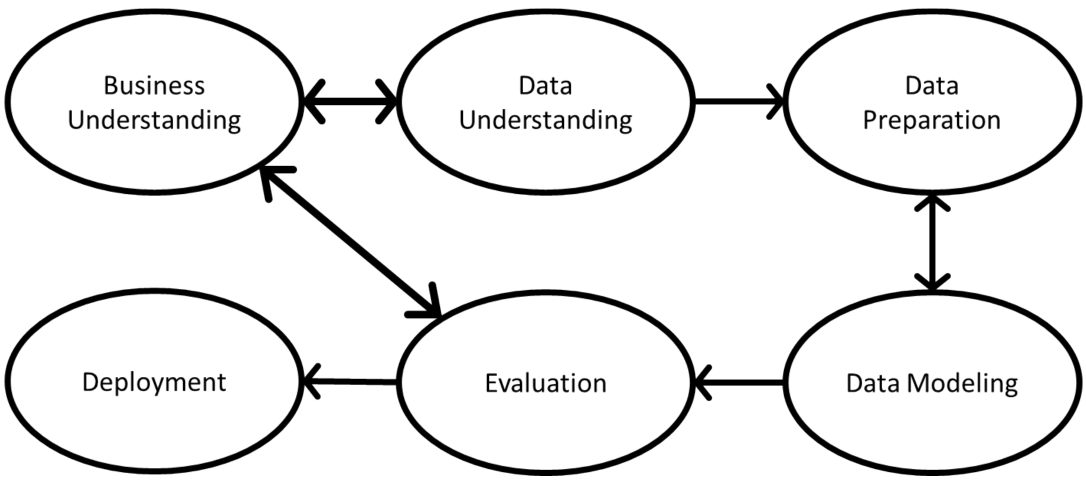
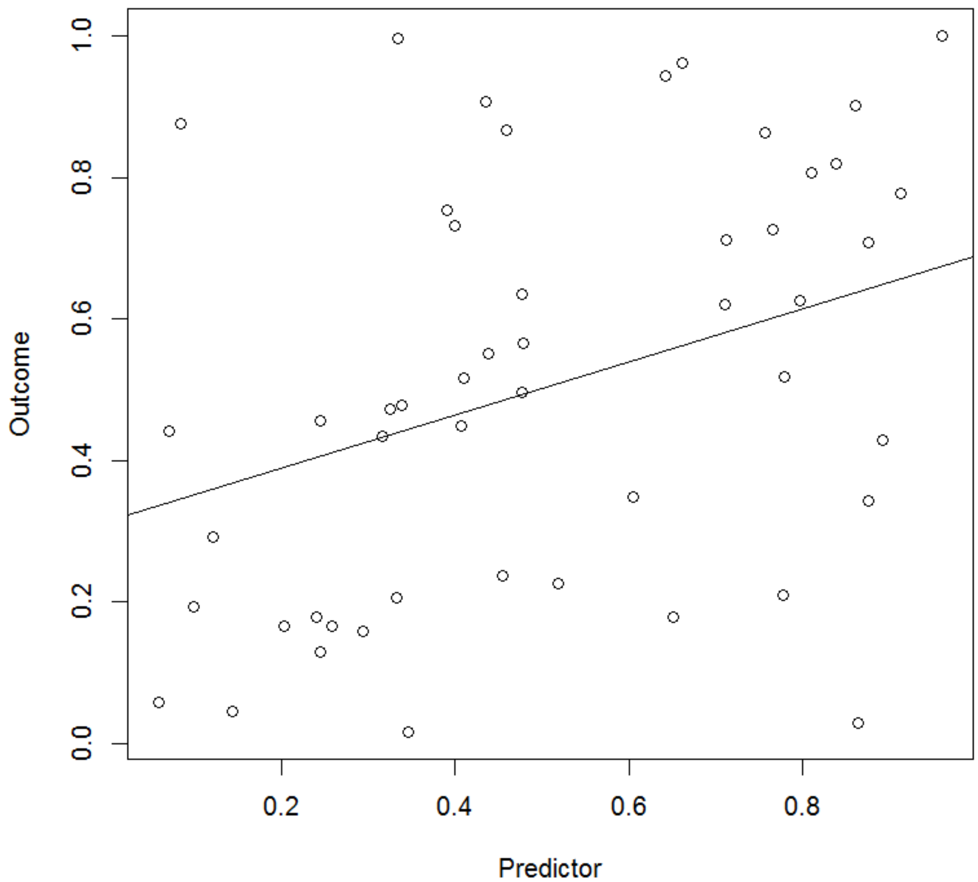
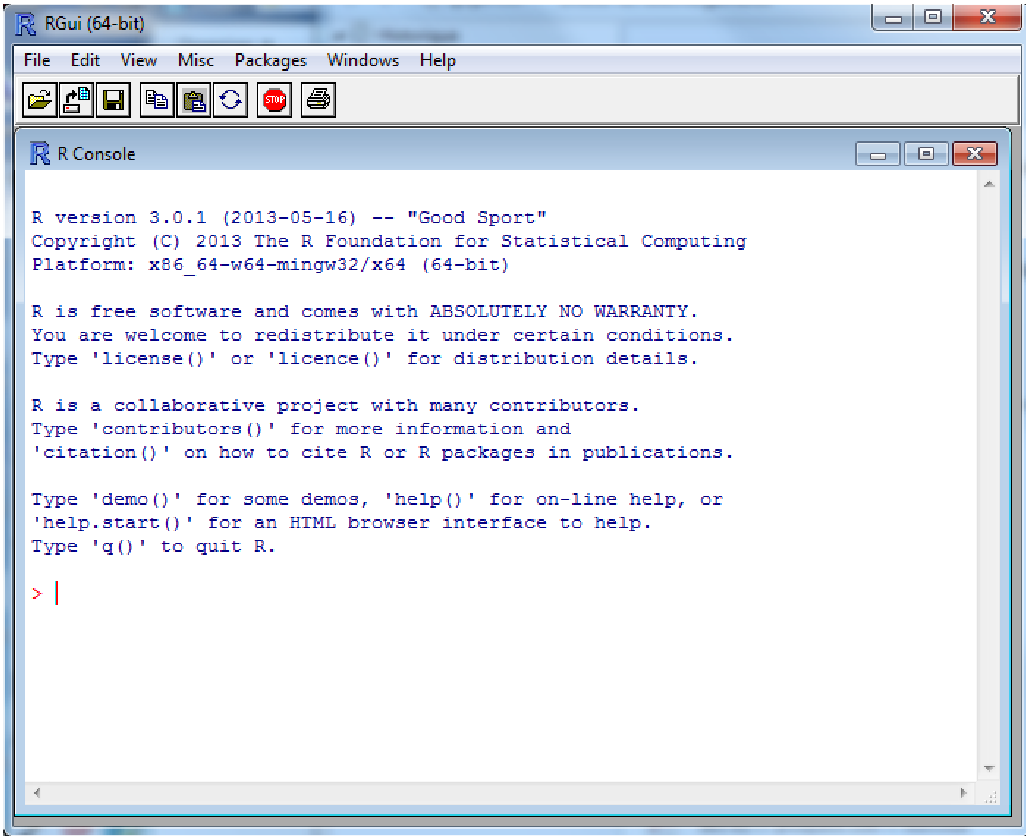
Learning Predictive Analytics with R

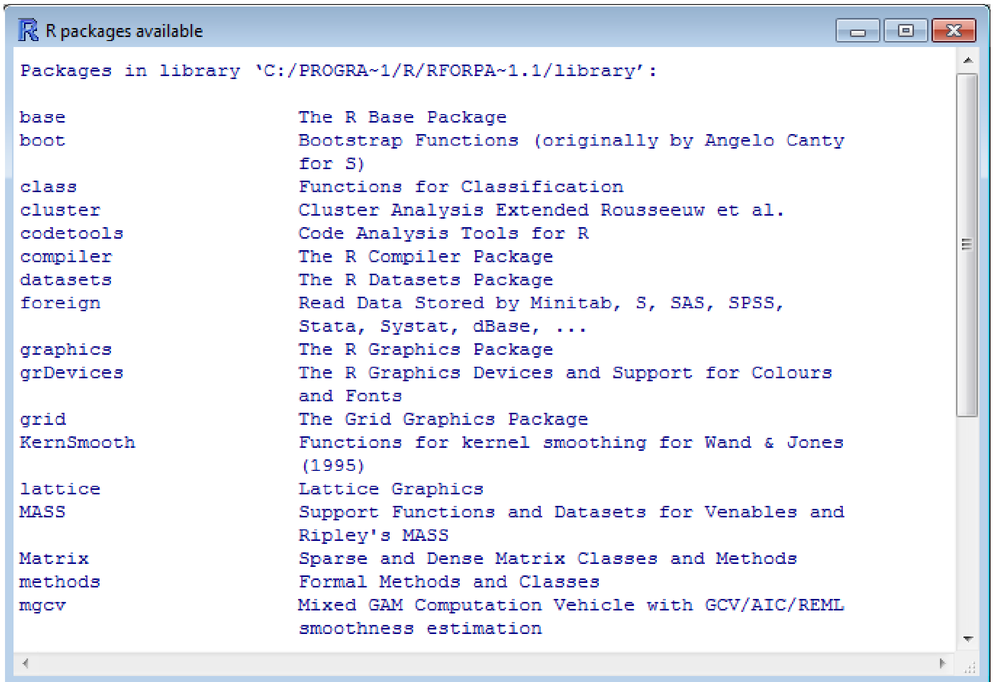
Preface:



