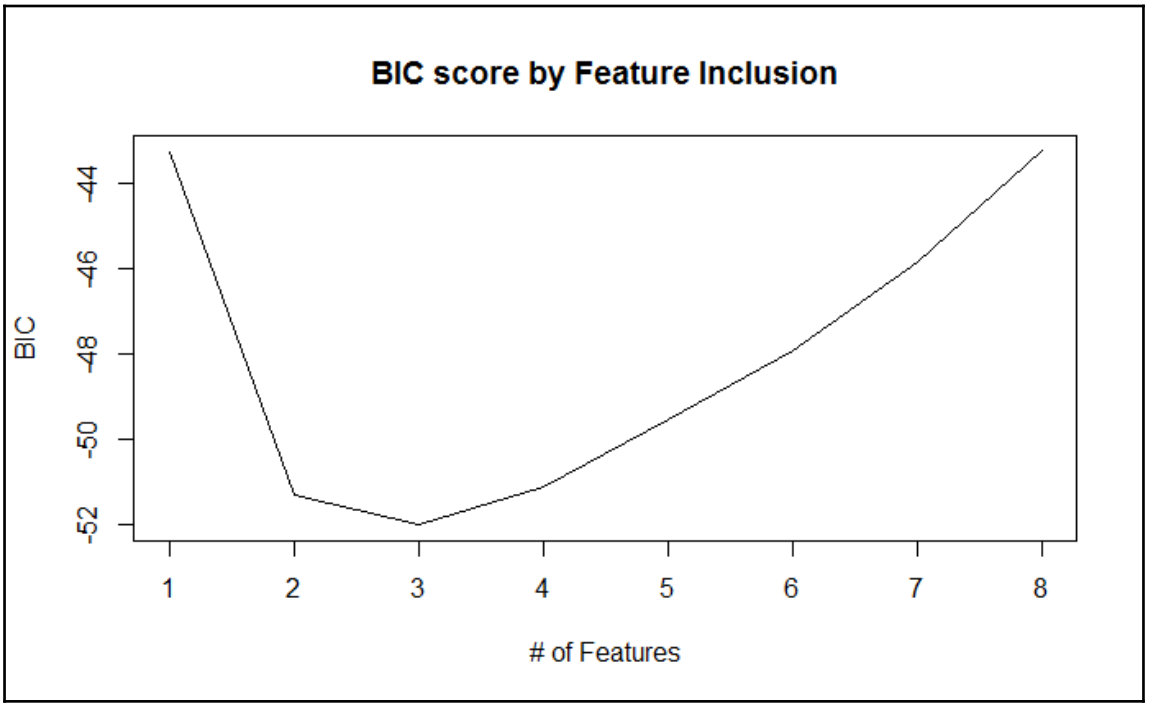
Chapter 4: Advanced Feature Selection in Linear Models

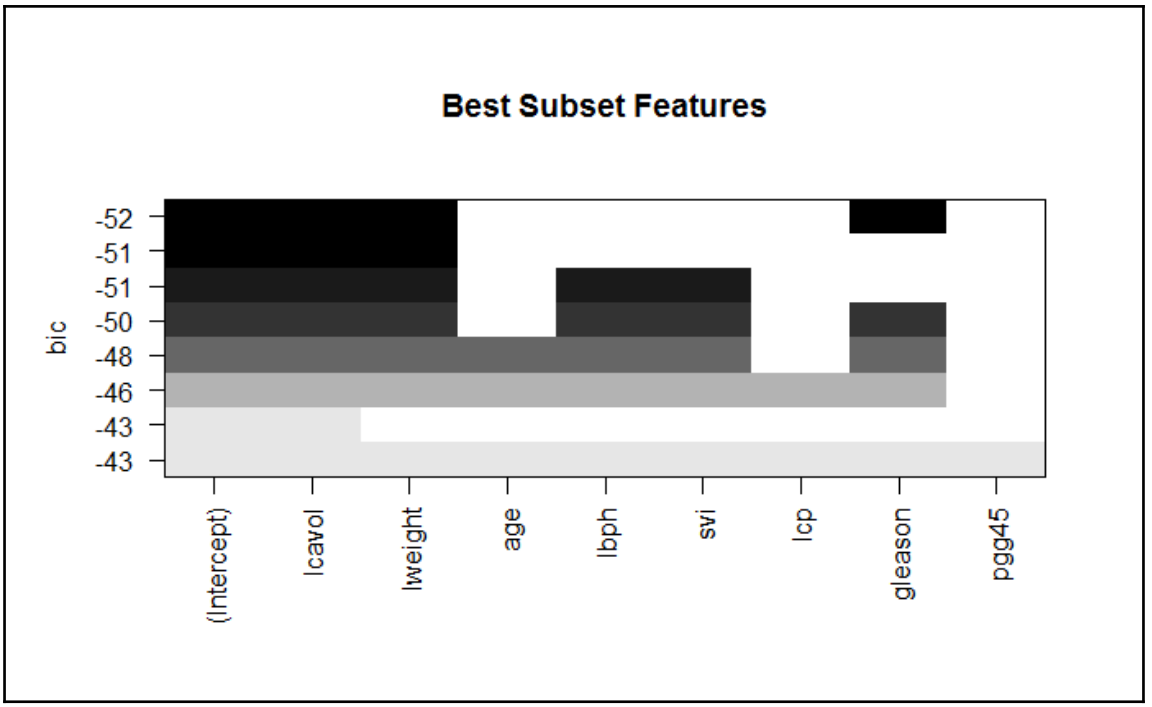


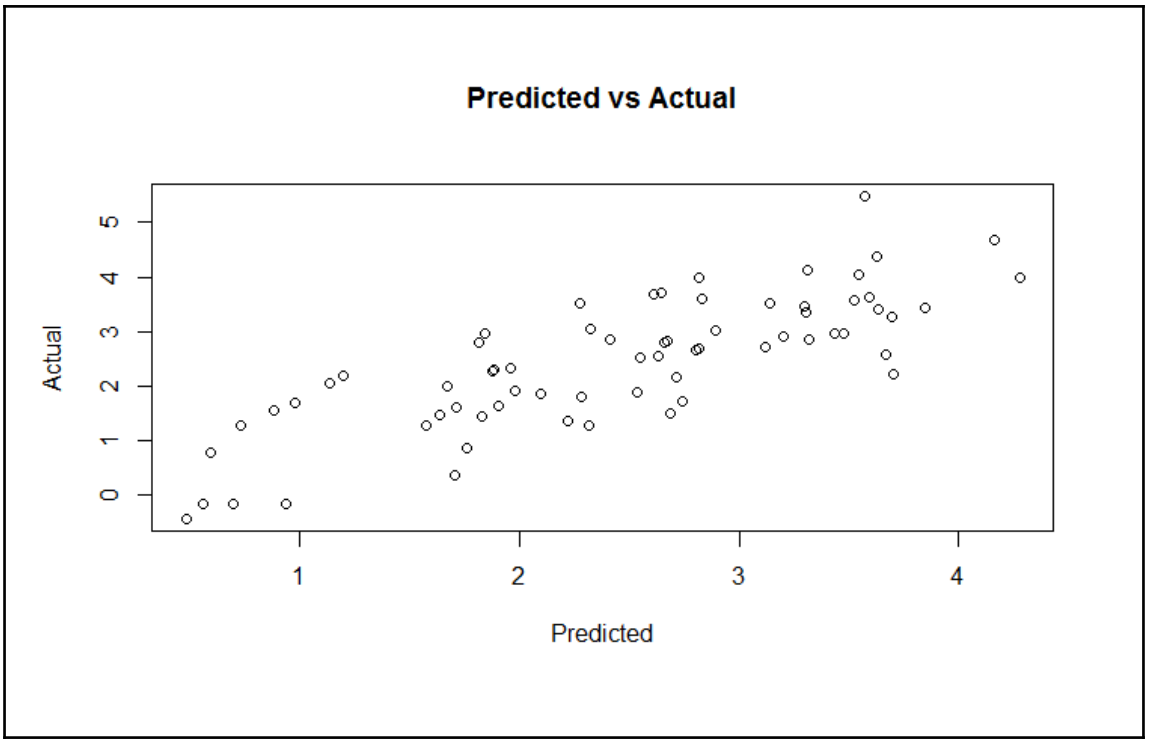

Gleason Score
6.0
6.5
7.0
7.5
8.0
8.5
9.0
0
20
40
60
80
100
Index

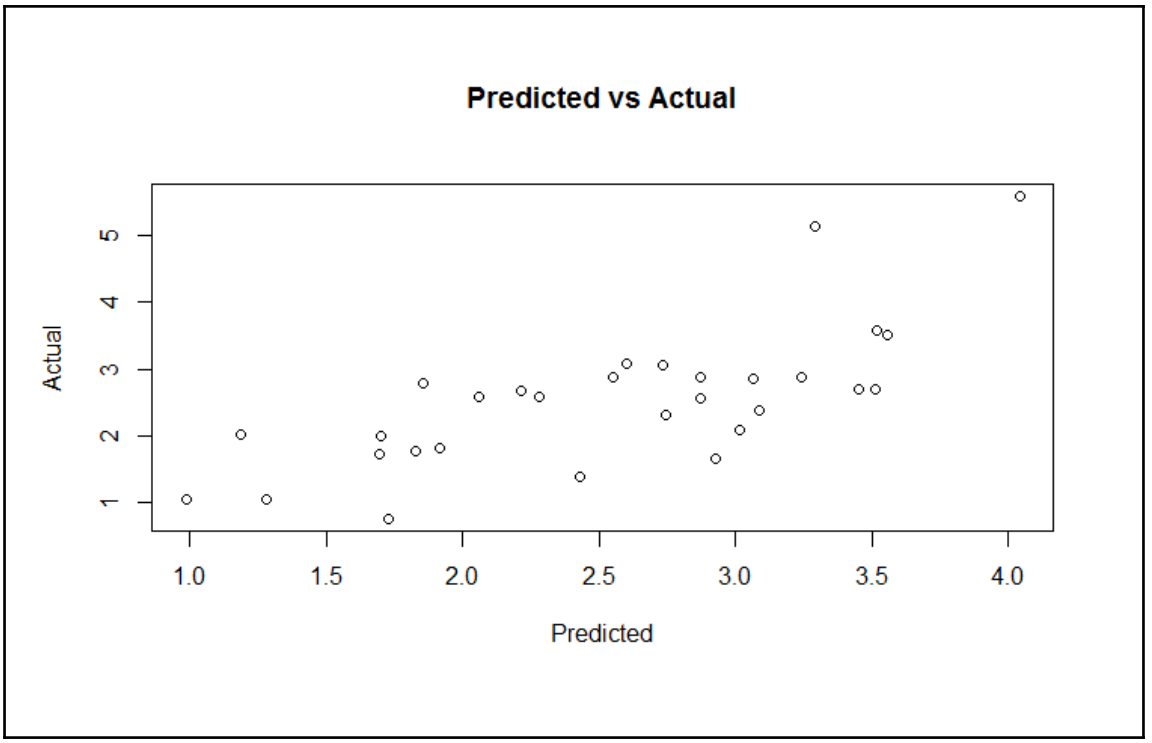

Log of PSA
0
1
2
3
4
5
6
7
8
9
Gleason Score

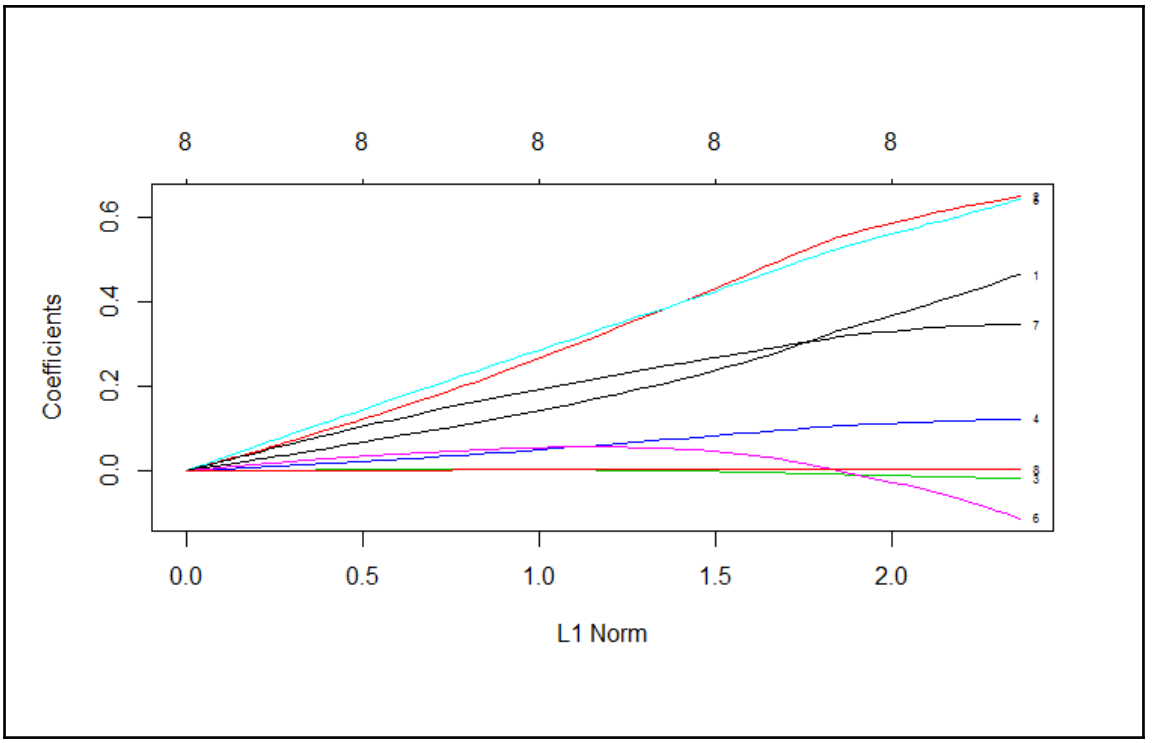

lcavol
0.28 lweight
0.22 0.35 age
0.44 0.35 lbph
0.54 0.16 0.12 -0.09 svi
0.68 0.16 0.13 0.67 lcp
0.5 0.12 0.28 0.12 0.39 0.54 gleason
0.43 0.11 0.28 0.46 0.63 0.65 pgg45
0.73 0.43 0.17 0.18 0.57 0.55 0.48 0.42 lpsa
0.18 train
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8

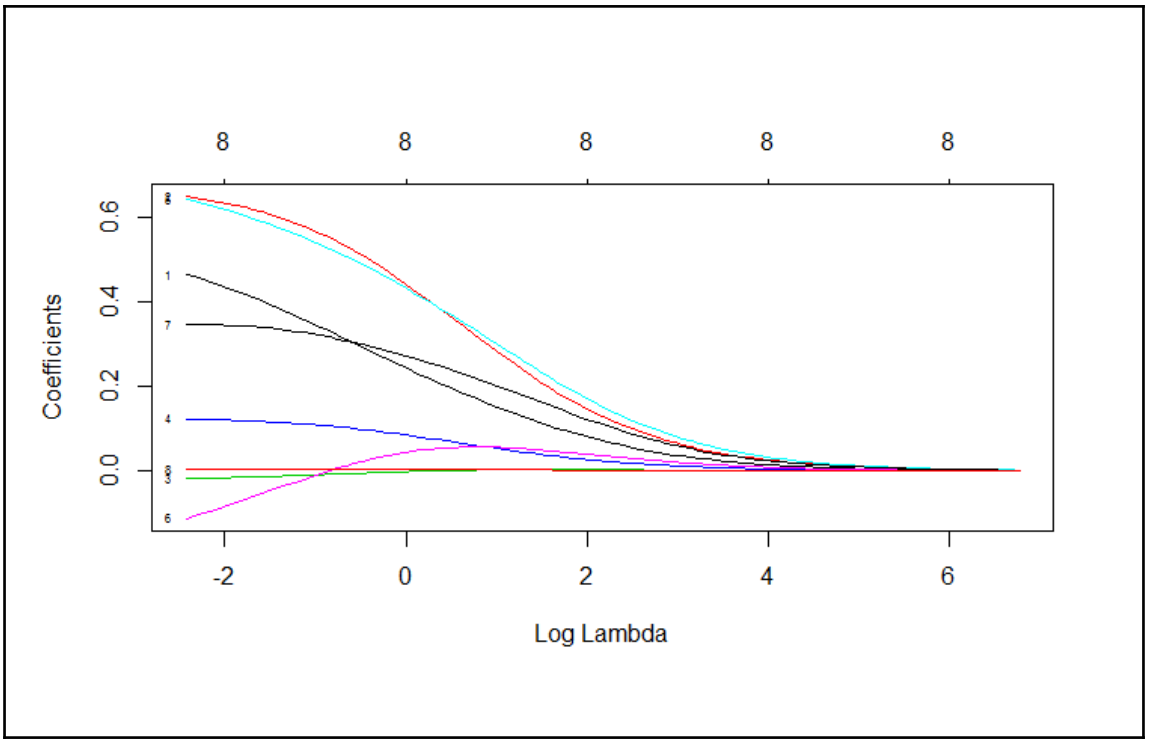

BIC score by Feature Inclusion
BIC
-44
-46
-48
-50
-52
1
2
3
4
5
6
7
8
# of Features

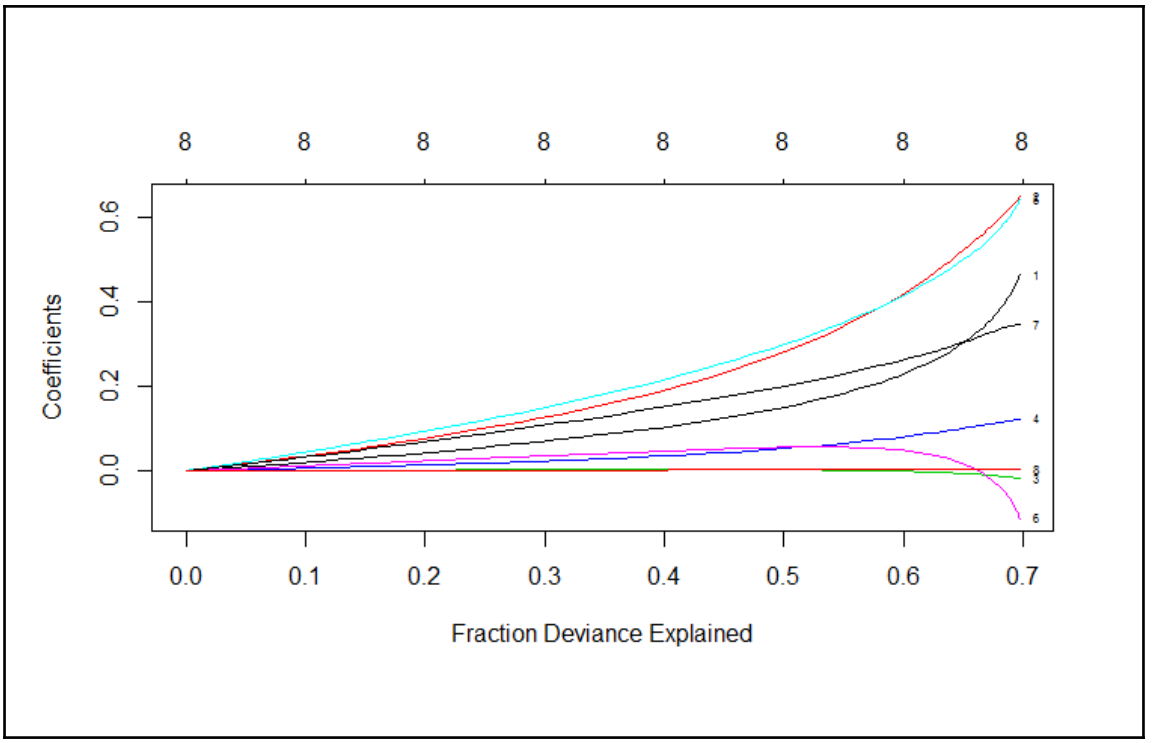

Best Subset Features
bic
-52
-51
-51
-50
-48
-46
-43
-43
(Intercept)
lcavol
lweight
age
lbph
svi
lcp
gleason
pgg45

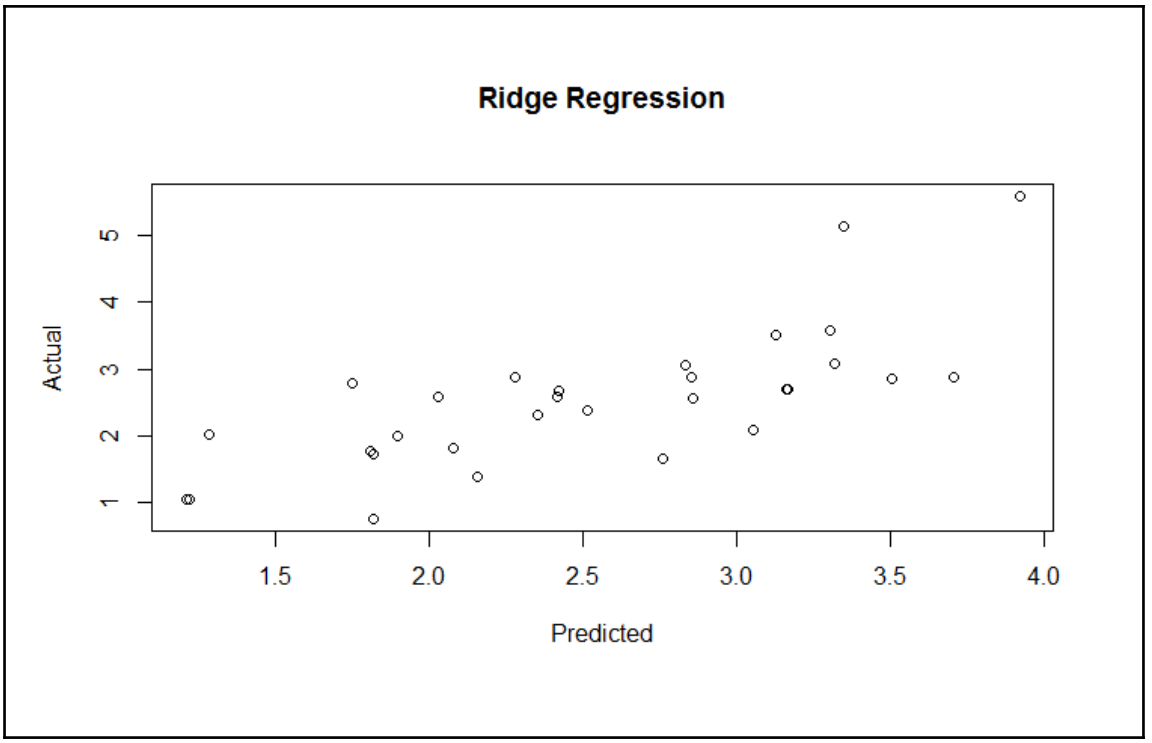

Predicted vs Actual
Actual
0
1
2
3
4
5
1
2
3
4
Predicted

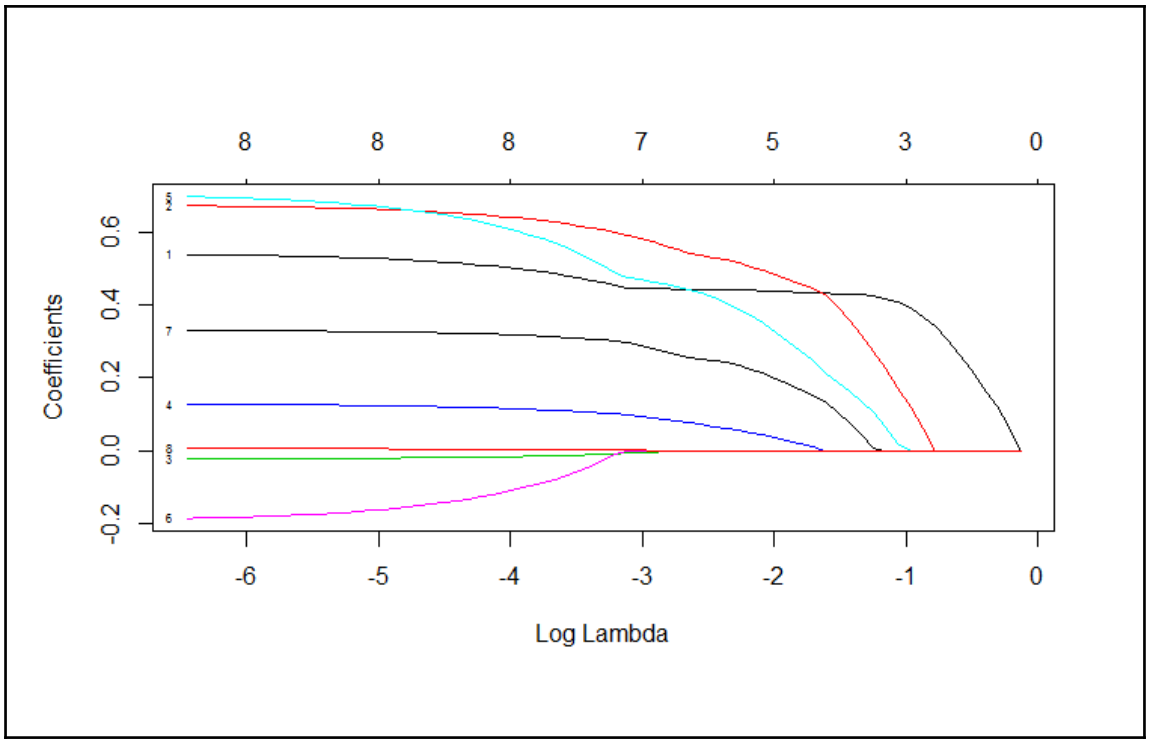

Predicted vs Actual
Actual
1
2
3
4
5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Predicted

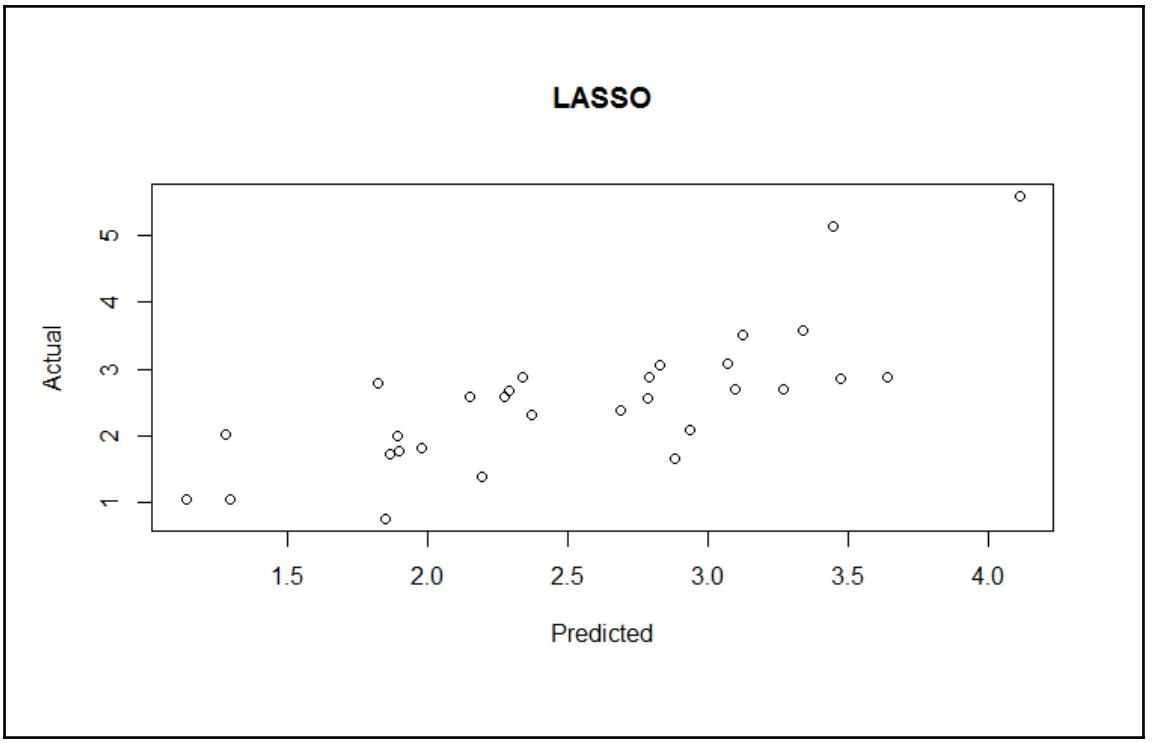

8
8
8
8
8
Coefficients
0.0
0.2
0.4
0.6
0.0
0.5
1.0
1.5
2.0
L1 Norm
1
7
4
6

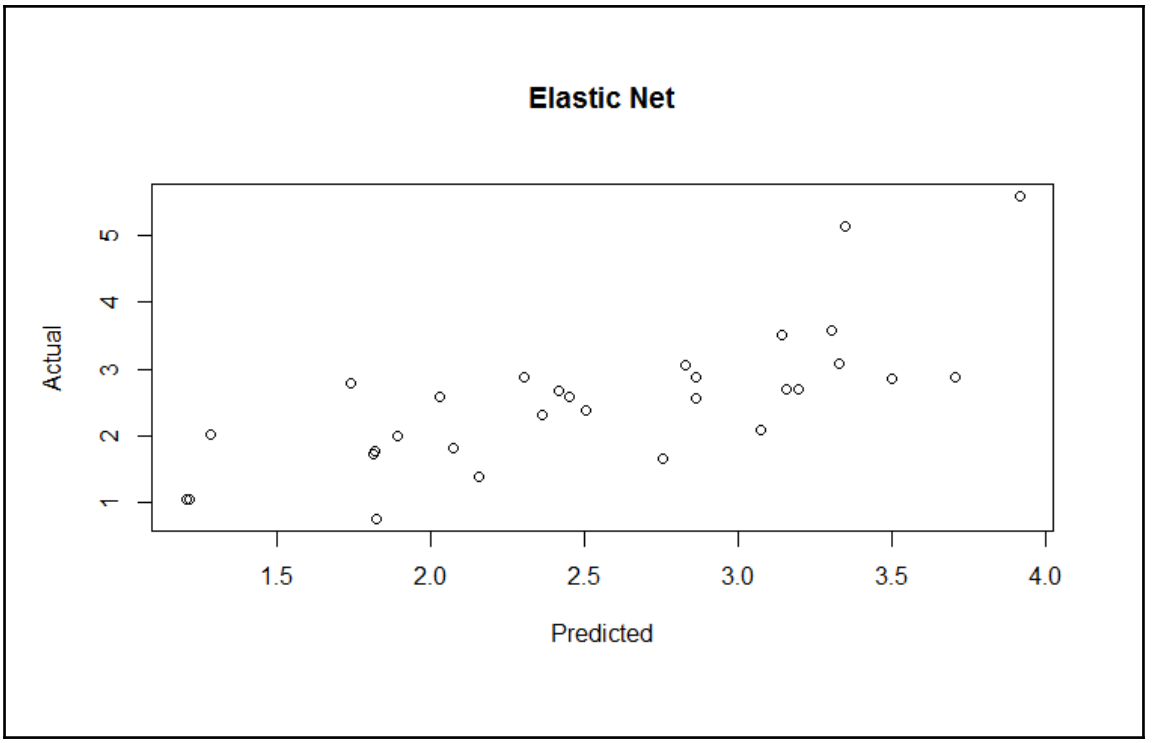

8
8
8
8
8
Coefficients
0.0
0.2
0.4
0.6
1
7
4
6
-2
0
2
4
6
Log Lambda

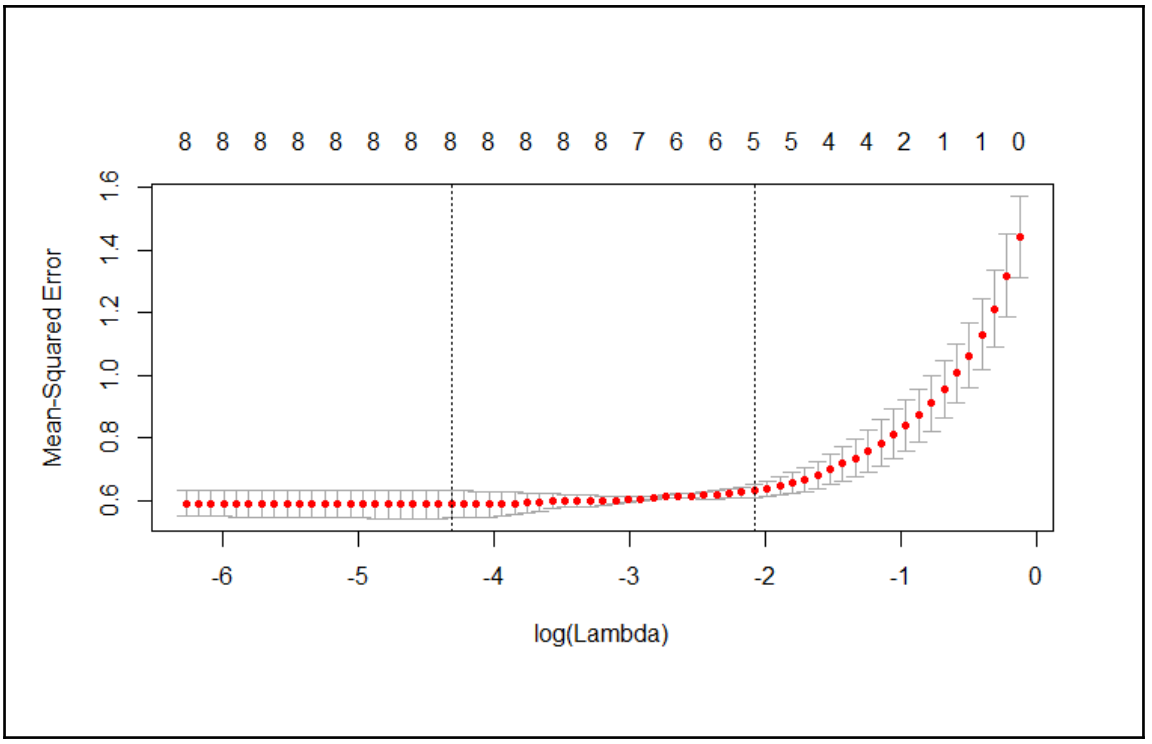

8
8
8
8
8
8
8
8
Coefficients
0.0
0.2
0.4
0.6
1
7
4
6
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Fraction Deviance Explained

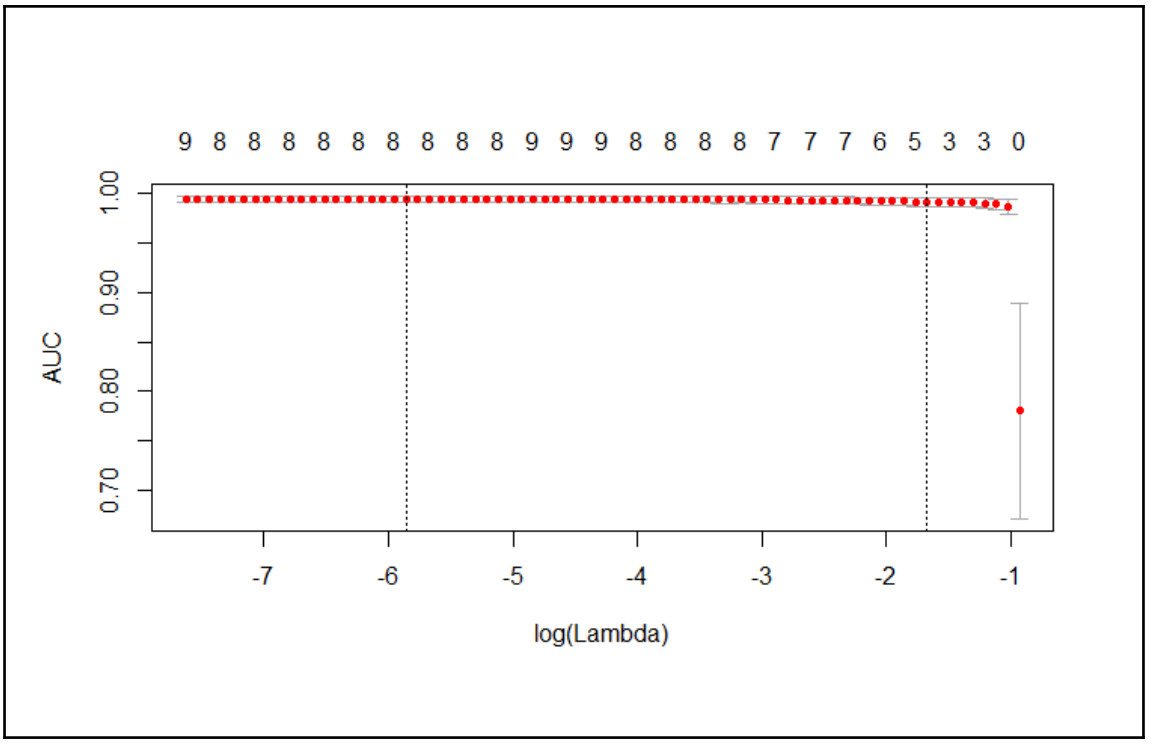

Ridge Regression
Actual
1
2
3
4
5
1.5
2.0
2.5
3.0
3.5
4.0
Predicted

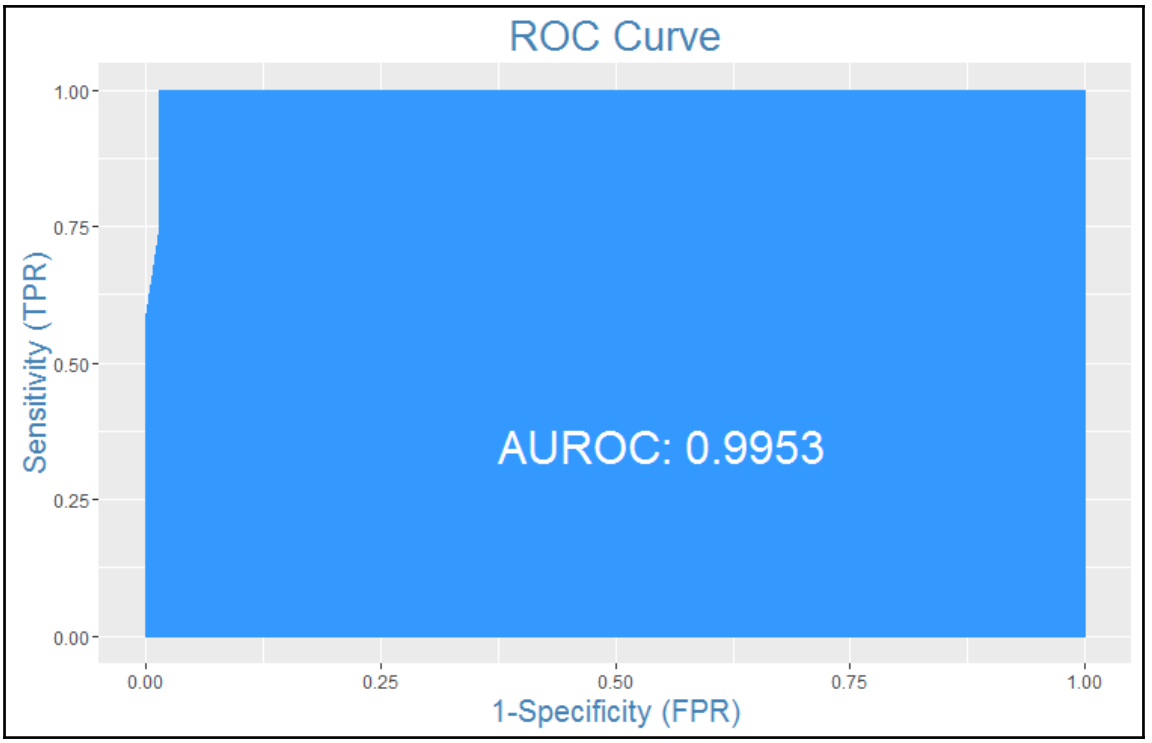

8
8
8
7
5
3
0
Coefficients
0.6
0.4
0.2
0.0
-0.2
-6
-5
-4
-3
-2
-1
0
Log Lambda


LASSO
Actual
1
2
3
4
5
1.5
2.0
2.5
3.0
3.5
4.0
Predicted

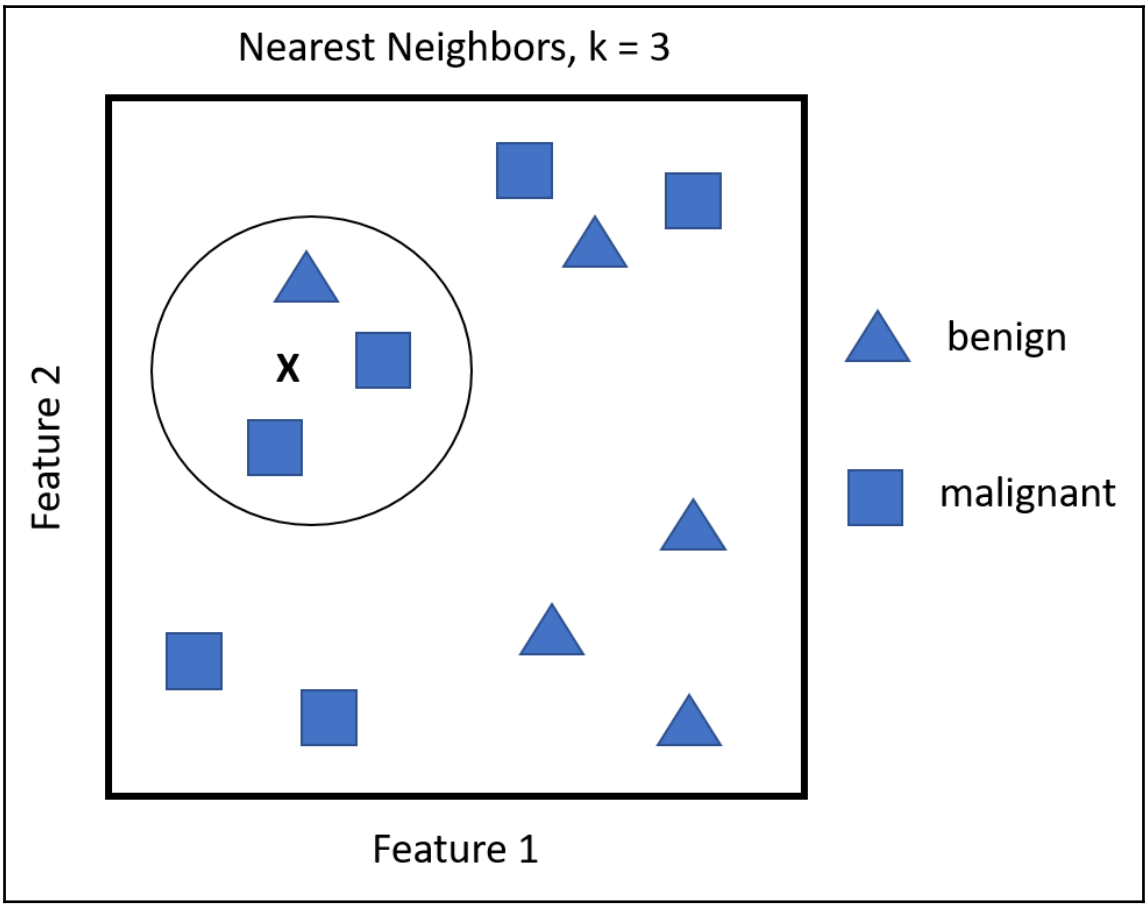

Elastic Net
Actual
1
2
3
4
5
1.5
2.0
2.5
3.0
3.5
4.0
Predicted

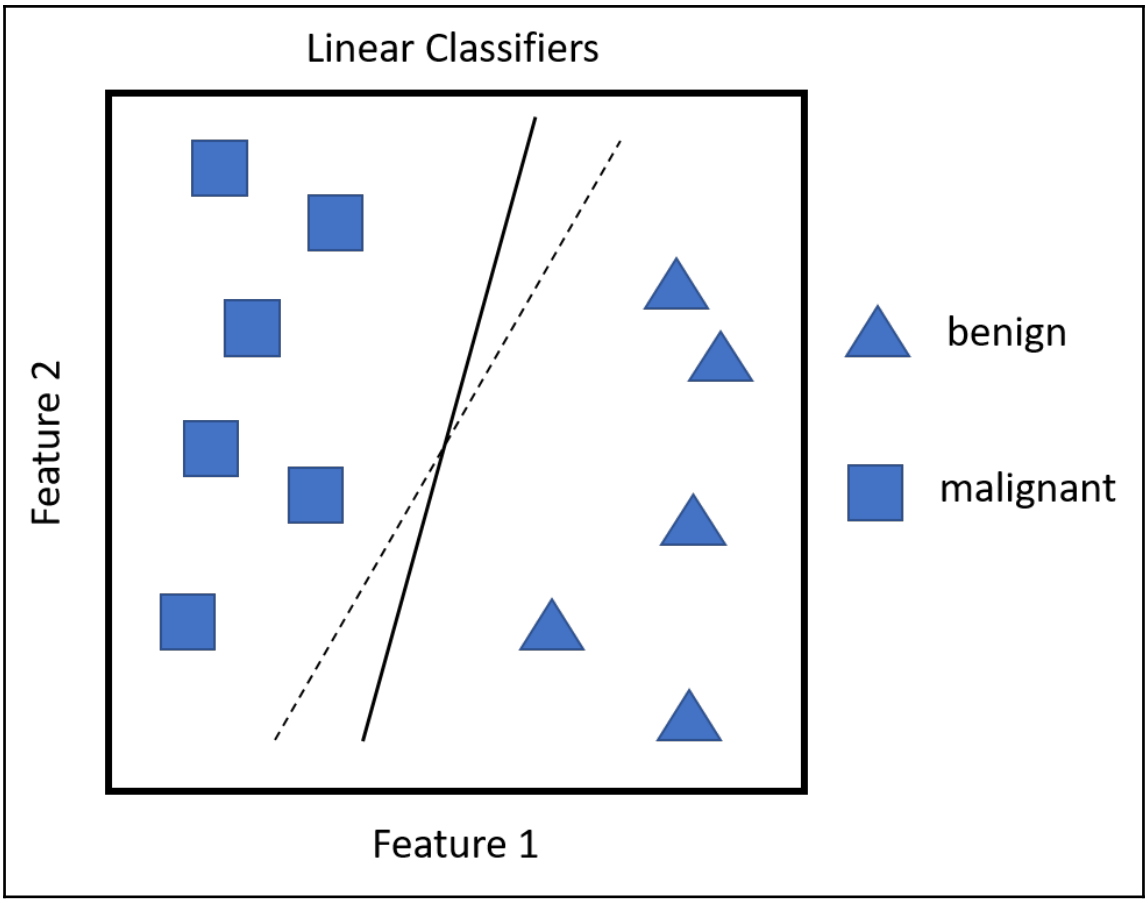

8 8 8 8 8 8 8 8 8 8 8 8 7 6 6 5 5 4 4 2 1 1 0
Mean-Squared Error
0.6 0.8 1.0 1.2 1.4 1.6
-6 -5 -4 -3 -2 -1 0
log(Lambda)