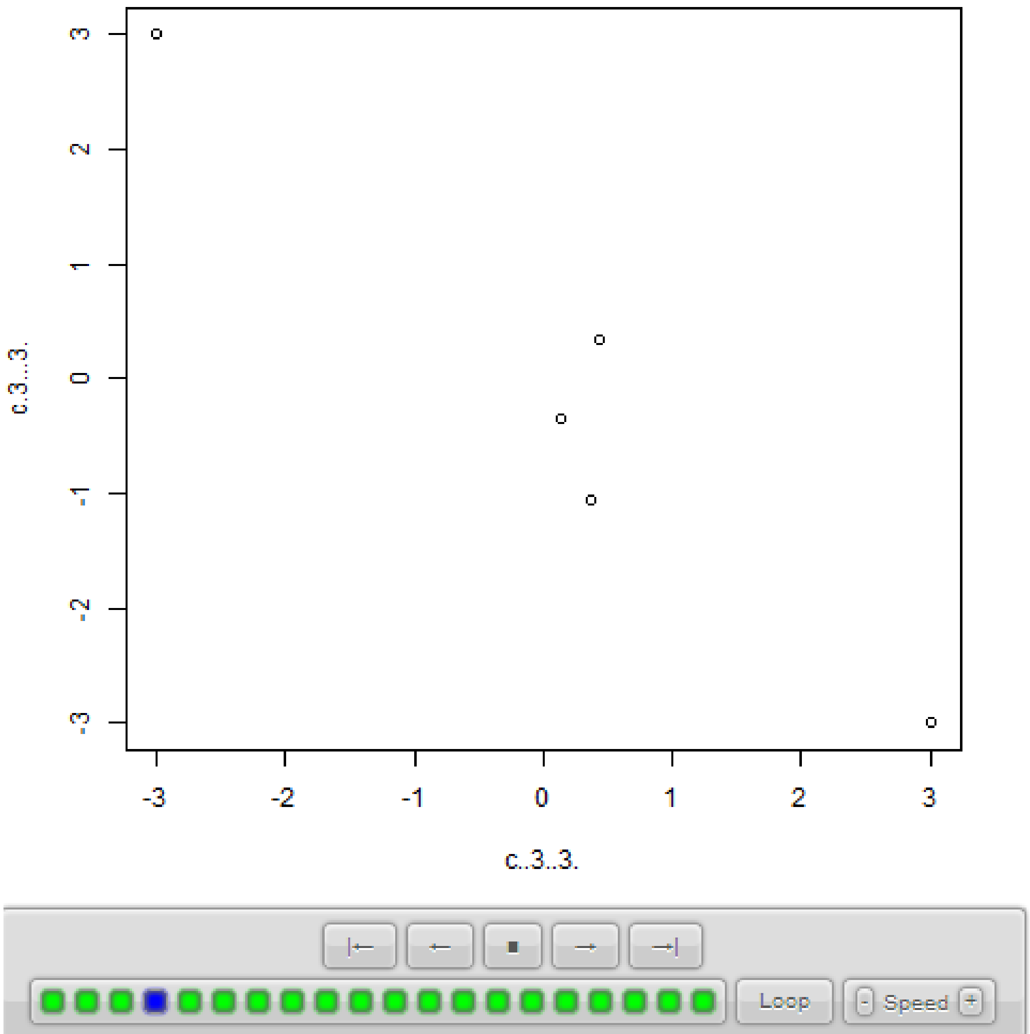


Chapter 1: Setting GNU R for Predictive Analytics


RGui (64-bit)
File Edit View Misc Packages Windows Help
R Console
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
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.
>

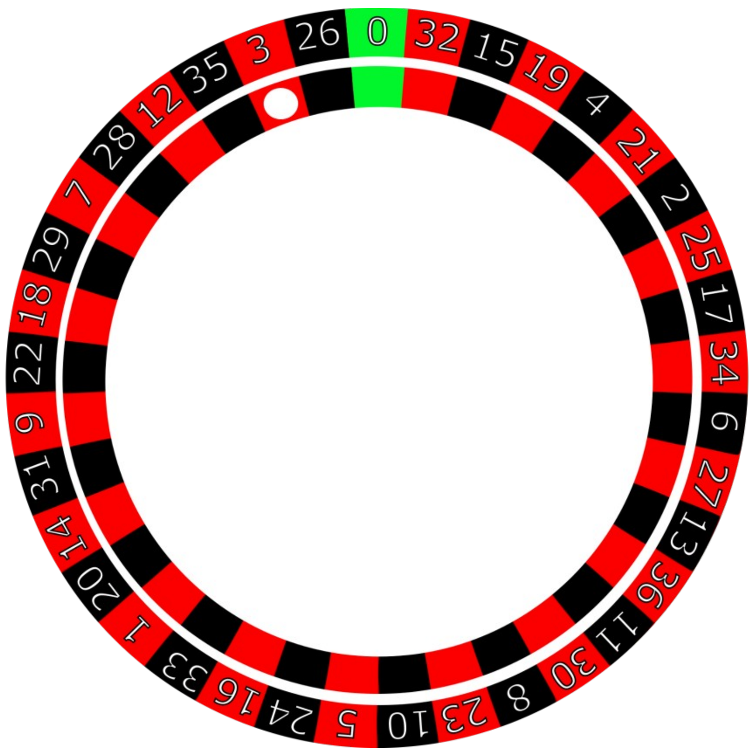


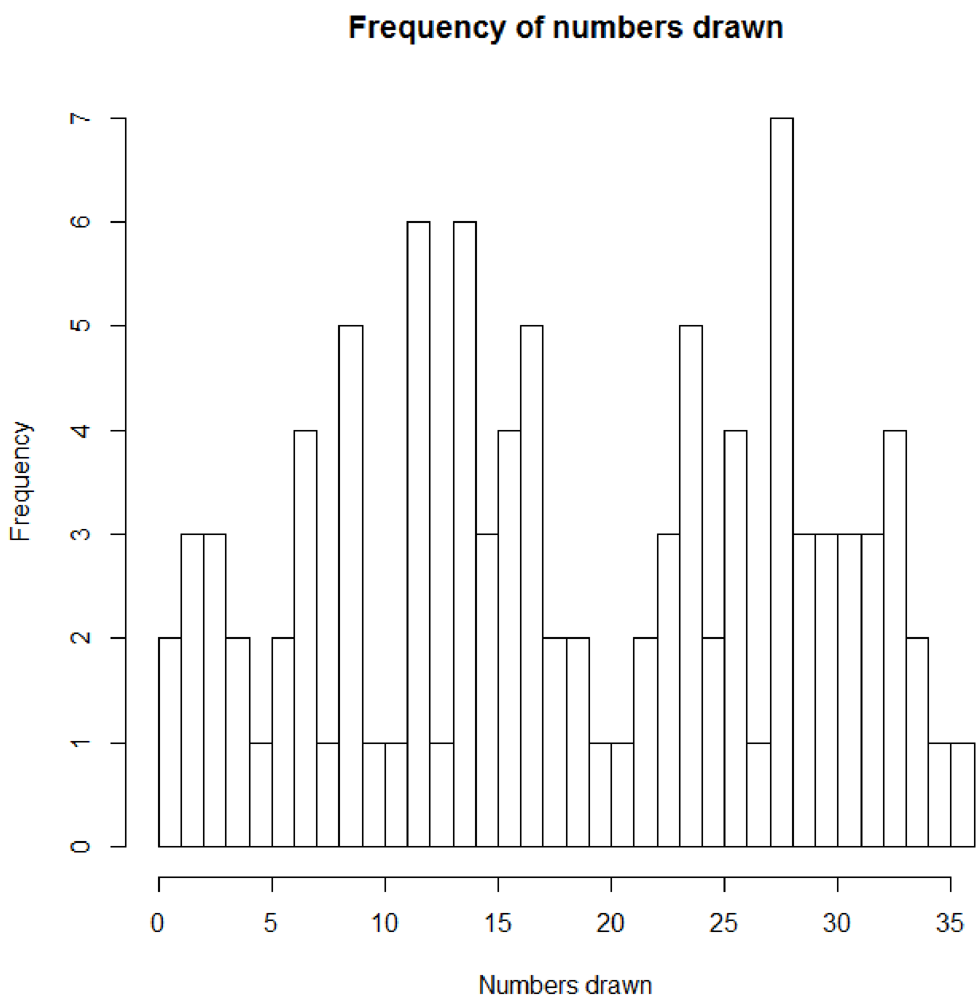

```
Packages in library 'C:/PROGRA~1/R/RFORPA~1.1/library':

base                    The R Base Package
boot                    Bootstrap Functions (originally by Angelo Canty
                        for S)
class                   Functions for Classification
cluster                 Cluster Analysis Extended Rousseeuw et al.
codetools               Code Analysis Tools for R
compiler                The R Compiler Package
datasets                The R Datasets Package
foreign                 Read Data Stored by Minitab, S, SAS, SPSS,
                        Stata, Systat, dBase, ...
graphics                The R Graphics Package
grDevices               The R Graphics Devices and Support for Colours
                        and Fonts
grid                    The Grid Graphics Package
KernSmooth              Functions for kernel smoothing for Wand & Jones
                        (1995)
lattice                 Lattice Graphics
MASS                    Support Functions and Datasets for Venables and
                        Ripley's MASS
Matrix                  Sparse and Dense Matrix Classes and Methods
methods                 Formal Methods and Classes
mgcv                    Mixed GAM Computation Vehicle with GCV/AIC/REML
                        smoothness estimation
```

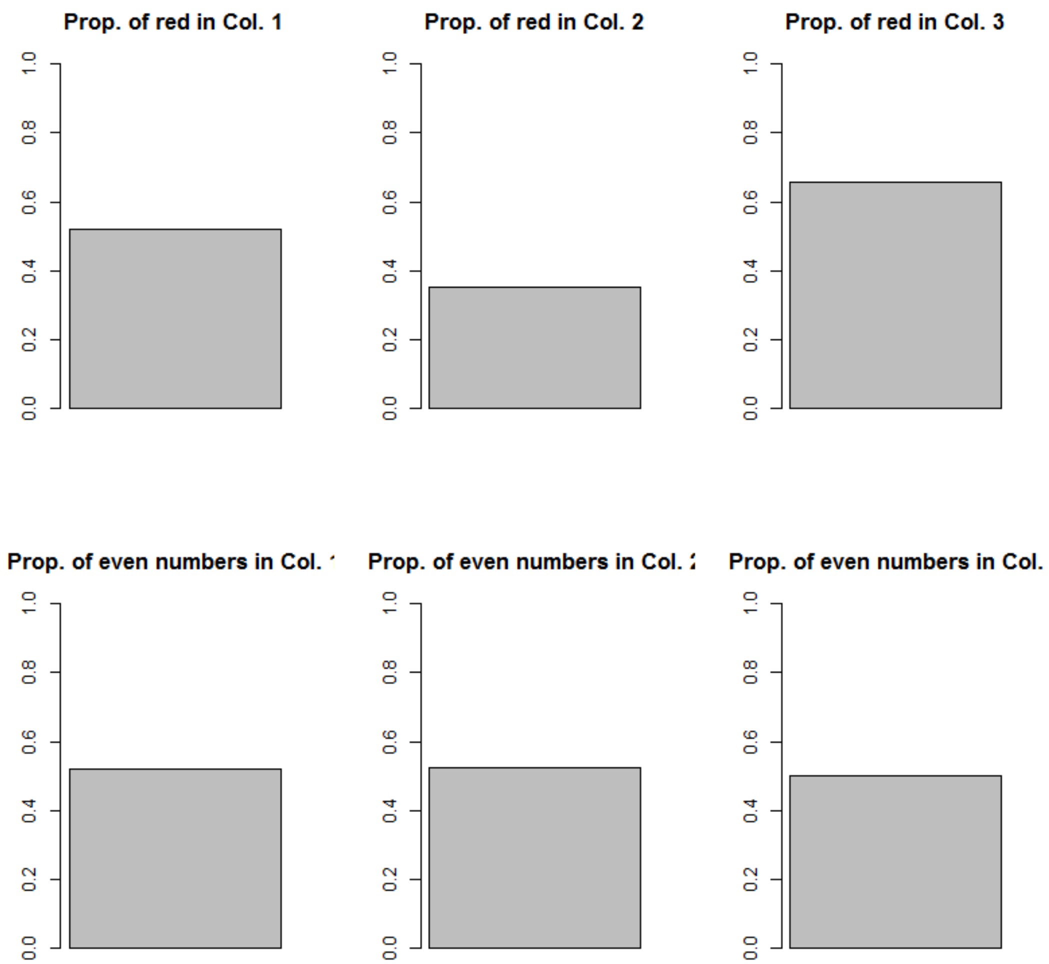


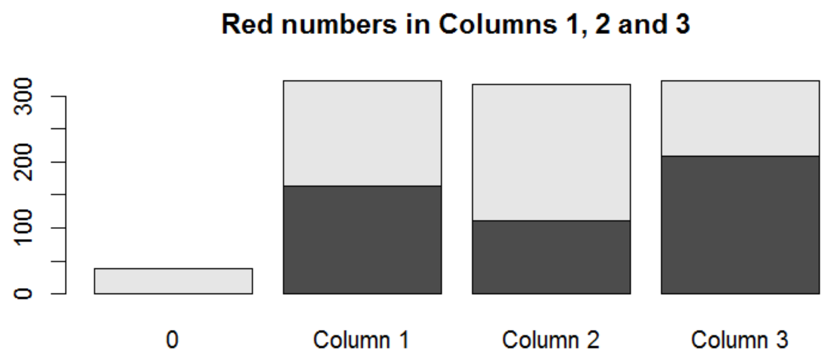

```r
## Testing the animation package for the first time.
library(animation)
library(foreign)
for (i in 1:20) {
     plot(df)
     df = rbind(df, c(rnorm(1), rnorm(1)))
}
## R version 3.0.1 (2013-05-16)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Other packages: animation 2.2, foreign 0.8-53
```

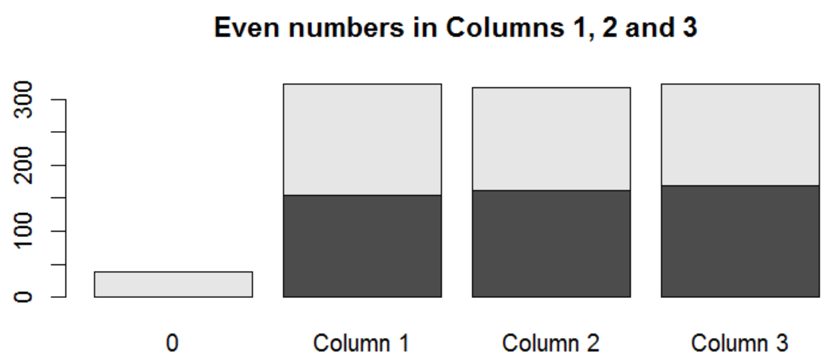
Chapter 2: Visualizing and Manipulating Data Using R

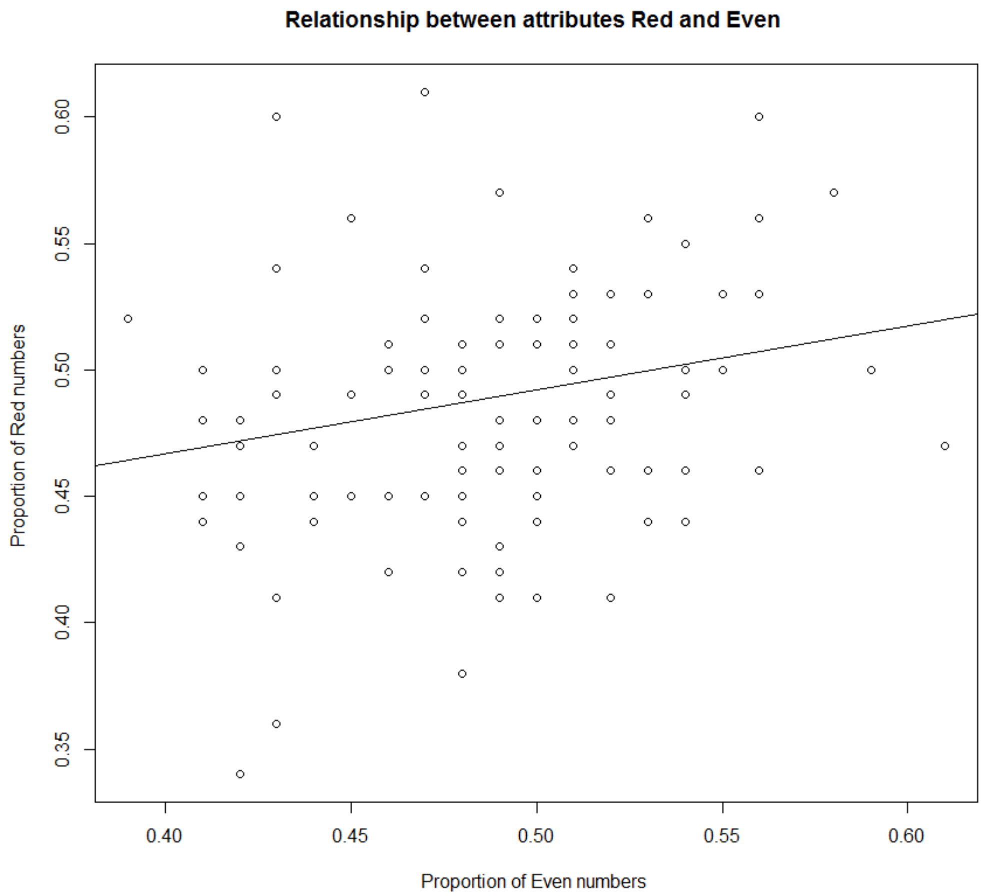



Frequency of numbers drawn
Frequency
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
Numbers drawn

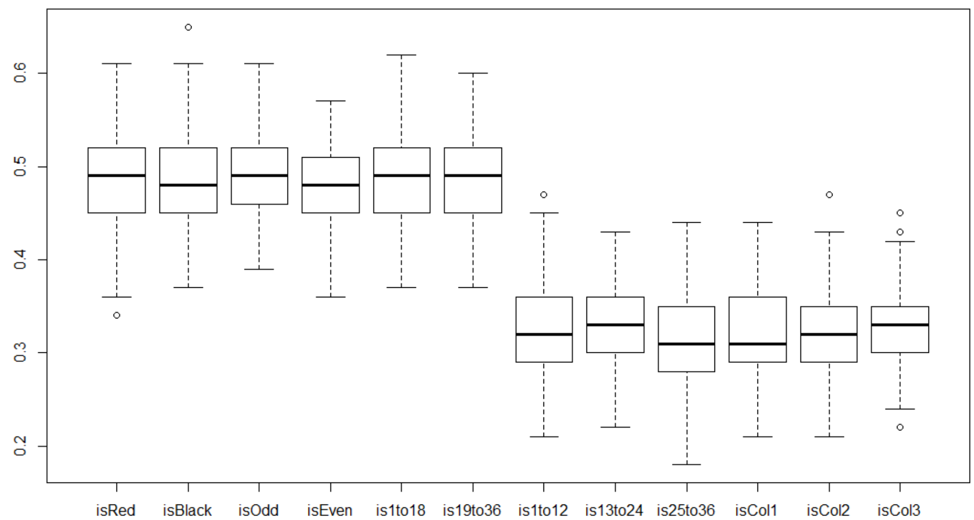

Prop. of red in Col. 1
Prop. of red in Col. 2
Prop. of red in Col. 3
Prop. of even numbers in Col.
Prop. of even numbers in Col.
Prop. of even numbers in Col.
0.0
0.2
0.4
0.6
0.8
1.0