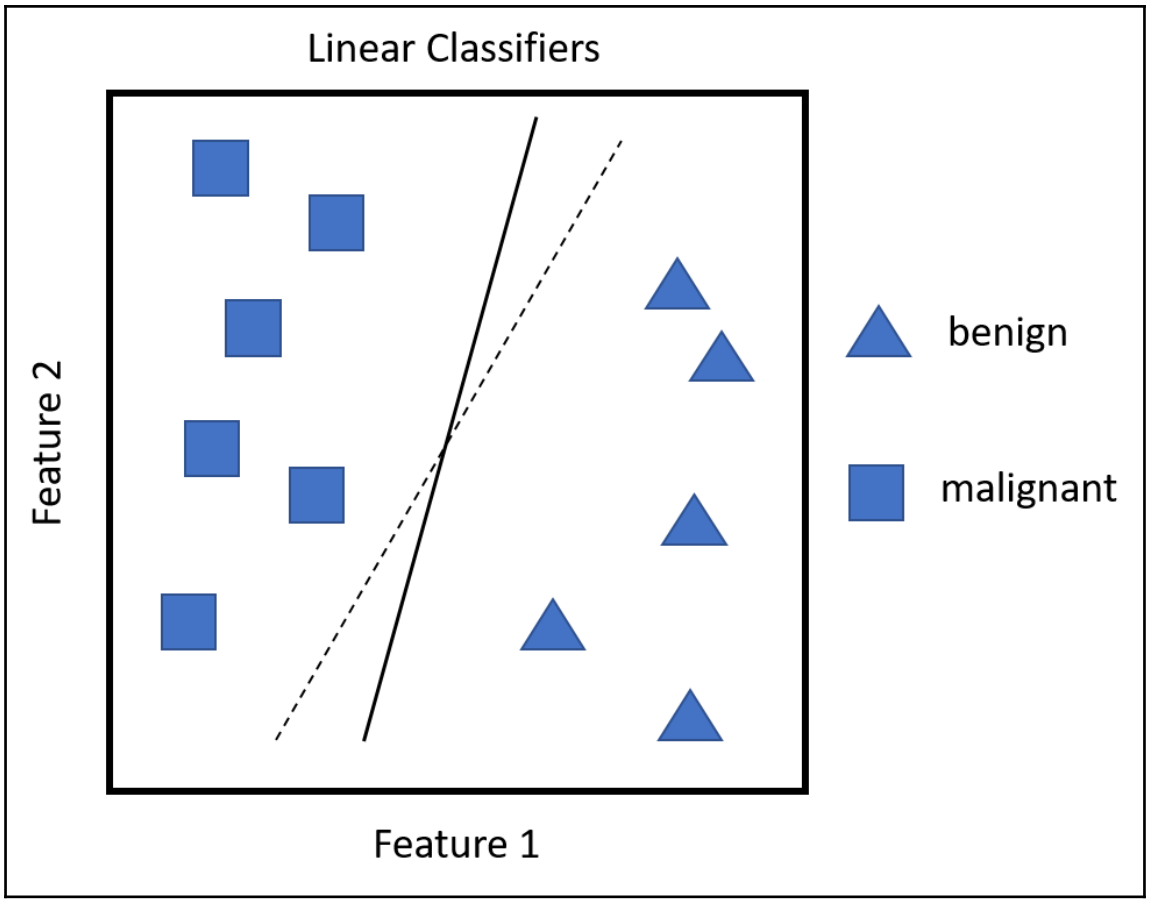

9 8 8 8 8 8 8 8 8 8 9 9 9 8 8 8 8 7 7 7 6 5 3 3 0
AUC
1.00
0.90
0.80
0.70
-7
-6
-5
-4
-3
-2
-1
log(Lambda)

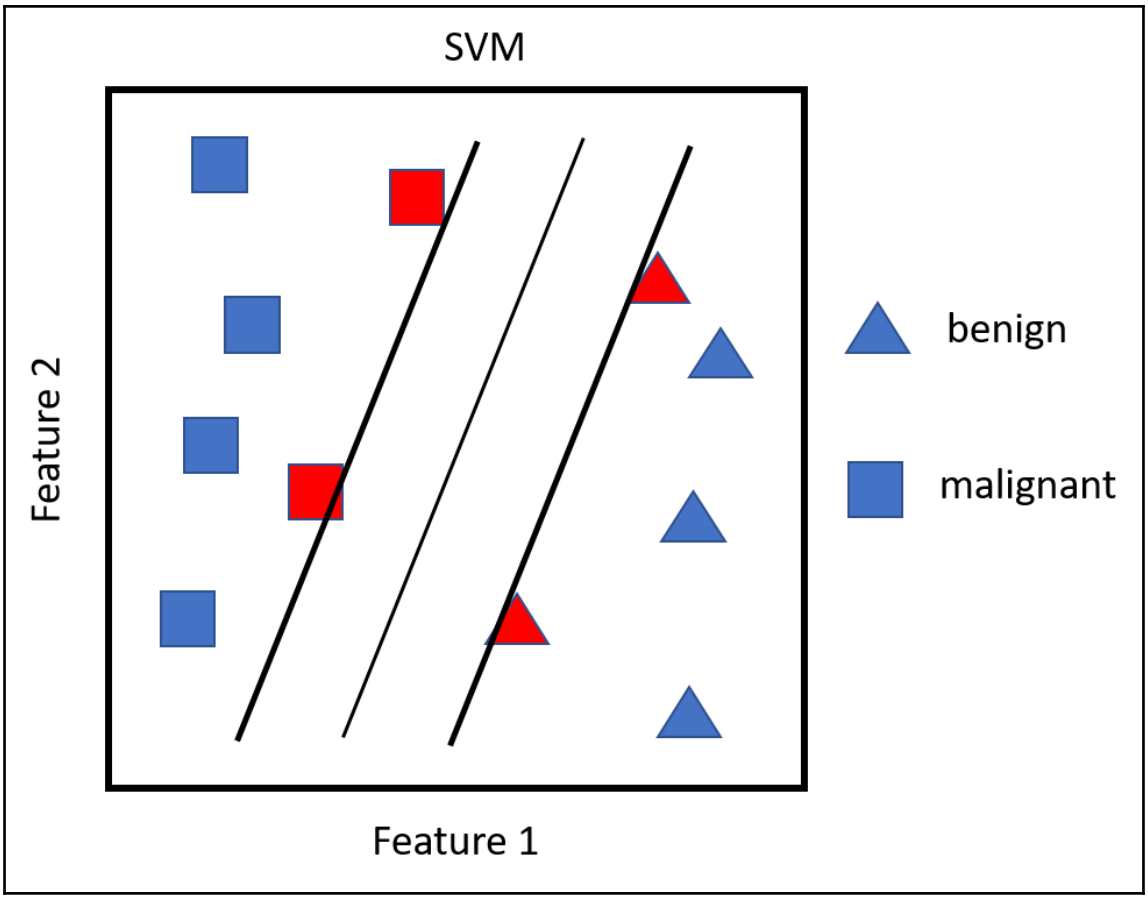

ROC Curve
Sensitivity (TPR)
1.00
0.75
0.50
0.25
0.00
AUROC: 0.9953
0.00
0.25
0.50
0.75
1.00
1-Specificity (FPR)

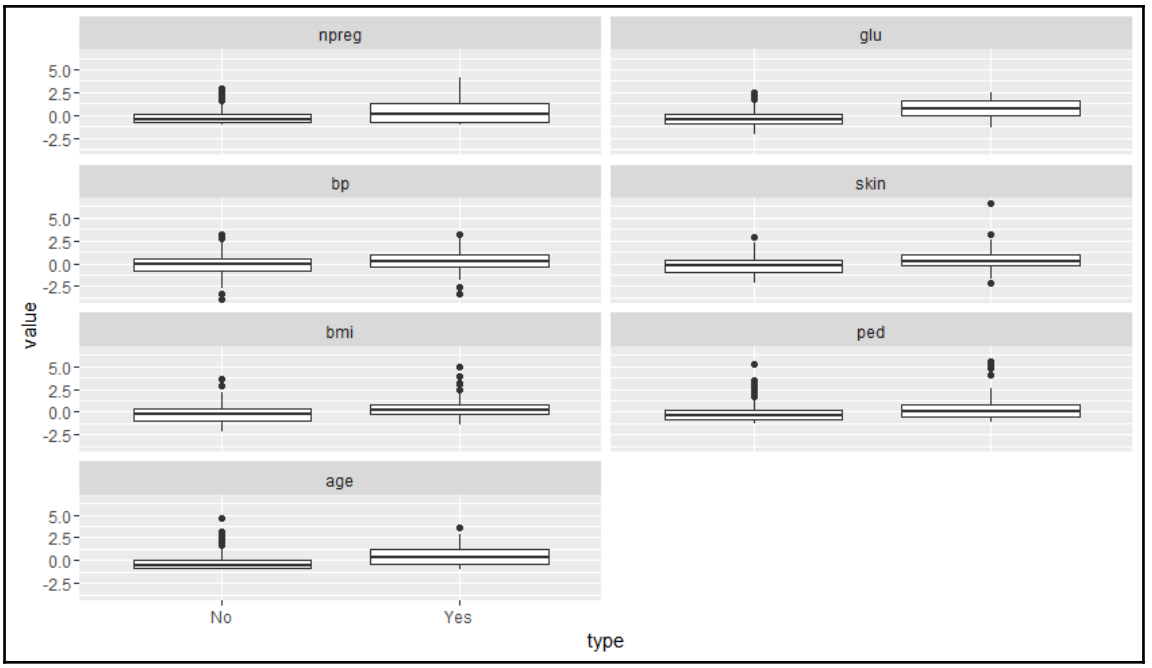
Chapter 5: More Classification Techniques - K-Nearest Neighbors and Support Vector Machines

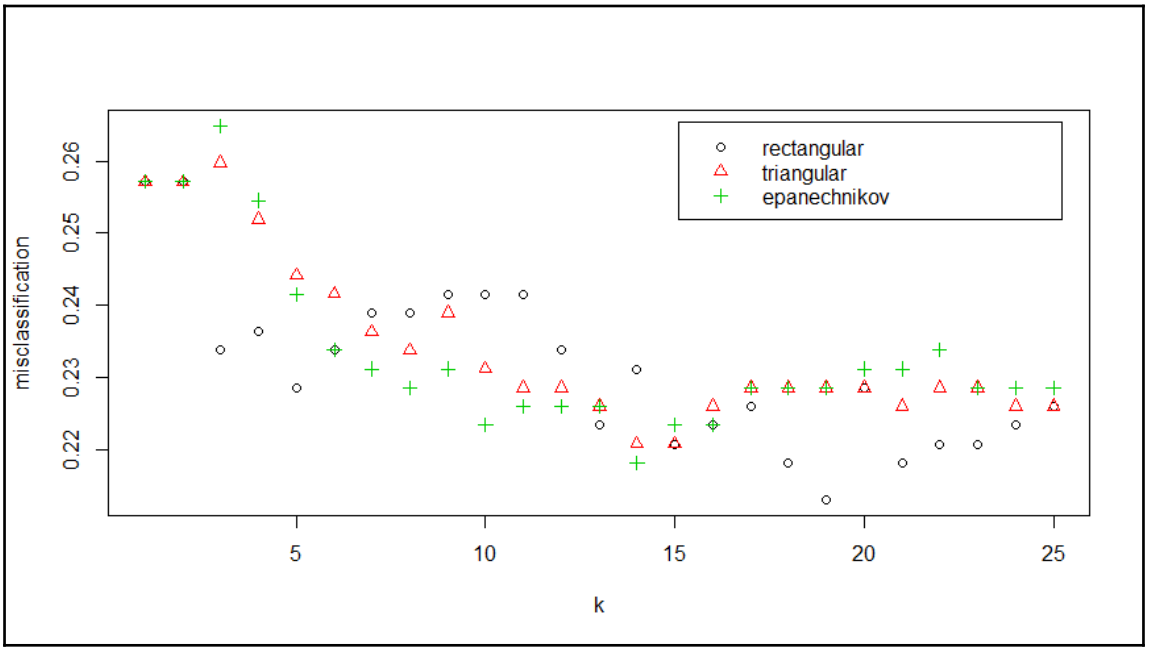



Linear Classifiers
Feature 2
Feature 1
benign
malignant

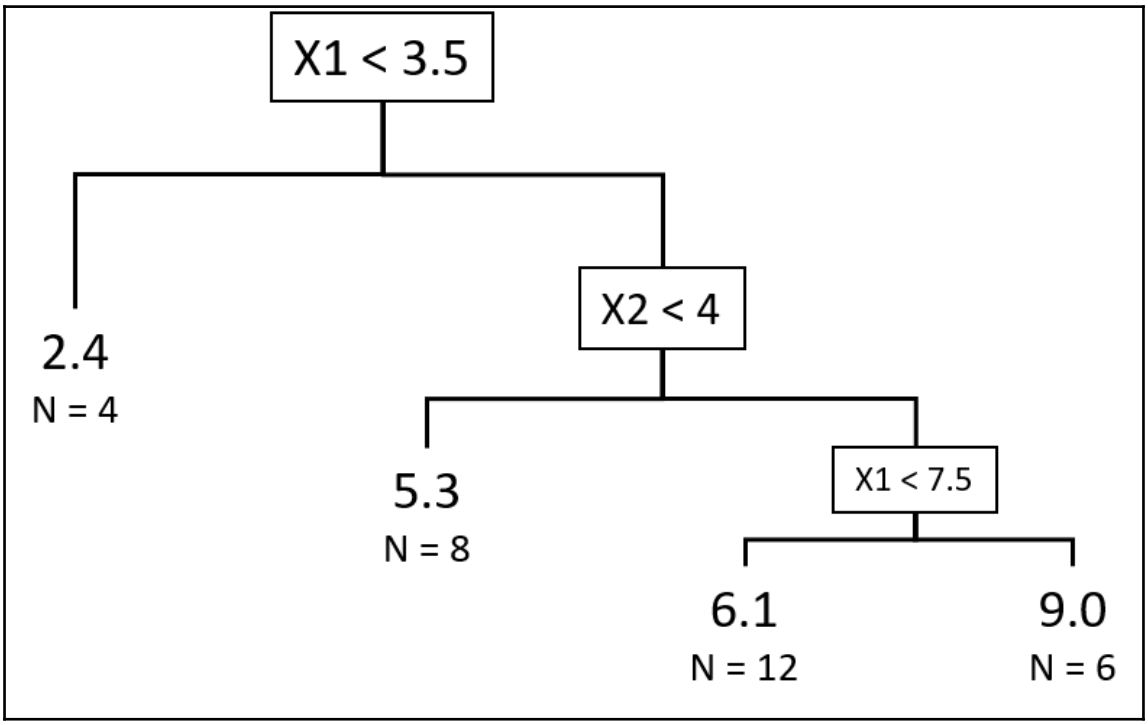

Linear Classifiers
Feature 2
Feature 1
benign
malignant

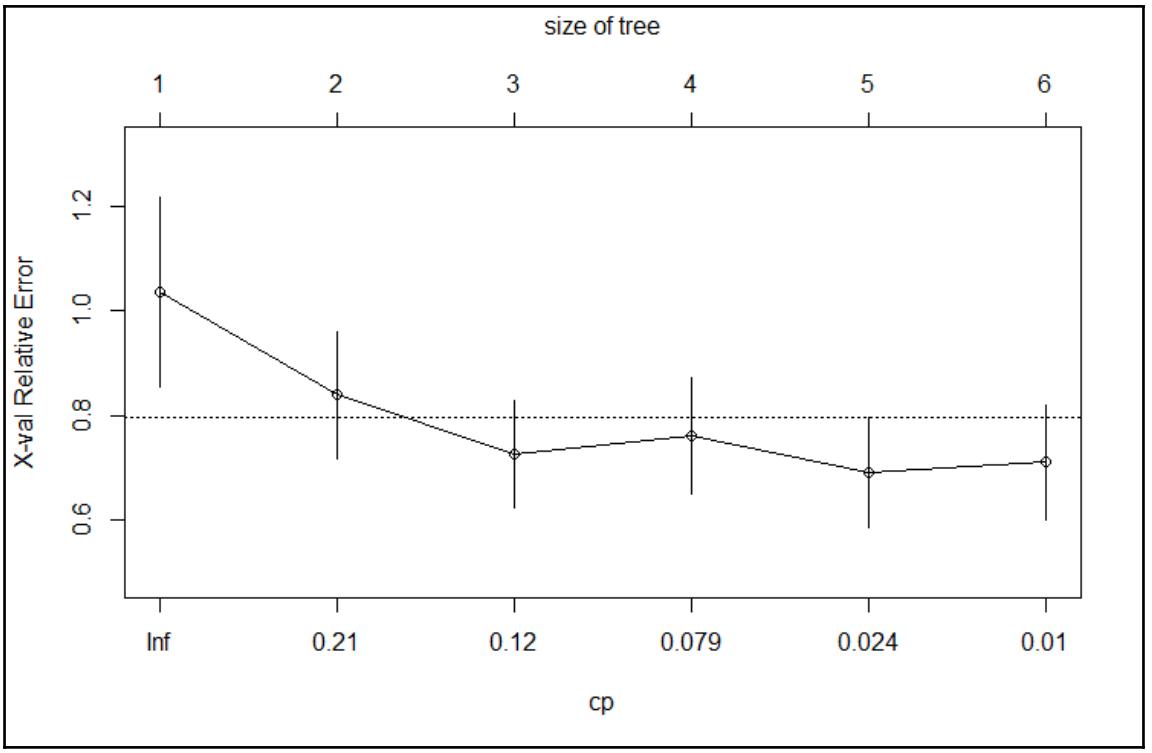

SVM
Feature 2
Feature 1
benign
malignant

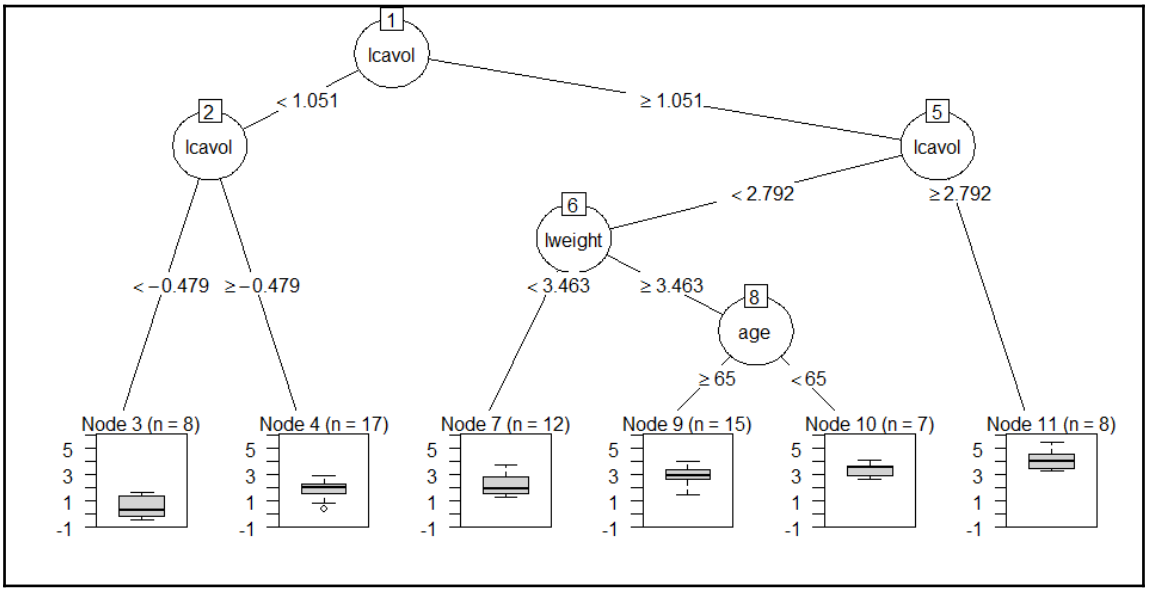

npreg
glu
bp
skin
bmi
ped
age
5.0
2.5
0.0
-2.5
value
No
Yes
type

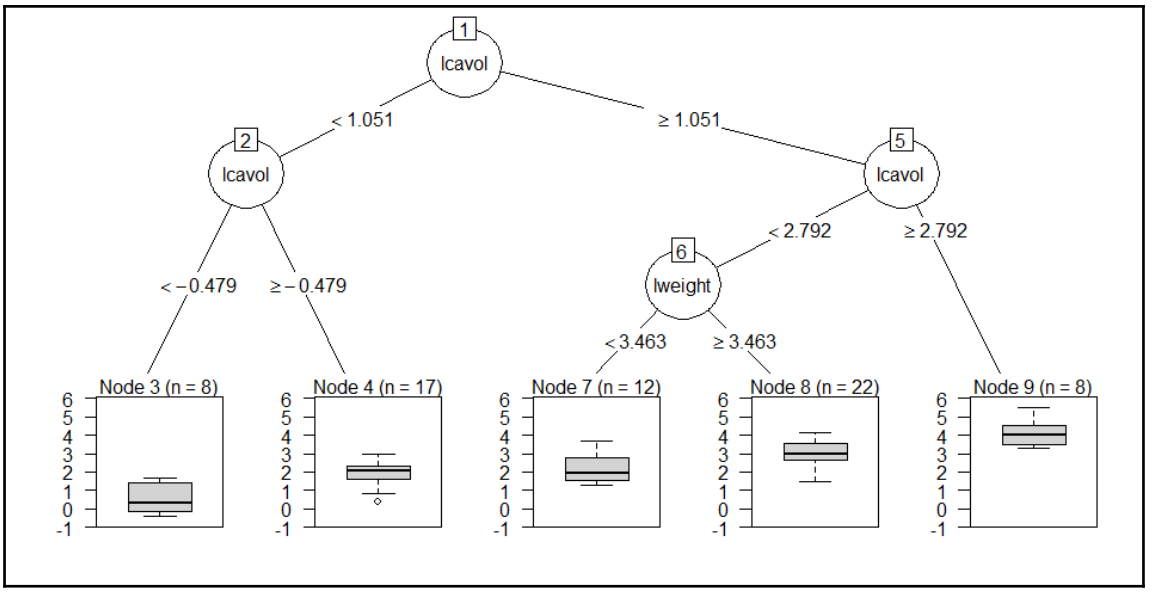

rectangular
triangular
epanechnikov
misclassification
0.22
0.23
0.24
0.25
0.26
5
10
15
20
25
k

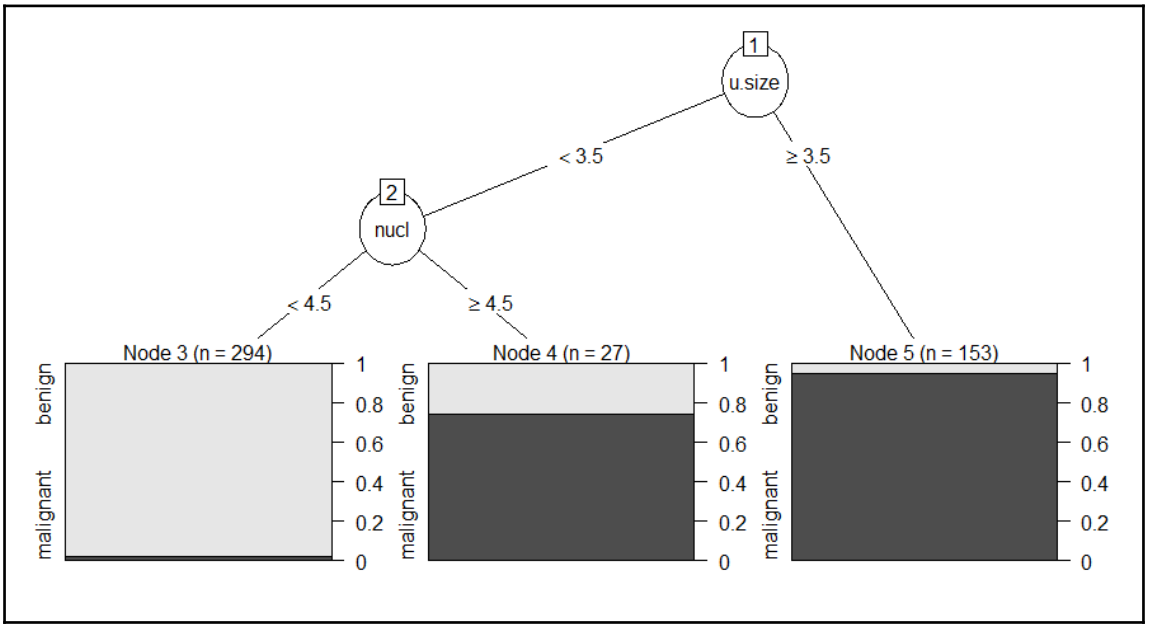
Chapter 6: Classification and Regression Trees

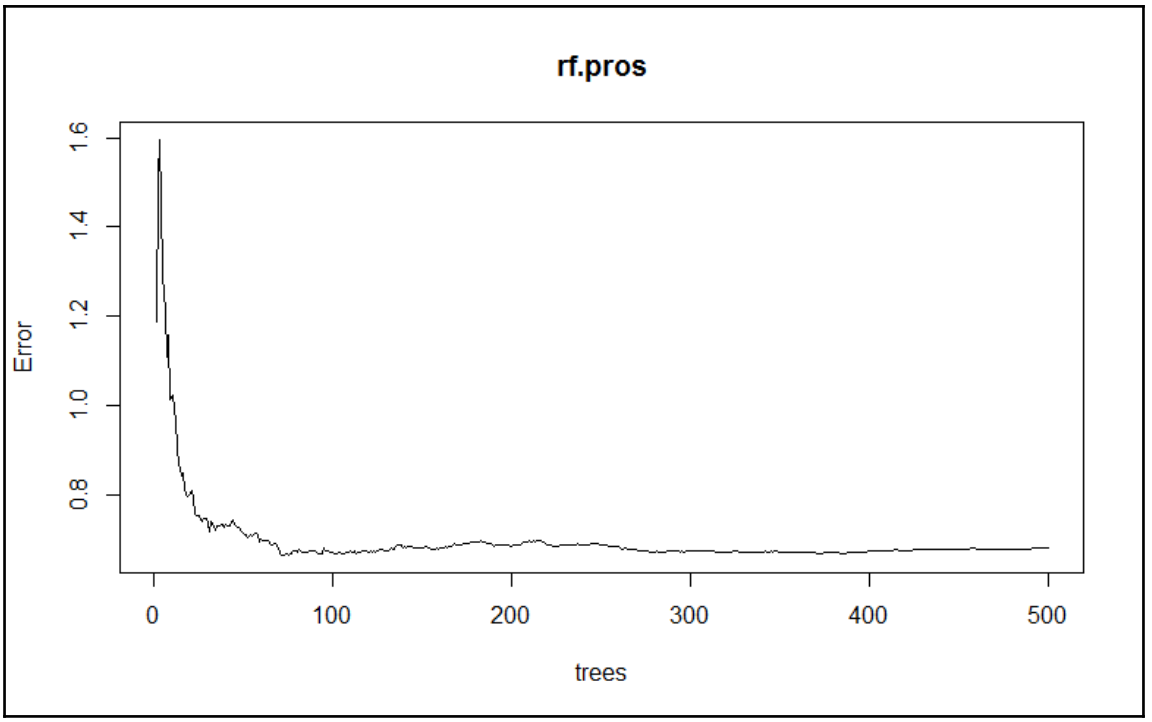


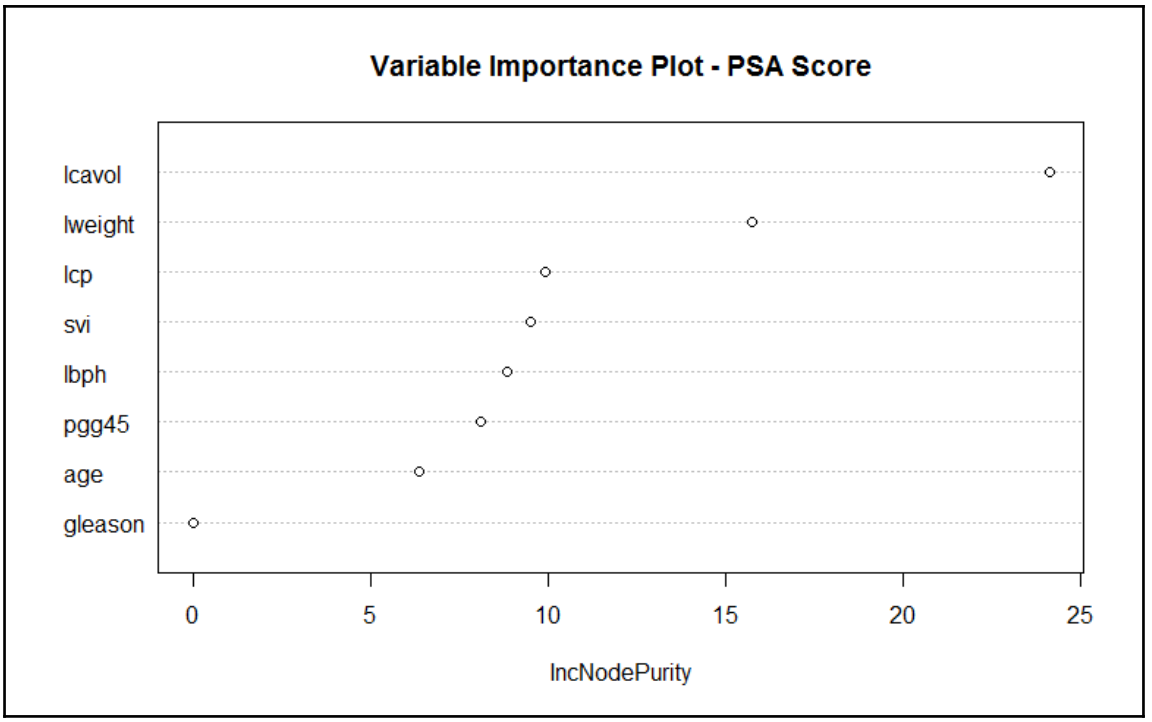

size of tree
1
2
3
4
5
6
X-val Relative Error
0.6
0.8
1.0
1.2
Inf
0.21
0.12
0.079
0.024
0.01
cp

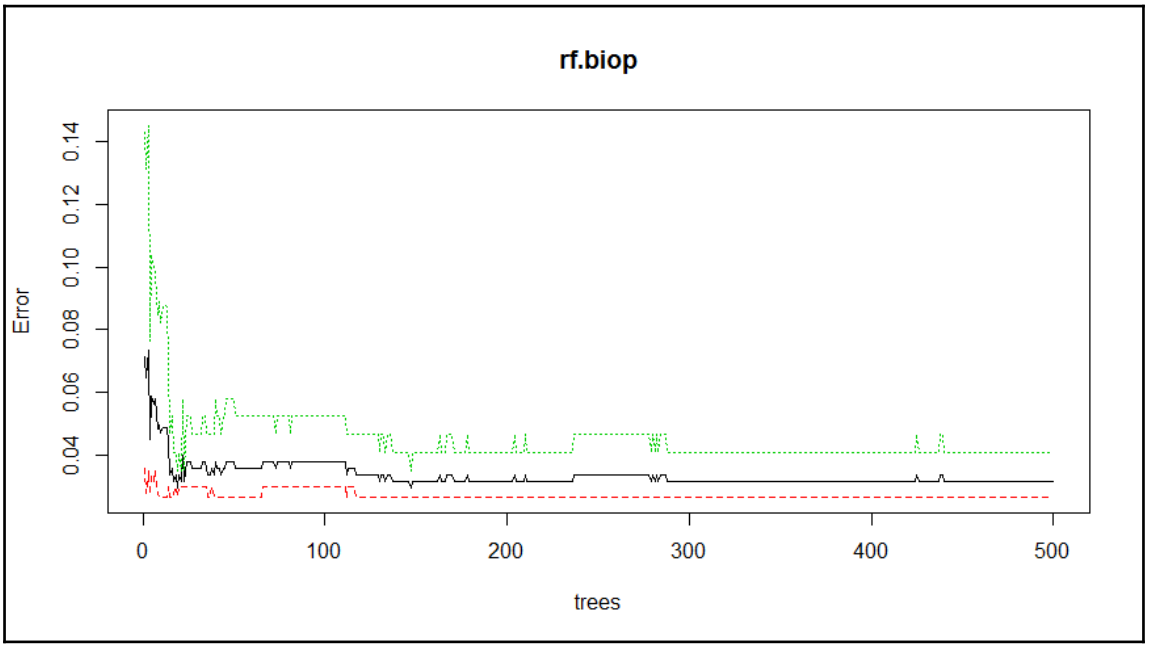

1
lcavol
< 1.051
≥ 1.051
2
lcavol
5
lcavol
< 2.792
≥ 2.792
6
lweight
< −0.479
≥ −0.479
< 3.463
≥ 3.463
8
age
≥ 65
< 65
Node 3 (n = 8)
Node 4 (n = 17)
Node 7 (n = 12)
Node 9 (n = 15)
Node 10 (n = 7)
Node 11 (n = 8)
5
3
1
-1

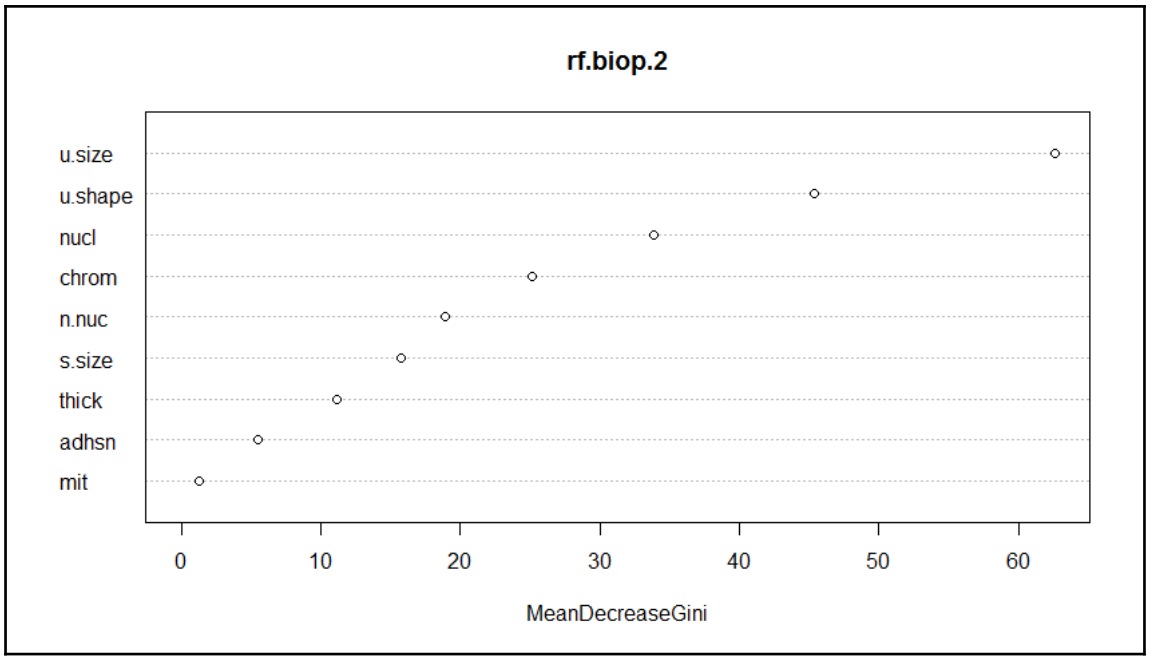

1
lcavol
< 1.051
≥ 1.051
2
lcavol
5
lcavol
< −0.479
≥ −0.479
< 2.792
≥ 2.792
6
lweight
< 3.463
≥ 3.463
Node 3 (n = 8)
Node 4 (n = 17)
Node 7 (n = 12)
Node 8 (n = 22)
Node 9 (n = 8)

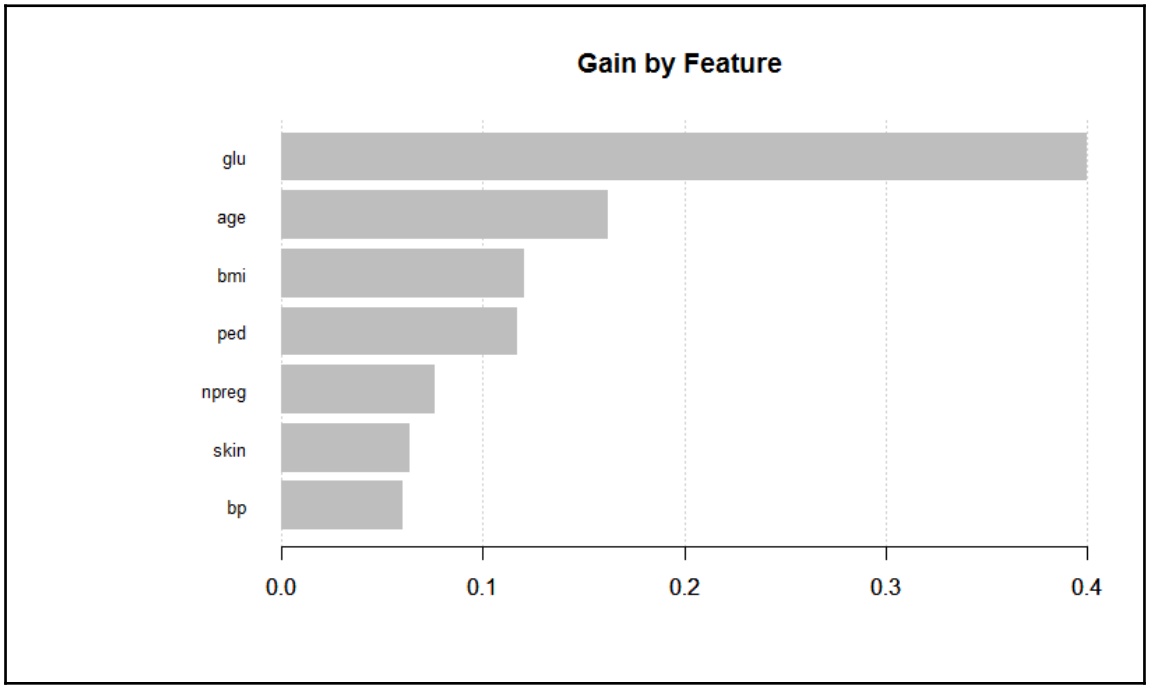

1
u.size
< 3.5
≥ 3.5
2
nucl
< 4.5
≥ 4.5
Node 3 (n = 294)
Node 4 (n = 27)
Node 5 (n = 153)
benign
malignant

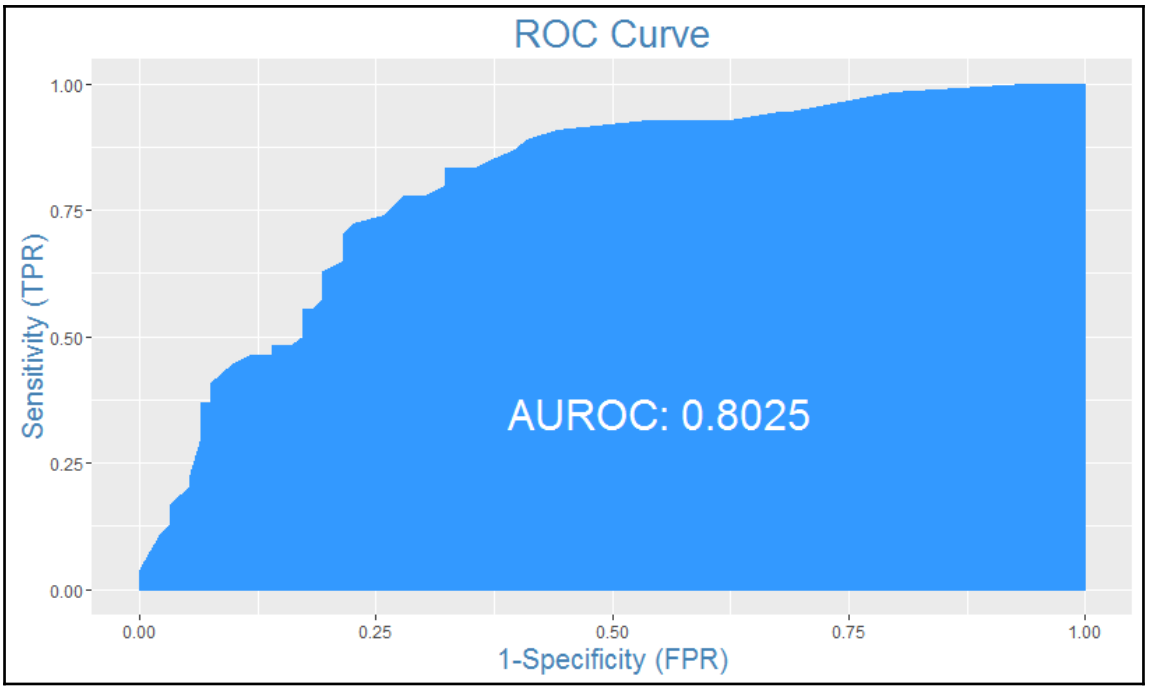

rf.pros
Error
0.8
1.0
1.2
1.4
1.6
0
100
200
300
400
500
trees


Variable Importance Plot - PSA Score
lcavol
lweight
lcp
svi
lbph
pgg45
age
gleason
0
5
10
15
20
25
IncNodePurity