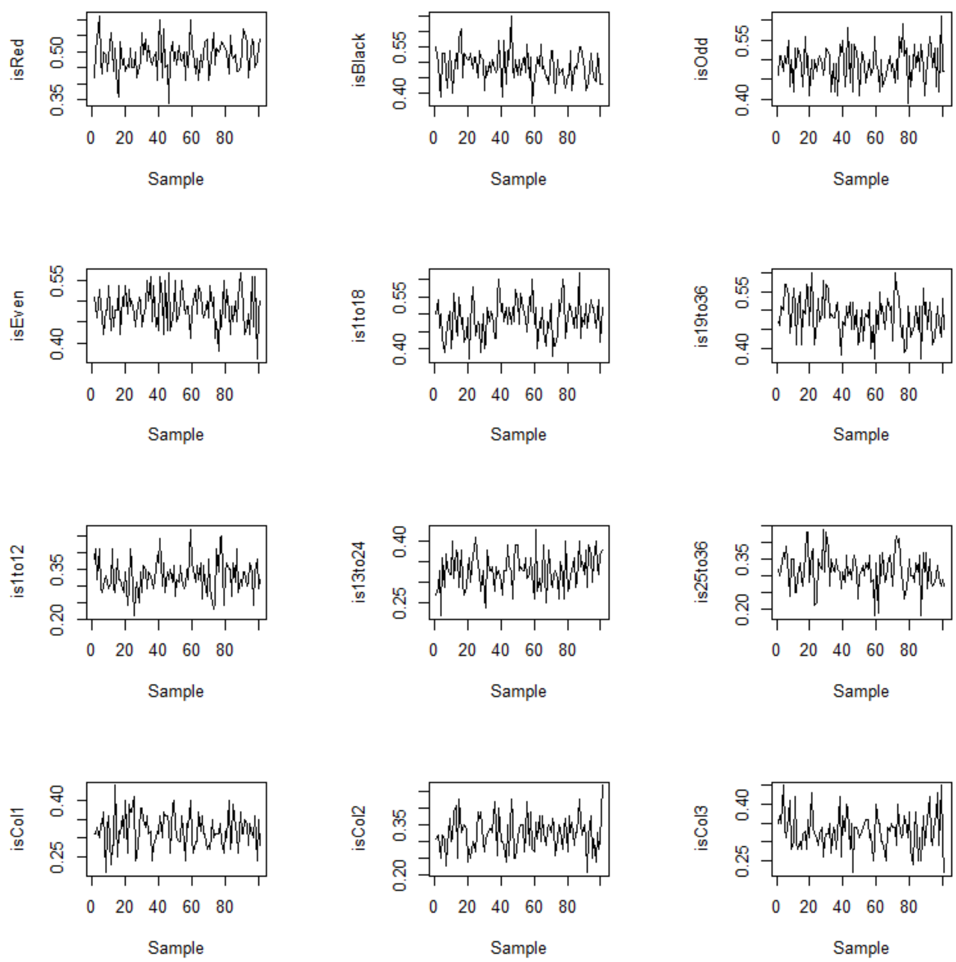

Red numbers in Columns 1, 2 and 3
0
100
200
300
0
Column 1
Column 2
Column 3


Even numbers in Columns 1, 2 and 3
0
100
200
300
0
Column 1
Column 2
Column 3

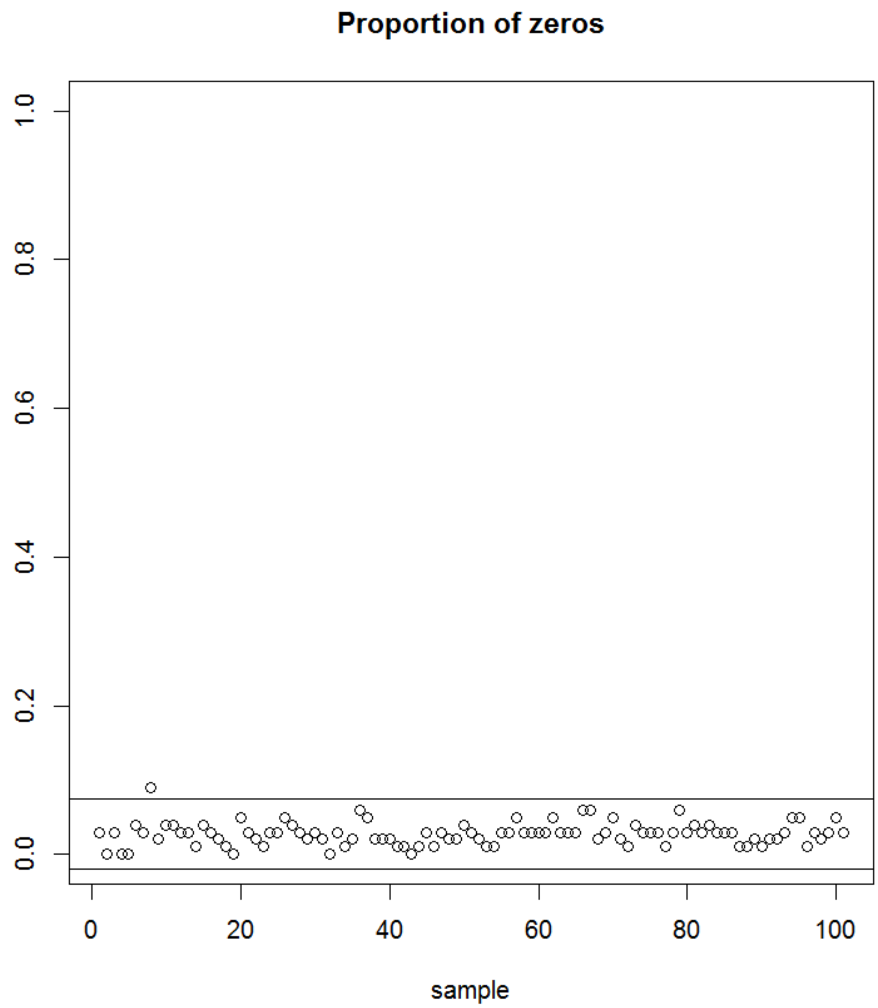


Relationship between attributes Red and Even
Proportion of Red numbers
0.35
0.40
0.45
0.50
0.55
0.60
Proportion of Even numbers
0.40
0.45
0.50
0.55
0.60


0.6
0.5
0.4
0.3
0.2
isRed
isBlack
isOdd
isEven
is1to18
is19to36
is1to12
is13to24
is25to36
isCol1
isCol2
isCol3

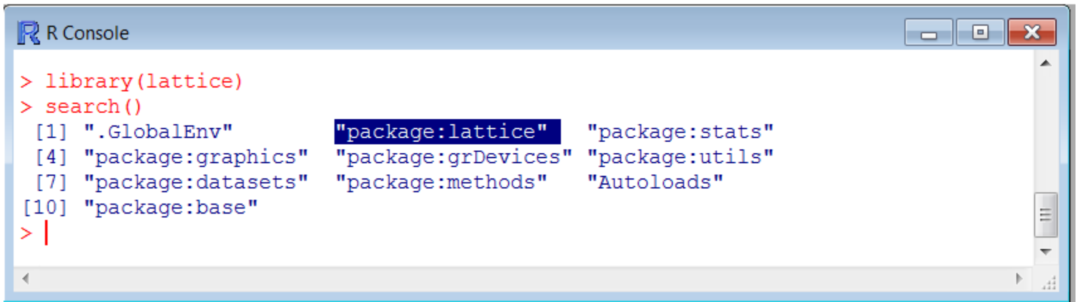


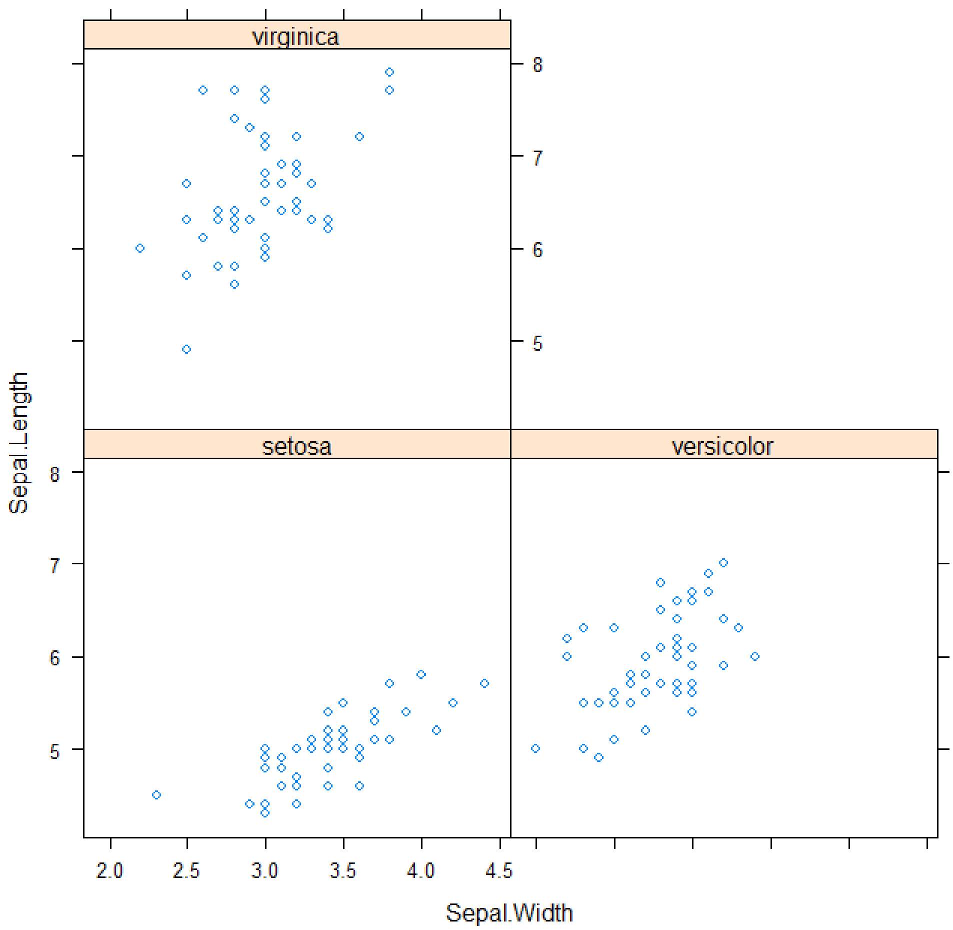
isRed
0.35
0.50
isBlack
0.40
0.55
isOdd
0.40
0.55
isEven
0.40
0.55
is1to18
0.40
0.55
is19to36
0.40
0.55
is1to12
0.20
0.35
is13to24
0.25
0.40
is25to36
0.20
0.35
isCol1
0.25
0.40
isCol2
0.20
0.35
isCol3
0.25
0.40
0 20 40 60 80
Sample

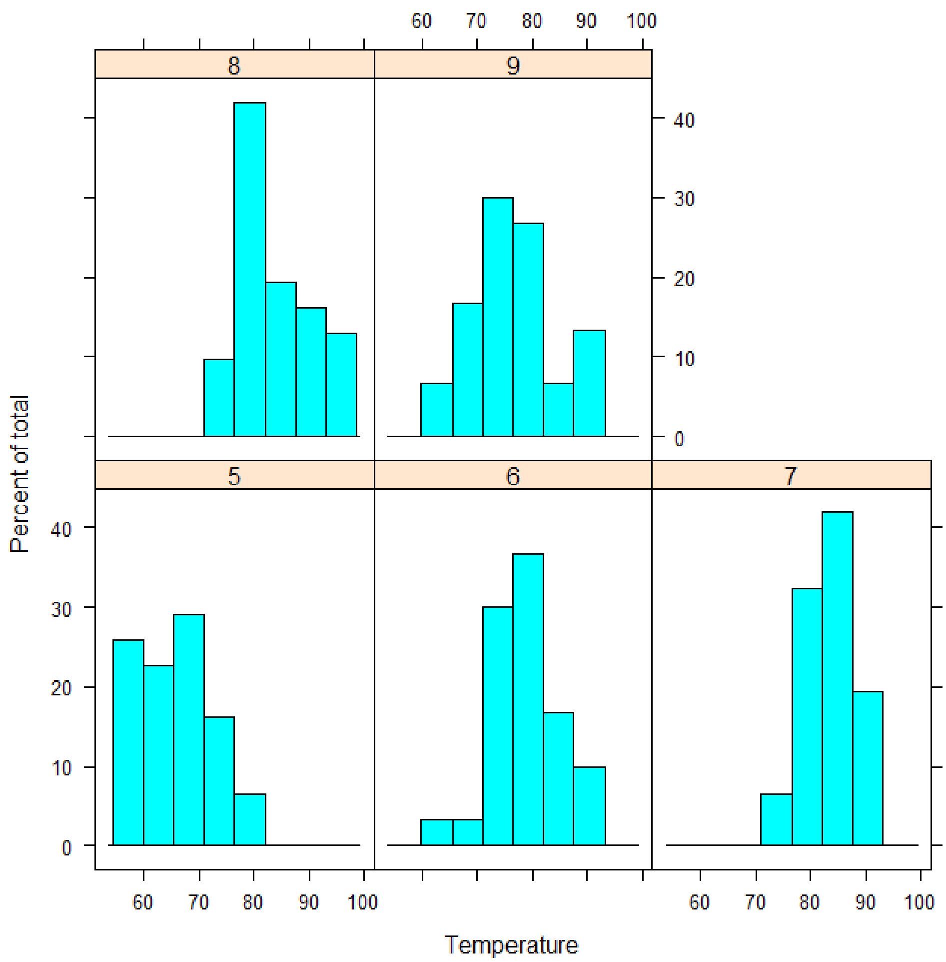

Proportion of zeros
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
sample

<table>
<tr><td></td><td></td><td colspan="3">0</td></tr>
<tr><td rowspan="2">1-18</td><td rowspan="4">1st 12</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>4</td><td>5</td><td>6</td></tr>
<tr><td rowspan="2">EVEN</td><td>7</td><td>8</td><td>9</td></tr>
<tr><td>10</td><td>11</td><td>12</td></tr>
<tr><td rowspan="2">RED</td><td rowspan="4">2nd 12</td><td>13</td><td>14</td><td>15</td></tr>
<tr><td>16</td><td>17</td><td>18</td></tr>
<tr><td rowspan="2">BLACK</td><td>19</td><td>20</td><td>21</td></tr>
<tr><td>22</td><td>23</td><td>24</td></tr>
<tr><td rowspan="2">ODD</td><td rowspan="4">3rd 12</td><td>25</td><td>26</td><td>27</td></tr>
<tr><td>28</td><td>29</td><td>30</td></tr>
<tr><td rowspan="2">19-36</td><td>31</td><td>32</td><td>33</td></tr>
<tr><td>34</td><td>35</td><td>36</td></tr>
<tr><td></td><td></td><td>2-1</td><td>2-1</td><td>2-1</td></tr>
</table>