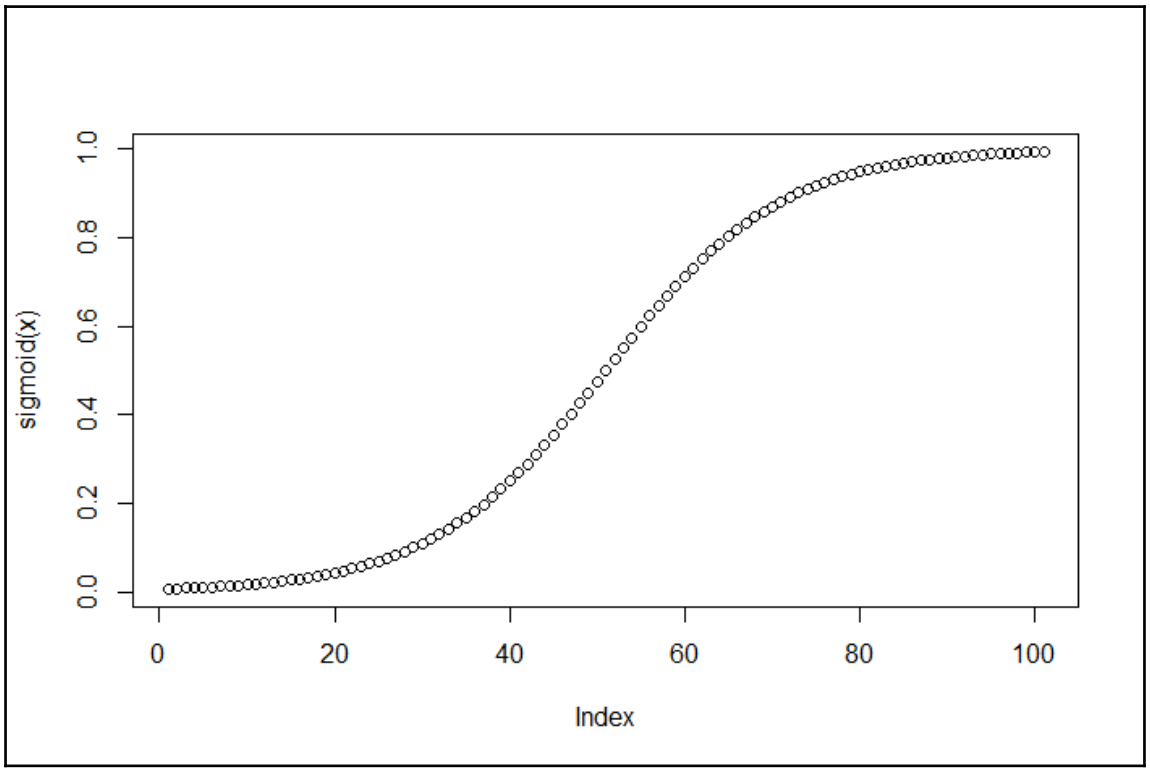

rf.biop
Error
0.04
0.06
0.08
0.10
0.12
0.14
0
100
200
300
400
500
trees

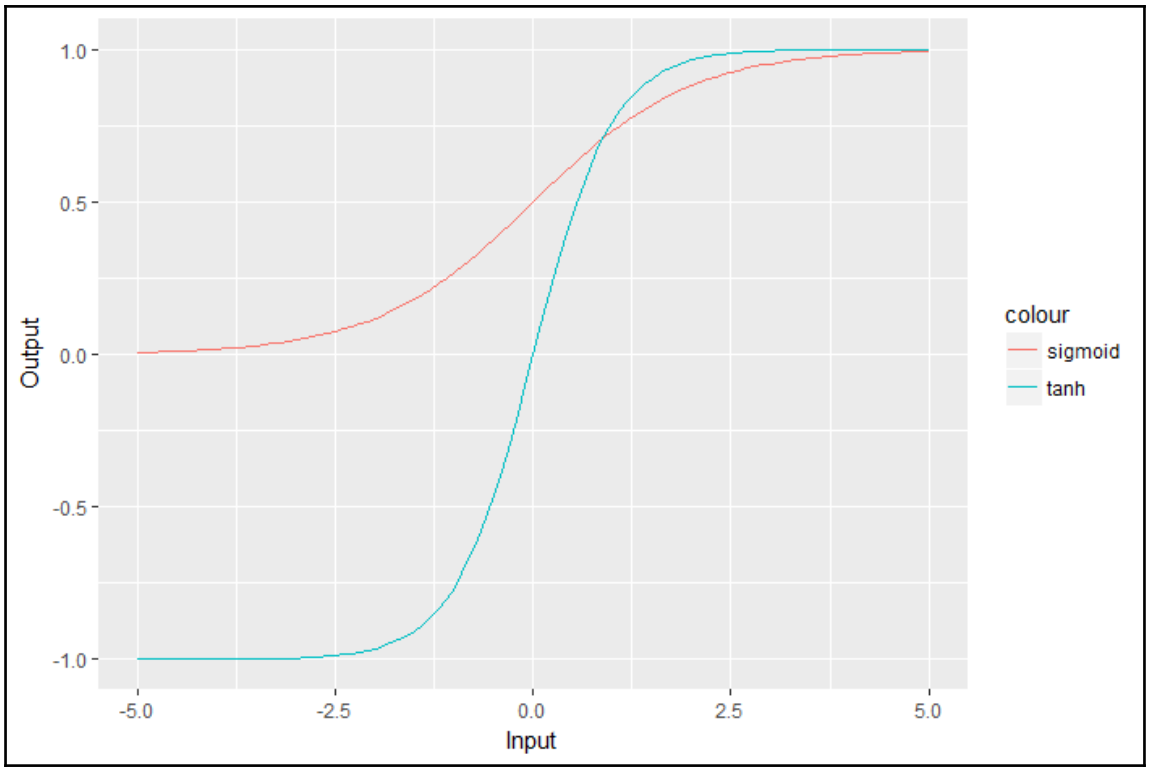

rf.biop.2
u.size
u.shape
nucl
chrom
n.nuc
s.size
thick
adhsn
mit
0
10
20
30
40
50
60
MeanDecreaseGini

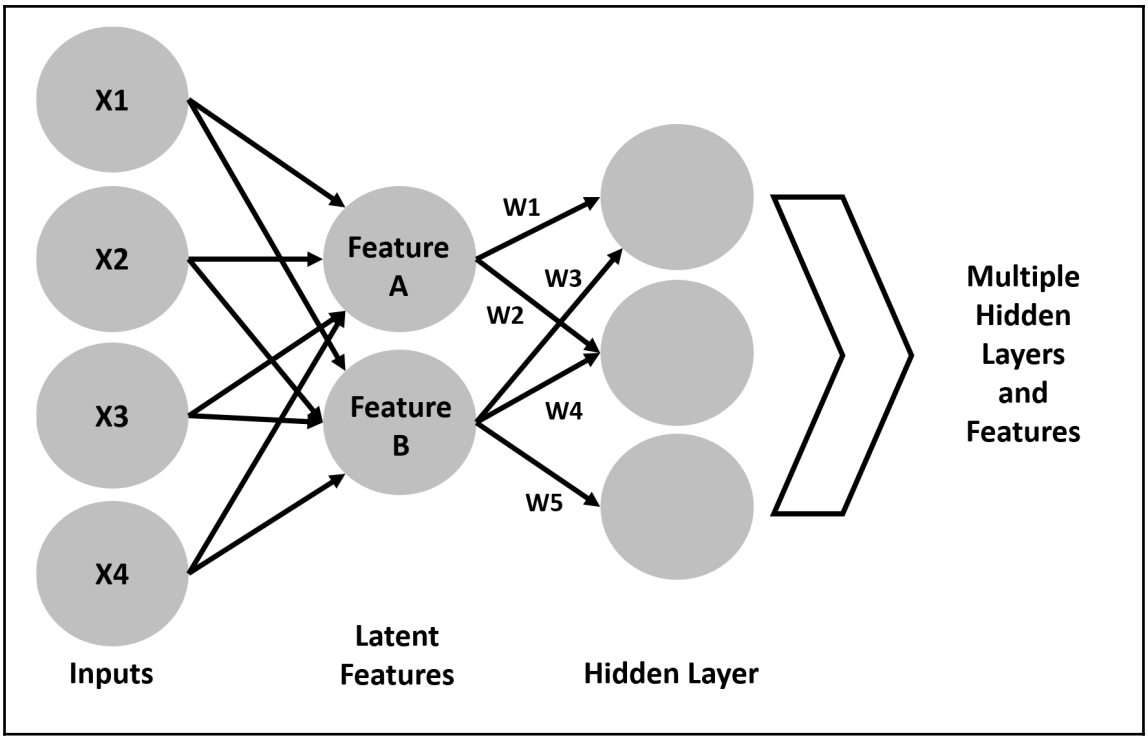

Gain by Feature
glu
age
bmi
ped
npreg
skin
bp
0.0
0.1
0.2
0.3
0.4

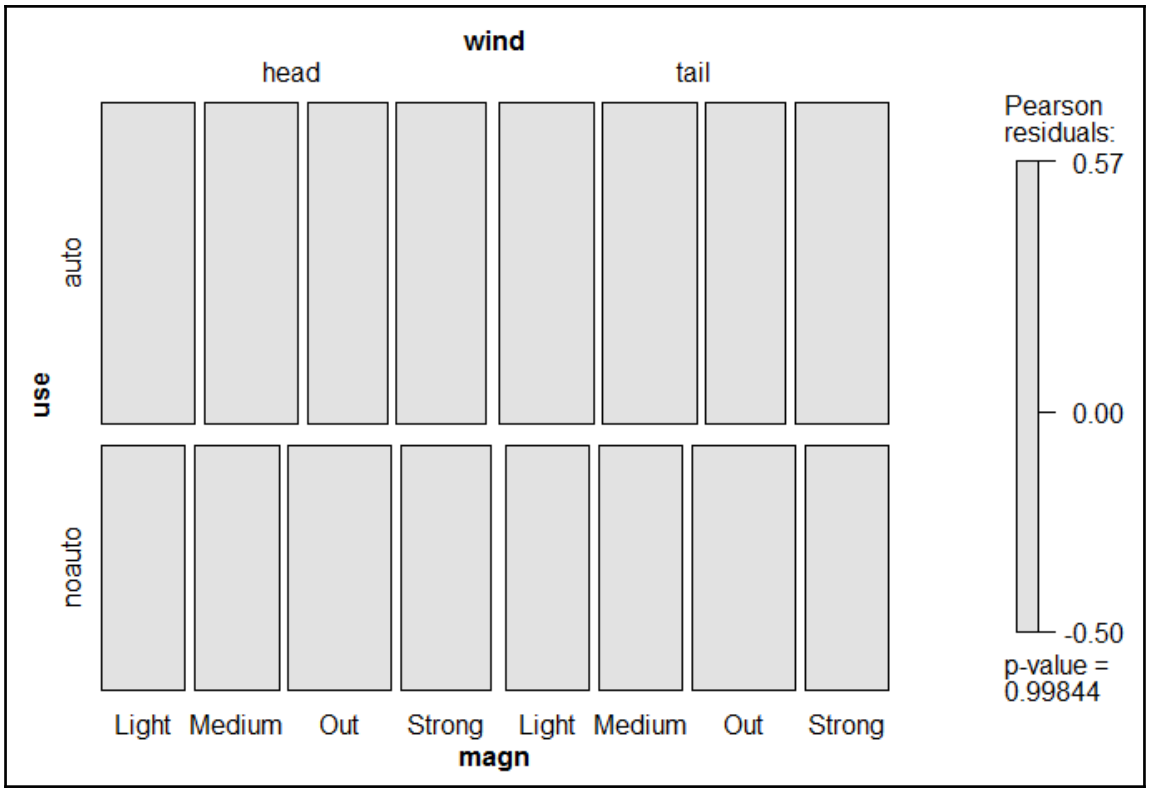

ROC Curve
Sensitivity (TPR)
1.00
0.75
0.50
0.25
0.00
AUROC: 0.8025
0.00
0.25
0.50
0.75
1.00
1-Specificity (FPR)

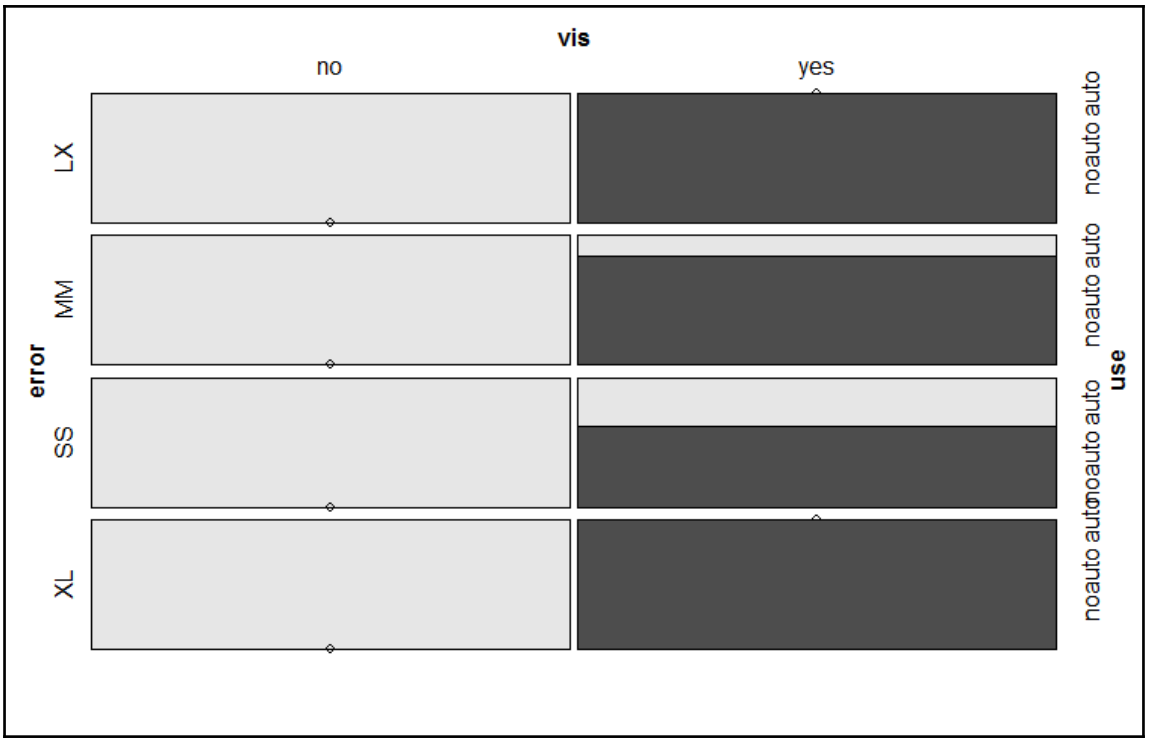
Chapter 7: Neural Networks and Deep Learning

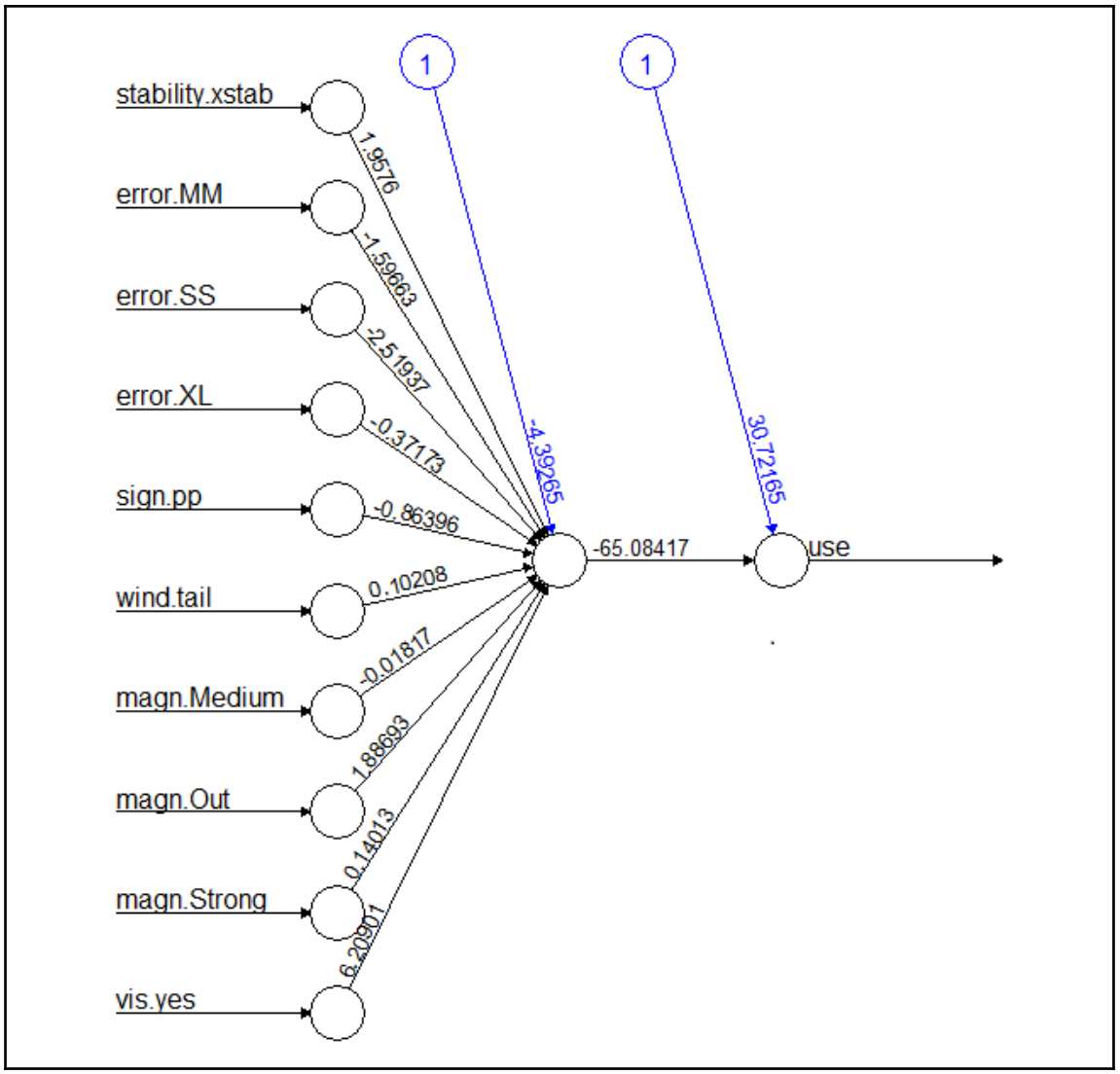


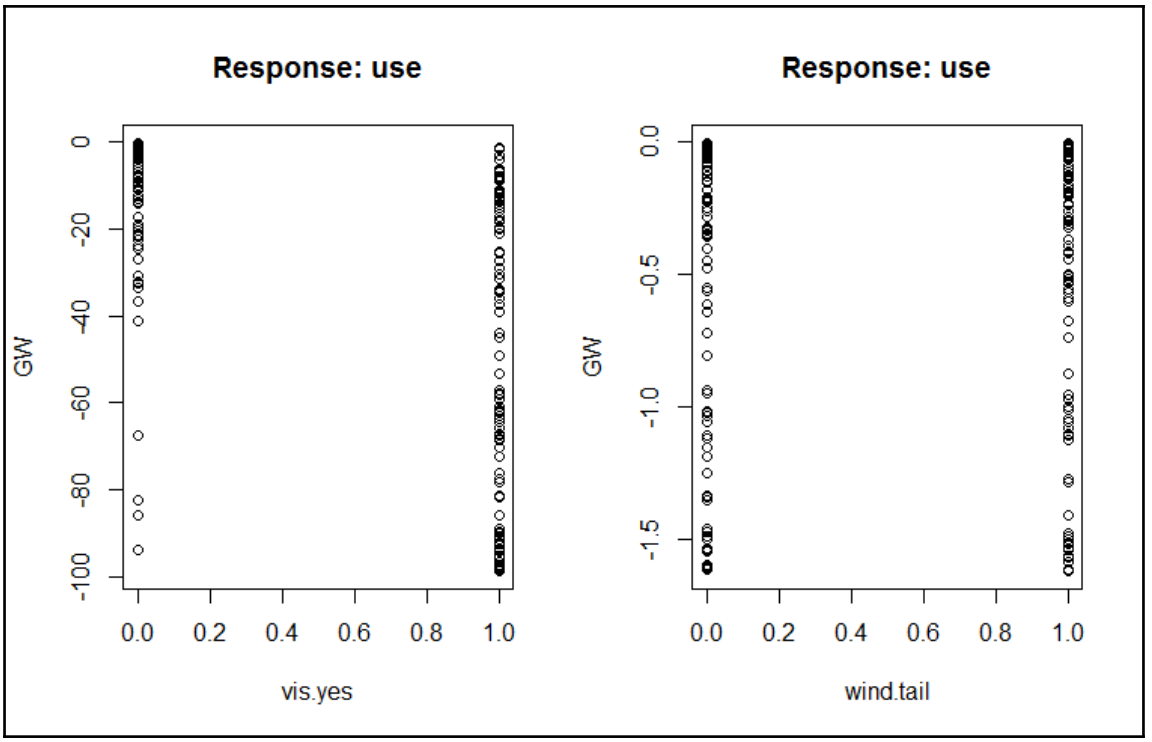

1.0
0.5
0.0
-0.5
-1.0
Output
-5.0
-2.5
0.0
2.5
5.0
Input
colour
sigmoid
tanh


X1
X2
X3
X4
Inputs
Feature A
Feature B
Latent Features
W1
W2
W3
W4
W5
Hidden Layer
Multiple Hidden Layers and Features

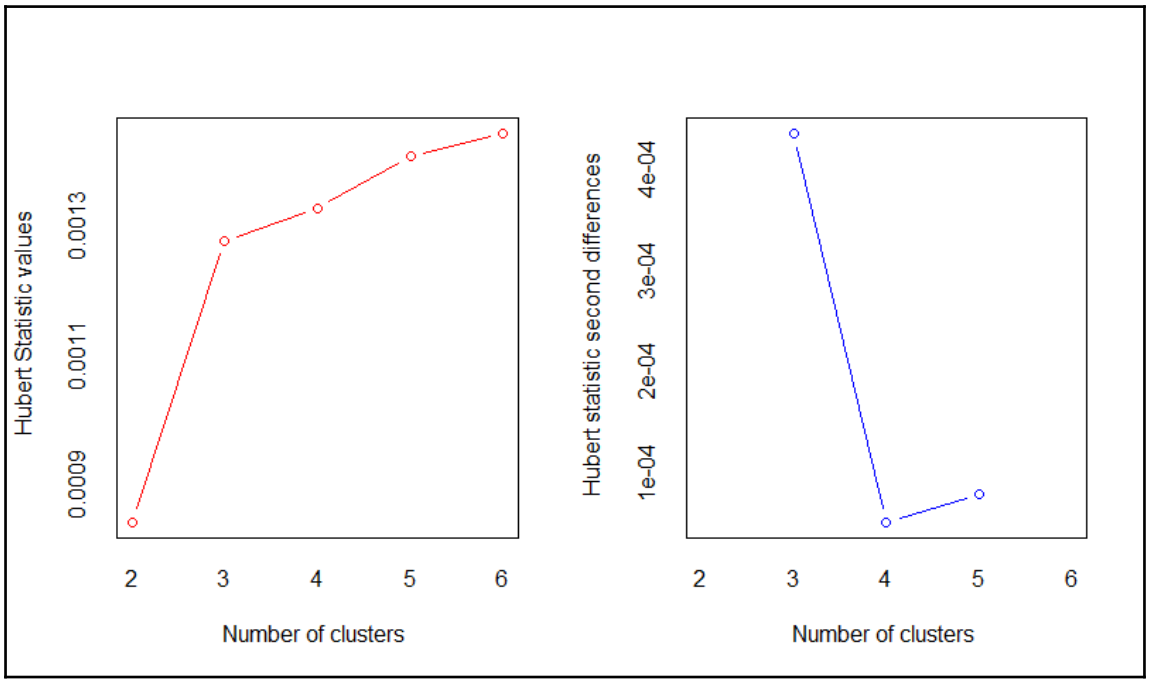

wind
head
tail
use
auto
noauto
Light
Medium
Out
Strong
Light
Medium
Out
Strong
magn
Pearson residuals:
0.57
0.00
-0.50
p-value = 0.99844

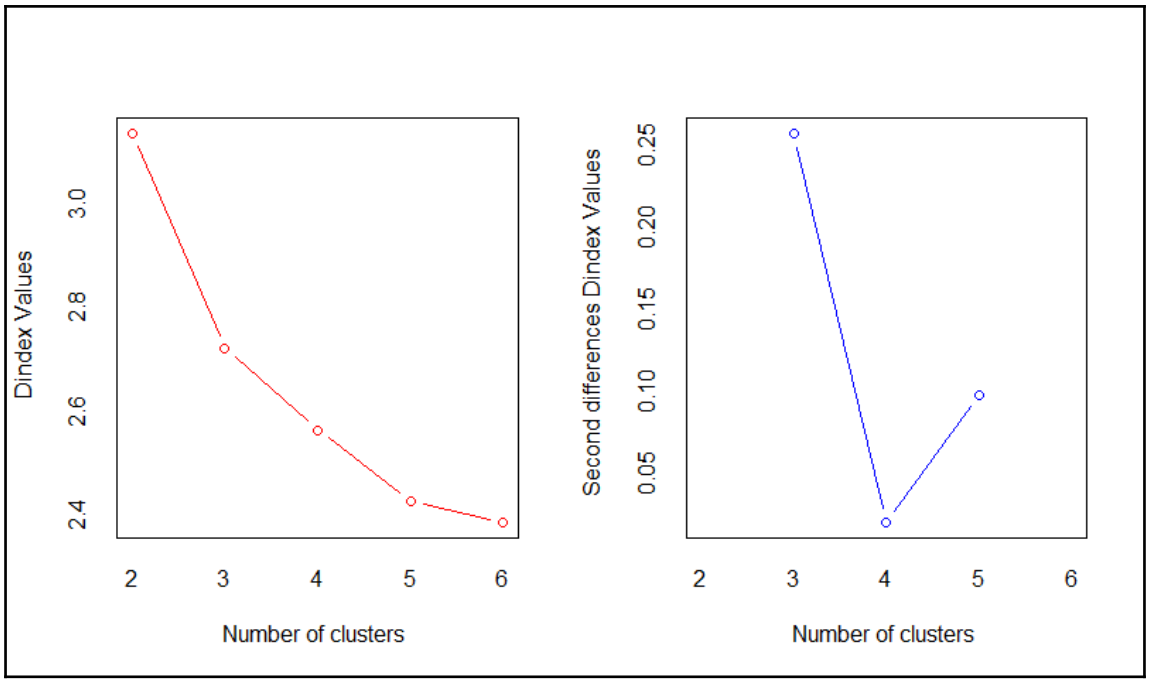

vis
no
yes
error
LX
MM
SS
XL
use
noauto auto
noauto auto
noauto auto
noauto auto

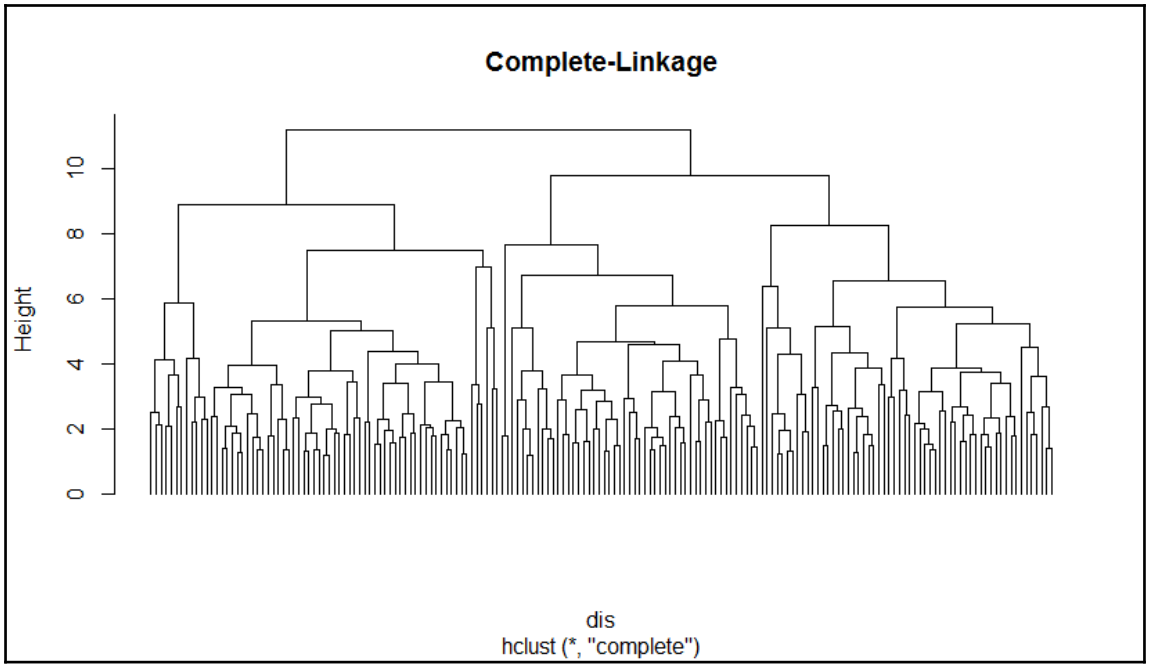

stability.xstab
error.MM
error.SS
error.XL
sign.pp
wind.tail
magn.Medium
magn.Out
magn.Strong
vis.yes
1
1
1.9576
-1.59663
-2.51937
-0.37173
-0.86396
0.10208
-0.01817
1.88693
0.14013
6.20907
-4.39265
30.72165
-65.08417
use

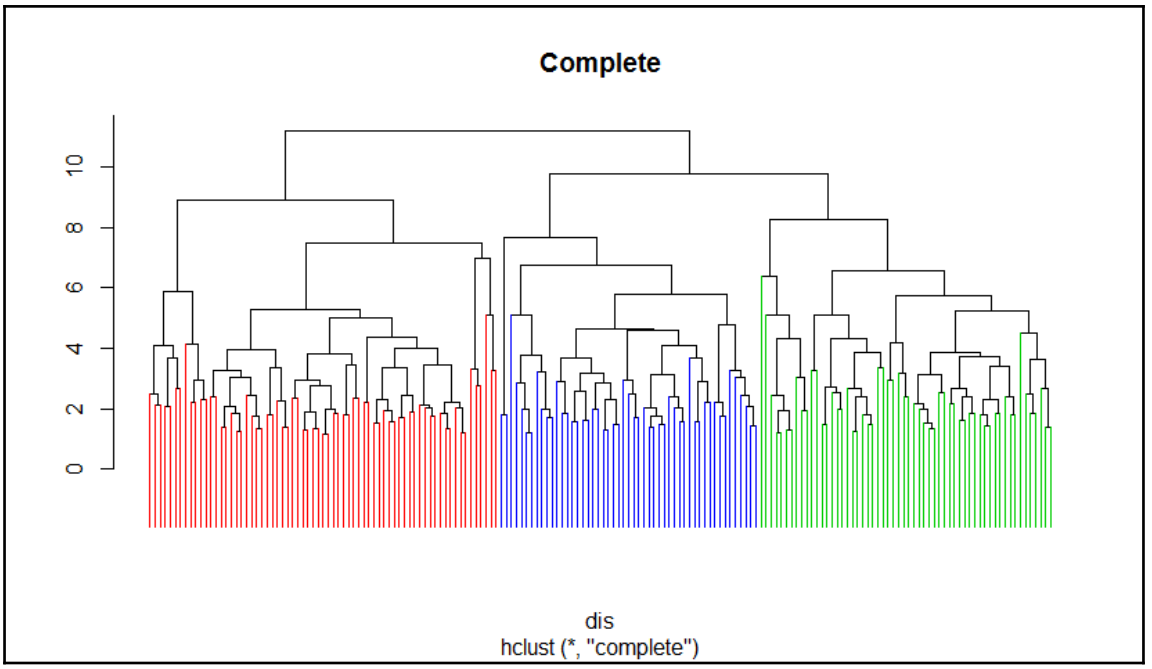

Response: use
GW
0
-20
-40
-60
-80
-100
0.0
0.2
0.4
0.6
0.8
1.0
vis.yes
Response: use
GW
0.0
-0.5
-1.0
-1.5
0.0
0.2
0.4
0.6
0.8
1.0
wind.tail

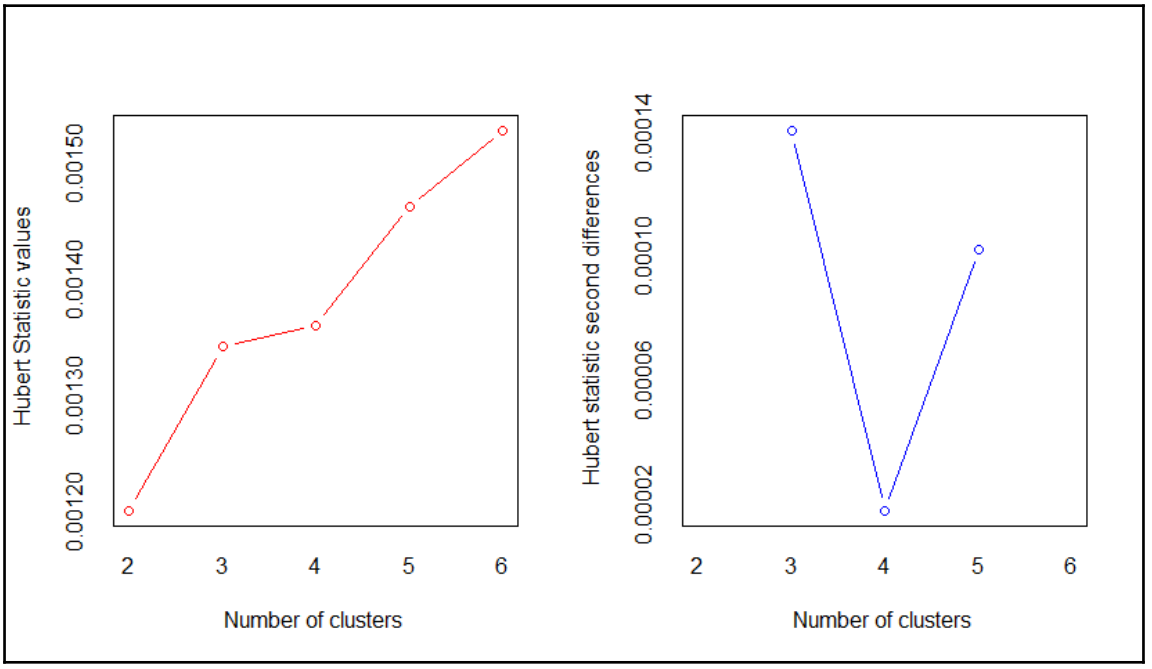
Chapter 8: Cluster Analysis

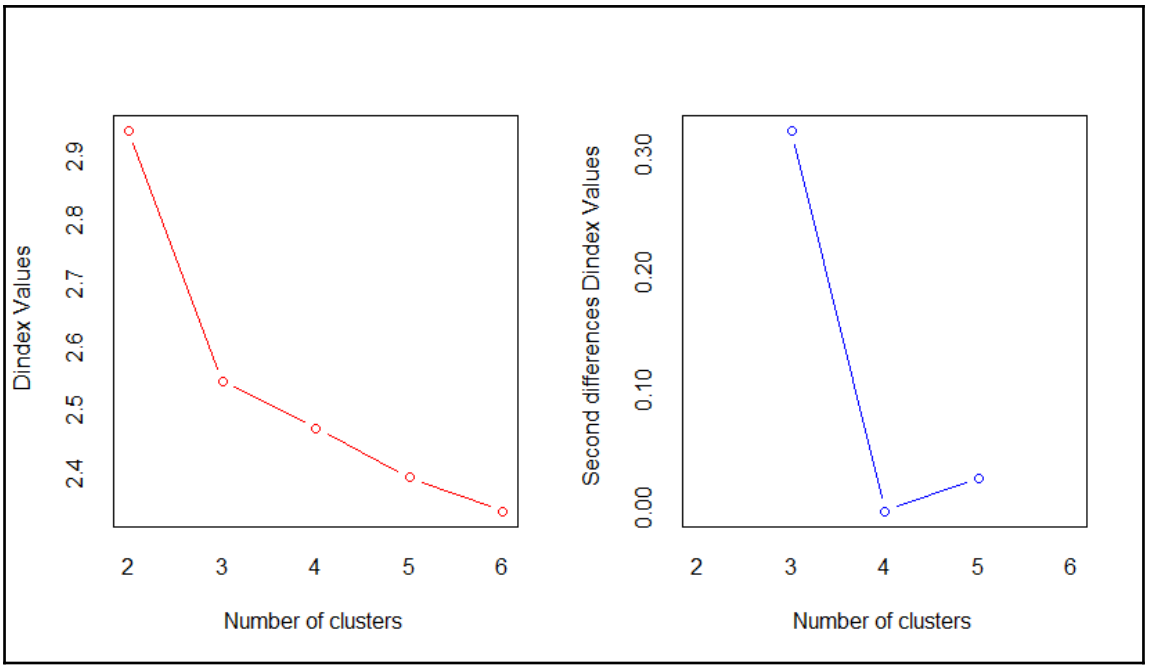


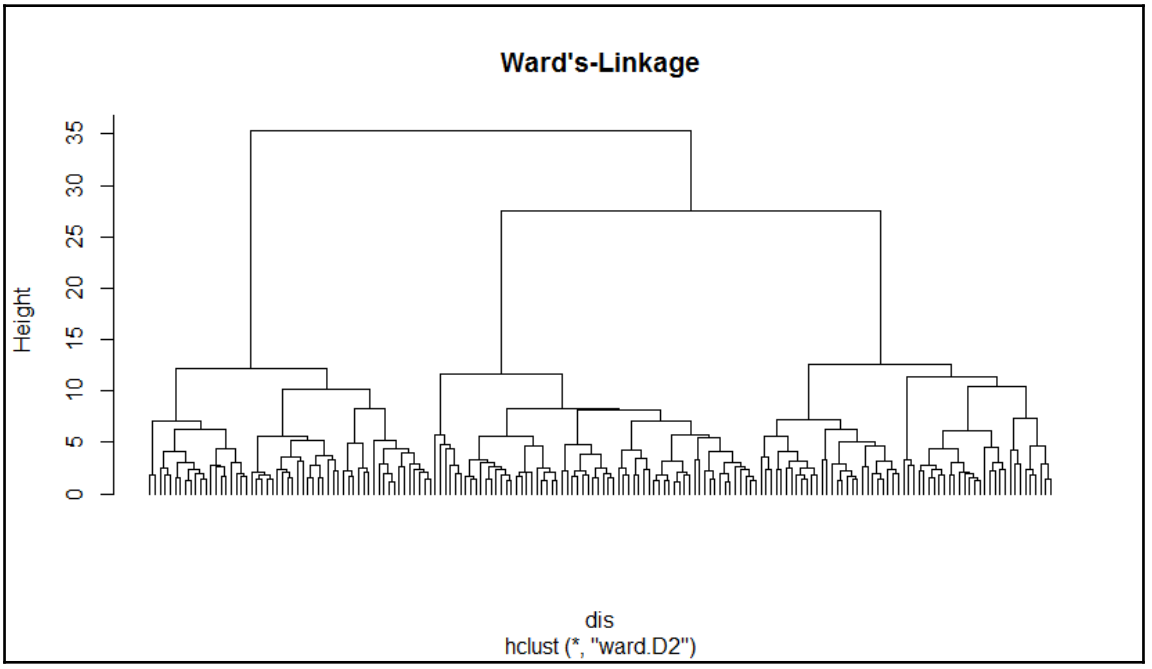

Dindex Values
2.4
2.6
2.8
3.0
2
3
4
5
6
Number of clusters
Second differences Dindex Values
0.05
0.10
0.15
0.20
0.25
2
3
4
5
6
Number of clusters

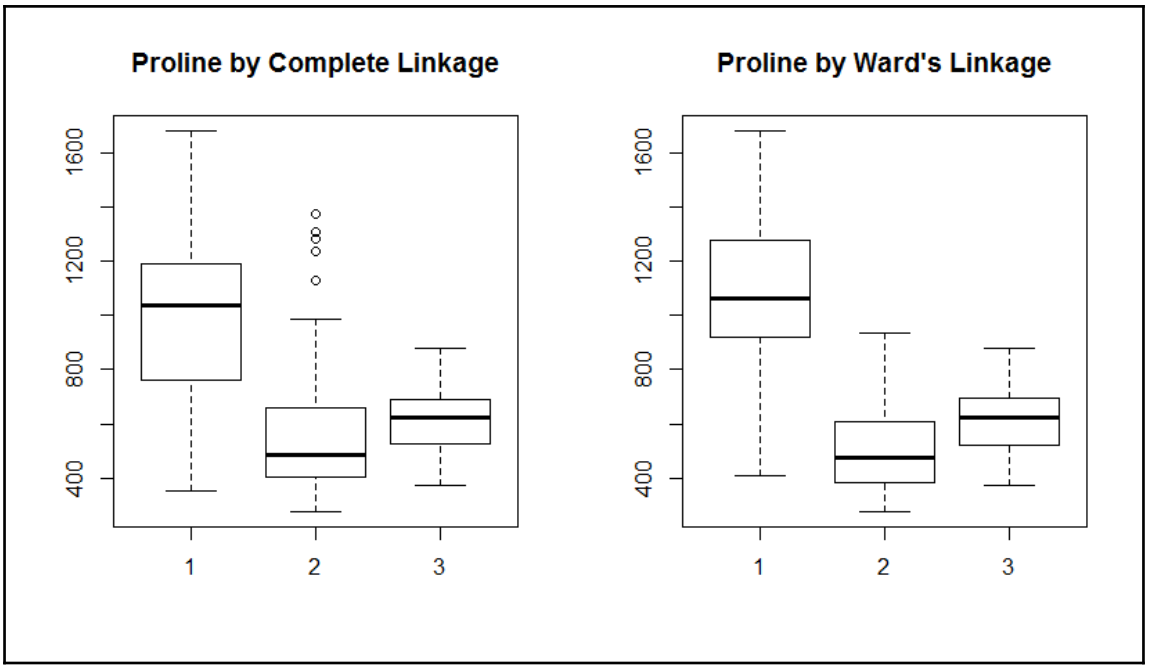

Complete-Linkage
Height
0
2
4
6
8
10
dis
hclust (*, "complete")

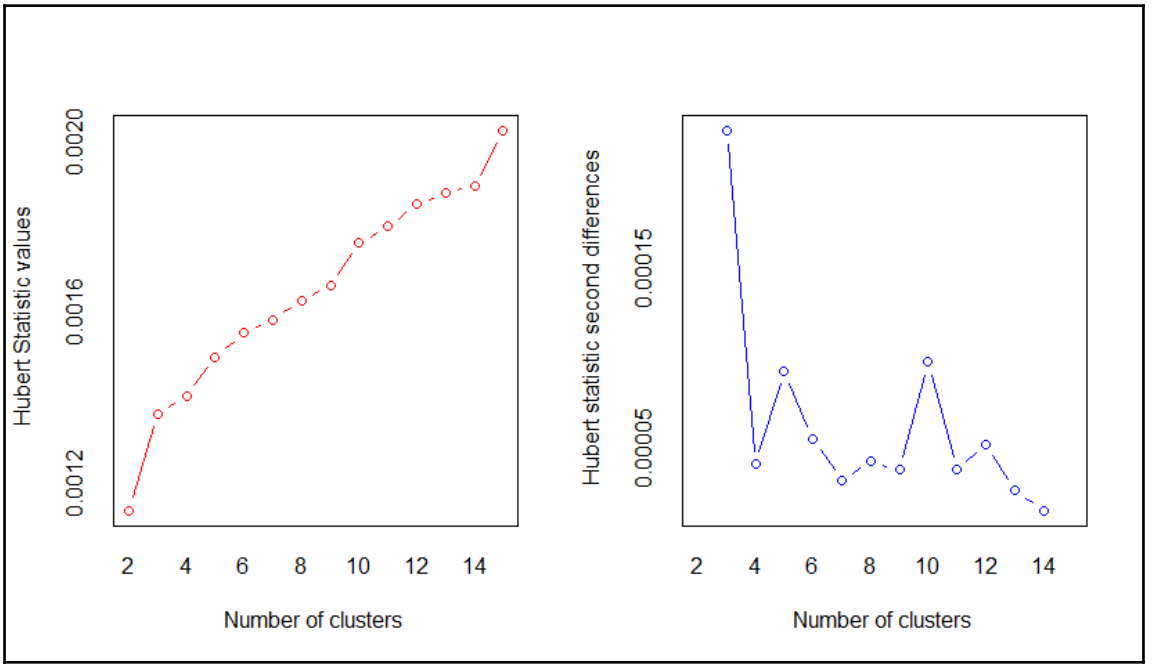

Complete
0
2
4
6
8
10
dis
hclust (*, "complete")

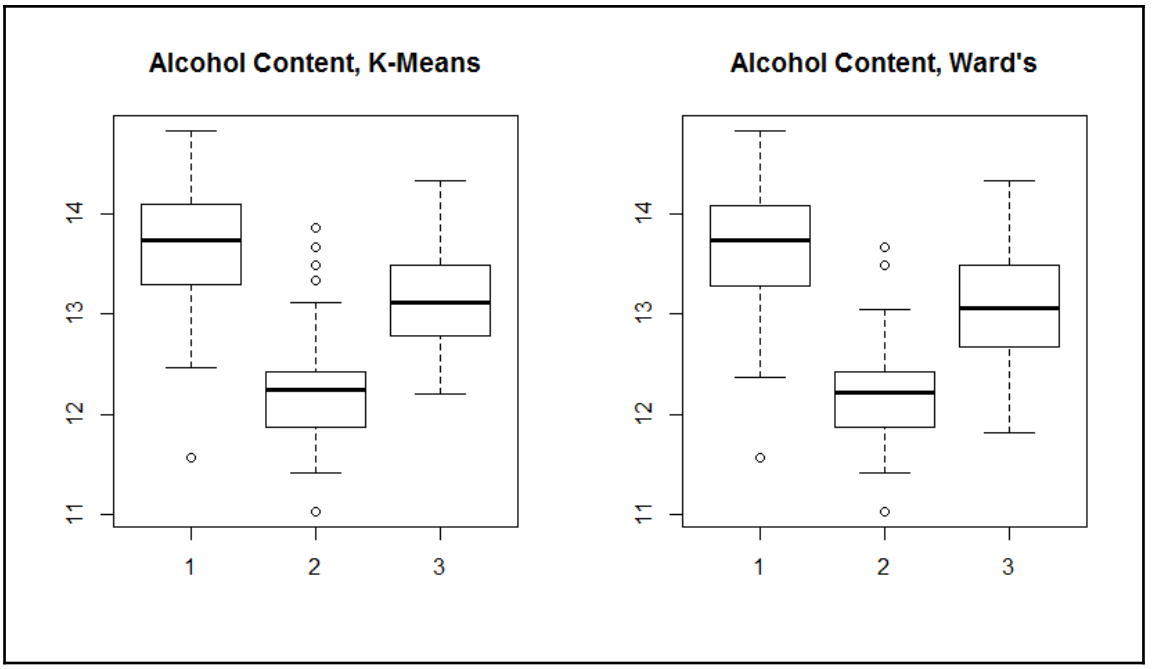

Hubert Statistic values
0.00120
0.00130
0.00140
0.00150
2
3
4
5
6
Number of clusters
Hubert statistic second differences
0.00002
0.00006
0.00010
0.00014
2
3
4
5
6
Number of clusters


Dindex Values
2.4
2.5
2.6
2.7
2.8
2.9
2
3
4
5
6
Number of clusters
Second differences Dindex Values
0.00
0.10
0.20
0.30
2
3
4
5
6
Number of clusters

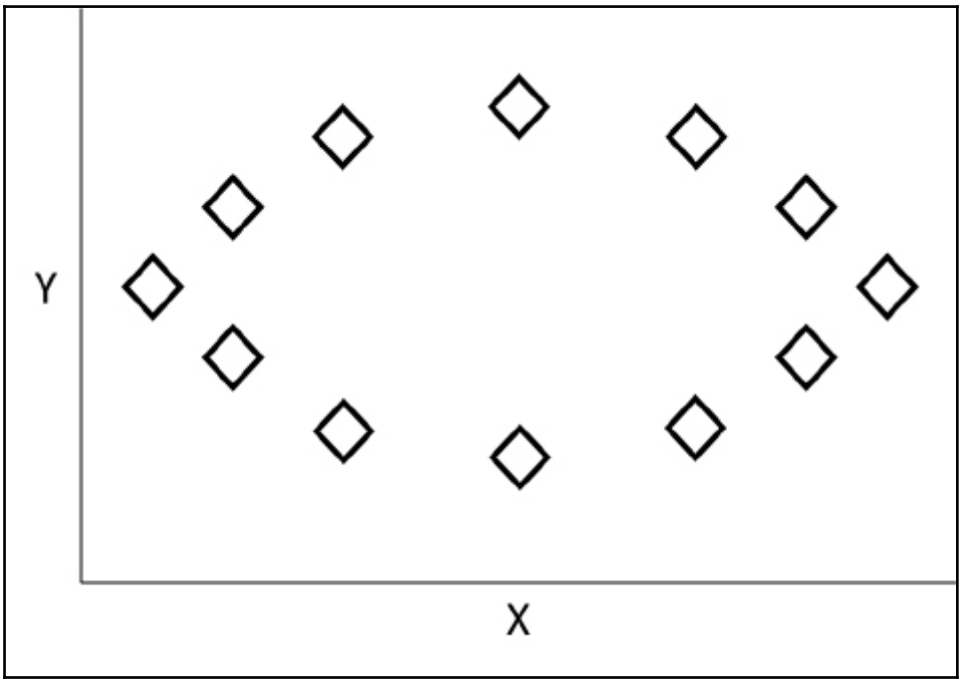

Ward's-Linkage
Height
0
5
10
15
20
25
30
35
dis
hclust (*, "ward.D2")

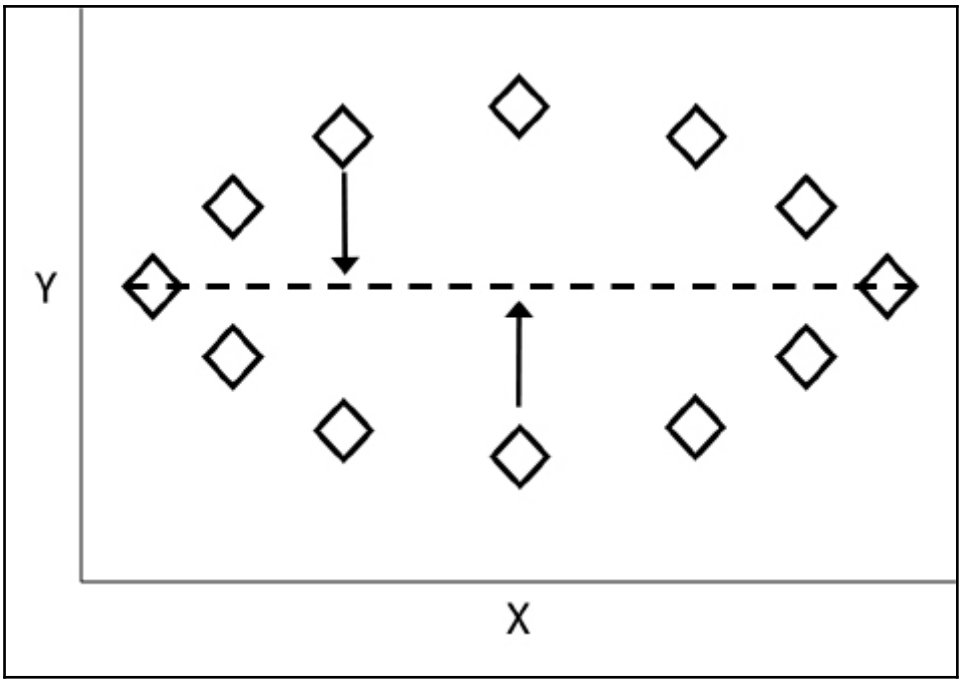

Proline by Complete Linkage
Proline by Ward's Linkage
400
800
1200
1600
1
2
3

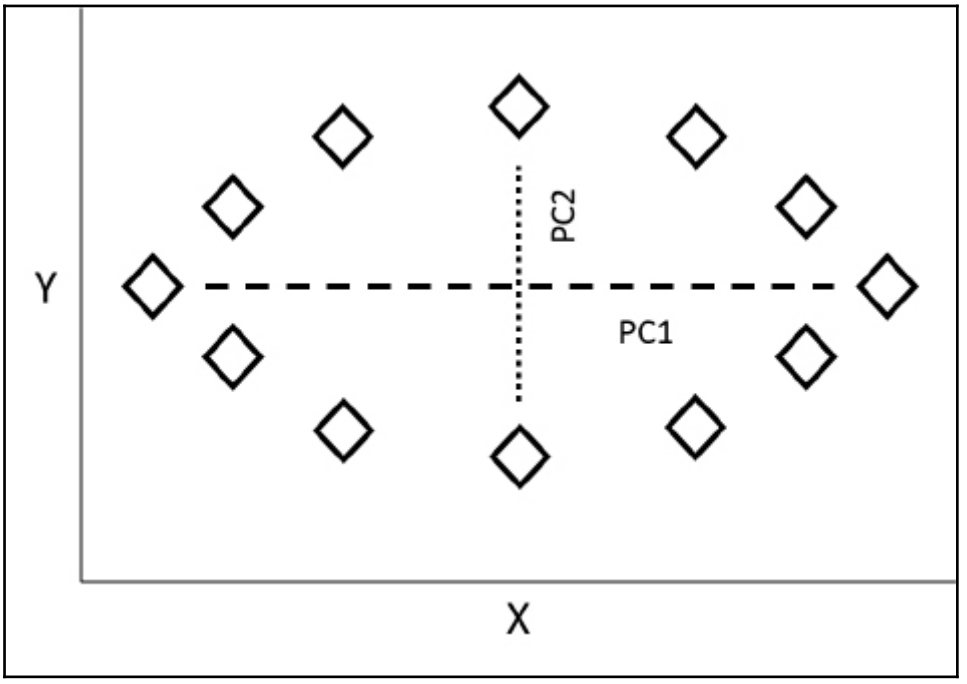

Hubert Statistic values
0.0012
0.0016
0.0020
Hubert statistic second differences
0.00005
0.00015
2
4
6
8
10
12
14
Number of clusters
Number of clusters

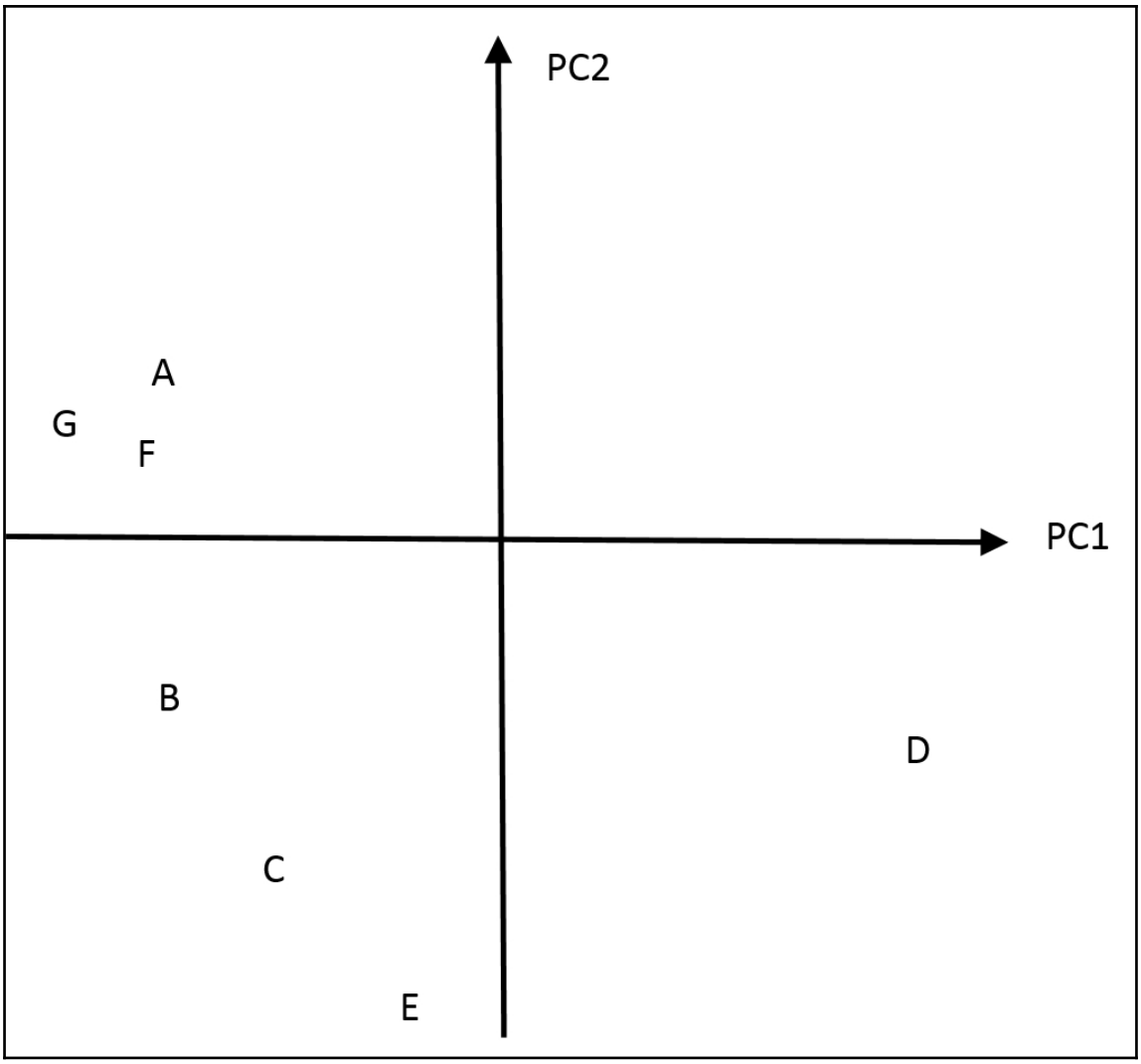

Alcohol Content, K-Means
Alcohol Content, Ward's
14
13
12
11
1
2
3

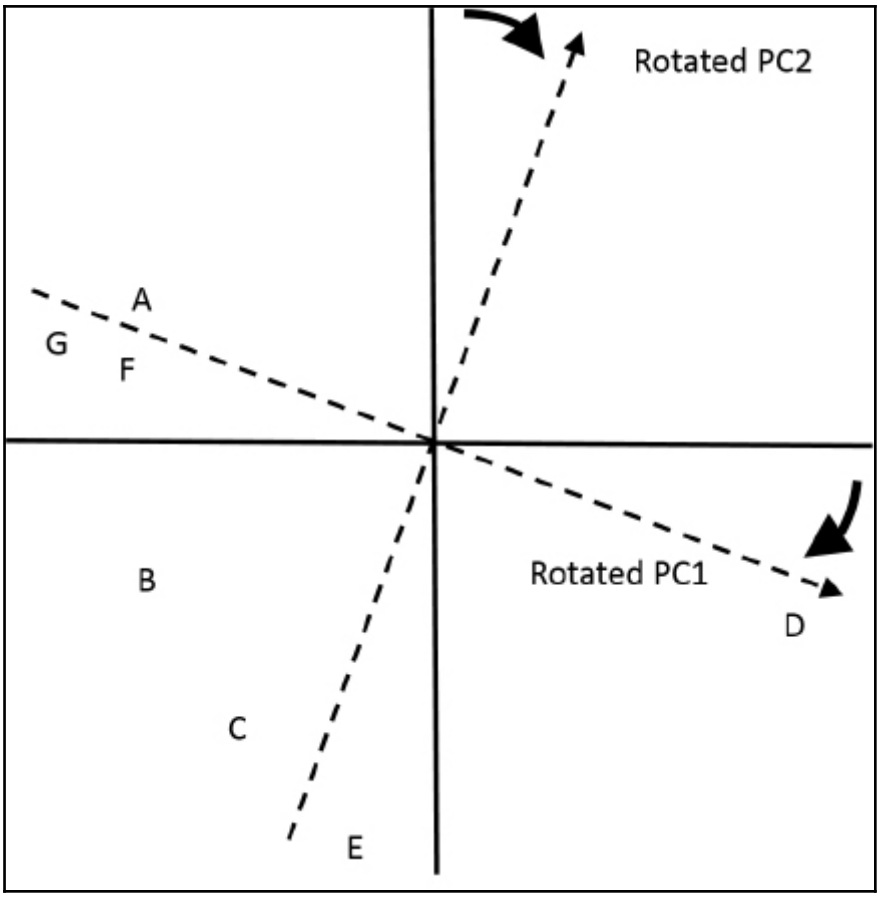
Chapter 9: Principal Components Analysis

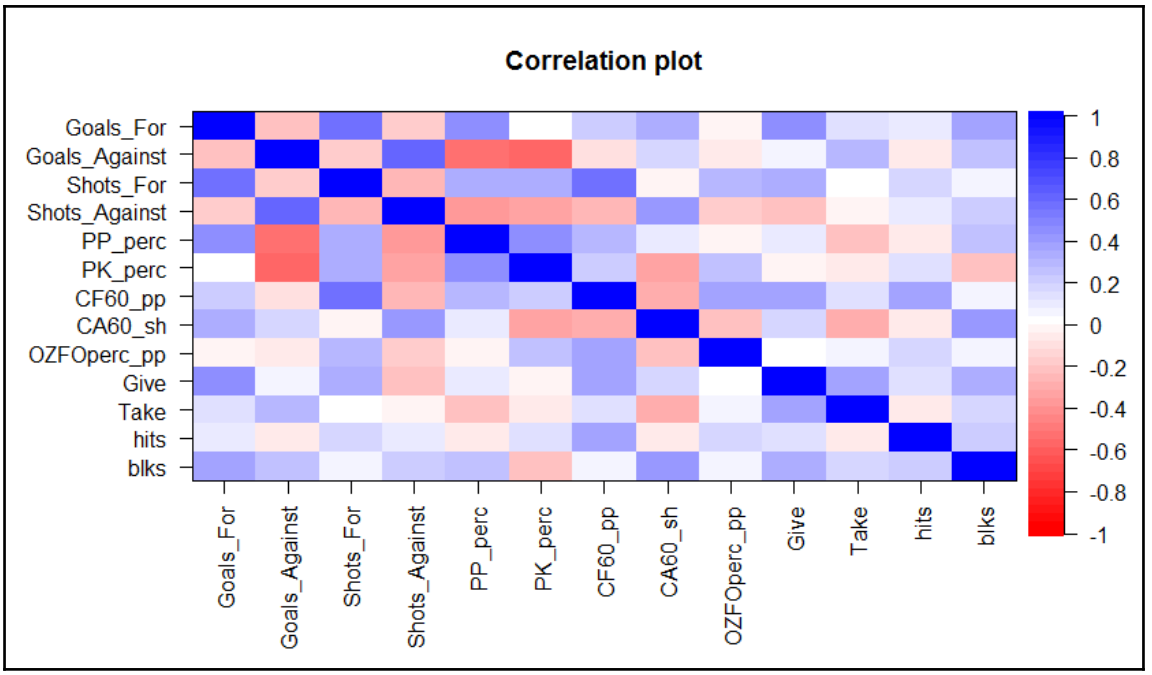


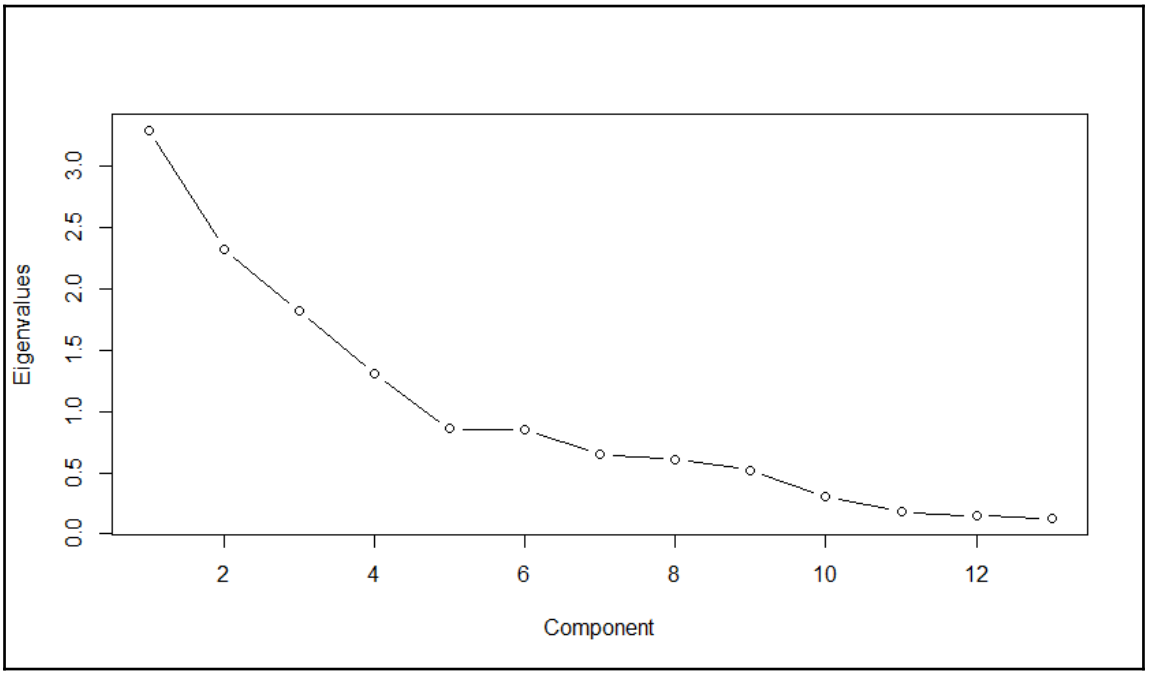

Y
X

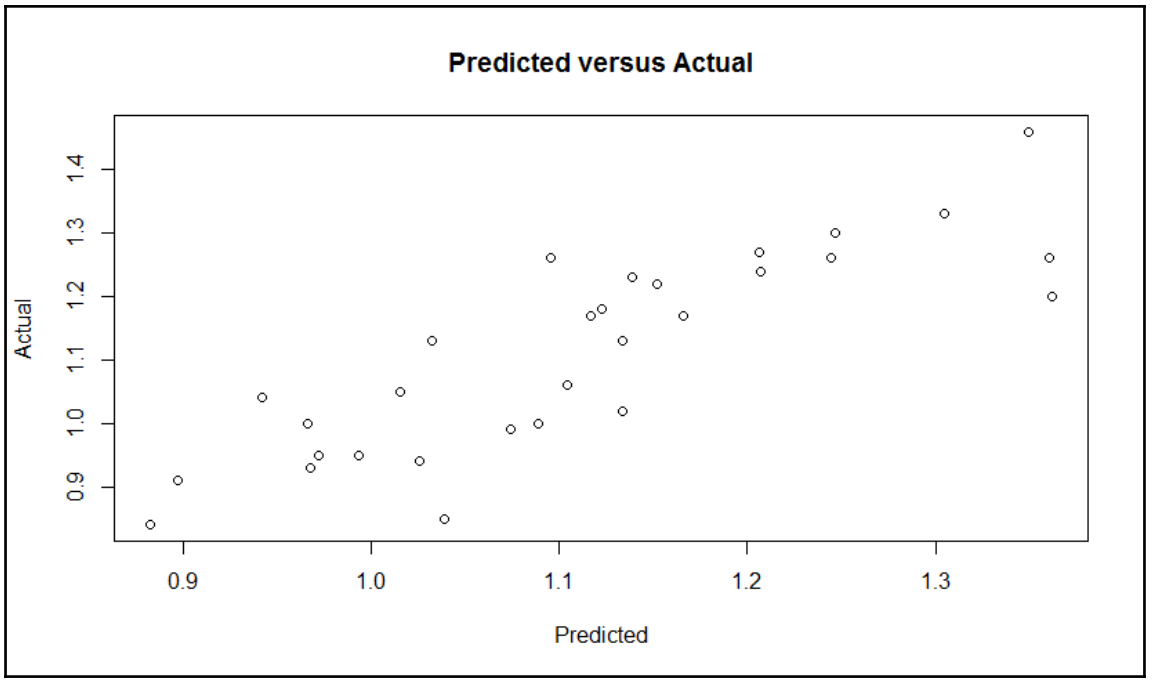

Y
PC2
PC1
X

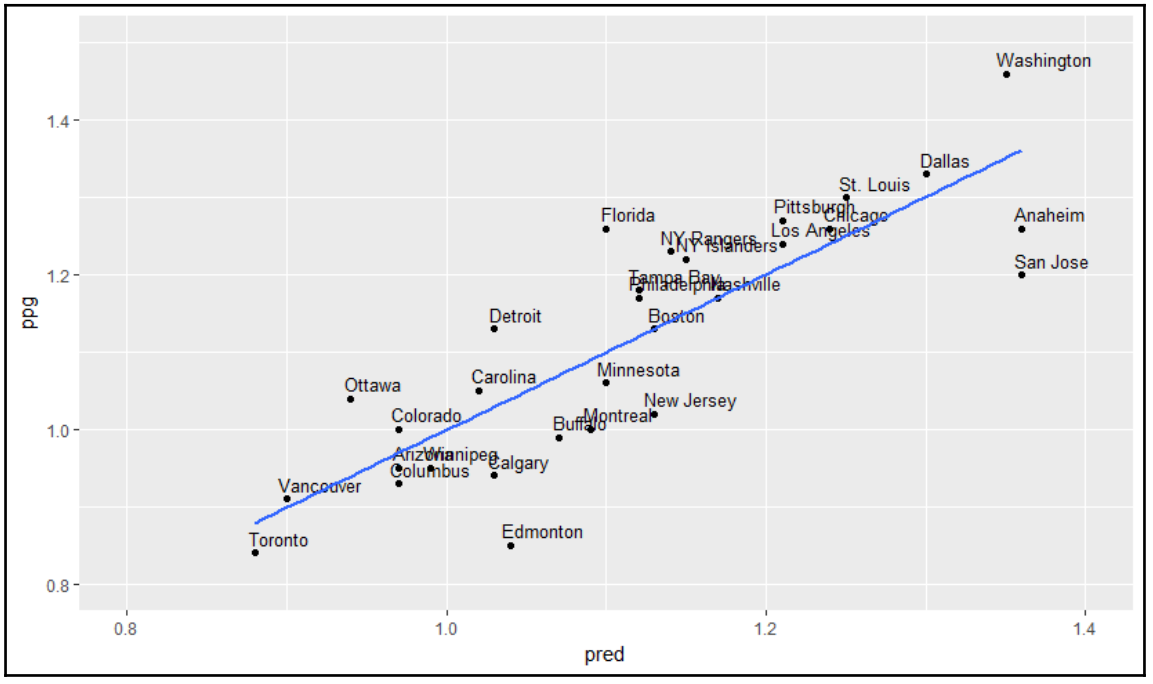

PC2
PC1
A
G
F
B
D
C
E

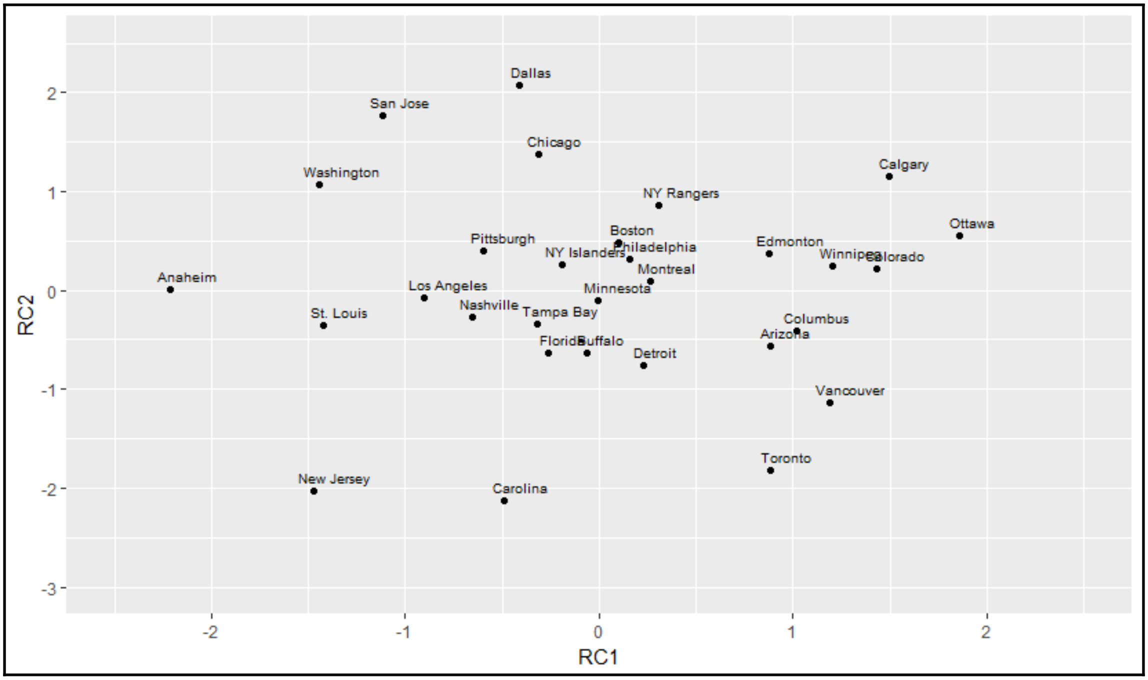

Rotated PC2
A
G
F
B
Rotated PC1
D
C
E

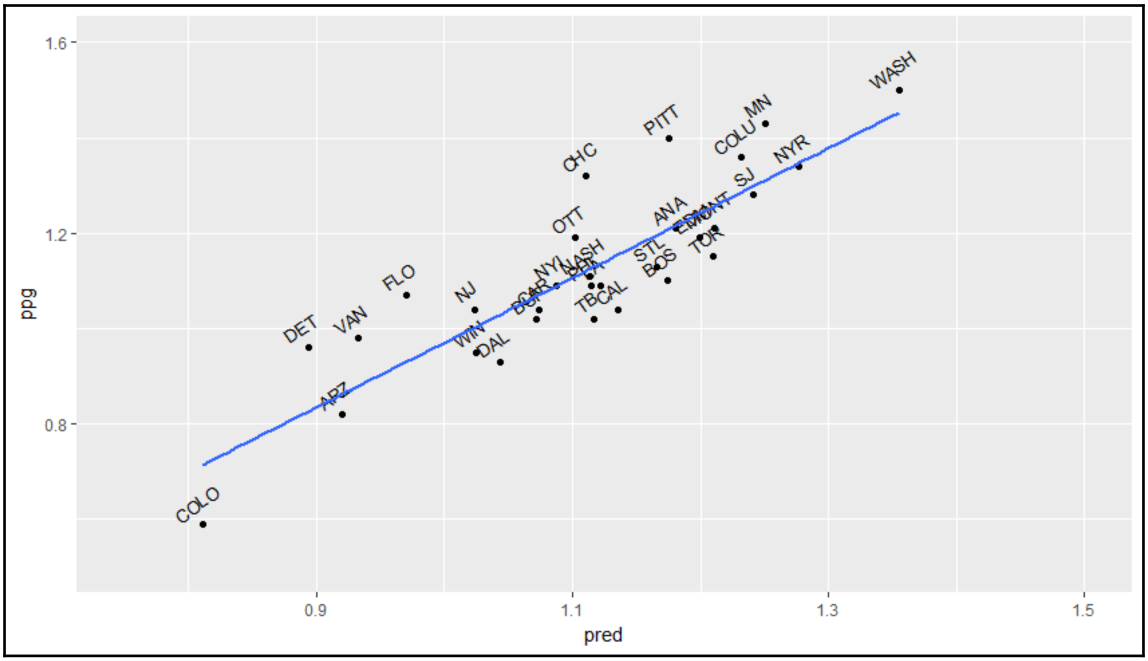

Correlation plot
Goals_For
Goals_Against
Shots_For
Shots_Against
PP_perc
PK_perc
CF60_pp
CA60_sh
OZFOperc_pp
Give
Take
hits
blks
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


Eigenvalues
0.0
0.5
1.0
1.5
2.0
2.5
3.0
2
4
6
8
10
12
Component

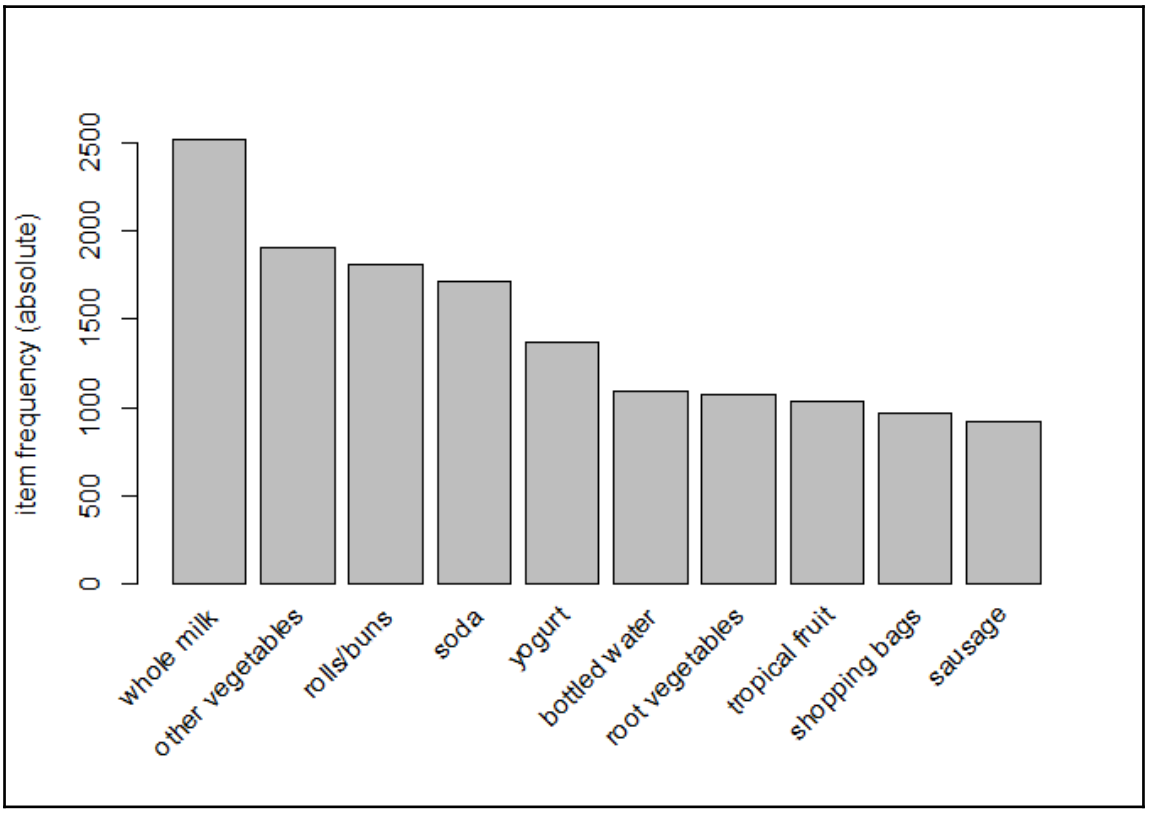

Predicted versus Actual
Actual
0.9
1.0
1.1
1.2
1.3
1.4
Predicted
0.9
1.0
1.1
1.2
1.3

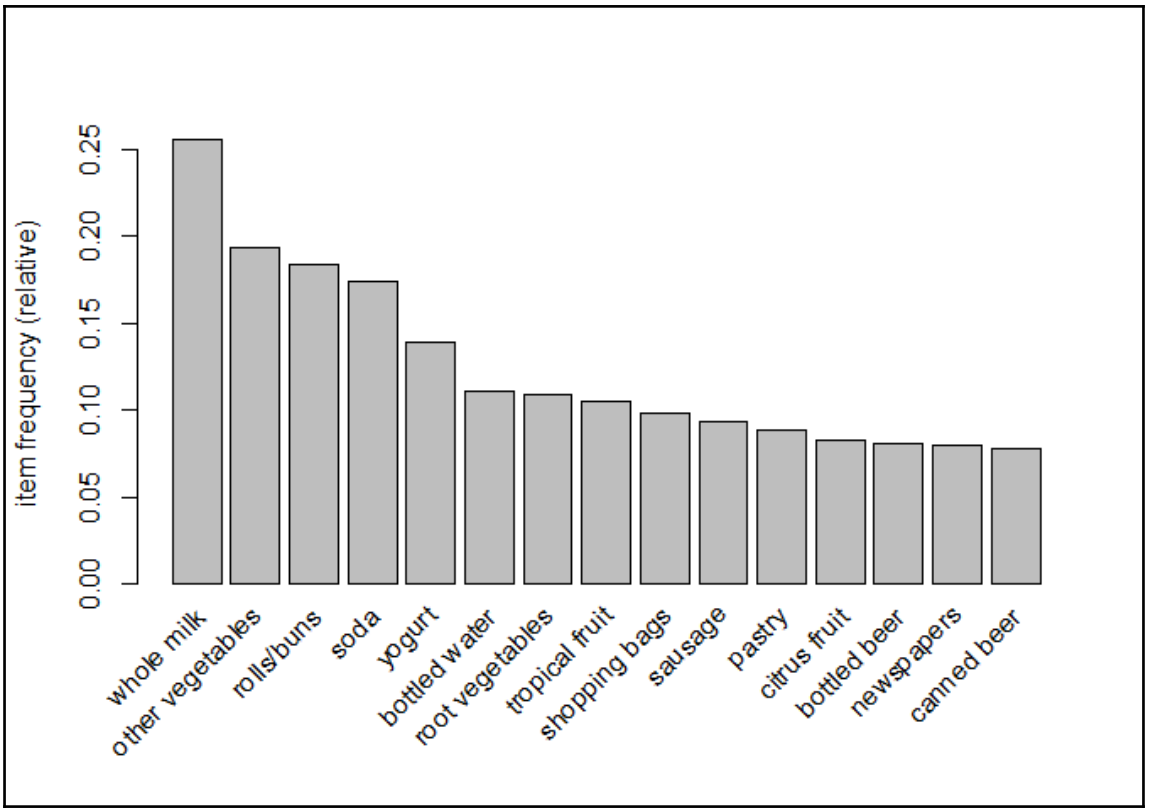

ppg
pred
0.8
1.0
1.2
1.4
Washington
Dallas
St. Louis
Pittsburgh
Chicago
Anaheim
Florida
NY Rangers
Los Angeles
NY Islanders
San Jose
Tampa Bay
Philadelphia
Nashville
Detroit
Boston
Carolina
Minnesota
Ottawa
New Jersey
Colorado
Montreal
Buffalo
Arizona
Winnipeg
Columbus
Calgary
Vancouver
Edmonton
Toronto

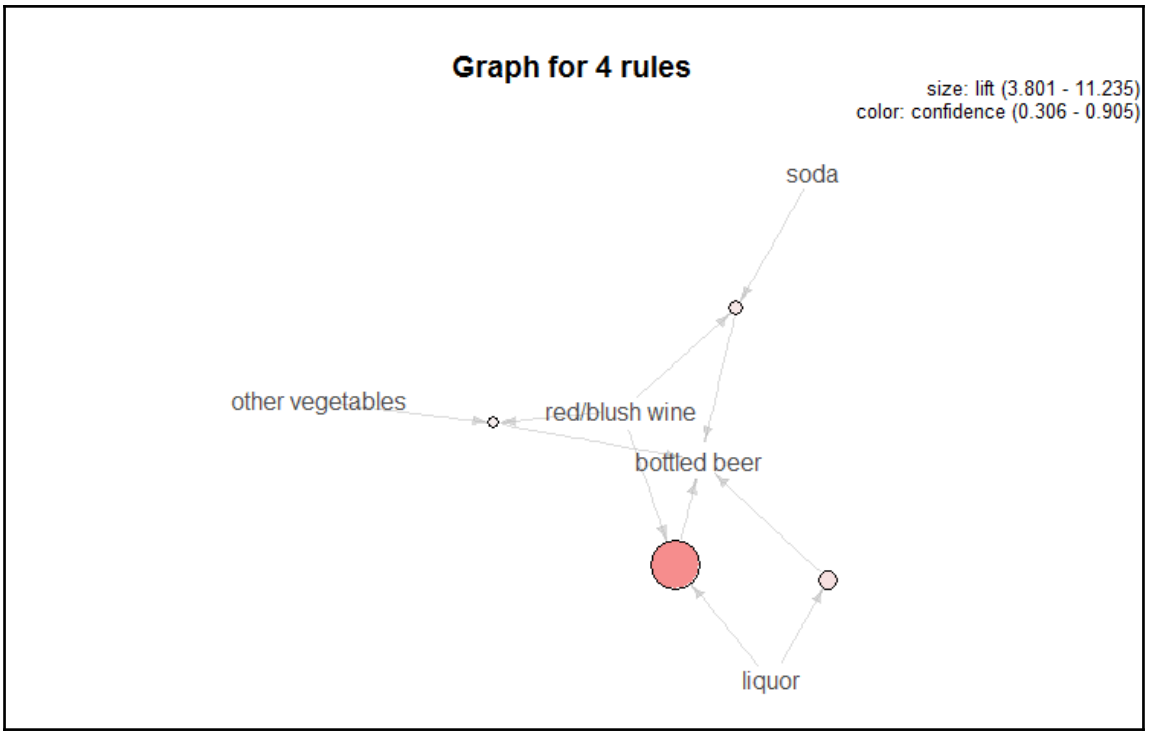

RC2
RC1
2
1
0
-1
-2
-3
Dallas
San Jose
Chicago
Washington
Calgary
NY Rangers
Ottawa
Boston
Edmonton
Pittsburgh
Philadelphia
NY Islanders
Winnipeg
Colorado
Montreal
Anaheim
Los Angeles
Minnesota
Nashville
St. Louis
Tampa Bay
Columbus
Arizona
Florida
Buffalo
Detroit
Vancouver
Toronto
New Jersey
Carolina


ppg
pred
1.6
1.2
0.8
0.9
1.1
1.3
1.5
WASH
MN
PITT
COLU
NYR
CHC
SJ
ANA
OTT
STL
BOS
TOR
NASH
FLO
NJ
DET
VAN
TB
CAL
WIN
DAL
ARZ
COLO

Chapter 10: Market Basket Analysis, Recommendation Engines, and Sequential Analysis