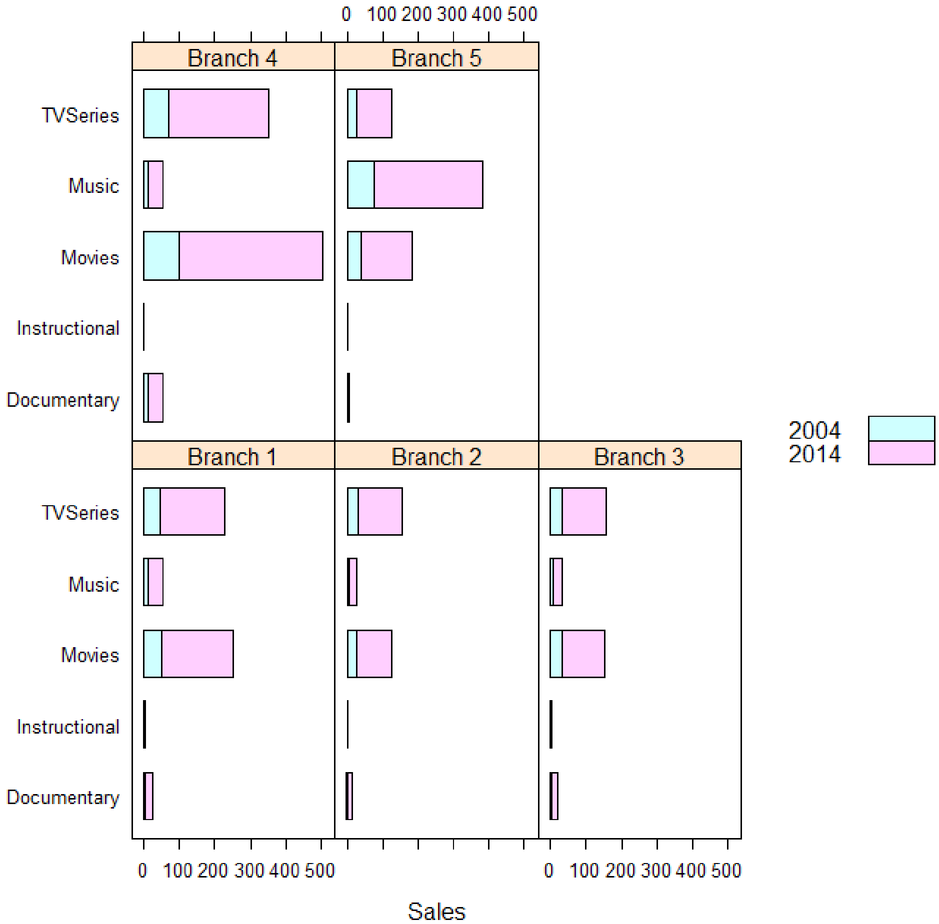
Chapter 3: Data Visualization with Lattice


R Console
> library(lattice)
> search()
[1] ".GlobalEnv" "package:lattice" "package:stats"
[4] "package:graphics" "package:grDevices" "package:utils"
[7] "package:datasets" "package:methods" "Autoloads"
[10] "package:base"
>

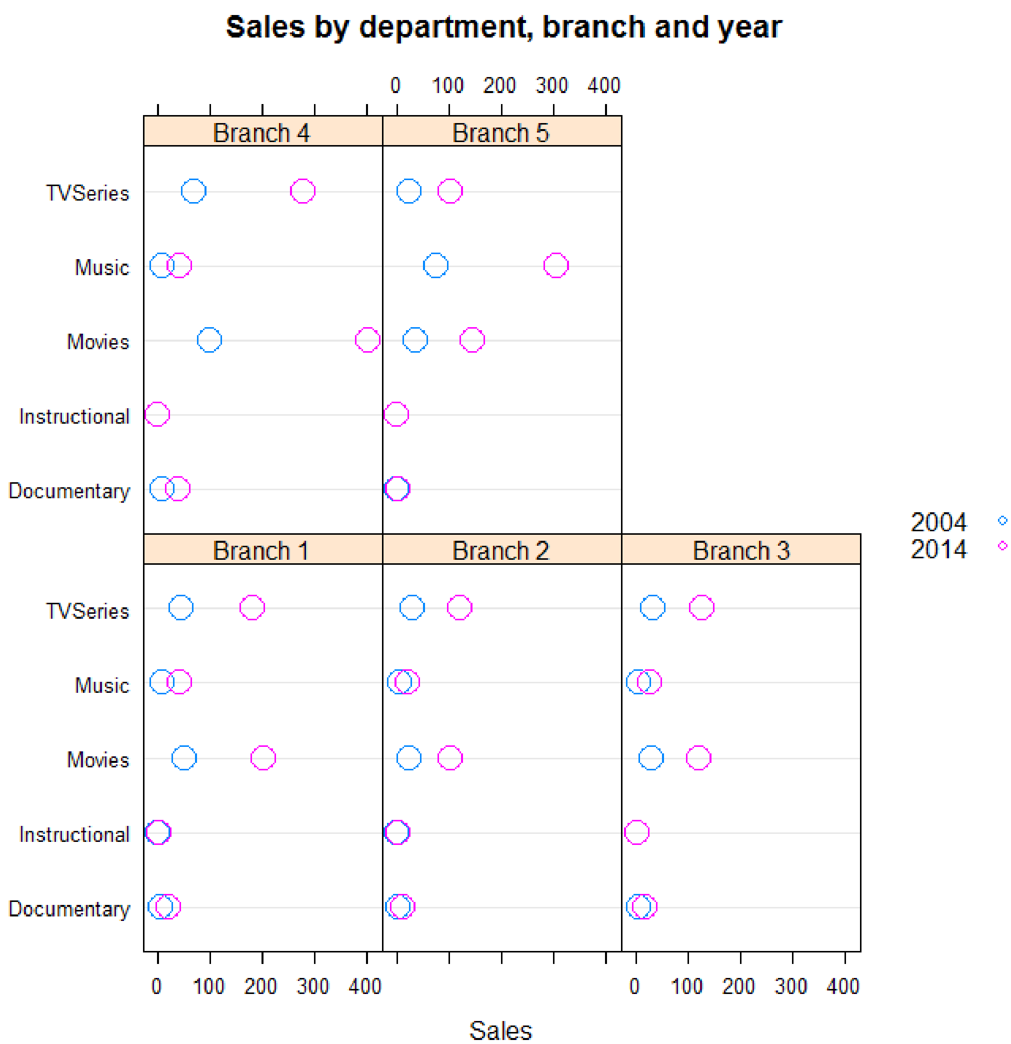

virginica
setosa
versicolor
Sepal.Length
Sepal.Width
8
7
6
5
2.0
2.5
3.0
3.5
4.0
4.5