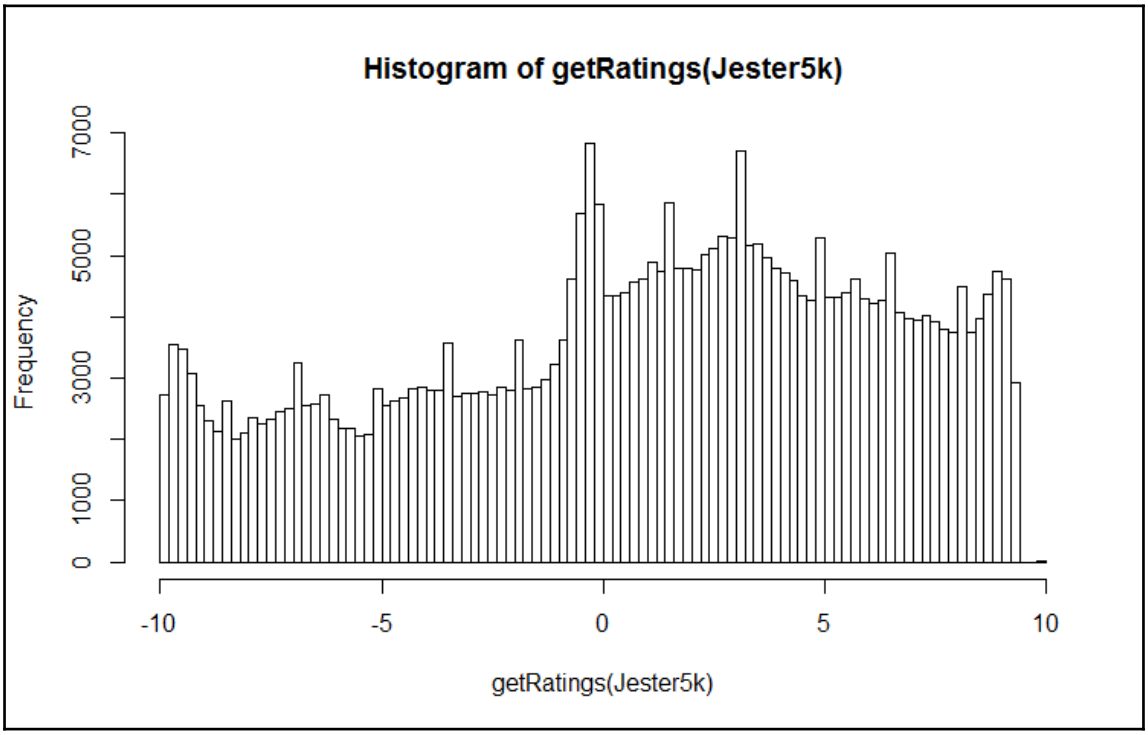


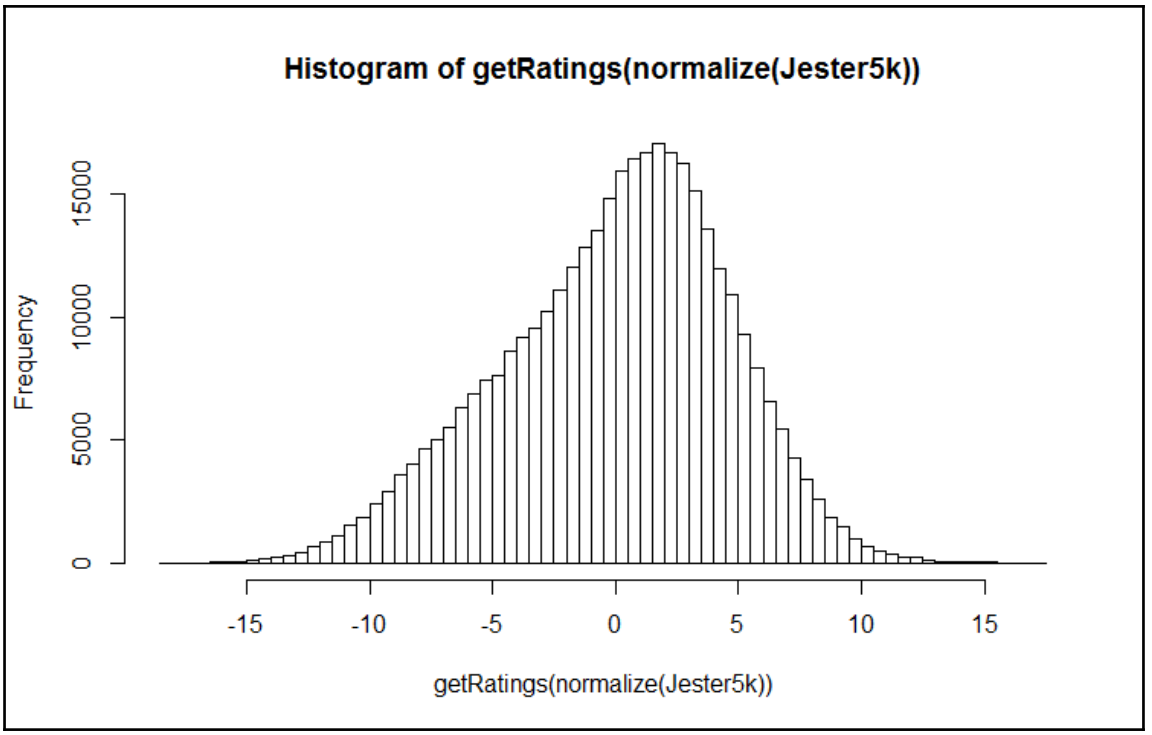

item frequency (relative)
0.00
0.05
0.10
0.15
0.20
0.25
whole milk
other vegetables
rolls/buns
soda
yogurt
bottled water
root vegetables
tropical fruit
shopping bags
sausage
pastry
citrus fruit
bottled beer
newspapers
canned beer

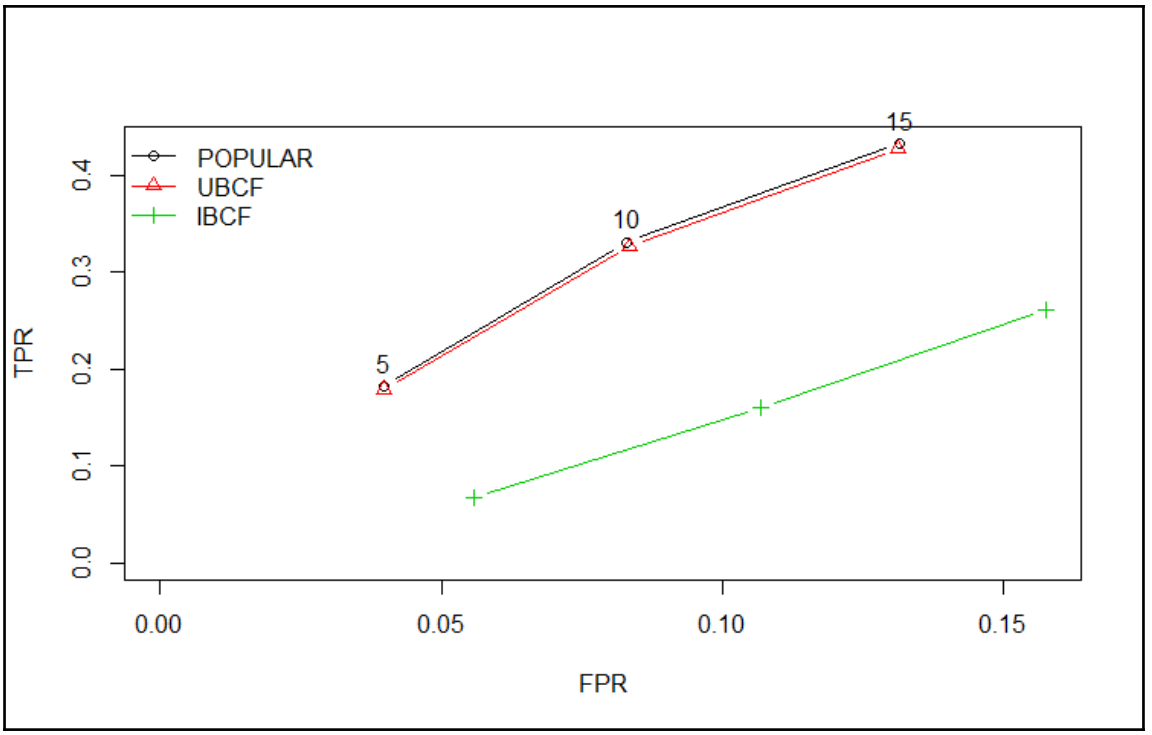


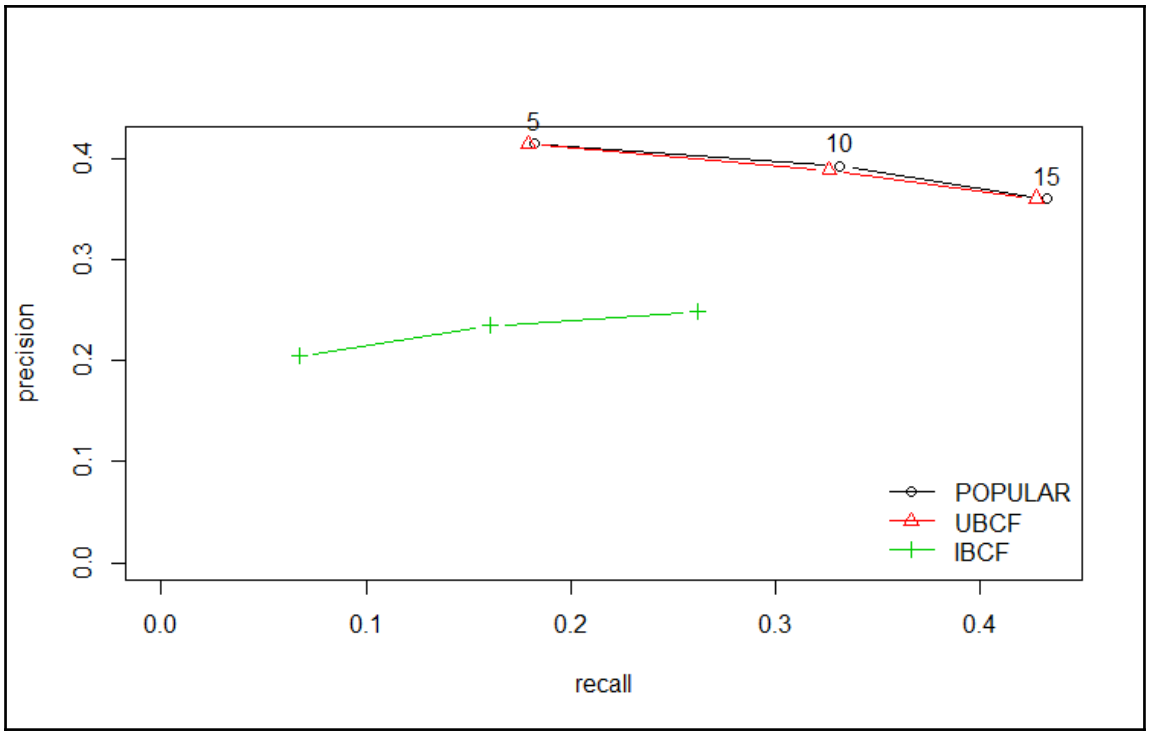
	Avengers	American Sniper	Les Miserable	Mad Max
Homer	3	5	3	4
Marge	5	2	5	3
Bart	5	5	1	4
Lisa	5	1	5	2
Flanders	1	1	4	1
Me	1	5	2	?

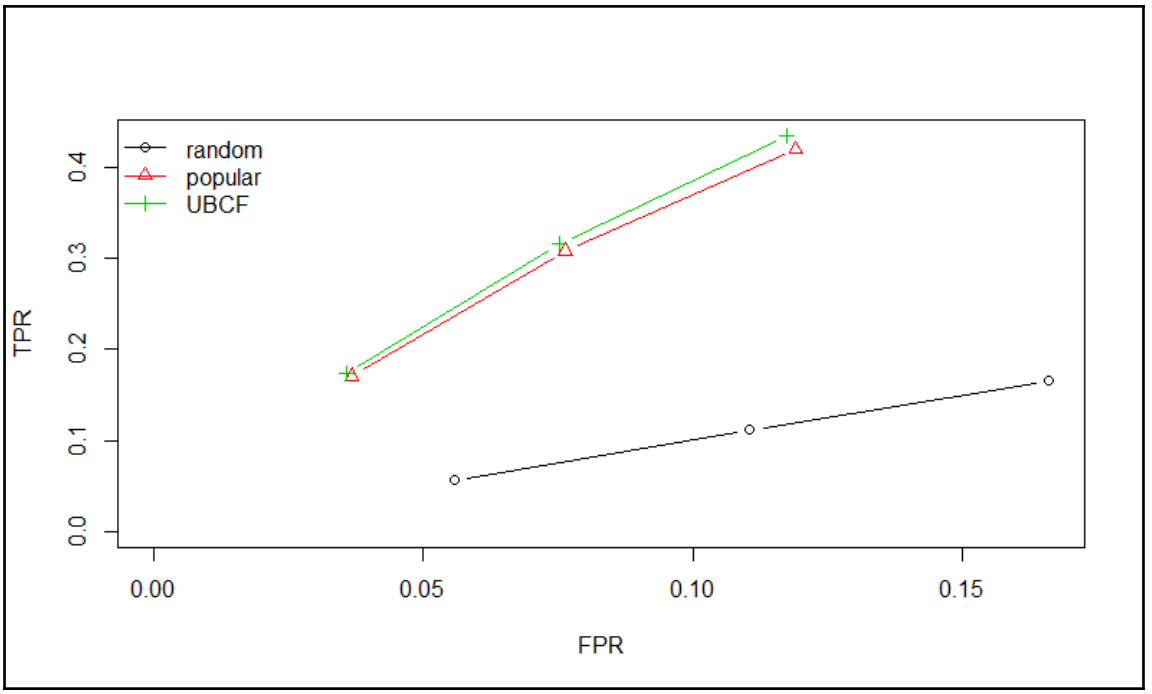
	Avengers	American Sniper	Les Miserable	Mad Max
Homer	3	5	3	4
Marge	5	2	5	3
Bart	5	5	1	4
Lisa	5	1	5	2
Flanders	1	1	4	1
Me	1	5	2	?

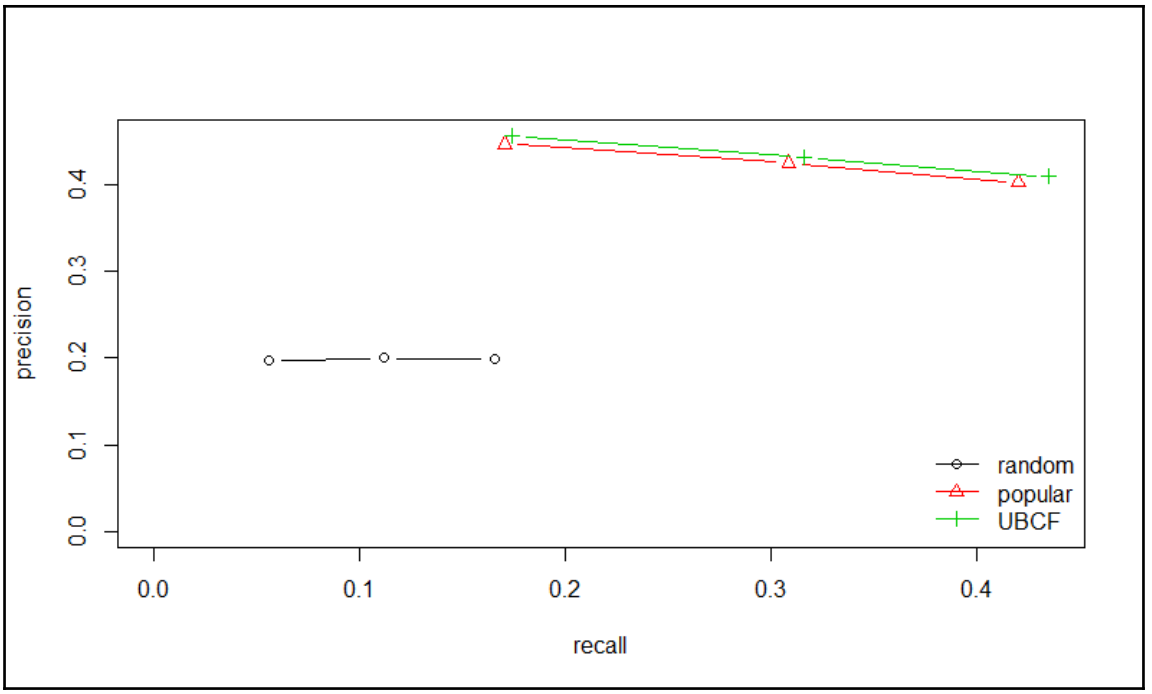


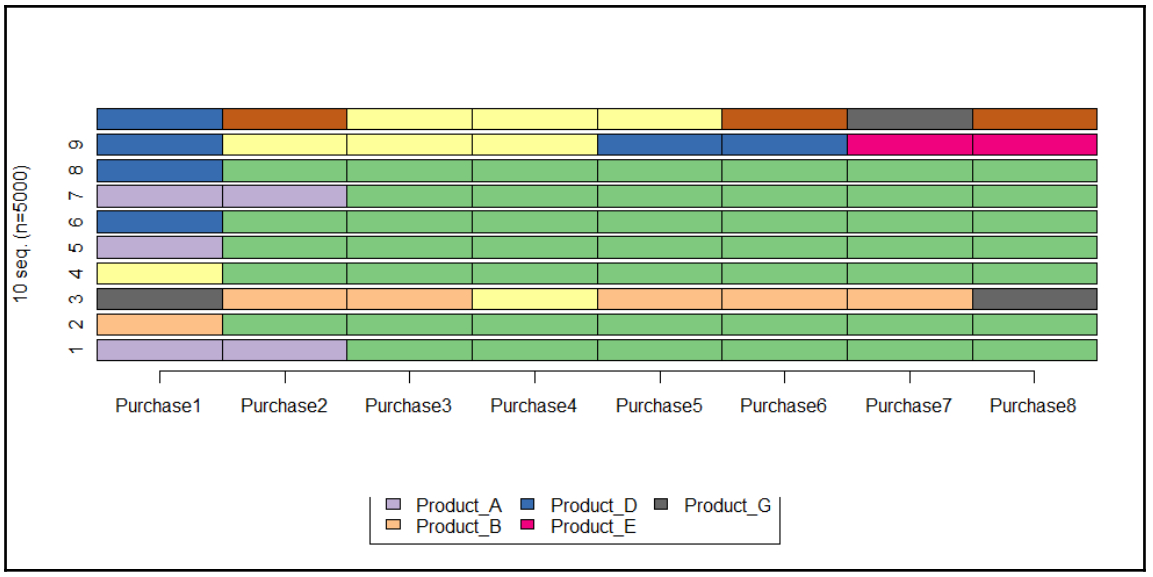

Histogram of getRatings(normalize(Jester5k))
Frequency
0
5000
10000
15000
-15
-10
-5
0
5
10
15
getRatings(normalize(Jester5k))

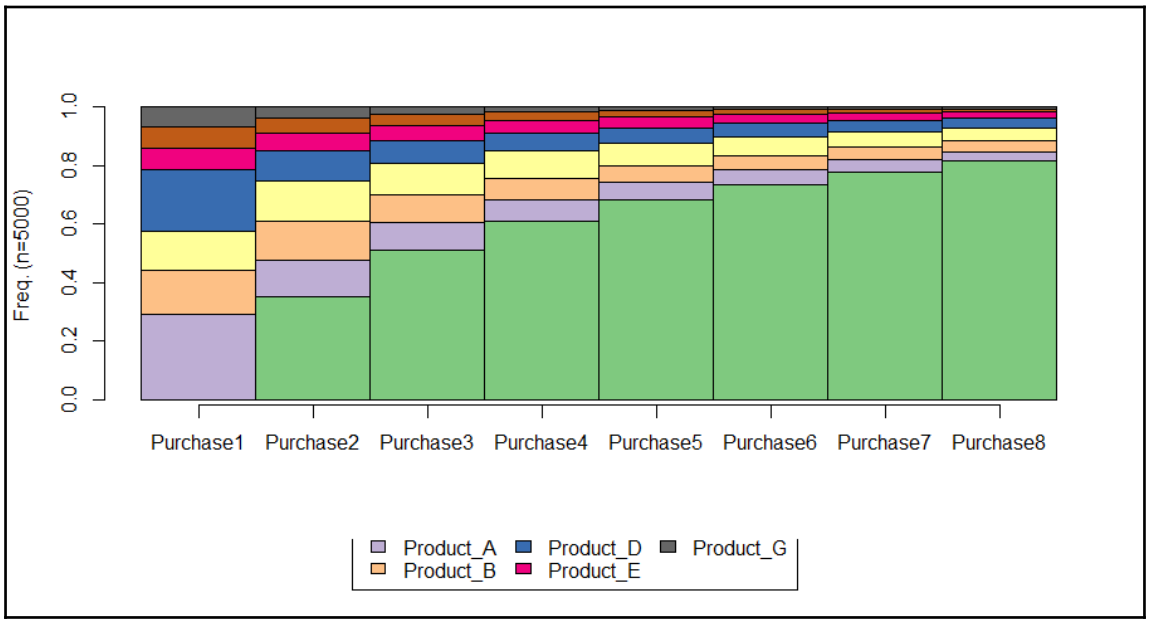

POPULAR
UBCF
IBCF
TPR
FPR
0.0
0.1
0.2
0.3
0.4
0.00
0.05
0.10
0.15
5
10
15

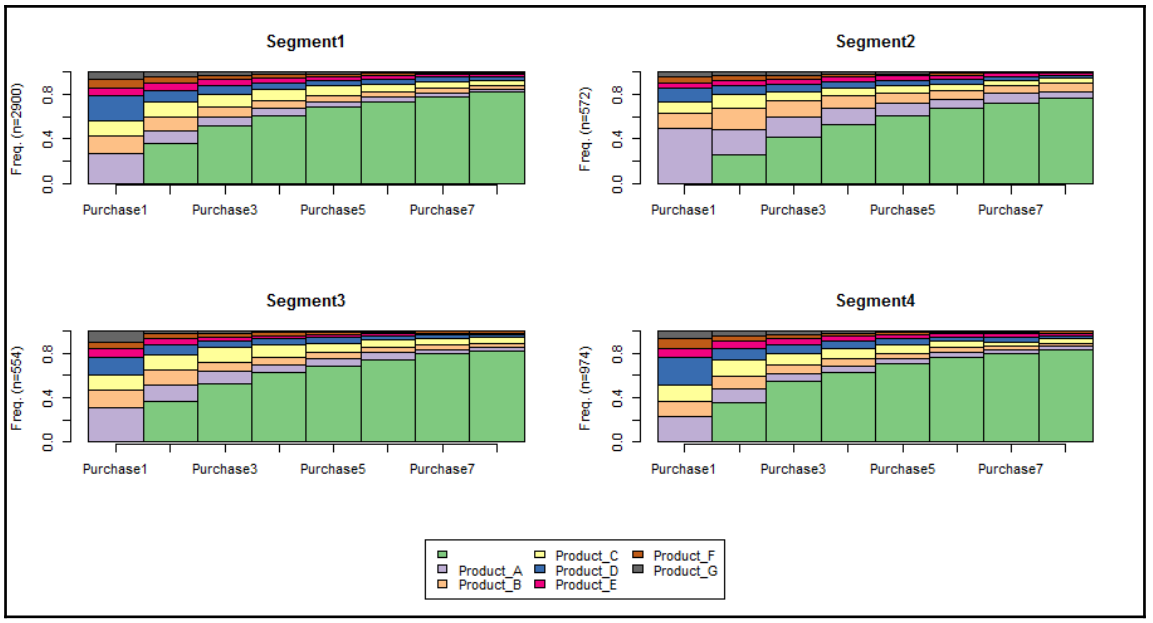

5
10
15
precision
0.0
0.1
0.2
0.3
0.4
recall
POPULAR
UBCF
IBCF

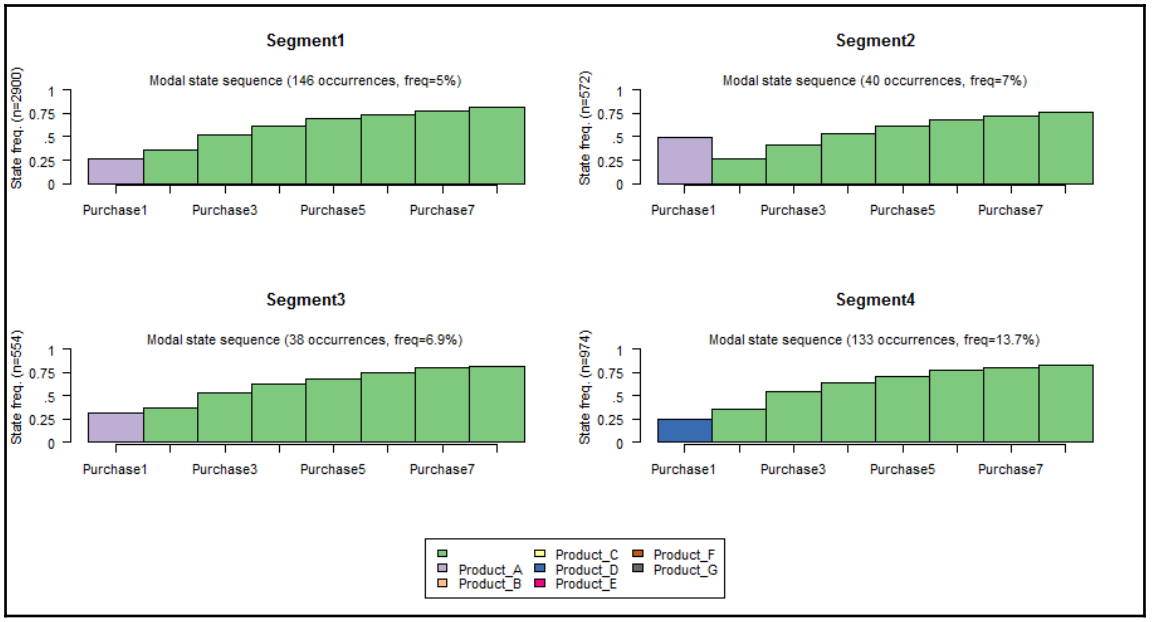

random
popular
UBCF
TPR
FPR
0.0
0.1
0.2
0.3
0.4
0.00
0.05
0.10
0.15

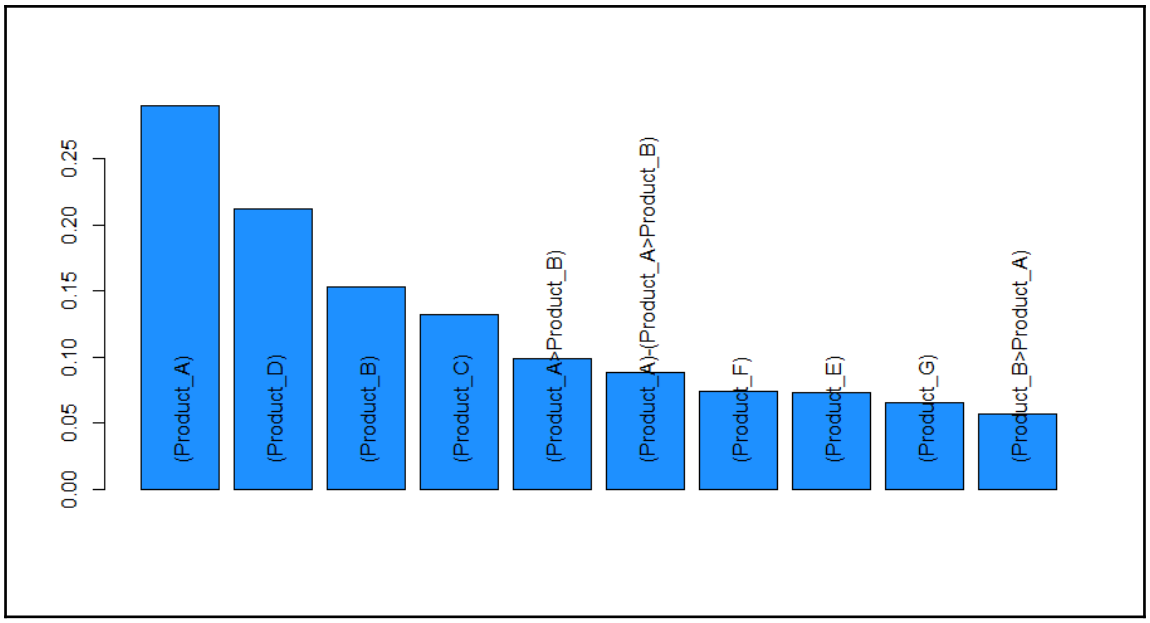

precision
recall
random
popular
UBCF
0.0
0.1
0.2
0.3
0.4


10 seq. (n=5000)
1
2
3
4
5
6
7
8
9
Purchase1
Purchase2
Purchase3
Purchase4
Purchase5
Purchase6
Purchase7
Purchase8
Product_A
Product_B
Product_D
Product_E
Product_G

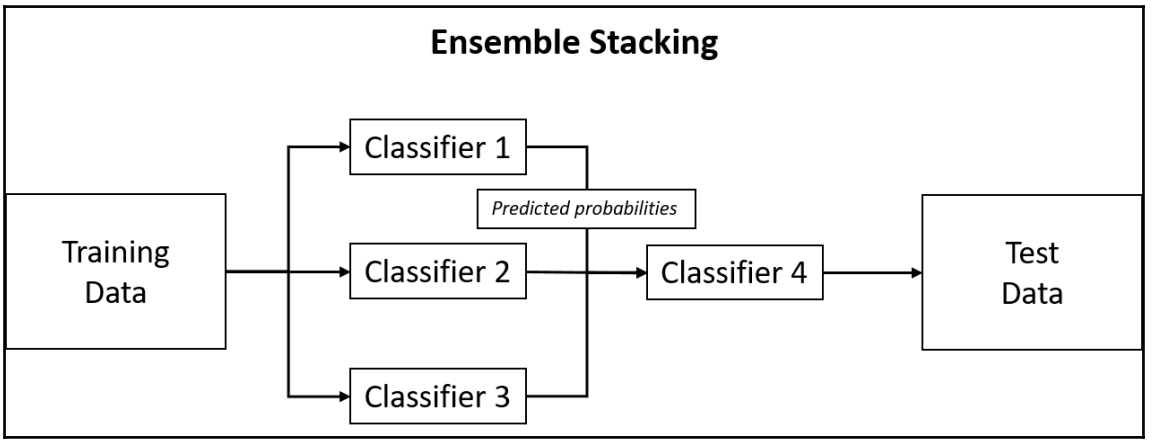

Freq. (n=5000)
0.0
0.2
0.4
0.6
0.8
1.0
Purchase1
Purchase2
Purchase3
Purchase4
Purchase5
Purchase6
Purchase7
Purchase8
Product_A
Product_B
Product_D
Product_E
Product_G

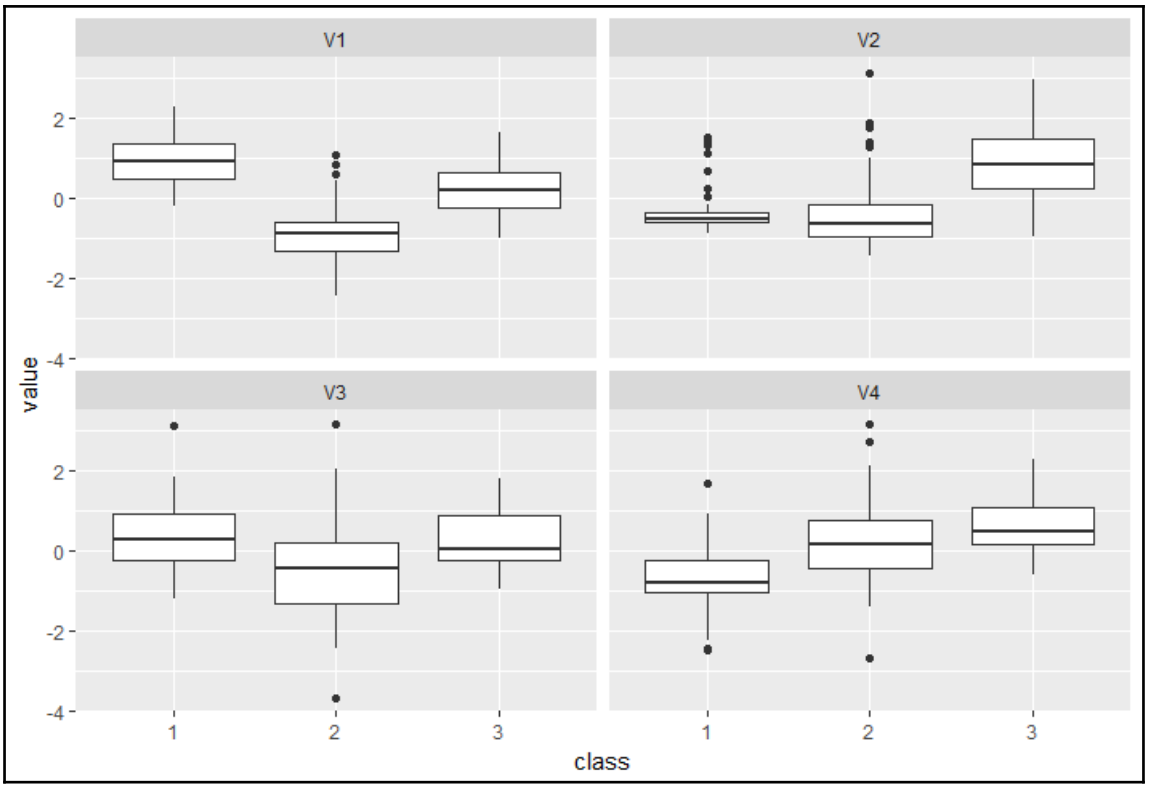

Segment1
Segment2
Segment3
Segment4
Freq. (n=2900)
Freq. (n=572)
Freq. (n=554)
Freq. (n=974)
0.0
0.4
0.8
Purchase1
Purchase3
Purchase5
Purchase7
Product_A
Product_B
Product_C
Product_D
Product_E
Product_F
Product_G

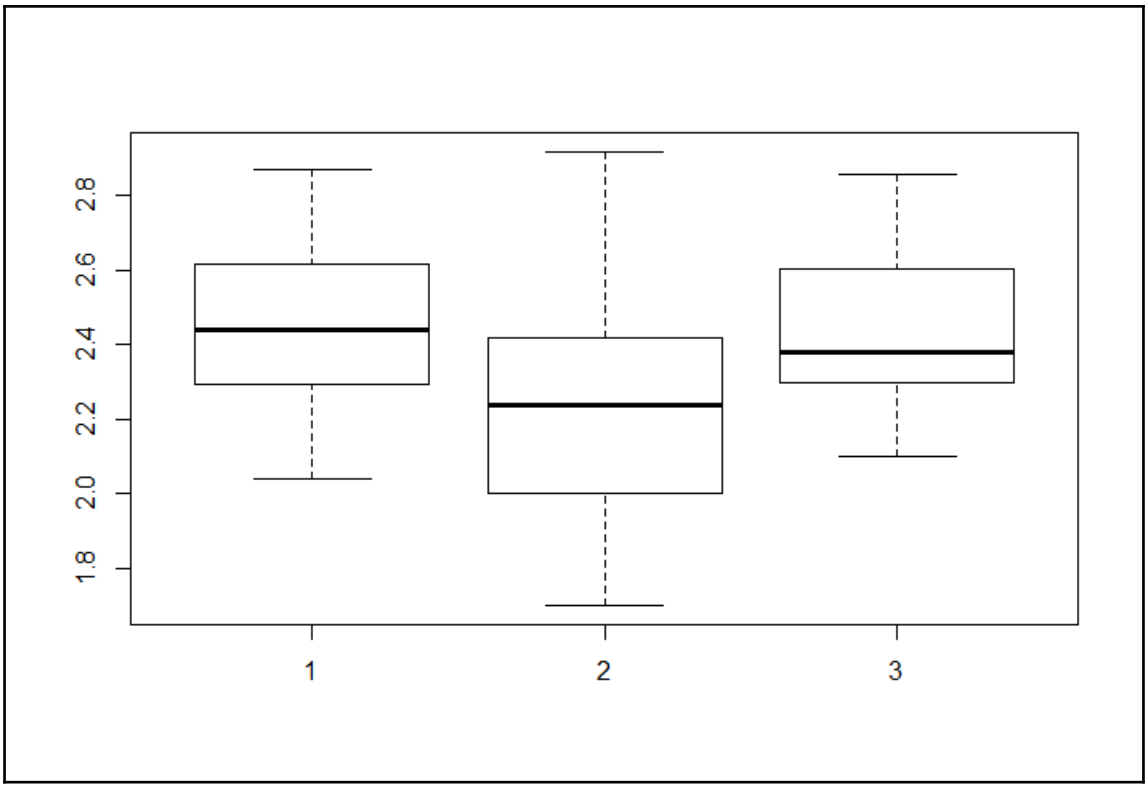

Segment1
Segment2
Segment3
Segment4
Modal state sequence (146 occurrences, freq=5%)
Modal state sequence (40 occurrences, freq=7%)
Modal state sequence (38 occurrences, freq=6.9%)
Modal state sequence (133 occurrences, freq=13.7%)
State freq. (n=2900)
State freq. (n=572)
State freq. (n=554)
State freq. (n=974)
0
0.25
.5
0.75
1
Purchase1
Purchase3
Purchase5
Purchase7
Product_A
Product_B
Product_C
Product_D
Product_E
Product_F
Product_G

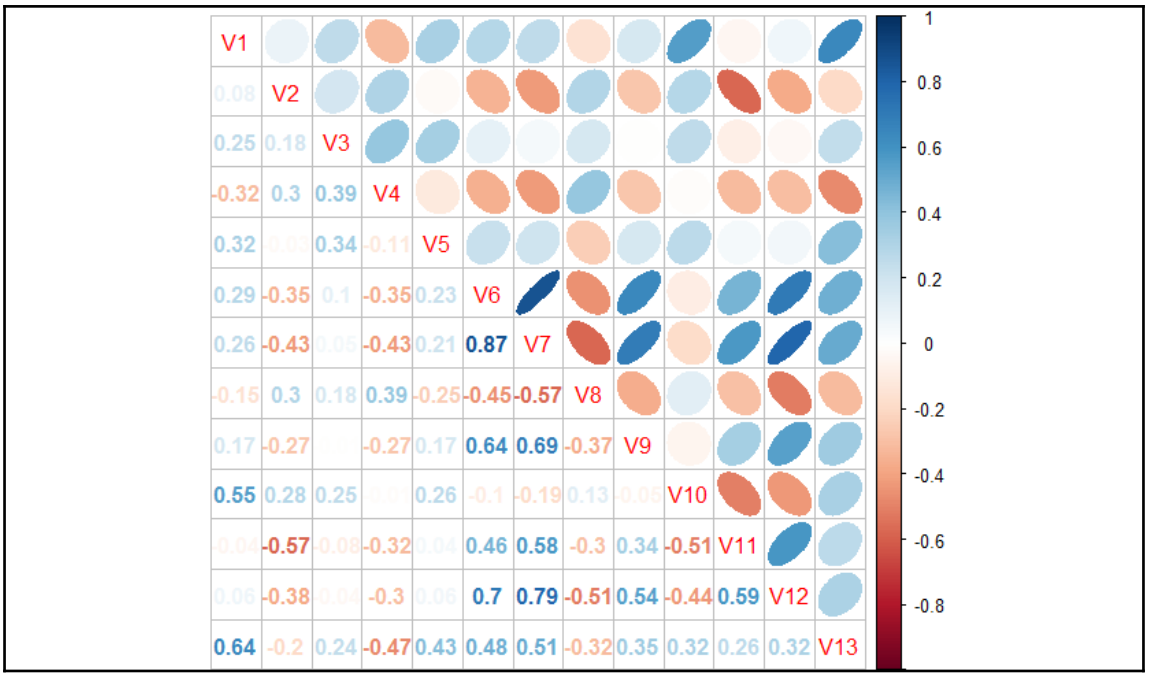

0.00
0.05
0.10
0.15
0.20
0.25
(Product_A)
(Product_D)
(Product_B)
(Product_C)
(Product_A>Product_B)
(Product_A)-(Product_A>Product_B)
(Product_F)
(Product_E)
(Product_G)
(Product_B>Product_A)