8
9
5
6
7
Percent of total
Temperature
0
10
20
30
40
60
70
80
90
100

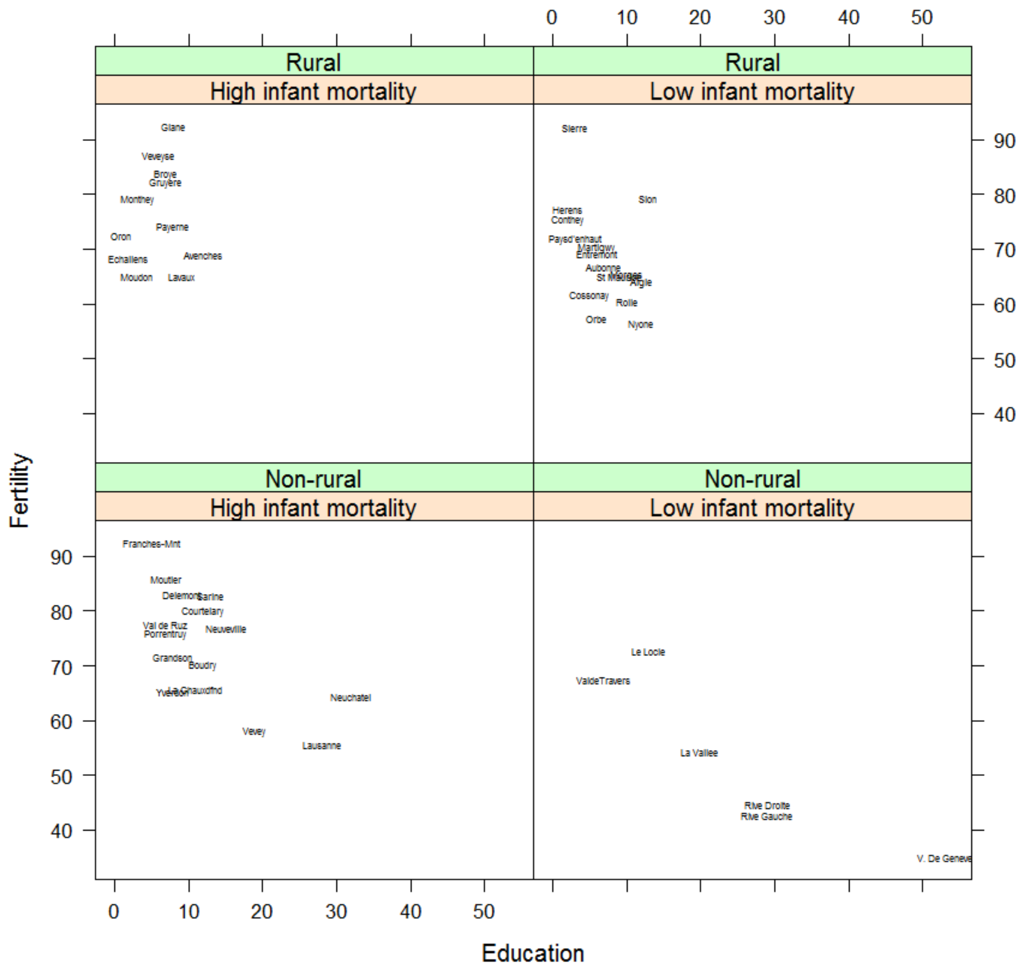


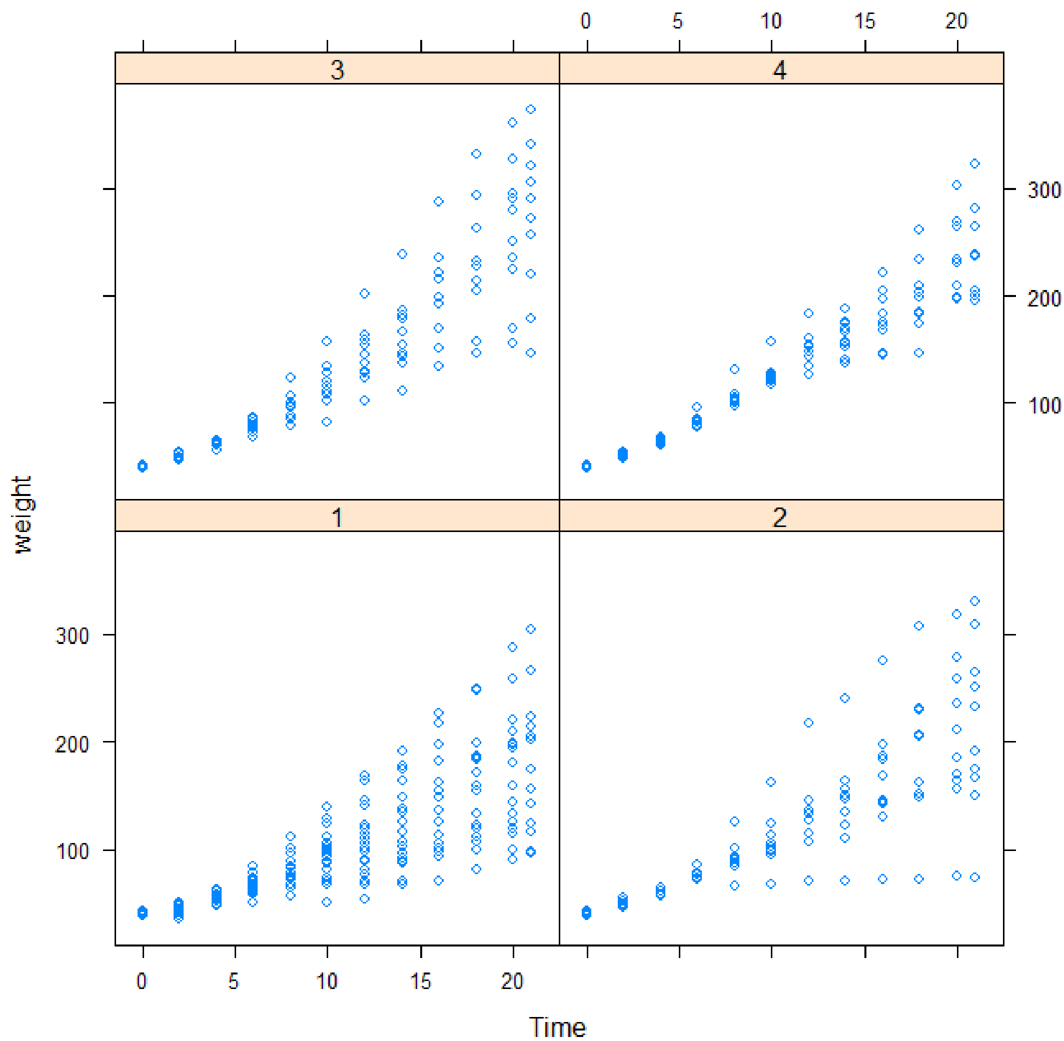
Branch 4
Branch 5
Branch 1
Branch 2
Branch 3
TVSeries
Music
Movies
Instructional
Documentary
0 100 200 300 400 500
2004
2014
Sales



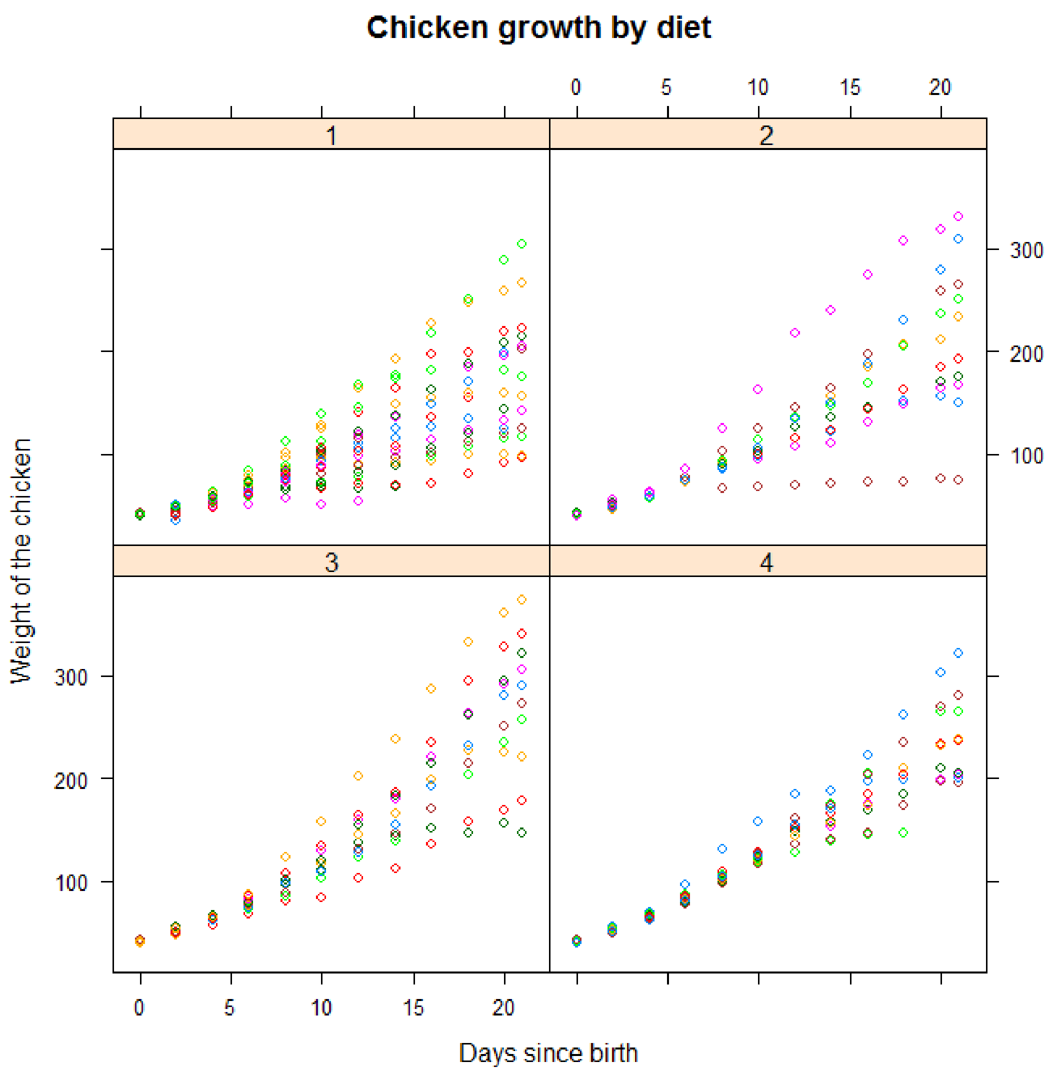
Sales by department, branch and year
Branch 4
Branch 5
Branch 1
Branch 2
Branch 3
TVSeries
Music
Movies
Instructional
Documentary
0
100
200
300
400
2004
2014
Sales

Fertility and education in 1888 Occidental Switzerland



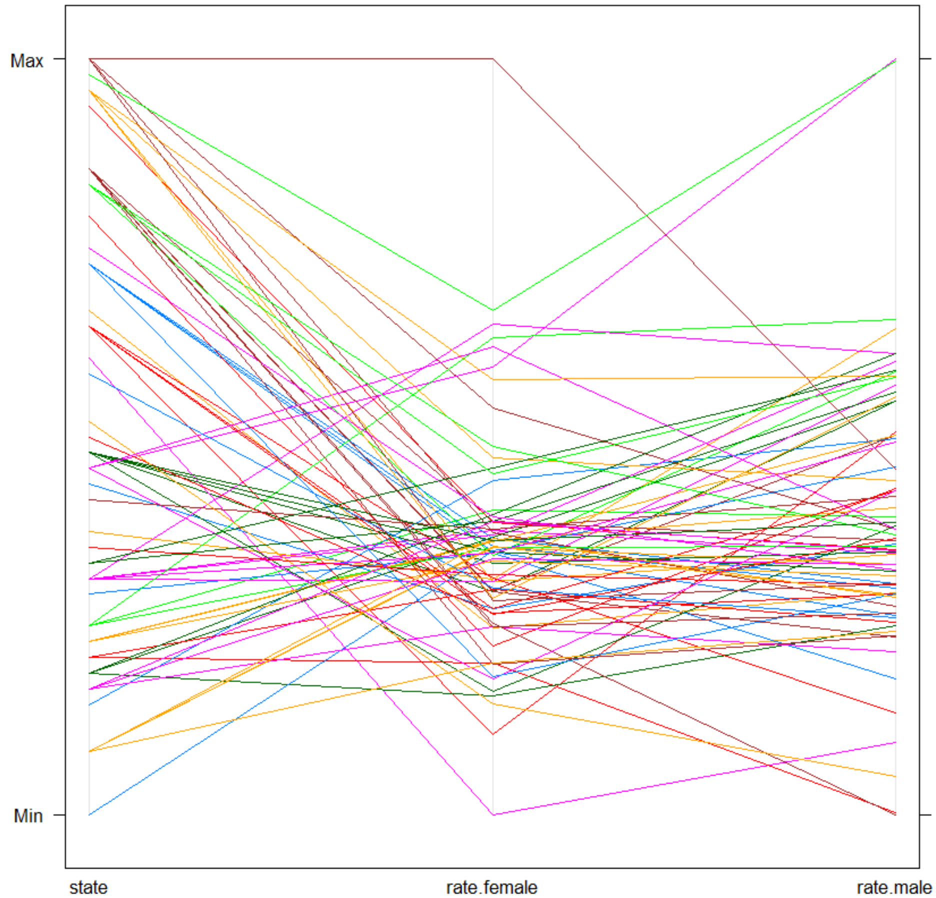


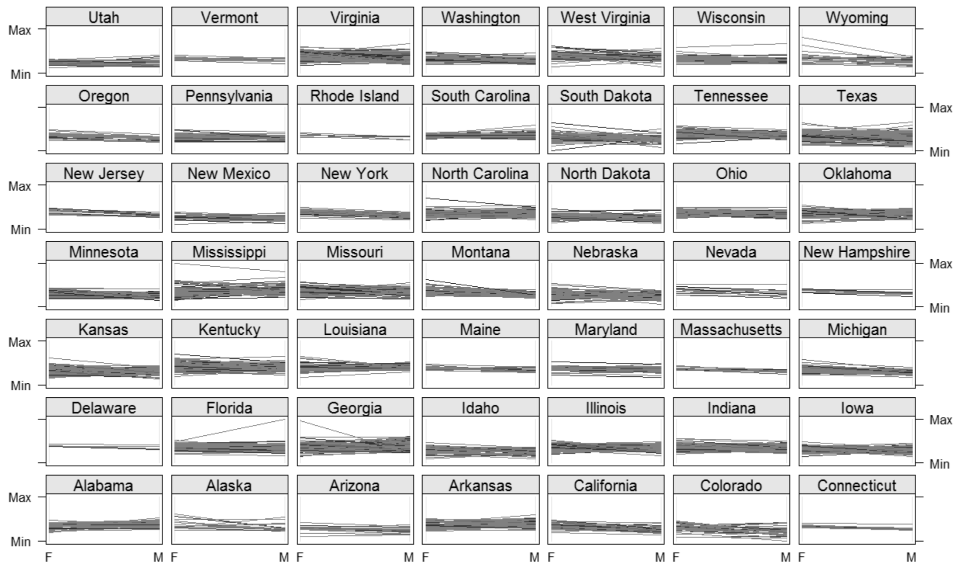
3
4
1
2
weight
Time
0
5
10
15
20
100
200
300

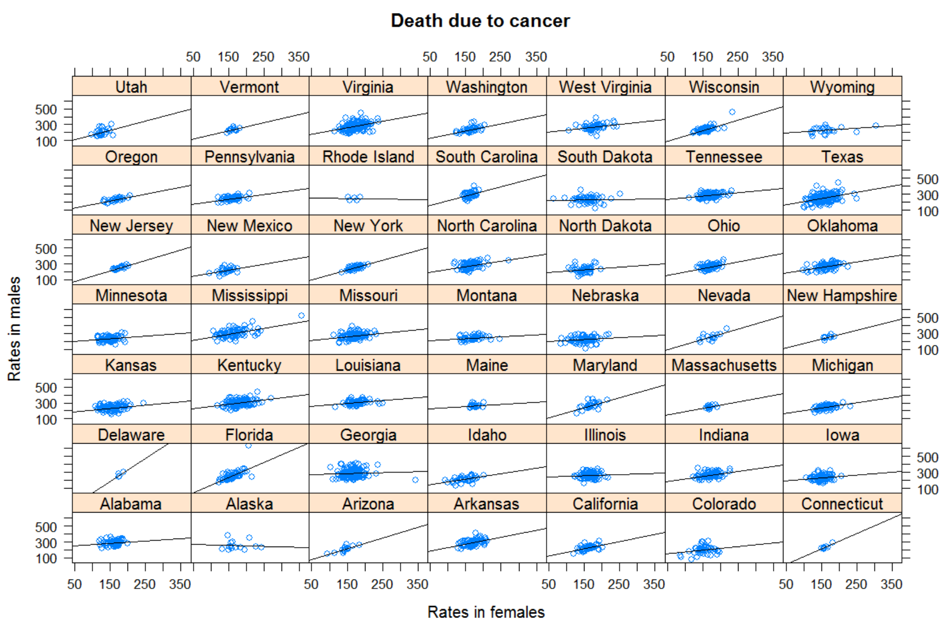

Chicken growth by diet
1
2
3
4
Weight of the chicken
Days since birth
0
5
10
15
20
100
200
300

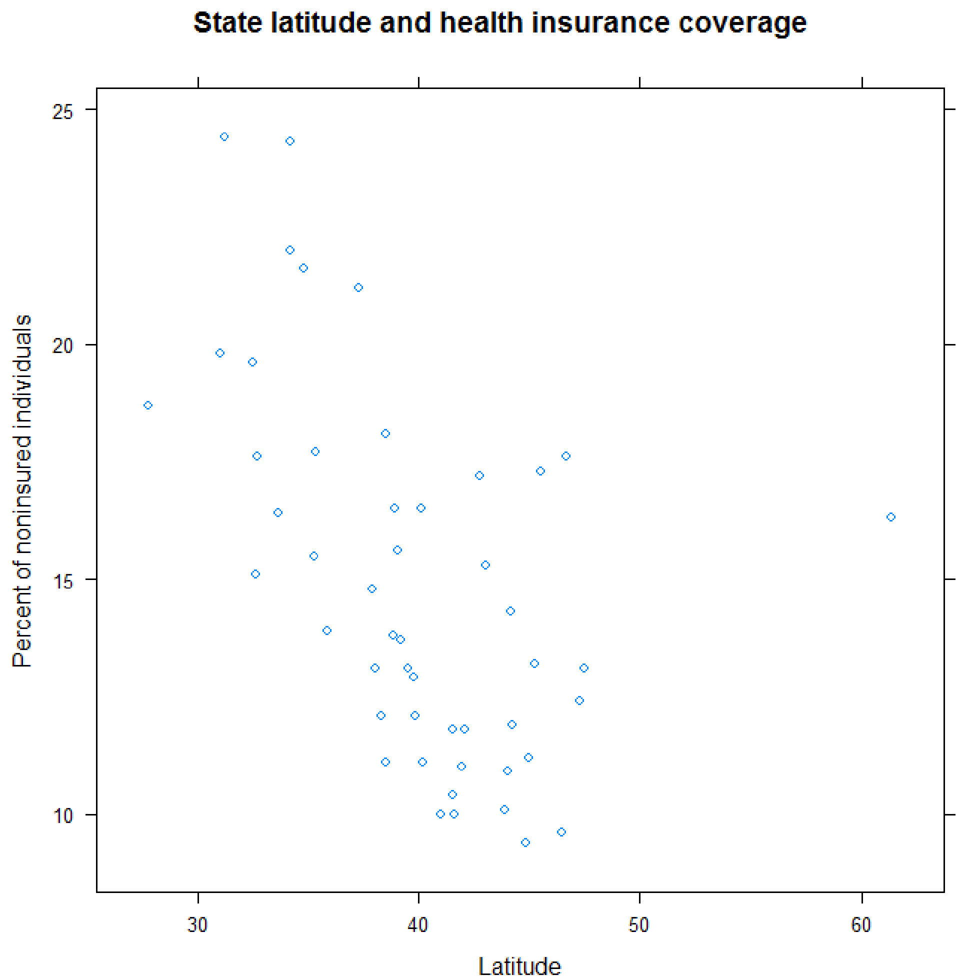
R Console

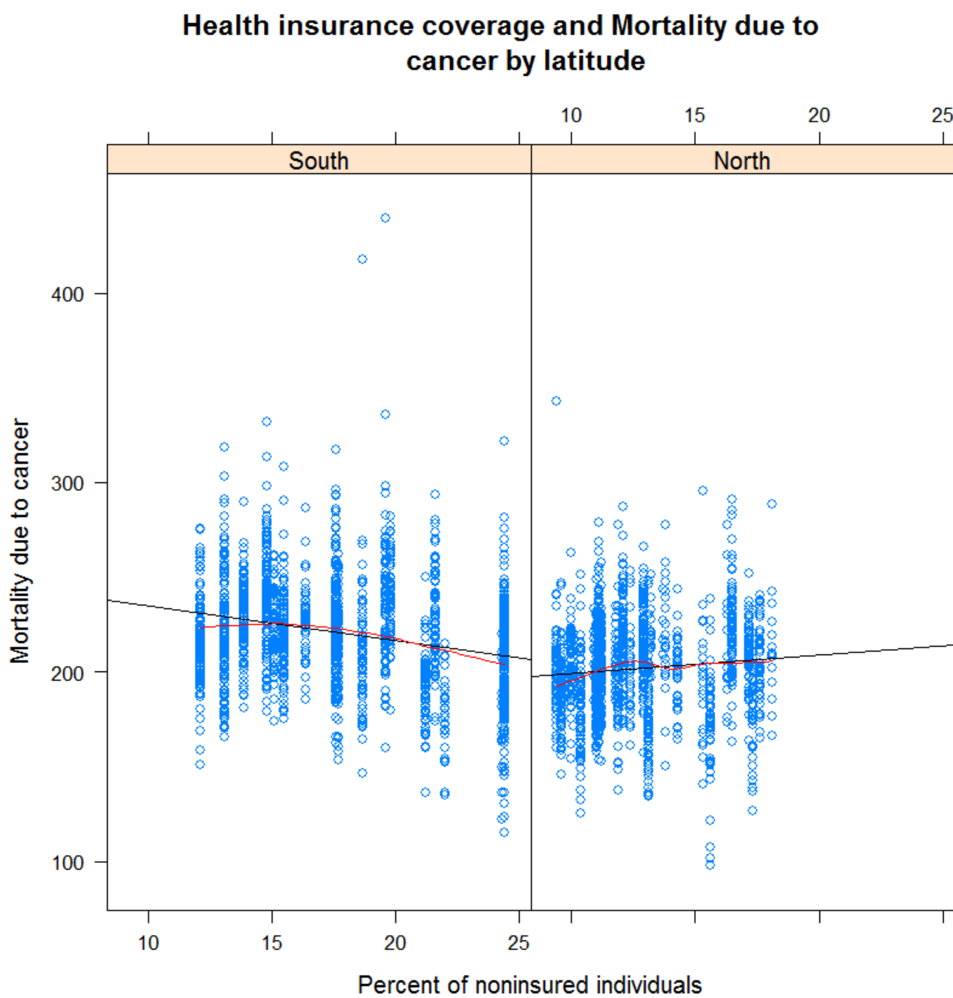
```
> head(USCancerRates, 3)
                 rate.male LCL95.male UCL95.male rate.female LCL95.female
alabama,pickens      363.7      311.1      423.2       151.0        123.6
alabama,bullock      345.7      274.2      431.4       140.5        102.8
alabama,russell      340.7      304.5      380.9       182.3        161.3
                 UCL95.female   state          county
alabama,pickens         183.6 Alabama Pickens County
alabama,bullock         189.7 Alabama Bullock County
alabama,russell         205.5 Alabama Russell County
> summary(USCancerRates)
   rate.male       LCL95.male      UCL95.male     rate.female
 Min.   : 76.5   Min.   : 34.6   Min.   :160.9   Min.   : 63.5
 1st Qu.:228.8   1st Qu.:189.3   1st Qu.:269.6   1st Qu.:150.5
 Median :254.8   Median :217.8   Median :301.3   Median :165.4
 Mean   :257.4   Mean   :215.2   Mean   :311.0   Mean   :165.0
 3rd Qu.:283.8   3rd Qu.:244.4   3rd Qu.:342.0   3rd Qu.:177.8
 Max.   :629.1   Max.   :528.8   Max.   :774.3   Max.   :357.2
 NA's   :10      NA's   :10      NA's   :10      NA's   :63
  LCL95.female    UCL95.female          state          county
 Min.   : 35.0   Min.   :112.9   Texas    : 235   Length:3041
 1st Qu.:120.4   1st Qu.:179.4   Georgia  : 157   Class :AsIs
 Median :139.1   Median :195.2   Virginia: 134   Mode  :character
 Mean   :136.4   Mean   :202.7   Kentucky: 119
 3rd Qu.:154.7   3rd Qu.:217.5   Missouri: 115
 Max.   :330.6   Max.   :786.8   Illinois: 102
 NA's   :63      NA's   :63      (Other) :2179
> 
```


Max
Min
state
rate.female
rate.male


Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Delaware
Florida
Georgia
Idaho
Illinois
Indiana
Iowa
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Max
Min
F
M


Death due to cancer
Rates in males
Rates in females
50
150
250
350
100
300
500
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Delaware
Florida
Georgia
Idaho
Illinois
Indiana
Iowa
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut


State latitude and health insurance coverage
Percent of noninsured individuals
25
20
15
10
30
40
50
60
Latitude