Chapter 11: Creating Ensembles and Multiclass Classification

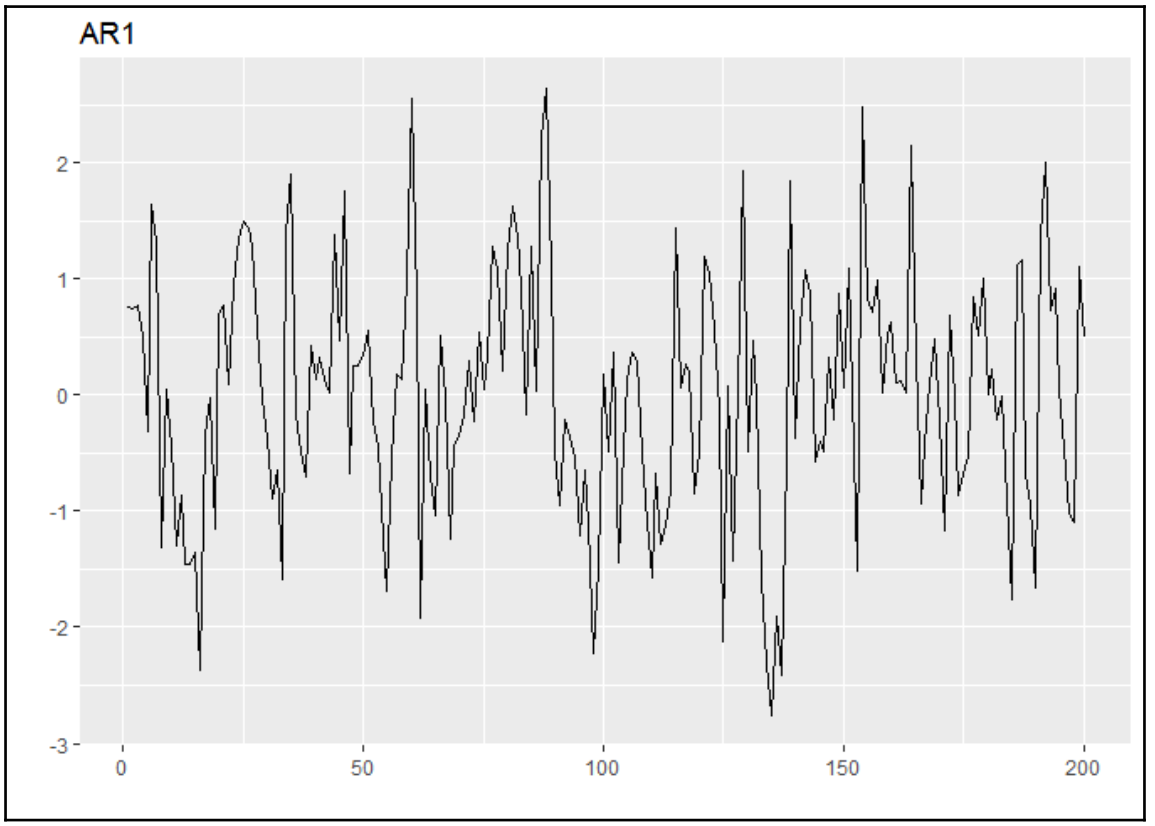


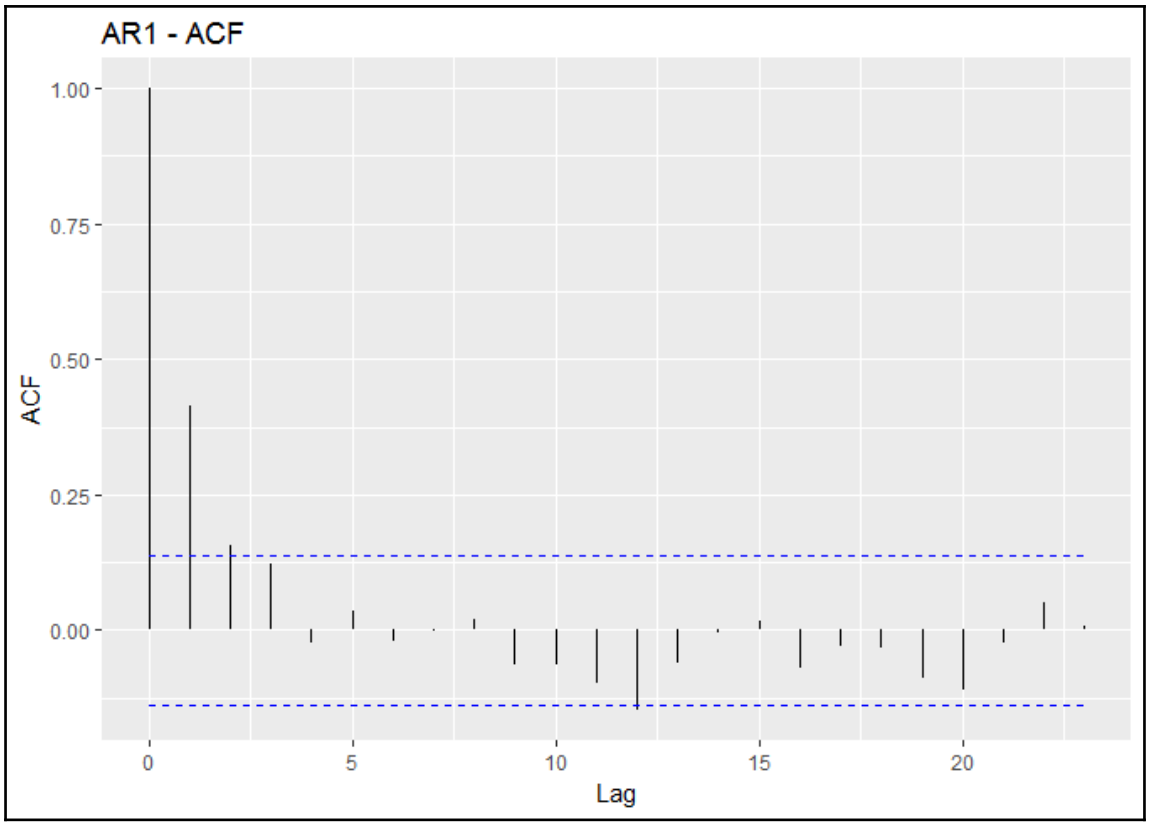

V1
V2
V3
V4
value
class
2
0
-2
-4
1
2
3

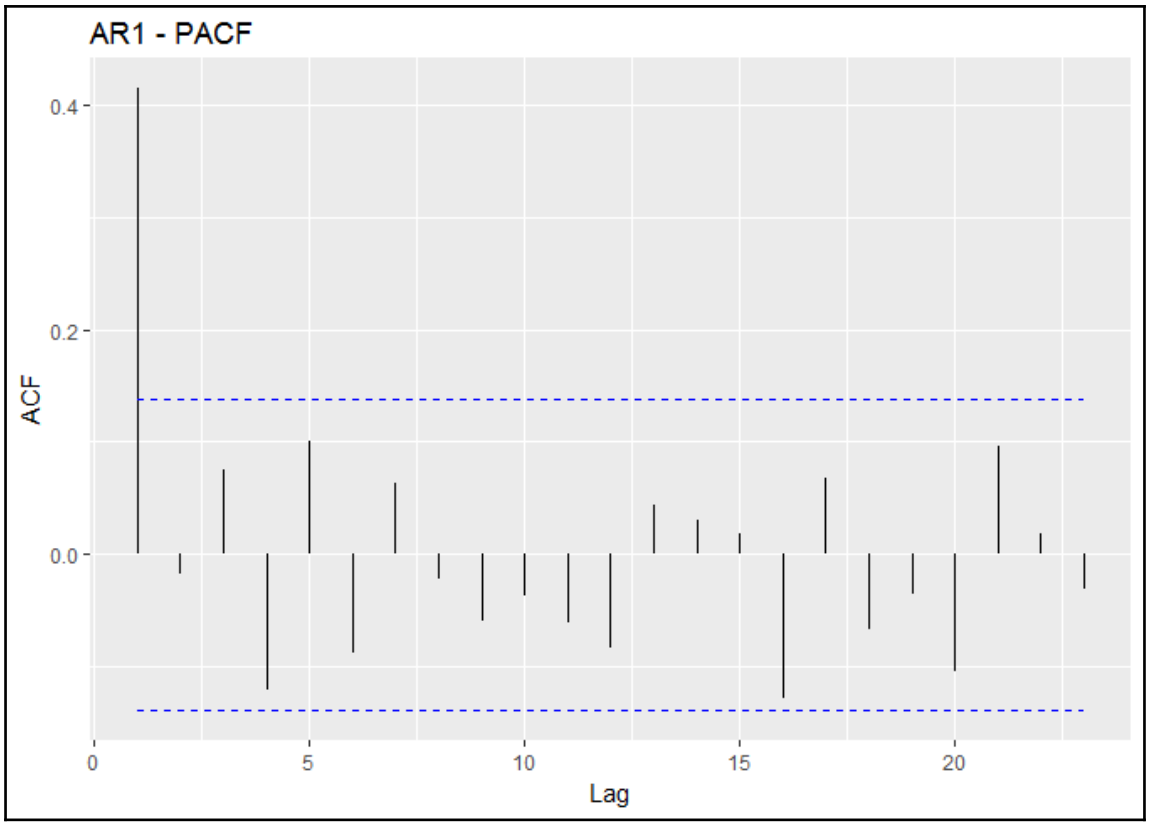

1.8
2.0
2.2
2.4
2.6
2.8
1
2
3

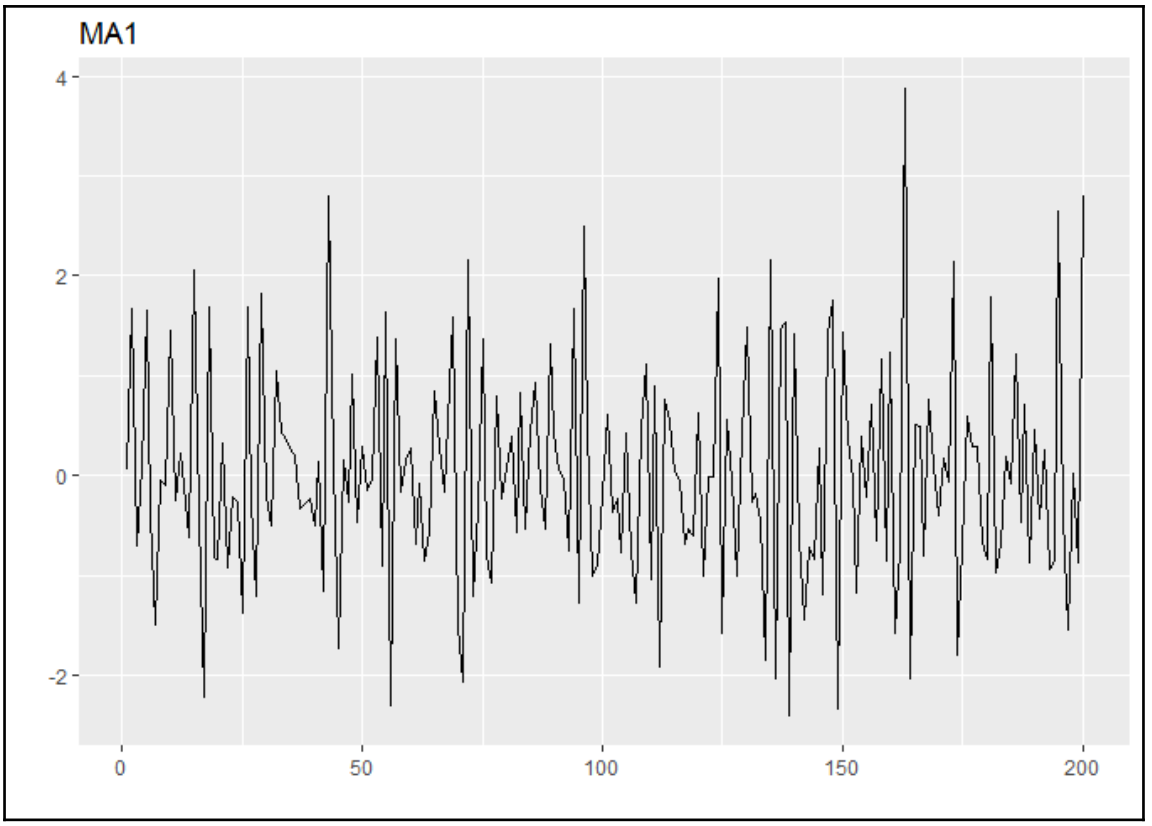

V1
0.08 V2
0.25 0.18 V3
-0.32 0.3 0.39 V4
0.32 0.34 -0.11 V5
0.29 -0.35 0.1 -0.35 0.23 V6
0.26 -0.43 -0.43 0.21 0.87 V7
-0.15 0.3 0.18 0.39 -0.25 -0.45 -0.57 V8
0.17 -0.27 -0.27 0.17 0.64 0.69 -0.37 V9
0.55 0.28 0.25 0.26 -0.1 -0.19 0.13 V10
-0.57 -0.08 -0.32 0.46 0.58 -0.3 0.34 -0.51 V11
-0.38 -0.3 0.7 0.79 -0.51 0.54 -0.44 0.59 V12
0.64 -0.2 0.24 -0.47 0.43 0.48 0.51 -0.32 0.35 0.32 0.26 0.32 V13
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8

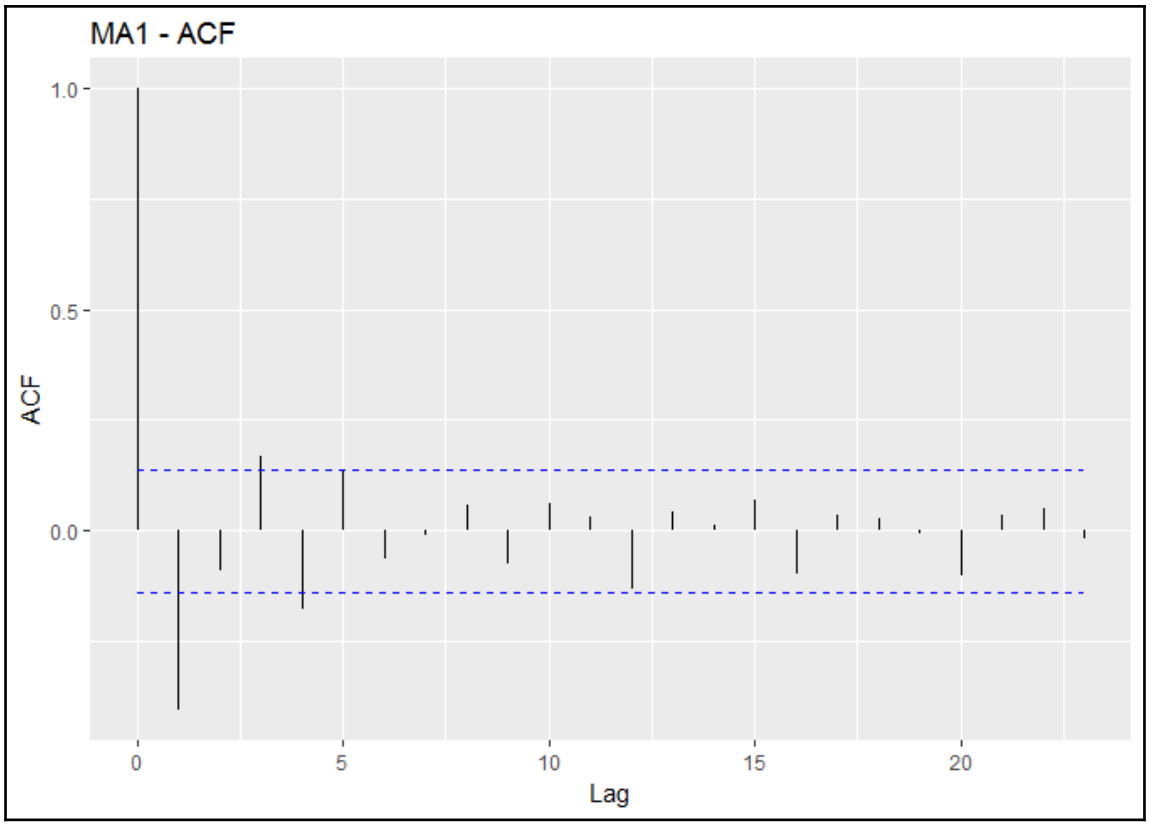
Chapter 12: Time Series and Causality

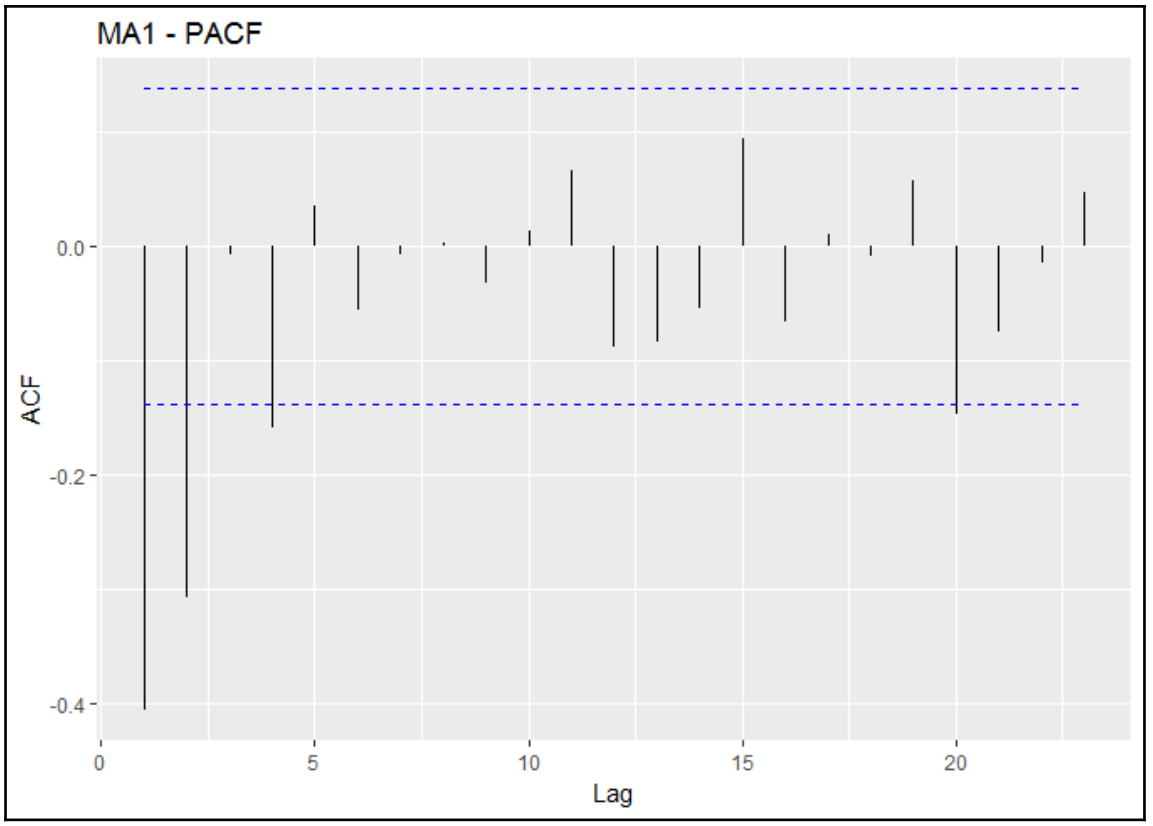



AR1 - ACF
1.00
0.75
0.50
0.25
0.00
ACF
0
5
10
15
20
Lag

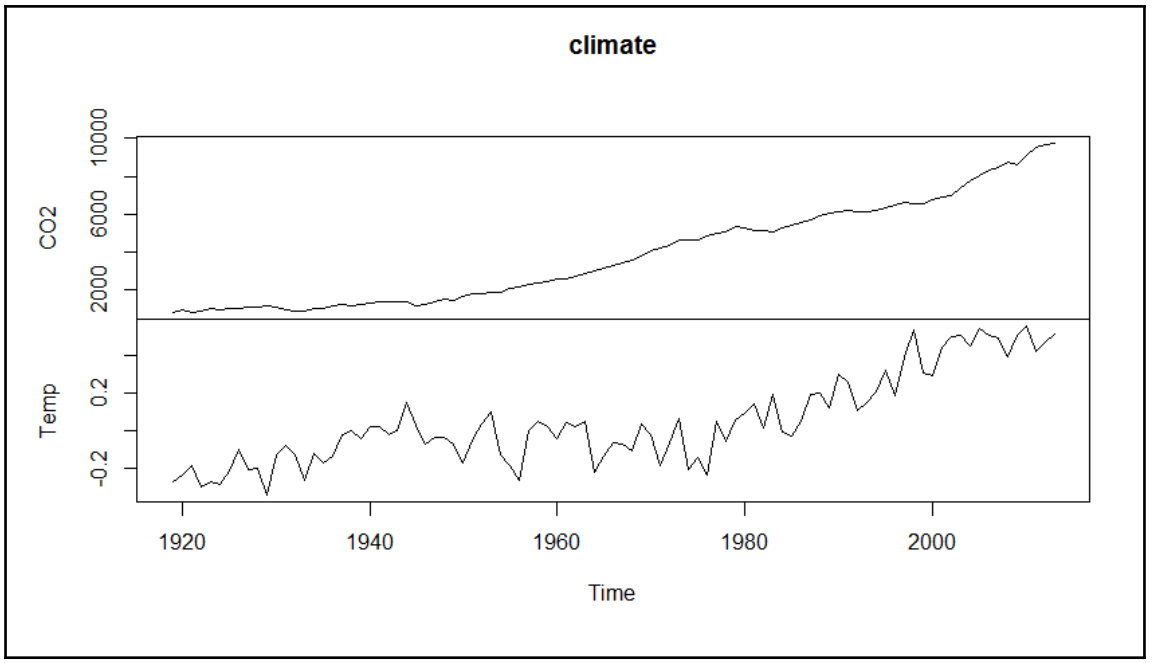

AR1 - PACF
0.4
0.2
0.0
ACF
0
5
10
15
20
Lag

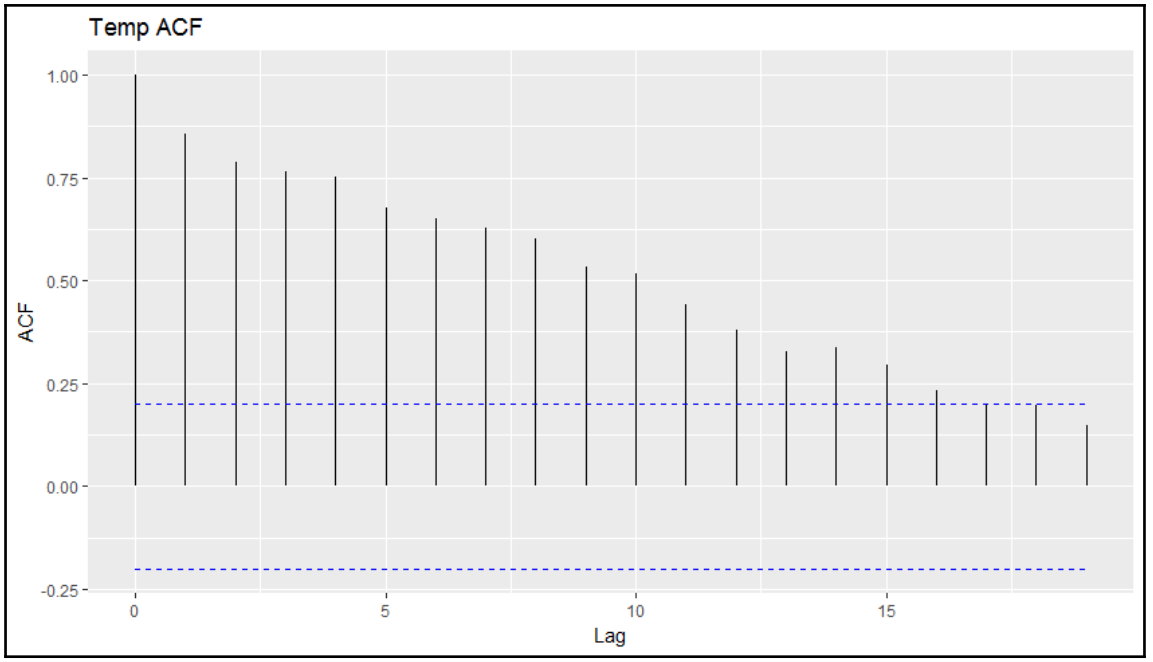

MA1
4
2
0
-2
0
50
100
150
200

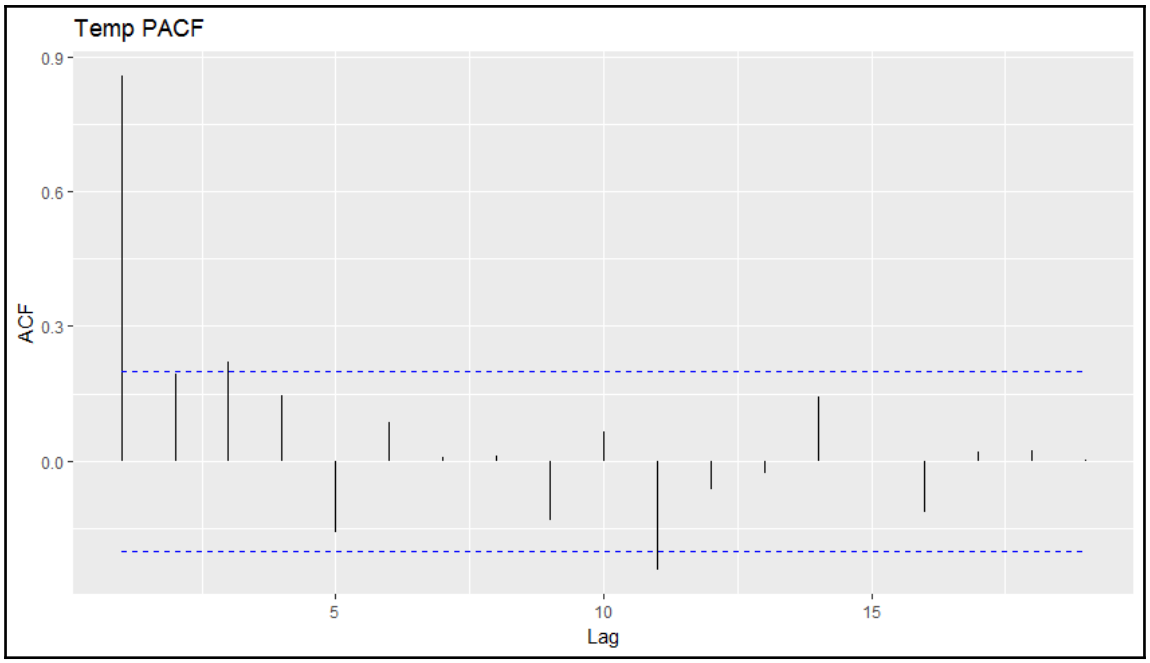

MA1 - ACF
1.0
0.5
0.0
ACF
0
5
10
15
20
Lag

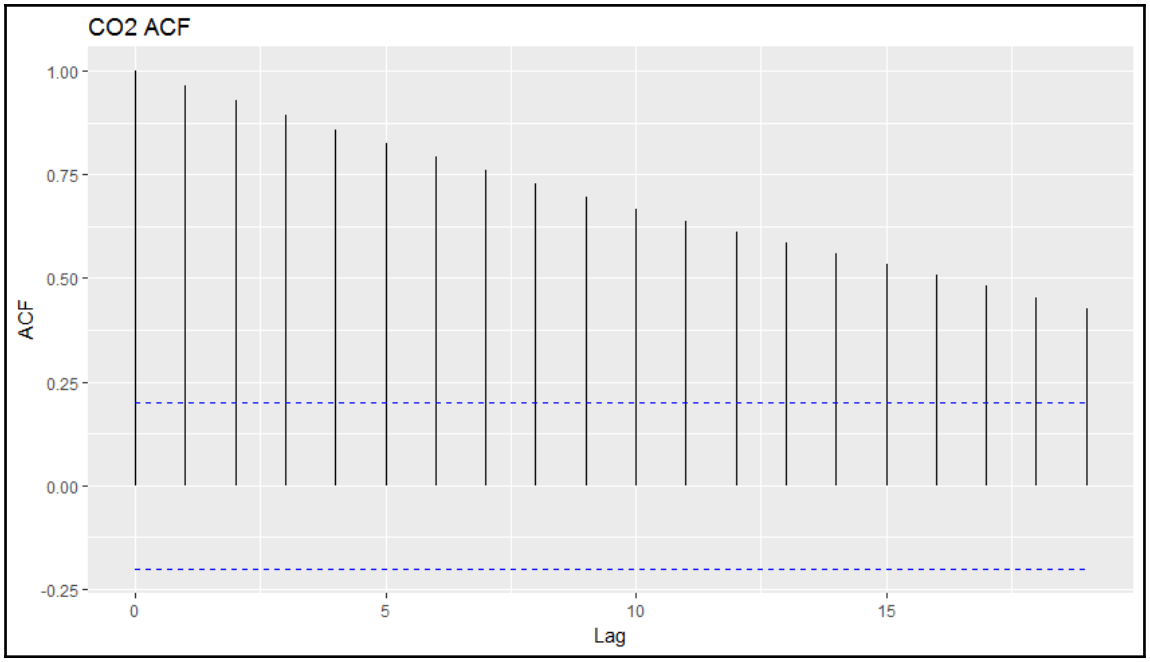


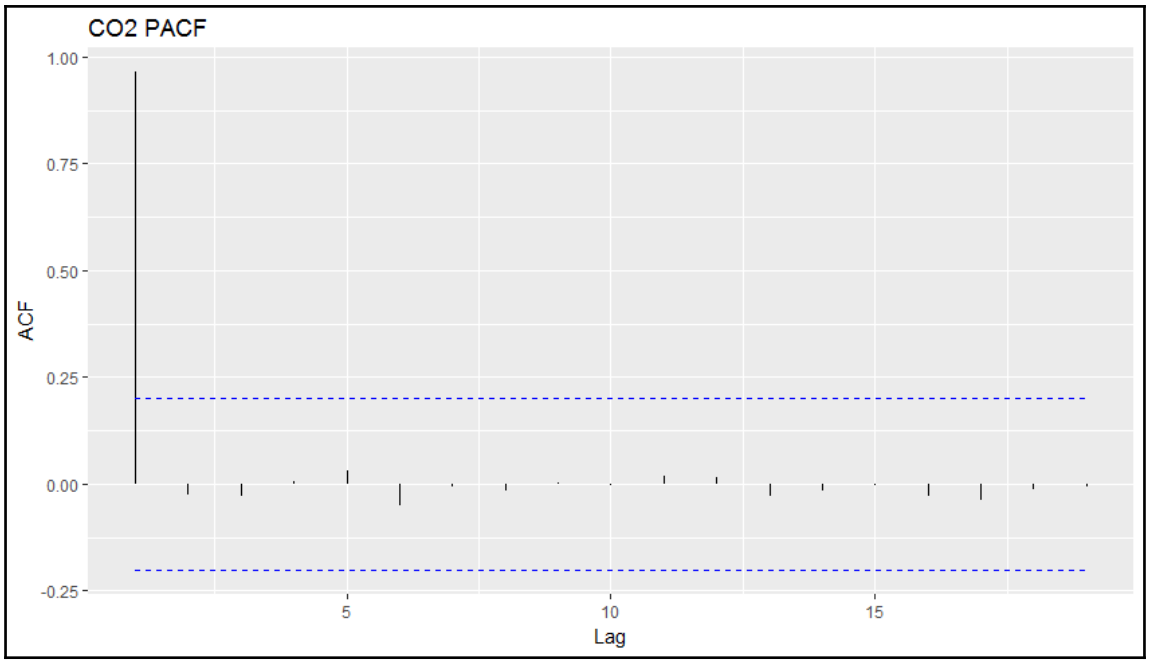
Let Ω = yt = β0 + β1yt-1 +…+ βkyt-k + ∈
and let π = yt = β0 + β1yt-1 +…+ βkyt-k + α1yt-1 +…+ αkyt-k + ∈

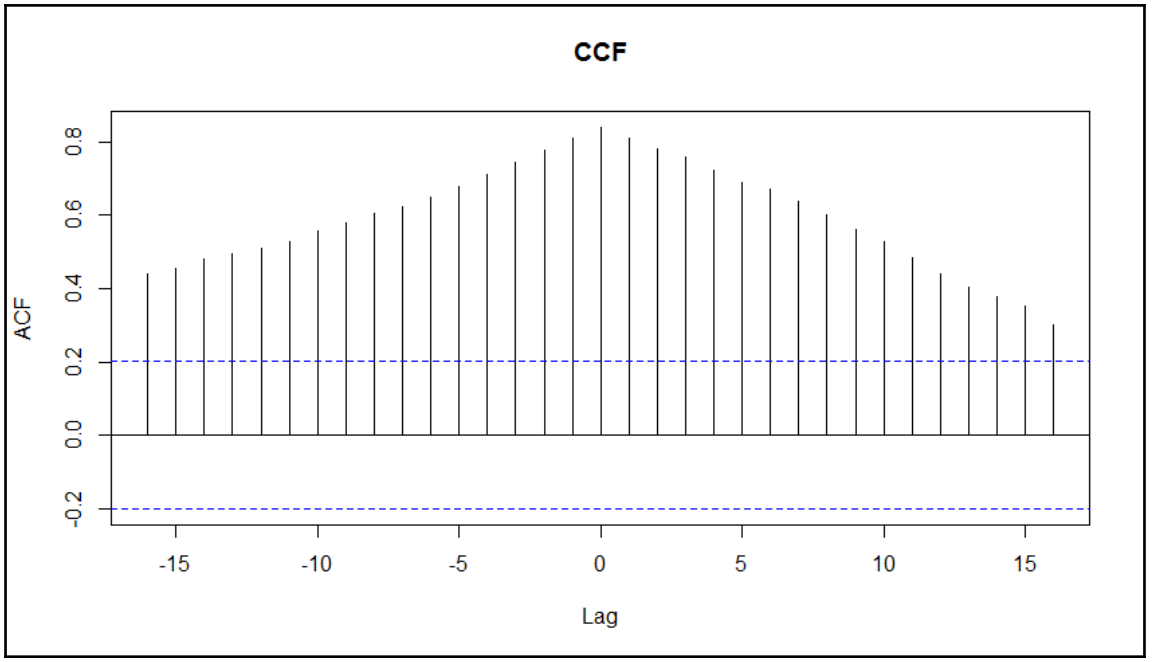

climate
CO2
2000
6000
10000
Temp
-0.2
0.2
1920
1940
1960
1980
2000
Time

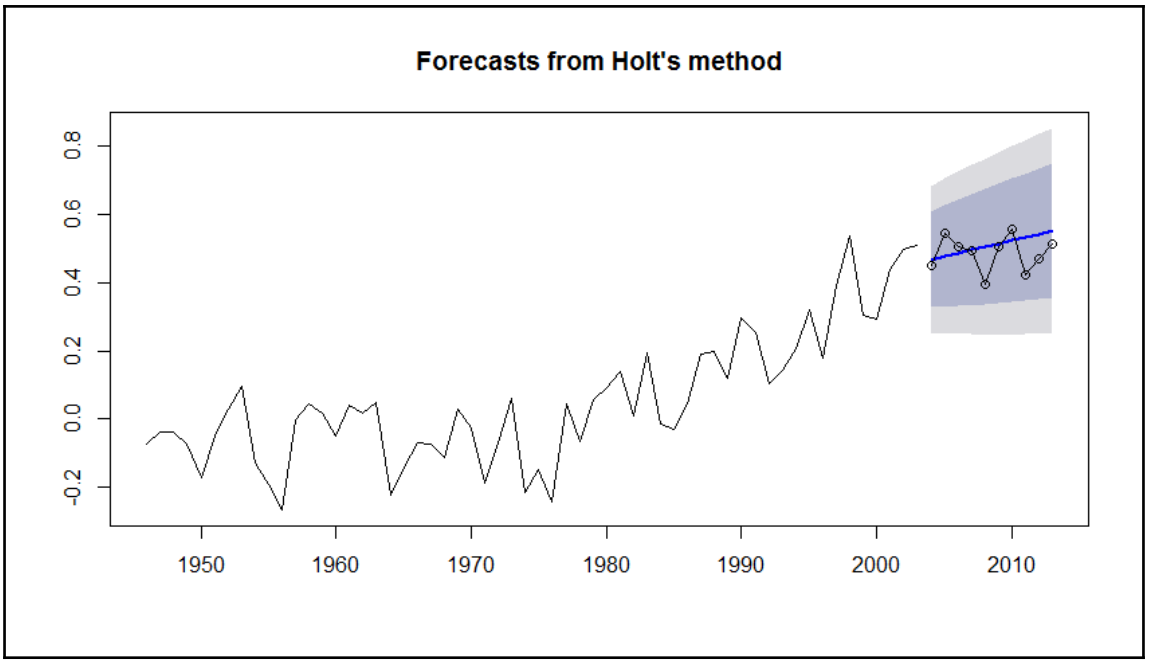

Temp ACF
ACF
1.00
0.75
0.50
0.25
0.00
-0.25
0
5
10
15
Lag

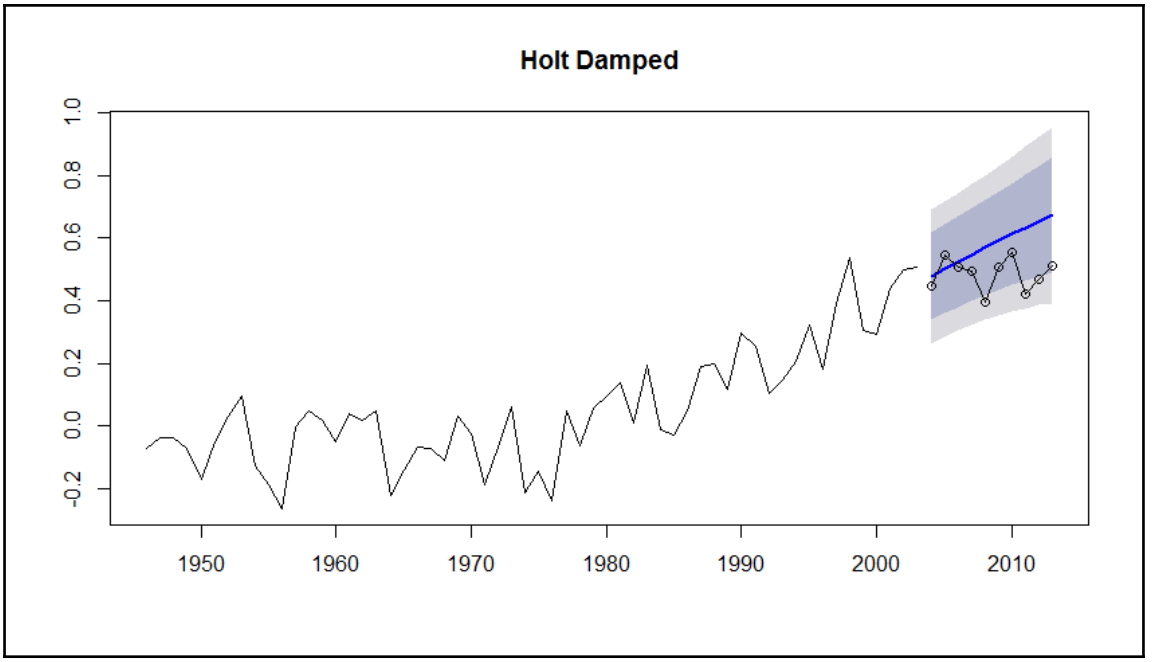

Temp PACF
ACF
0.9
0.6
0.3
0.0
5
10
15
Lag

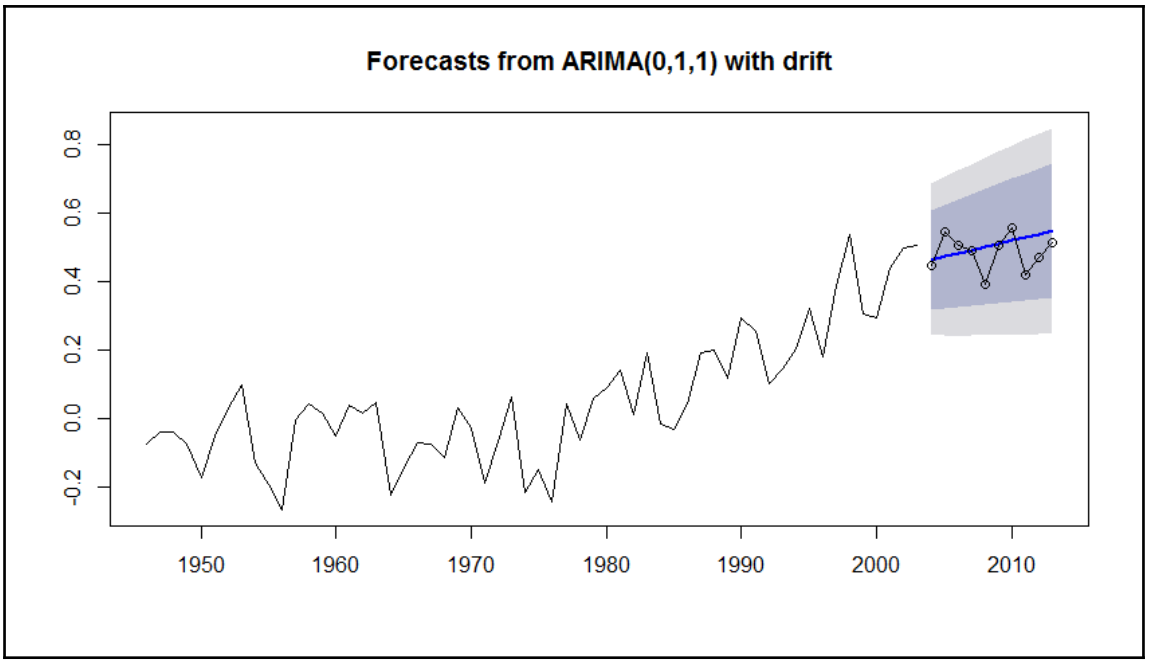

CO2 ACF
ACF
1.00
0.75
0.50
0.25
0.00
-0.25
0
5
10
15
Lag

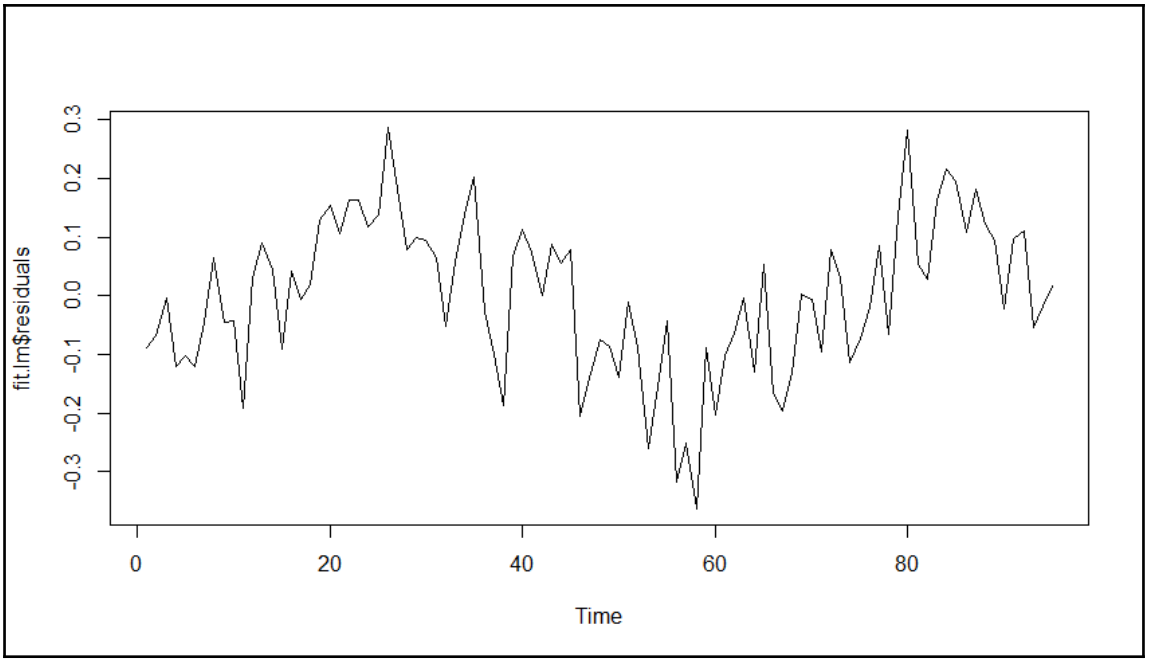

CO2 PACF
ACF
1.00
0.75
0.50
0.25
0.00
-0.25
5
10
15
Lag

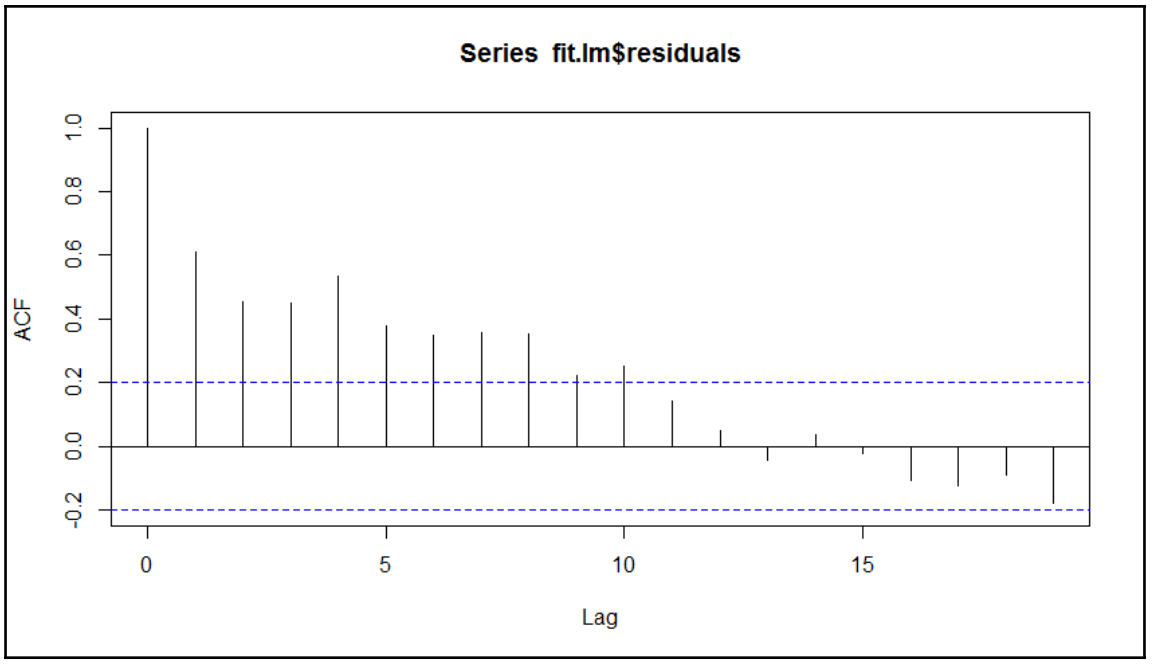

CCF
ACF
0.8
0.6
0.4
0.2
0.0
-0.2
-15
-10
-5
0
5
10
15
Lag

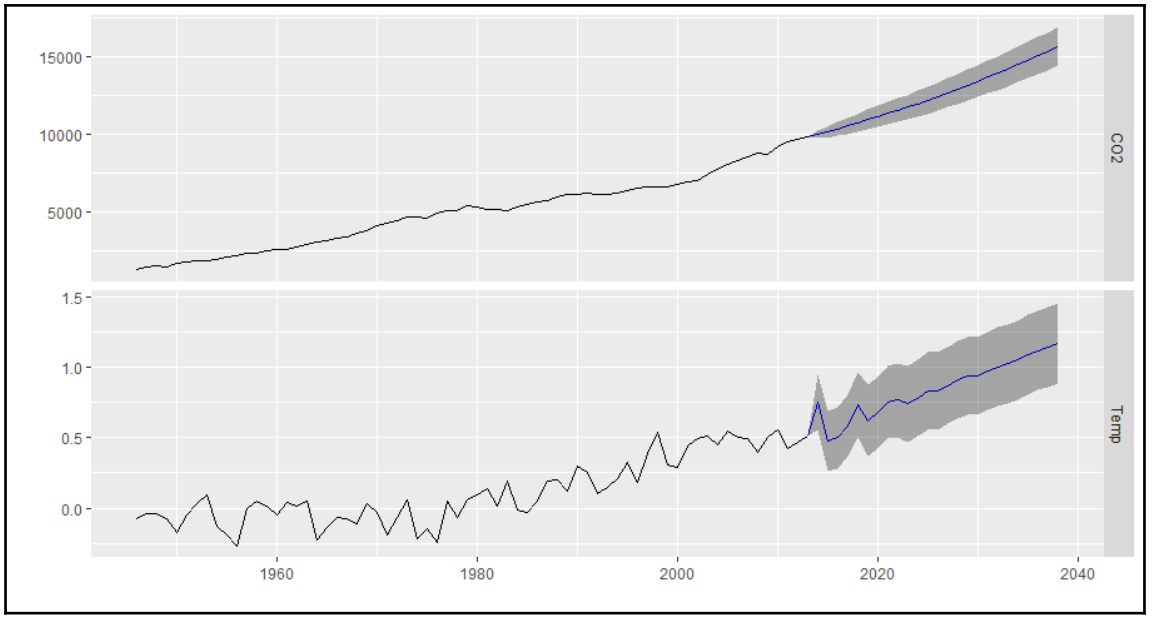

Forecasts from Holt's method
0.8
0.6
0.4
0.2
0.0
-0.2
1950
1960
1970
1980
1990
2000
2010


Holt Damped
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
1950
1960
1970
1980
1990
2000
2010

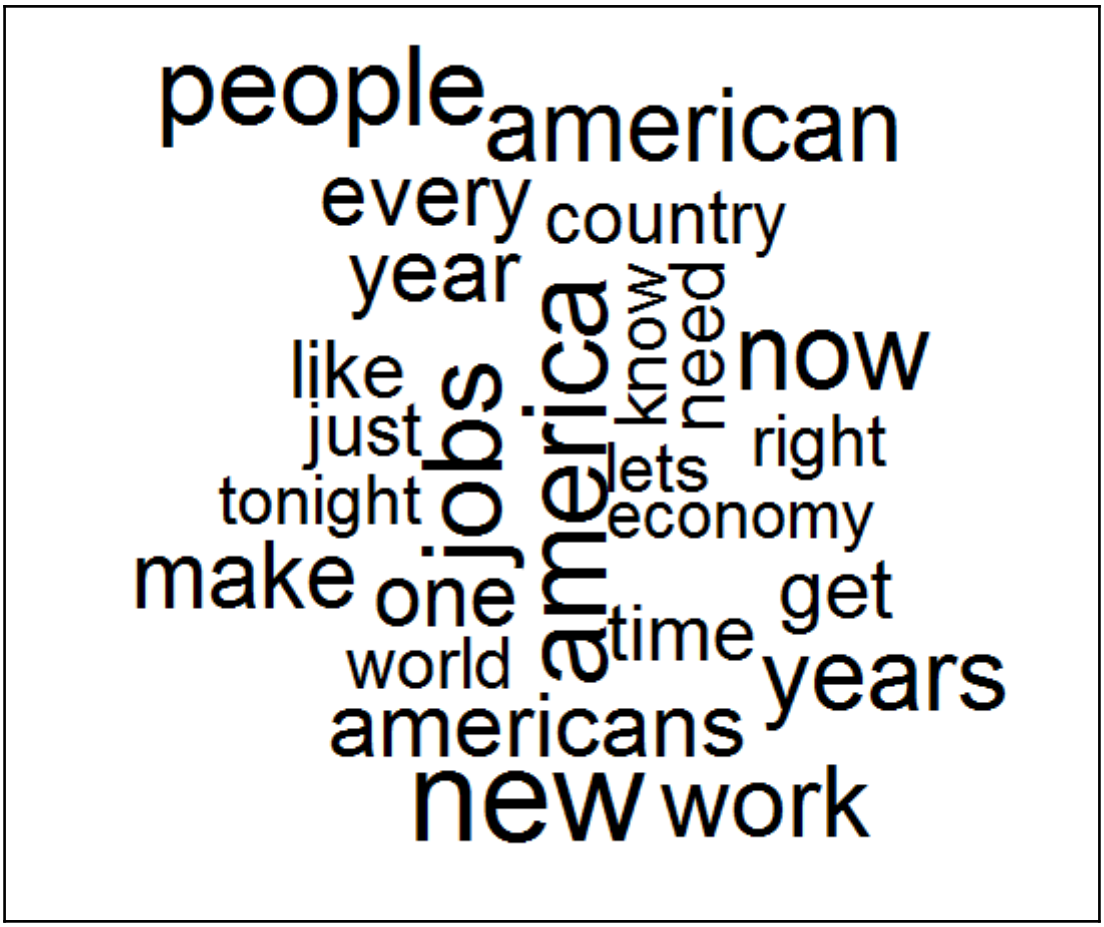

Forecasts from ARIMA(0,1,1) with drift
-0.2
0.0
0.2
0.4
0.6
0.8
1950
1960
1970
1980
1990
2000
2010

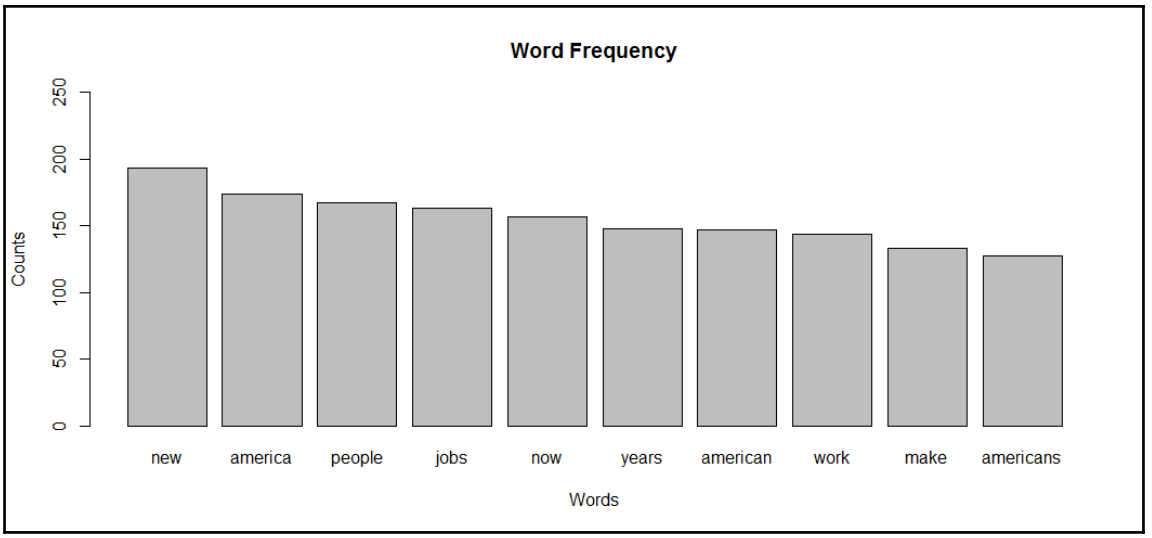

fit.lm$residuals
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0
20
40
60
80
Time

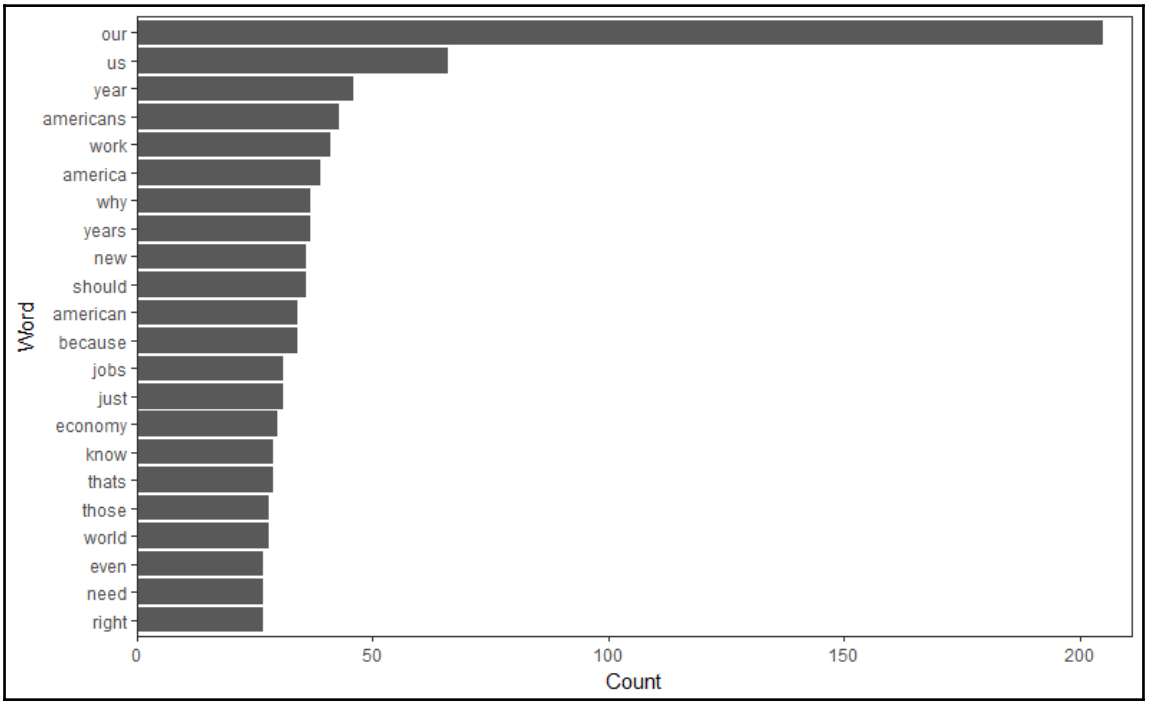

Series fit.lm$residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

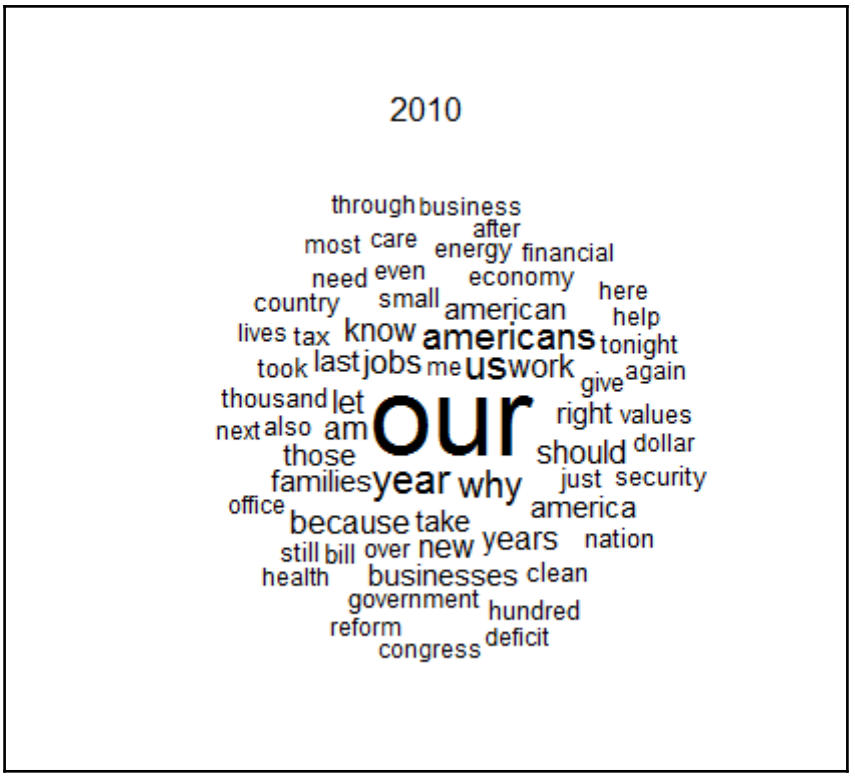

15000
10000
5000
CO2
1.5
1.0
0.5
0.0
Temp
1960
1980
2000
2020
2040

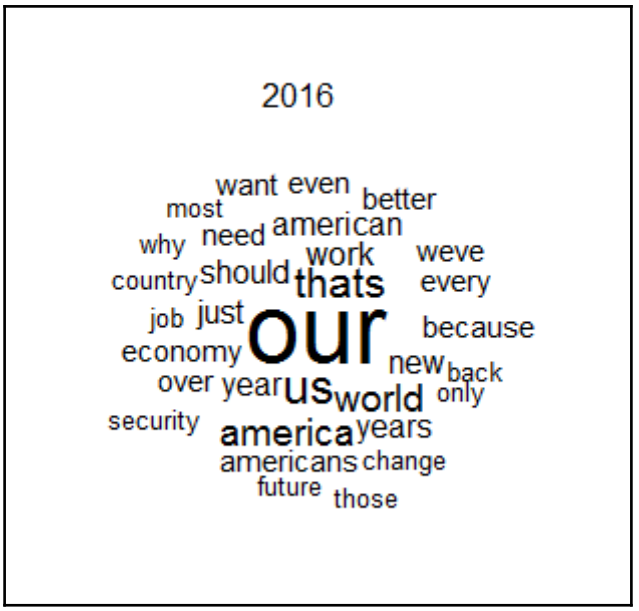
Chapter 13: Text Mining

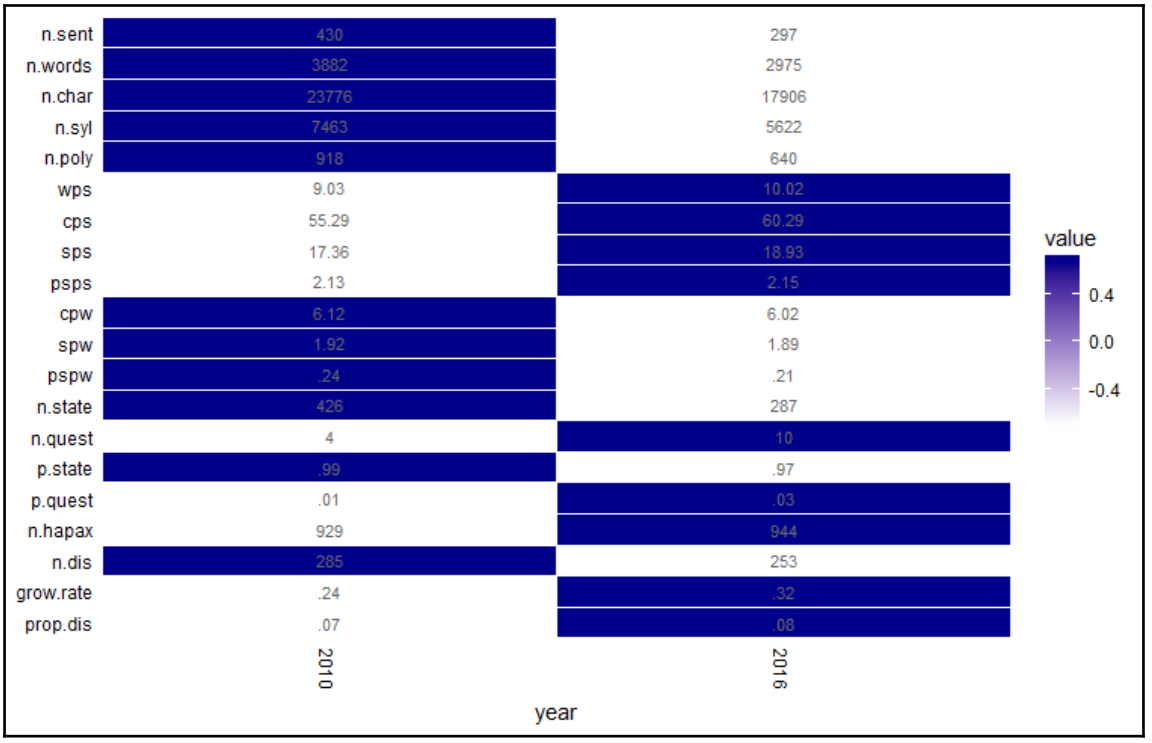

people
american
every
country
year
know
need
now
like
jobs
america
just
right
lets
tonight
economy
make
one
get
time
world
years
americans
new
work


Word Frequency
Counts
0
50
100
150
200
250
new
america
people
jobs
now
years
american
work
make
americans
Words

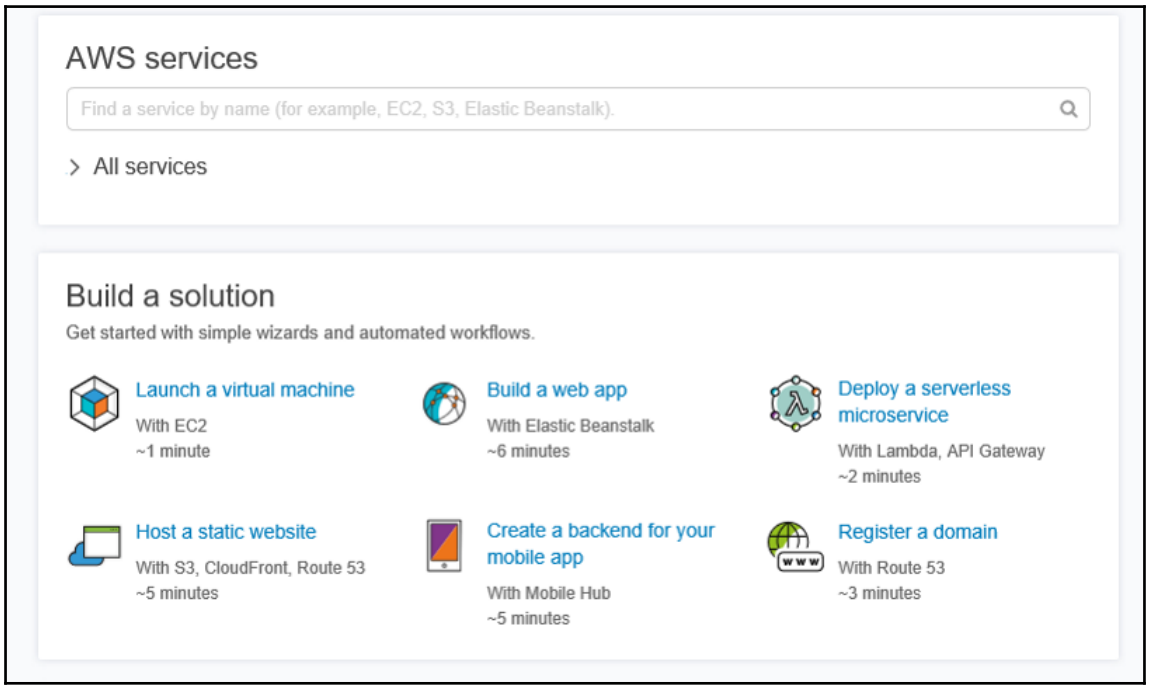

Word
our
us
year
americans
work
america
why
years
new
should
american
because
jobs
just
economy
know
thats
those
world
even
need
right
0
50
100
150
200
Count

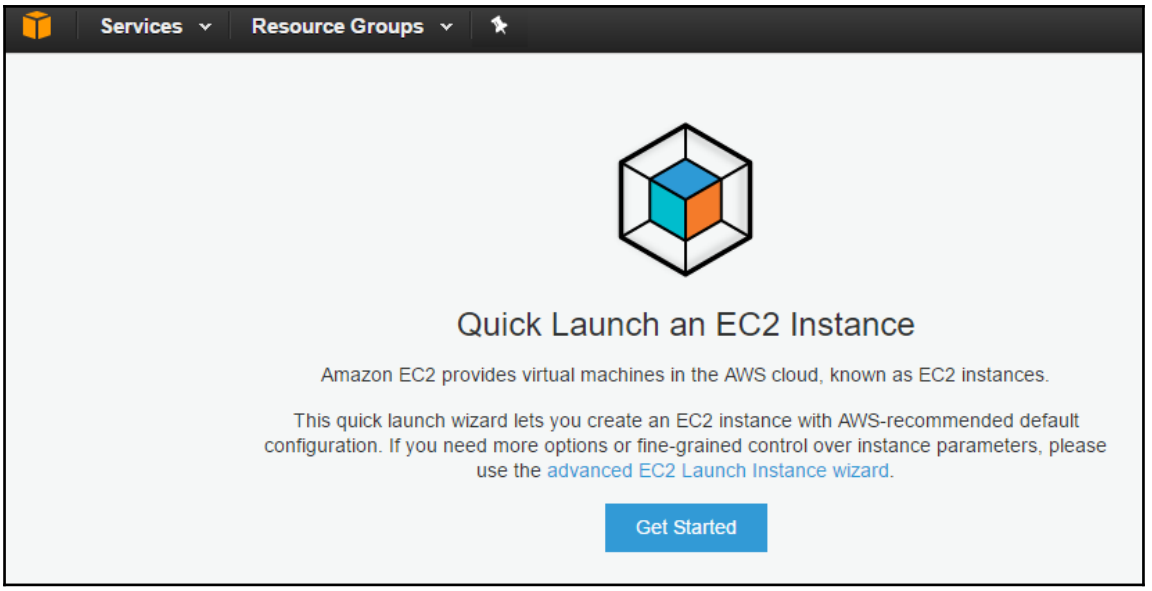

2010
through
business
after
most
care
energy
financial
need
even
economy
here
country
small
american
help
lives
tax
know
americans
tonight
took
last
jobs
me
us
work
give
again
thousand
let
our
right
values
next
also
am
should
dollar
those
families
year
why
just
security
office
america
because
take
years
nation
still
bill
over
new
businesses
clean
health
government
hundred
reform
deficit
congress

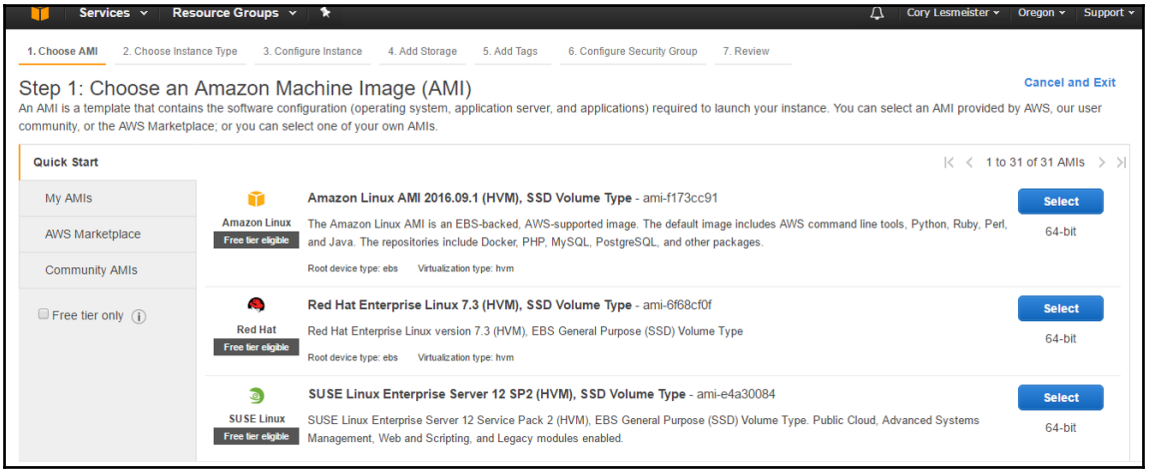

2016
want
even
better
most
american
need
why
work
weve
country
should
thats
every
job
just
our
because
economy
new
back
over
year
us
world
only
security
america
years
americans
change
future
those

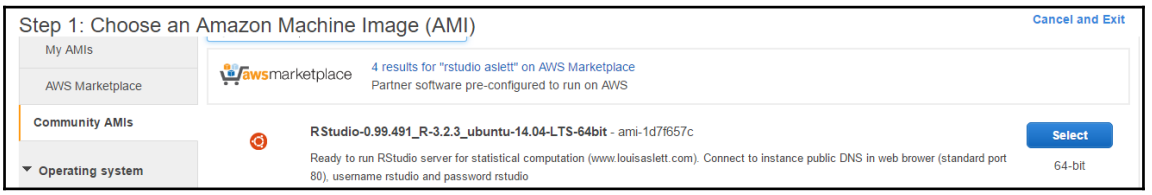

n.sent
n.words
n.char
n.syl
n.poly
wps
cps
sps
psps
cpw
spw
pspw
n.state
n.quest
p.state
p.quest
n.hapax
n.dis
grow.rate
prop.dis
430
3882
23776
7463
918
9.03
55.29
17.36
2.13
6.12
1.92
.24
426
4
.99
.01
929
285
.24
.07
297
2975
17906
5622
640
10.02
60.29
18.93
2.15
6.02
1.89
.21
287
10
.97
.03
944
253
.32
.08
2010
2016
year
value
0.4
0.0
-0.4

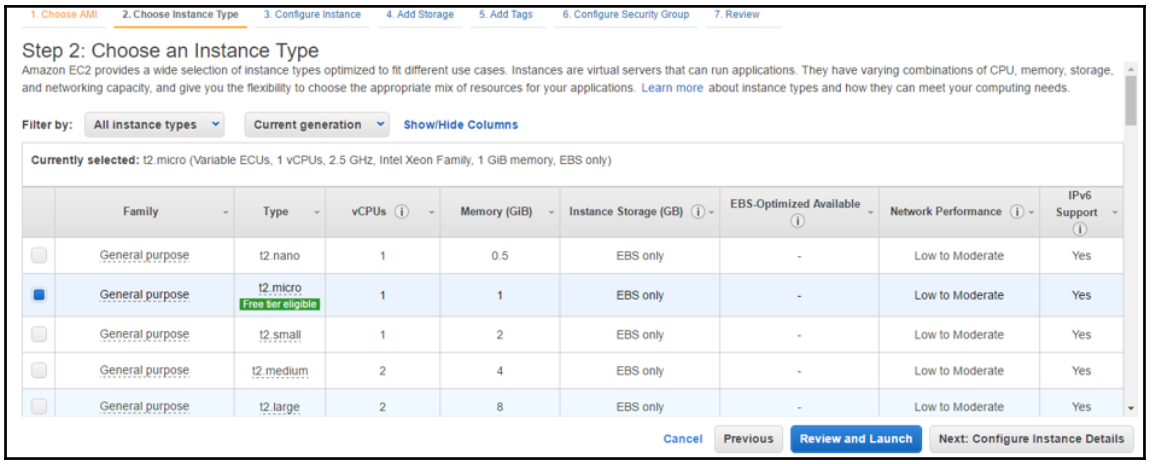
Chapter 14: R on the Cloud

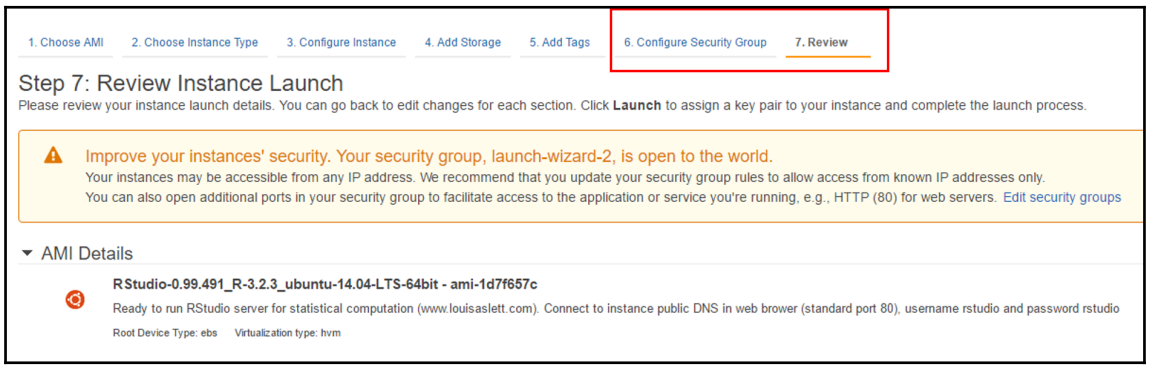


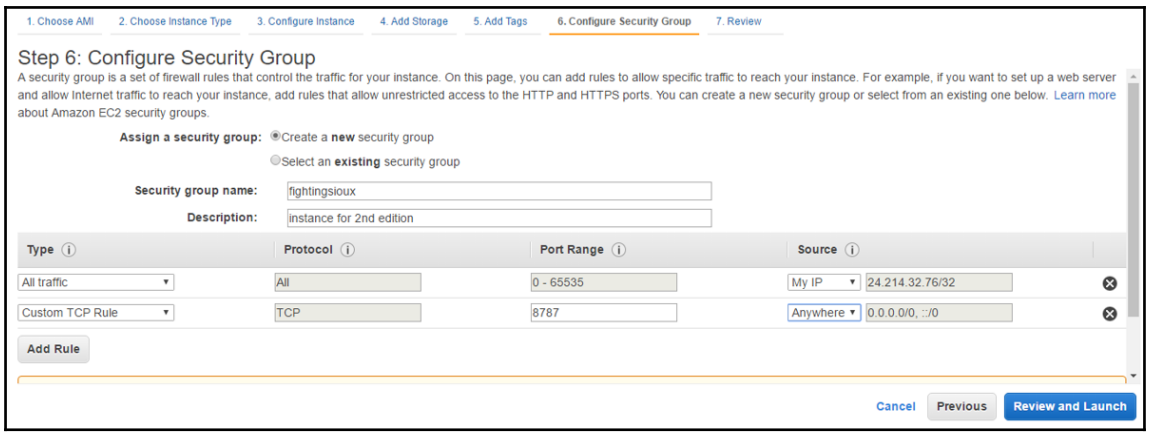

Services
Resource Groups
Quick Launch an EC2 Instance
Amazon EC2 provides virtual machines in the AWS cloud, known as EC2 instances.
This quick launch wizard lets you create an EC2 instance with AWS-recommended default configuration. If you need more options or fine-grained control over instance parameters, please use the advanced EC2 Launch Instance wizard.
Get Started

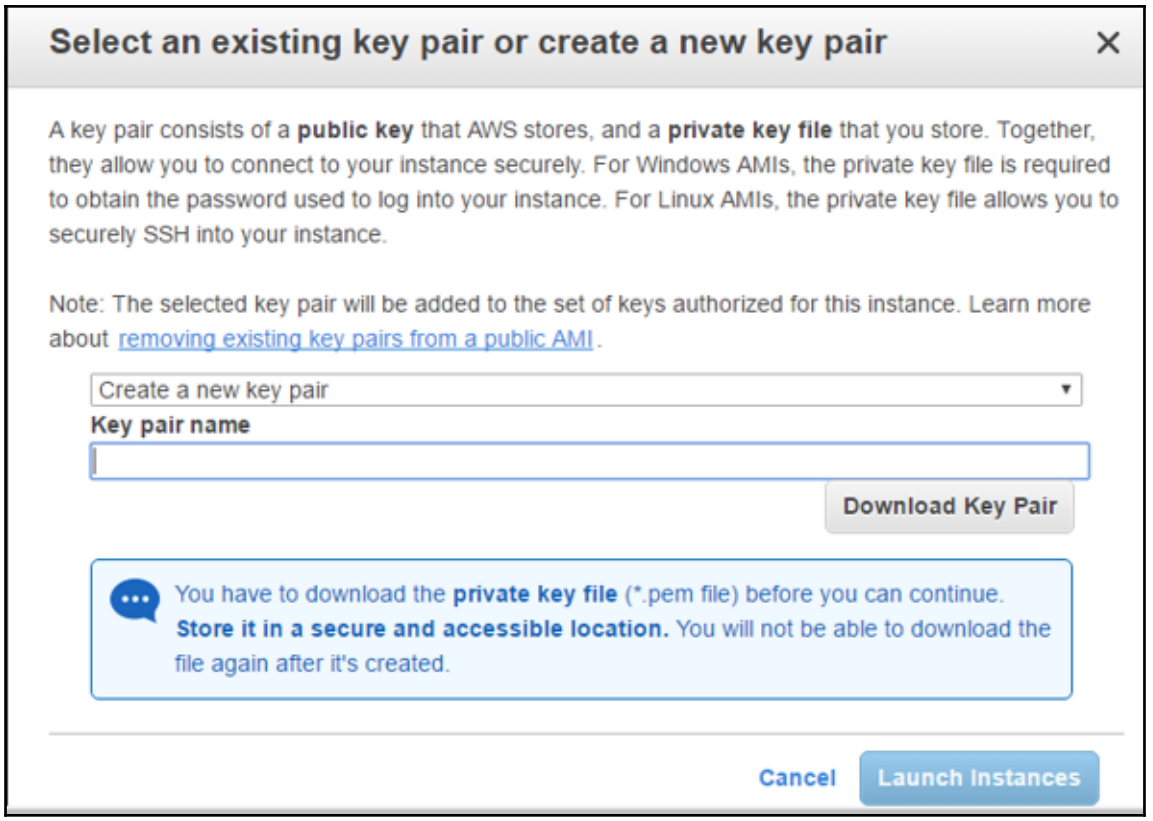

Services
Resource Groups
Cory Lesmeister
Oregon
Support
1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Add Tags
6. Configure Security Group
7. Review
Step 1: Choose an Amazon Machine Image (AMI)
Cancel and Exit
An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.
Quick Start
1 to 31 of 31 AMIs
My AMIs
AWS Marketplace
Community AMIs
Free tier only
Amazon Linux
Free tier eligible
Amazon Linux AMI 2016.09.1 (HVM), SSD Volume Type - ami-f173cc91
The Amazon Linux AMI is an EBS-backed, AWS-supported image. The default image includes AWS command line tools, Python, Ruby, Perl, and Java. The repositories include Docker, PHP, MySQL, PostgreSQL, and other packages.
Root device type: ebs
Virtualization type: hvm
Select
64-bit
Red Hat
Free tier eligible
Red Hat Enterprise Linux 7.3 (HVM), SSD Volume Type - ami-6f68cf0f
Red Hat Enterprise Linux version 7.3 (HVM), EBS General Purpose (SSD) Volume Type
Root device type: ebs
Virtualization type: hvm
Select
64-bit
SUSE Linux
Free tier eligible
SUSE Linux Enterprise Server 12 SP2 (HVM), SSD Volume Type - ami-e4a30084
SUSE Linux Enterprise Server 12 Service Pack 2 (HVM), EBS General Purpose (SSD) Volume Type. Public Cloud, Advanced Systems Management, Web and Scripting, and Legacy modules enabled.
Select
64-bit

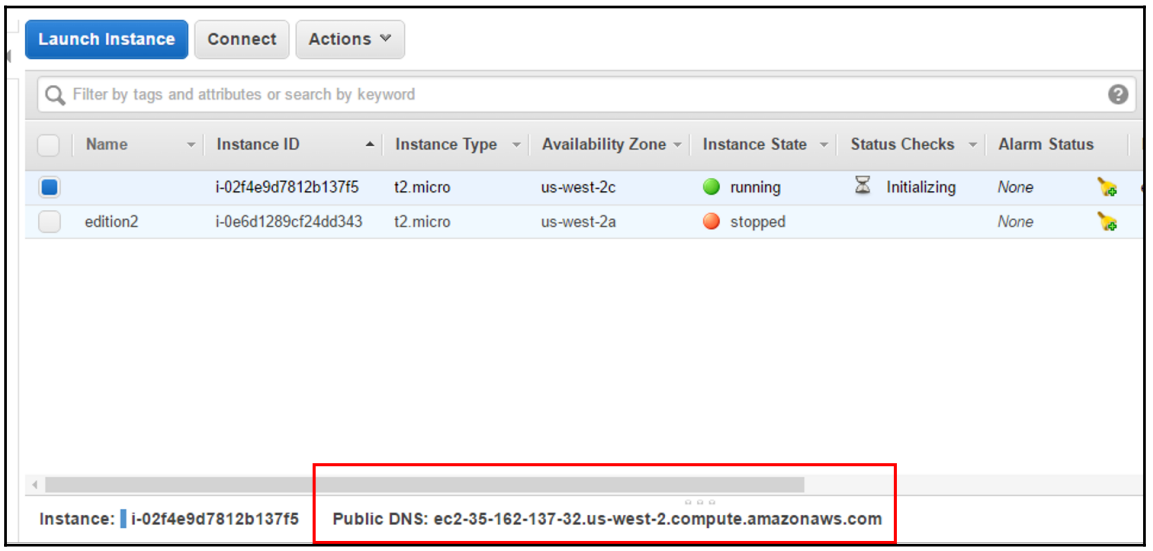

Step 1: Choose an Amazon Machine Image (AMI)
Cancel and Exit
My AMIs
AWS Marketplace
Community AMIs
Operating system
aws marketplace
4 results for "rstudio aslett" on AWS Marketplace
Partner software pre-configured to run on AWS
RStudio-0.99.491_R-3.2.3_ubuntu-14.04-LTS-64bit - ami-1d7f657c
Ready to run RStudio server for statistical computation (www.louisaslett.com). Connect to instance public DNS in web brower (standard port 80), username rstudio and password rstudio
Select
64-bit

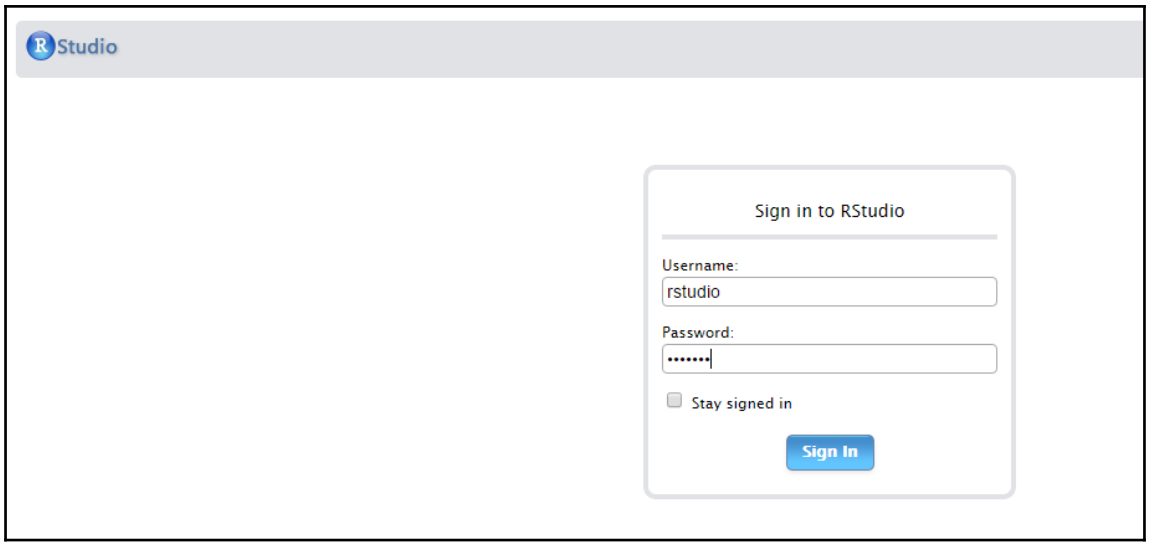

1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Add Tags
6. Configure Security Group
7. Review
Step 2: Choose an Instance Type
Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. Learn more about instance types and how they can meet your computing needs.
Filter by:
All instance types
Current generation
Show/Hide Columns
Currently selected: t2.micro (Variable ECUs, 1 vCPUs, 2.5 GHz, Intel Xeon Family, 1 GiB memory, EBS only)
Family
Type
vCPUs
Memory (GiB)
Instance Storage (GB)
EBS-Optimized Available
Network Performance
IPv6 Support
General purpose
t2.nano
1
0.5
EBS only
-
Low to Moderate
Yes
General purpose
t2.micro
Free tier eligible
1
1
EBS only
-
Low to Moderate
Yes
General purpose
t2.small
1
2
EBS only
-
Low to Moderate
Yes
General purpose
t2.medium
2
4
EBS only
-
Low to Moderate
Yes
General purpose
t2.large
2
8
EBS only
-
Low to Moderate
Yes
Cancel
Previous
Review and Launch
Next: Configure Instance Details

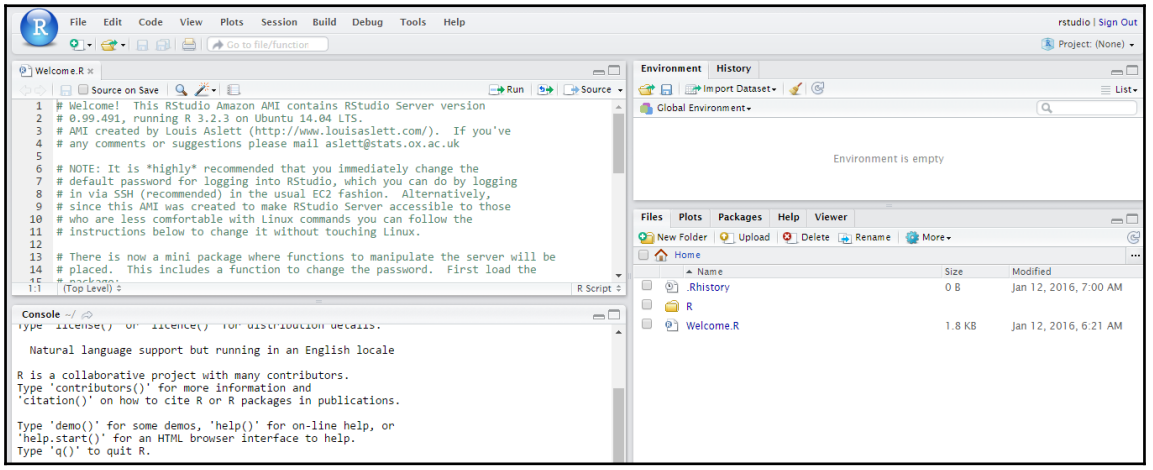

1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Add Tags
6. Configure Security Group
7. Review
Step 7: Review Instance Launch
Please review your instance launch details. You can go back to edit changes for each section. Click Launch to assign a key pair to your instance and complete the launch process.
Improve your instances' security. Your security group, launch-wizard-2, is open to the world.
Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only.
You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. Edit security groups
AMI Details
RStudio-0.99.491_R-3.2.3_ubuntu-14.04-LTS-64bit - ami-1d7f657c
Ready to run RStudio server for statistical computation (www.louisaslett.com). Connect to instance public DNS in web brower (standard port 80), username rstudio and password rstudio
Root Device Type: ebs
Virtualization type: hvm


1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Add Tags
6. Configure Security Group
7. Review
Step 6: Configure Security Group
A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. Learn more about Amazon EC2 security groups.
Assign a security group:
Create a new security group
Select an existing security group
Security group name:
fightingsioux
Description:
instance for 2nd edition
Type
Protocol
Port Range
Source
All traffic
All
0 - 65535
My IP
24.214.32.76/32
Custom TCP Rule
TCP
8787
Anywhere
0.0.0.0/0, ::/0
Add Rule
Cancel
Previous
Review and Launch

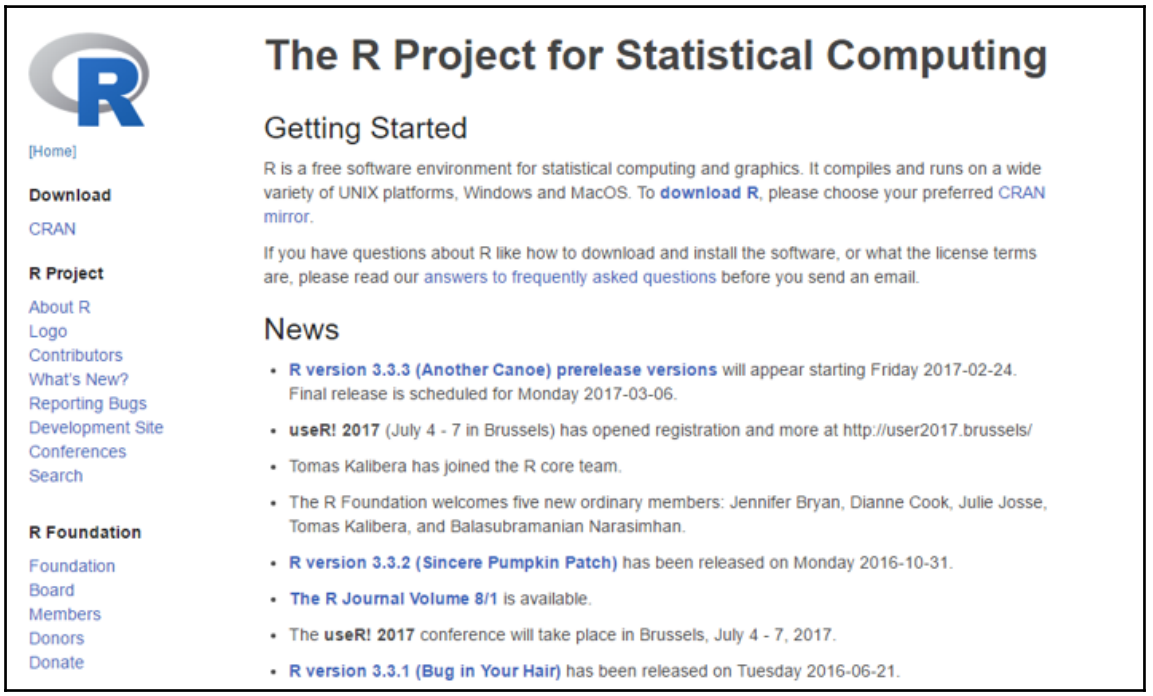

Select an existing key pair or create a new key pair
A key pair consists of a public key that AWS stores, and a private key file that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.
Note: The selected key pair will be added to the set of keys authorized for this instance. Learn more about removing existing key pairs from a public AMI.
Create a new key pair
Key pair name
Download Key Pair
You have to download the private key file (*.pem file) before you can continue. Store it in a secure and accessible location. You will not be able to download the file again after it's created.
Cancel
Launch Instances


Launch Instance
Connect
Actions
Filter by tags and attributes or search by keyword
Name
Instance ID
Instance Type
Availability Zone
Instance State
Status Checks
Alarm Status
i-02f4e9d7812b137f5
t2.micro
us-west-2c
running
Initializing
None
edition2
i-0e6d1289cf24dd343
t2.micro
us-west-2a
stopped
None
Instance: i-02f4e9d7812b137f5
Public DNS: ec2-35-162-137-32.us-west-2.compute.amazonaws.com


RStudio
Sign in to RStudio
Username:
rstudio
Password:
Stay signed in
Sign In


File Edit Code View Plots Session Build Debug Tools Help
rstudio | Sign Out
Welcome.R
Source on Save
Run
Source
Welcome! This RStudio Amazon AMI contains RStudio Server version
0.99.491, running R 3.2.3 on Ubuntu 14.04 LTS.
AMI created by Louis Aslett (http://www.louisaslett.com/). If you've
any comments or suggestions please mail aslett@stats.ox.ac.uk
NOTE: It is *highly* recommended that you immediately change the
default password for logging into RStudio, which you can do by logging
in via SSH (recommended) in the usual EC2 fashion. Alternatively,
since this AMI was created to make RStudio Server accessible to those
who are less comfortable with Linux commands you can follow the
instructions below to change it without touching Linux.
There is now a mini package where functions to manipulate the server will be
placed. This includes a function to change the password. First load the
(Top Level)
R Script
Console
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Environment History
Import Dataset
Global Environment
Environment is empty
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Home
Name Size Modified
.Rhistory 0 B Jan 12, 2016, 7:00 AM
R
Welcome.R 1.8 KB Jan 12, 2016, 6:21 AM

Appendix A: R Fundamentals



[Home]

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Reporting Bugs
Development Site
Conferences
Search

R Foundation

Foundation
Board
Members
Donors
Donate

The R Project for Statistical Computing

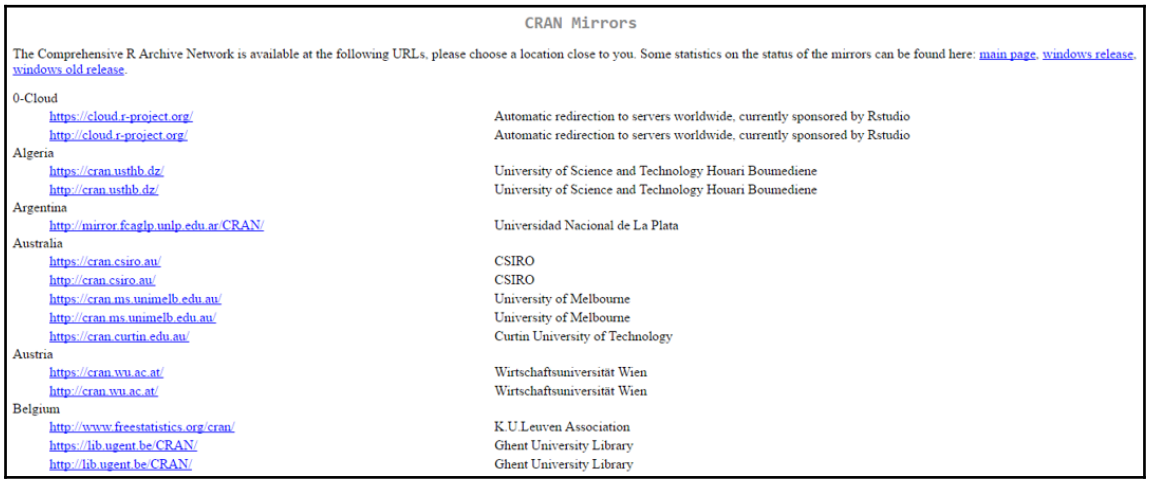
Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- **R version 3.3.3 (Another Canoe) prerelease versions** will appear starting Friday 2017-02-24. Final release is scheduled for Monday 2017-03-06.
- **useR! 2017** (July 4 - 7 in Brussels) has opened registration and more at http://user2017.brussels/
- Tomas Kalibera has joined the R core team.
- The R Foundation welcomes five new ordinary members: Jennifer Bryan, Dianne Cook, Julie Josse, Tomas Kalibera, and Balasubramanian Narasimhan.
- **R version 3.3.2 (Sincere Pumpkin Patch)** has been released on Monday 2016-10-31.
- **The R Journal Volume 8/1** is available.
- The **useR! 2017** conference will take place in Brussels, July 4 - 7, 2017.
- **R version 3.3.1 (Bug in Your Hair)** has been released on Tuesday 2016-06-21.



CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: main page, windows release, windows old release.

0-Cloud	
https://cloud.r-project.org/	Automatic redirection to servers worldwide, currently sponsored by Rstudio
http://cloud.r-project.org/	Automatic redirection to servers worldwide, currently sponsored by Rstudio
Algeria	
https://cran.usthb.dz/	University of Science and Technology Houari Boumediene
http://cran.usthb.dz/	University of Science and Technology Houari Boumediene
Argentina	
http://mirror.fcaglp.unlp.edu.ar/CRAN/	Universidad Nacional de La Plata
Australia	
https://cran.csiro.au/	CSIRO
http://cran.csiro.au/	CSIRO
https://cran.ms.unimelb.edu.au/	University of Melbourne
http://cran.ms.unimelb.edu.au/	University of Melbourne
https://cran.curtin.edu.au/	Curtin University of Technology
Austria	
https://cran.wu.ac.at/	Wirtschaftsuniversität Wien
http://cran.wu.ac.at/	Wirtschaftsuniversität Wien
Belgium	
http://www.freestatistics.org/cran/	K.U.Leuven Association
https://lib.ugent.be/CRAN/	Ghent University Library
http://lib.ugent.be/CRAN/	Ghent University Library

USA

https://cran.cnr.berkeley.edu/	University of California, Berkeley, CA
http://cran.cnr.berkeley.edu/	University of California, Berkeley, CA
http://cran.stat.ucla.edu/	University of California, Los Angeles, CA
https://mirror.las.iastate.edu/CRAN/	Iowa State University, Ames, IA
http://mirror.las.iastate.edu/CRAN/	Iowa State University, Ames, IA
https://ftp.ussg.iu.edu/CRAN/	Indiana University
http://ftp.ussg.iu.edu/CRAN/	Indiana University
https://rweb.crmda.ku.edu/cran/	University of Kansas, Lawrence, KS
http://rweb.crmda.ku.edu/cran/	University of Kansas, Lawrence, KS
https://cran.mtu.edu/	Michigan Technological University, Houghton, MI
http://cran.mtu.edu/	Michigan Technological University, Houghton, MI
http://cran.wustl.edu/	Washington University, St. Louis, MO
http://archive.linux.duke.edu/cran/	Duke University, Durham, NC
http://cran.case.edu/	Case Western Reserve University, Cleveland, OH
http://iis.stat.wright.edu/CRAN/	Wright State University, Dayton, OH

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

R-3.3.2 for Windows (32/64 bit)

Download R 3.3.2 for Windows (62 megabytes, 32/64 bit)

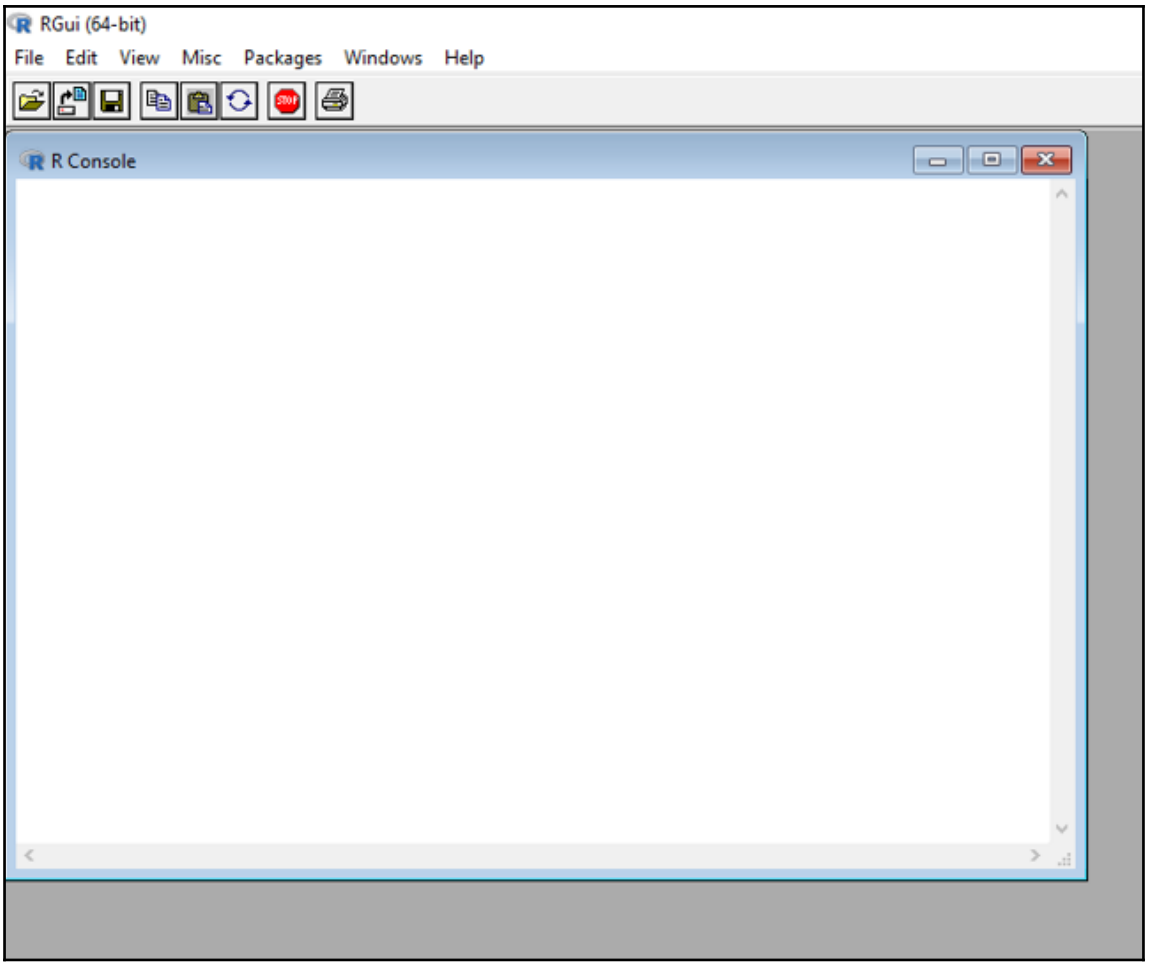
Installation and other instructions
New features in this version

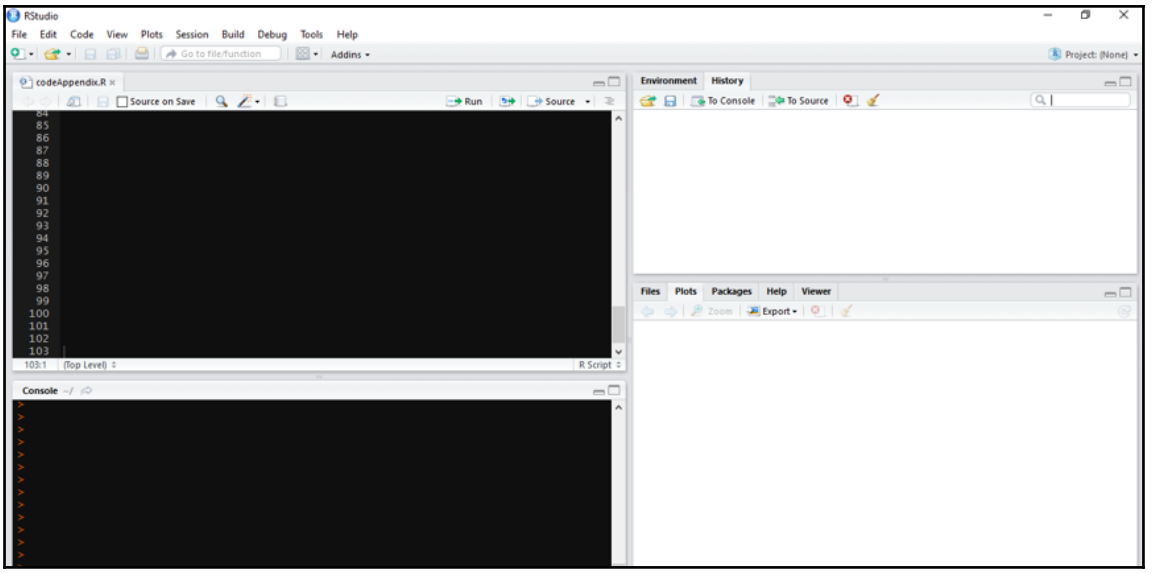
If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the md5sum of the .exe to the true fingerprint. You will need a version of md5sum for windows: both graphical and command line versions are available.

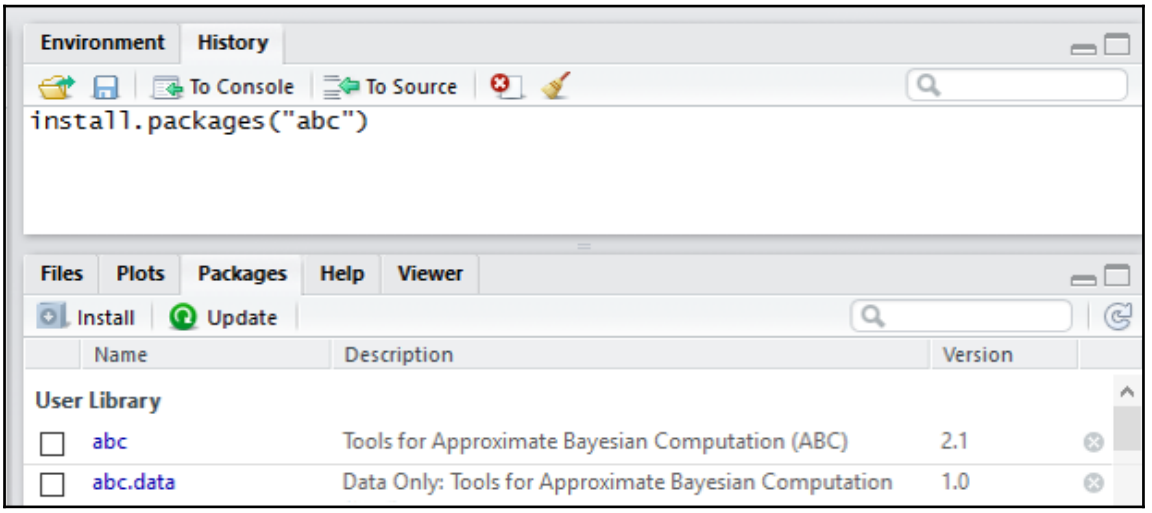
Frequently asked questions

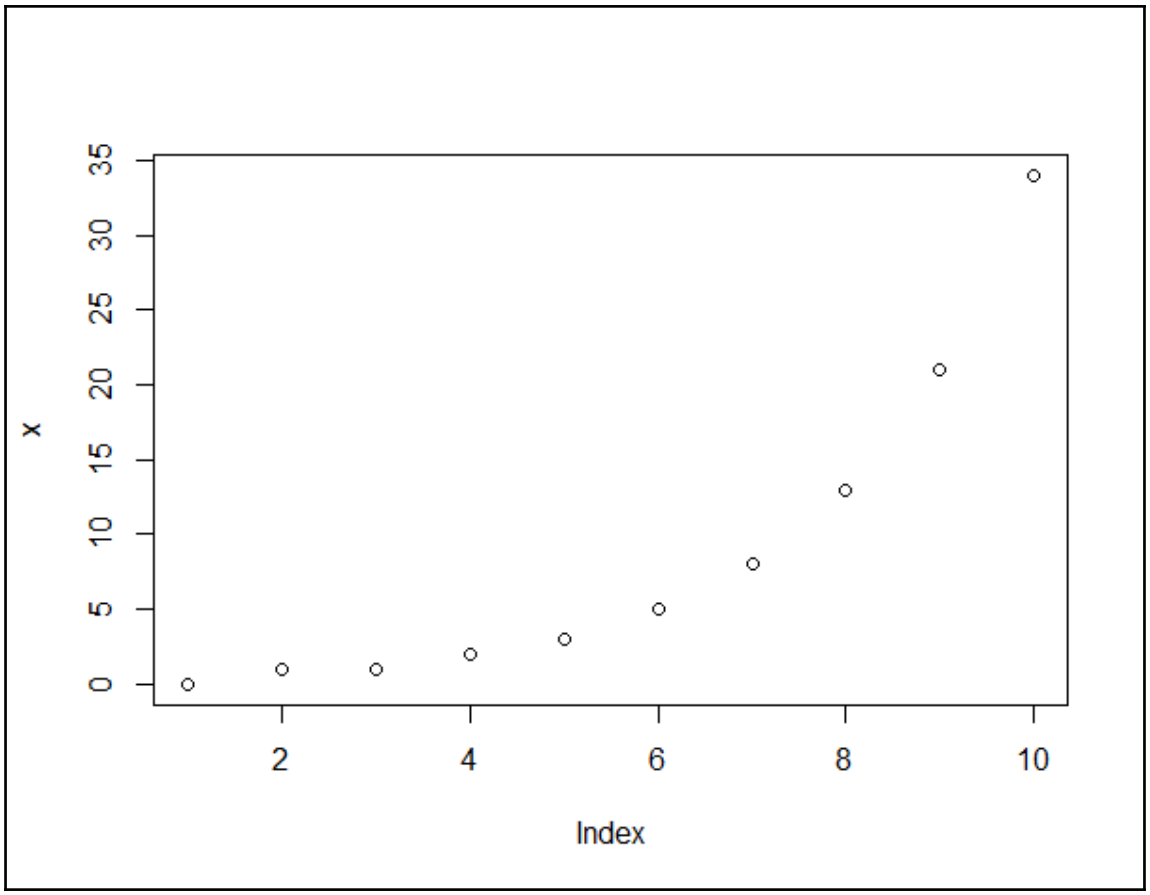
- Does R run under my version of Windows?
- How do I update packages in my previous version of R?
- Should I run 32-bit or 64-bit R?

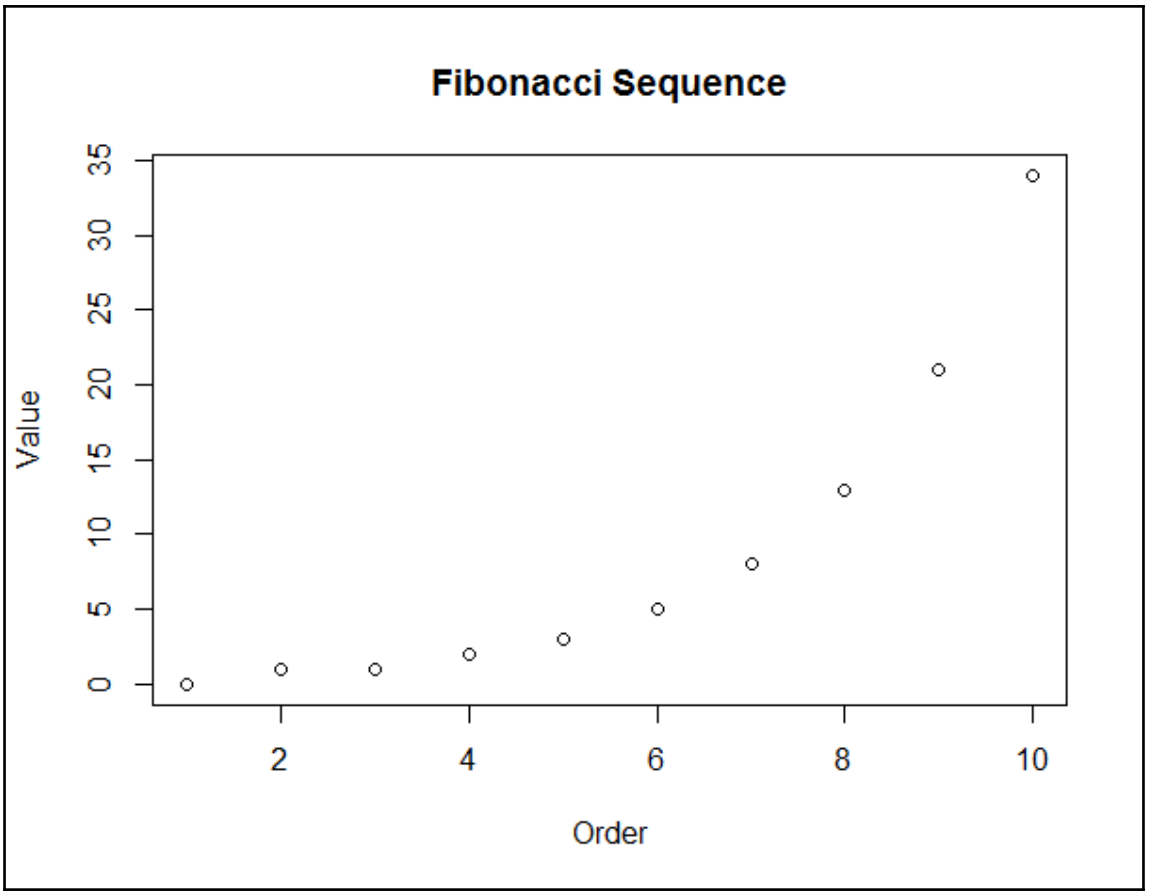
Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.

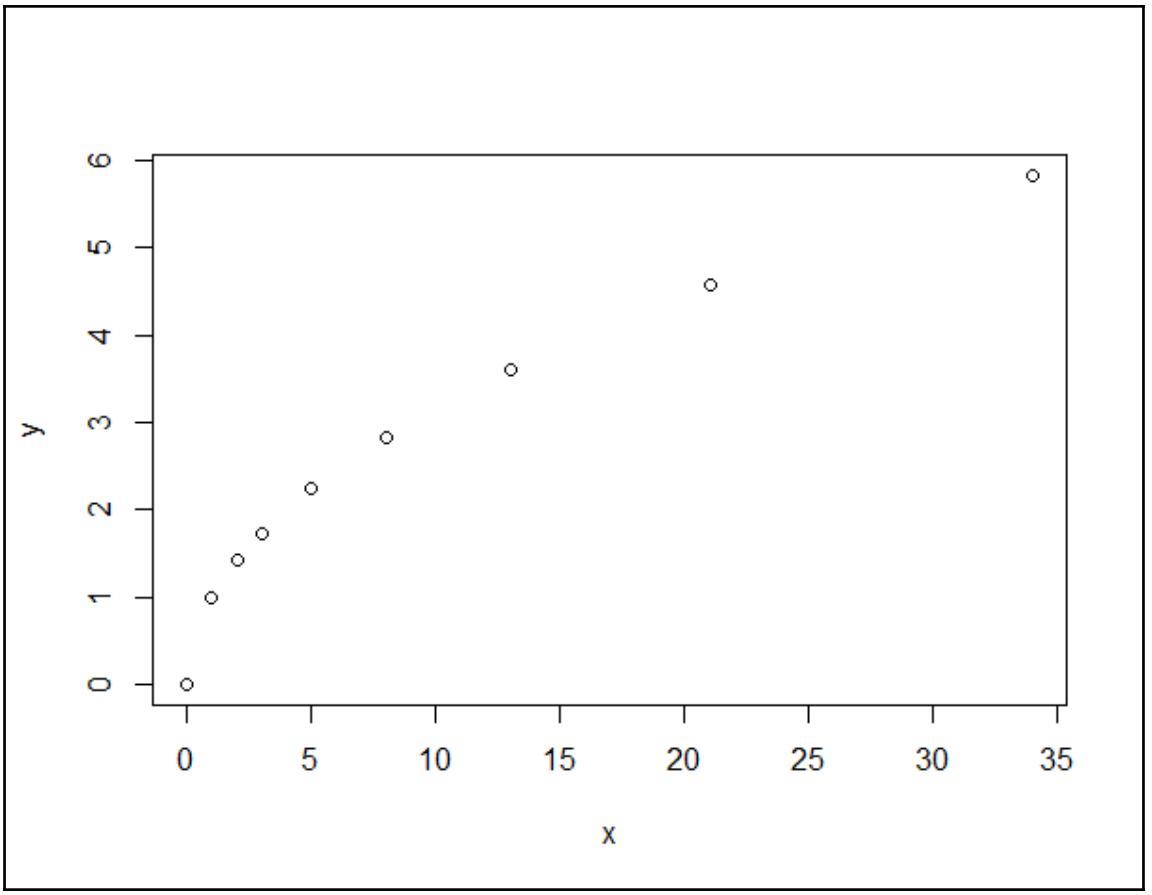

RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console

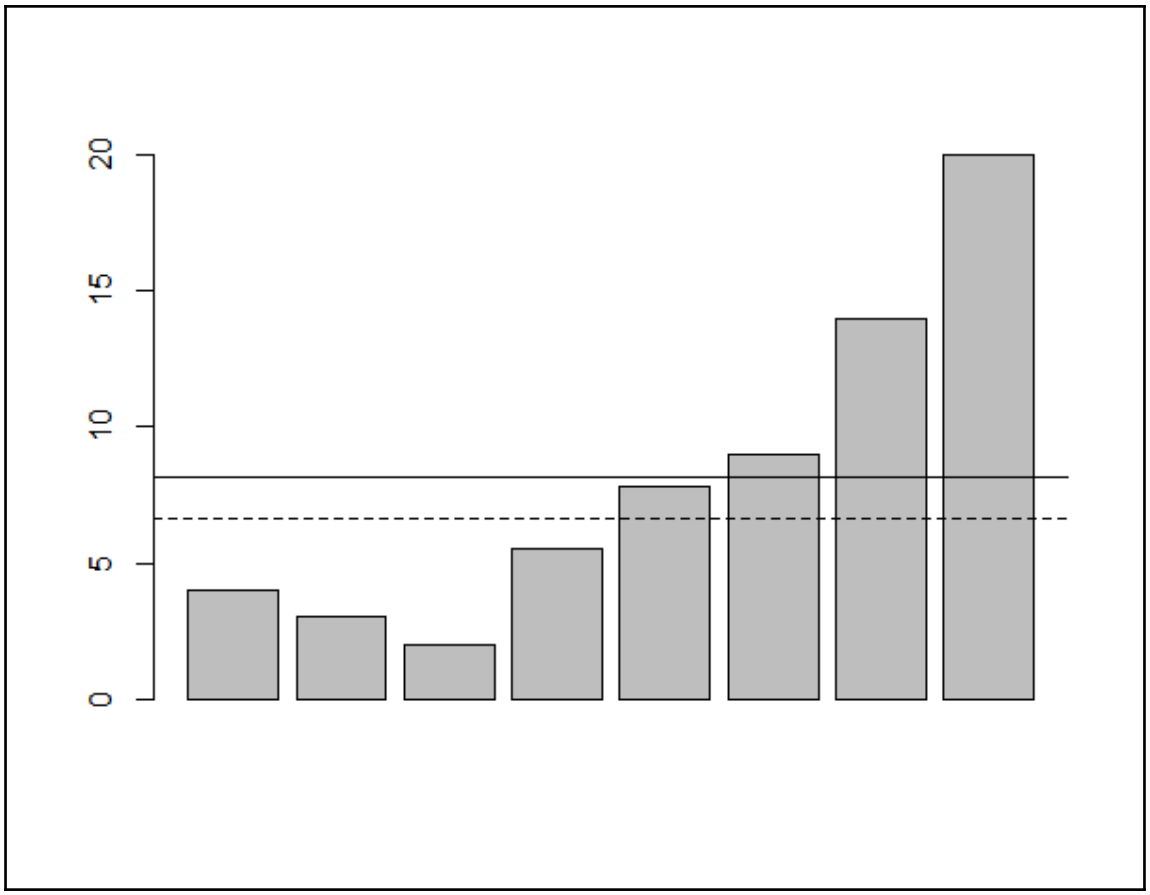

RStudio
File Edit Code View Plots Session Build Debug Tools Help
codeAppendix.R
Source on Save
Run
Source
Environment
History
To Console
To Source
Files
Plots
Packages
Help
Viewer
Export
Console

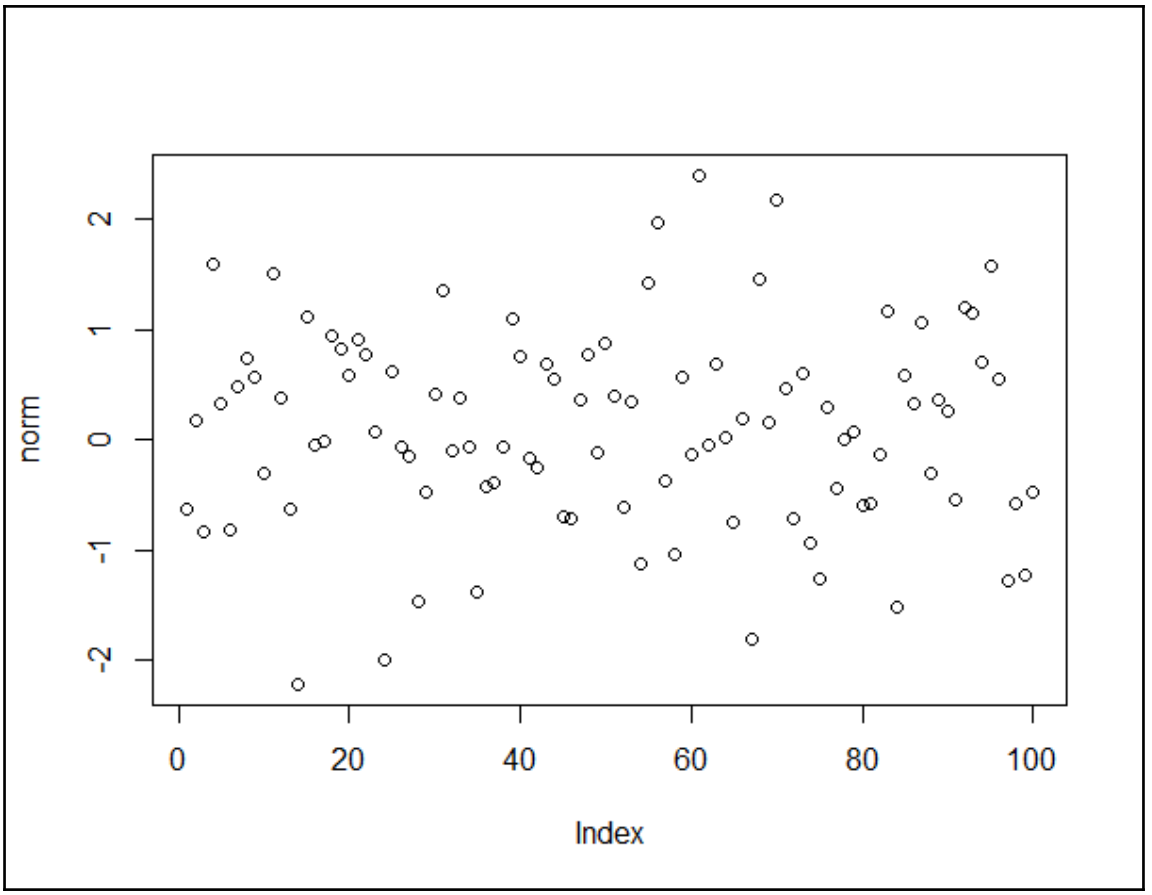

Environment
History
To Console
To Source
install.packages("abc")
Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
User Library
abc
Tools for Approximate Bayesian Computation (ABC)
2.1
abc.data
Data Only: Tools for Approximate Bayesian Computation
1.0

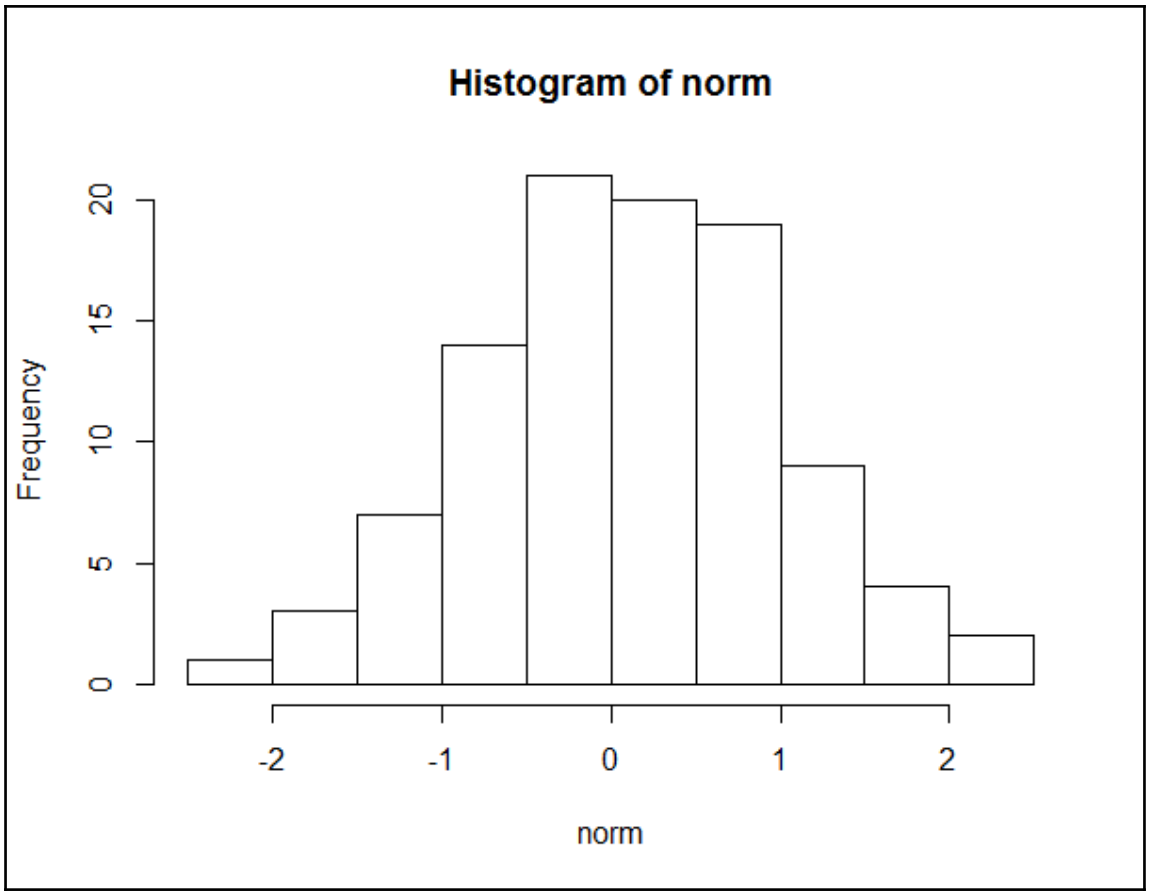

x
0
5
10
15
20
25
30
35
2
4
6
8
10
Index

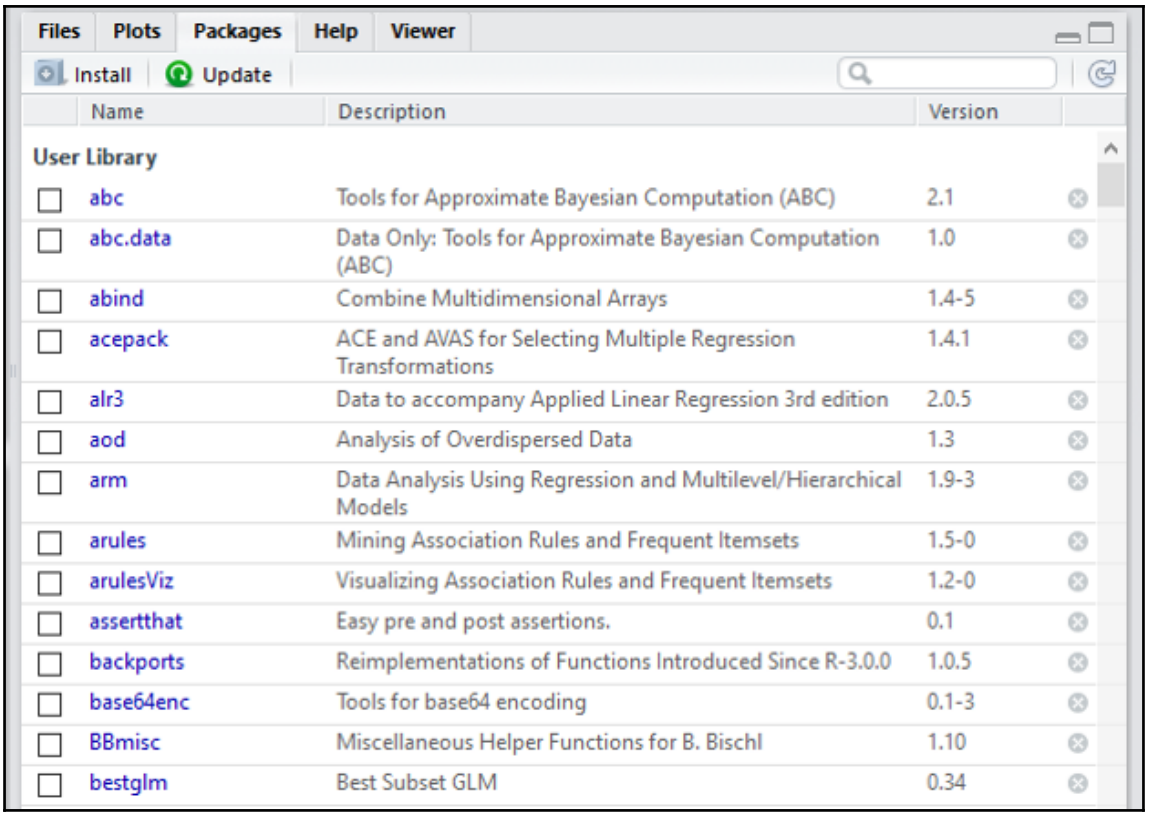

Fibonacci Sequence
Value
0
5
10
15
20
25
30
35
2
4
6
8
10
Order

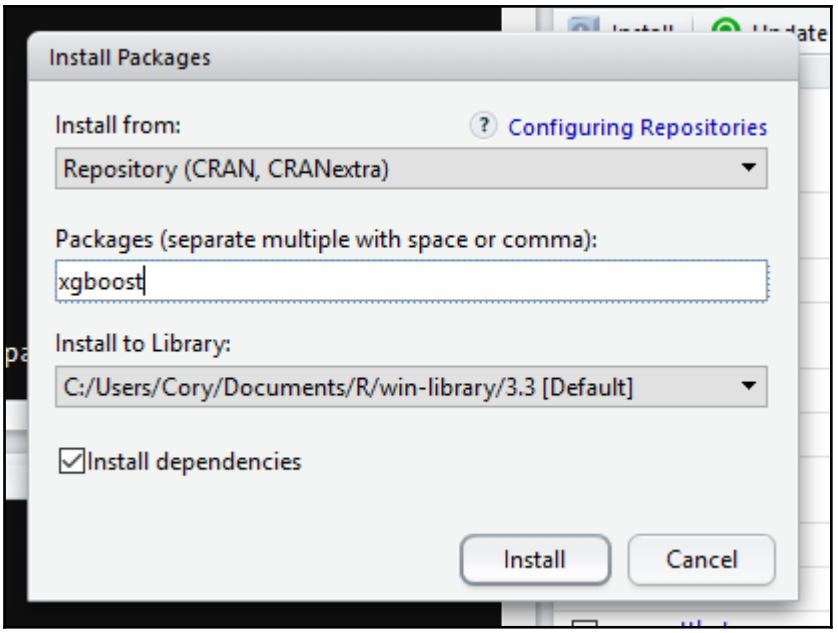

y
0
1
2
3
4
5
6
0
5
10
15
20
25
30
35
x


0
5
10
15
20


norm
2
1
0
-1
-2
0
20
40
60
80
100
Index


Histogram of norm
Frequency
0
5
10
15
20
-2
-1
0
1
2
norm


Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
User Library
abc
Tools for Approximate Bayesian Computation (ABC)
2.1
abc.data
Data Only: Tools for Approximate Bayesian Computation (ABC)
1.0
abind
Combine Multidimensional Arrays
1.4-5
acepack
ACE and AVAS for Selecting Multiple Regression Transformations
1.4.1
alr3
Data to accompany Applied Linear Regression 3rd edition
2.0.5
aod
Analysis of Overdispersed Data
1.3
arm
Data Analysis Using Regression and Multilevel/Hierarchical Models
1.9-3
arules
Mining Association Rules and Frequent Itemsets
1.5-0
arulesViz
Visualizing Association Rules and Frequent Itemsets
1.2-0
assertthat
Easy pre and post assertions.
0.1
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.0.5
base64enc
Tools for base64 encoding
0.1-3
BBmisc
Miscellaneous Helper Functions for B. Bischl
1.10
bestglm
Best Subset GLM
0.34


Install Packages
Install from:
Configuring Repositories
Repository (CRAN, CRANextra)
Packages (separate multiple with space or comma):
xgboost
Install to Library:
C:/Users/Cory/Documents/R/win-library/3.3 [Default]
Install dependencies
Install
Cancel