Chapter 4: Cluster Analysis

$$\frac{x - \min(x)}{\max(x) - \min(x)}$$

$$\frac{x - \bar{x}}{\sigma(x)}$$

$$tf(t,d) = 1 + \log\big(f(t,d)\big)$$

$$idf(t,D) = \log \frac{N}{\{d \ \in D: t \ \in d\}}$$

$$tfidf(t,d,D) = \ tf(t,d)\, idf(t,D)$$

$$\sum_{i=1}^{n} |p_i - q_i|$$

$$\sqrt{\left(\sum_{i=1}^{n} p_i - q_i\right)^2}$$

$$\frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

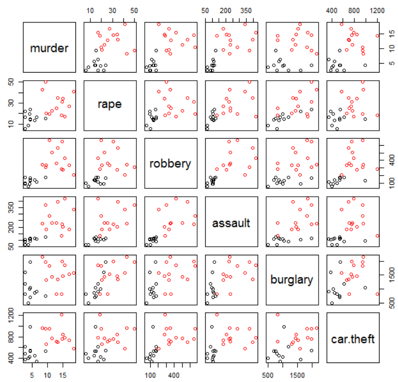
$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$\frac{|A \cap B|}{|A \cup B|}$$

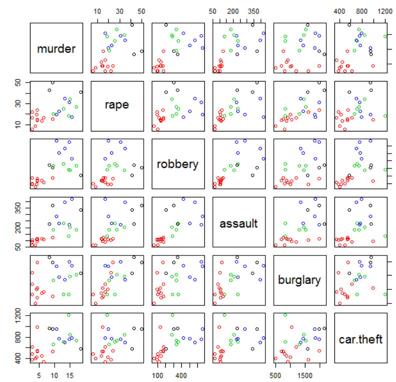
1. Initiate centroids randomly
2. Compute distance from each case to each centroid
3. Assign case to the closest cluster (smaller distance between case and centroid)
4. Update centroids using mean value of observations pertaining to given cluster

(repeat until convergence)

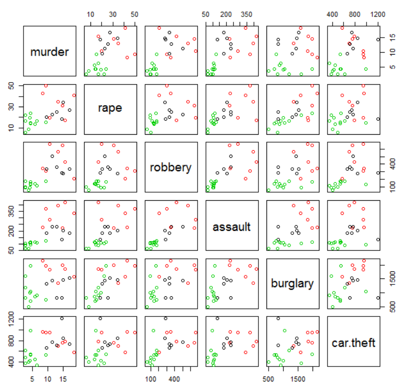
2-cluster solution



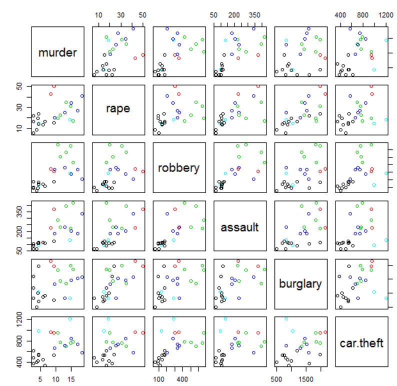
4-cluster solution

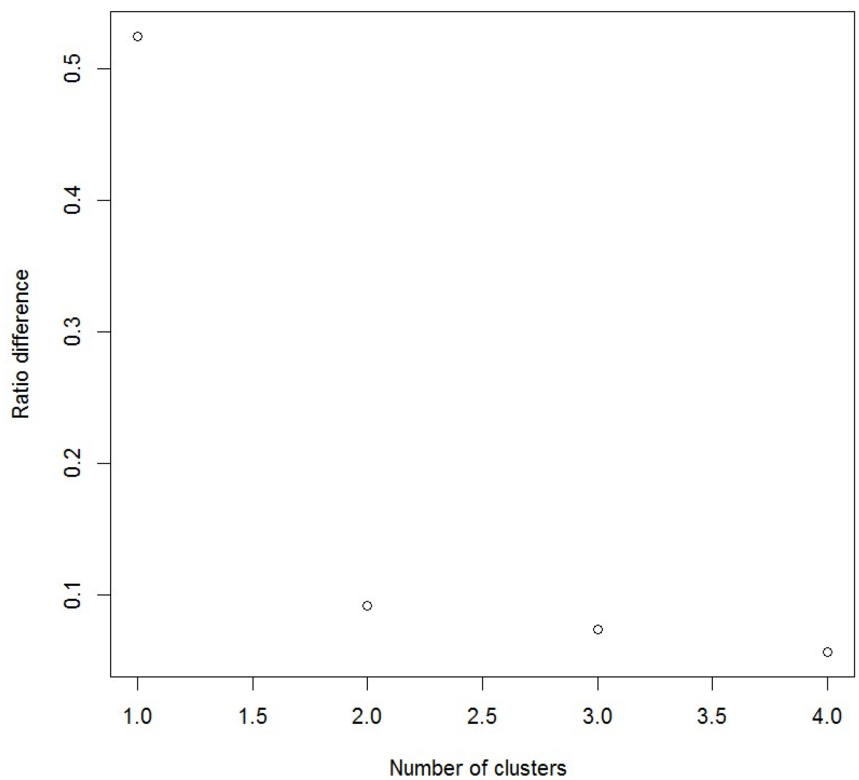


3-cluster solution

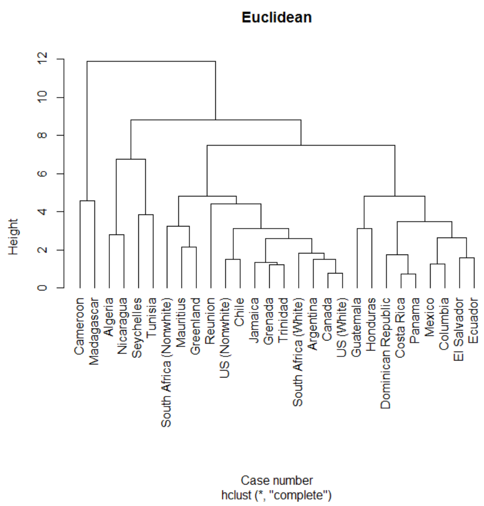


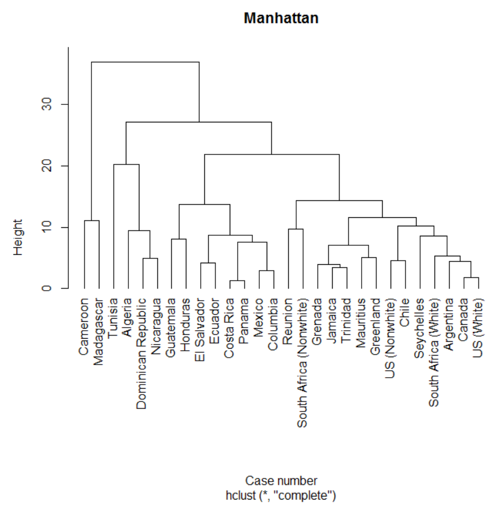
5-cluster solution

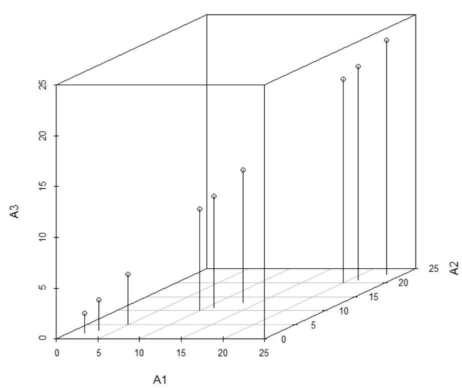


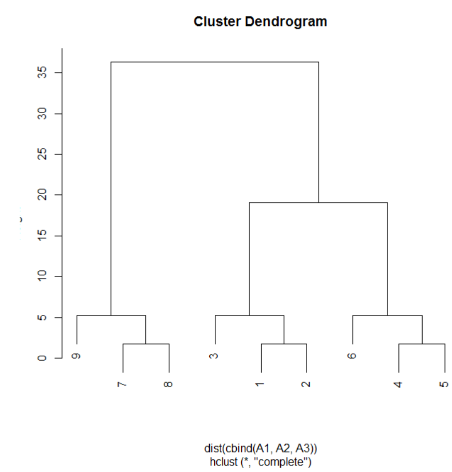


Chapter 5: Agglomerative Clustering Using hclust()

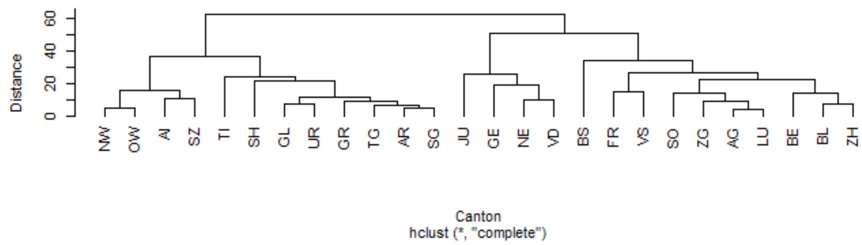

Euclidean
Height
0
2
4
6
8
10
12
Cameroon
Madagascar
Algeria
Nicaragua
Seychelles
Tunisia
South Africa (Nonwhite)
Mauritius
Greenland
Reunion
US (Nonwhite)
Chile
Jamaica
Grenada
Trinidad
South Africa (White)
Argentina
Canada
US (White)
Guatemala
Honduras
Dominican Republic
Costa Rica
Panama
Mexico
Columbia
El Salvador
Ecuador
Case number
hclust (*, "complete")


Manhattan
Height
0
10
20
30
Cameroon
Madagascar
Tunisia
Algeria
Dominican Republic
Nicaragua
Guatemala
Honduras
El Salvador
Ecuador
Costa Rica
Panama
Mexico
Columbia
Reunion
South Africa (Nonwhite)
Grenada
Jamaica
Trinidad
Mauritius
Greenland
US (Nonwhite)
Chile
Seychelles
South Africa (White)
Argentina
Canada
US (White)
Case number
hclust (*, "complete")

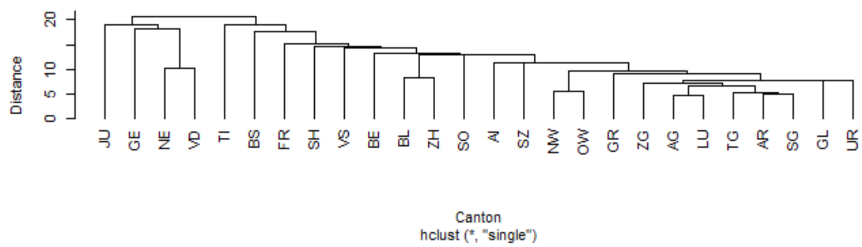

A3
A1
A2
0
5
10
15
20
25


Cluster Dendrogram
0
5
10
15
20
25
30
35
9
7
8
3
1
2
6
4
5
dist(cbind(A1, A2, A3))
hclust (*, "complete")

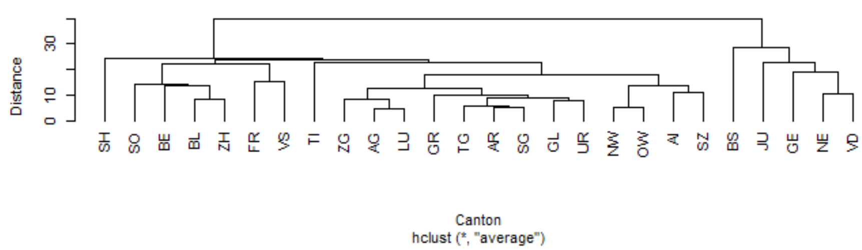
Complete linkage

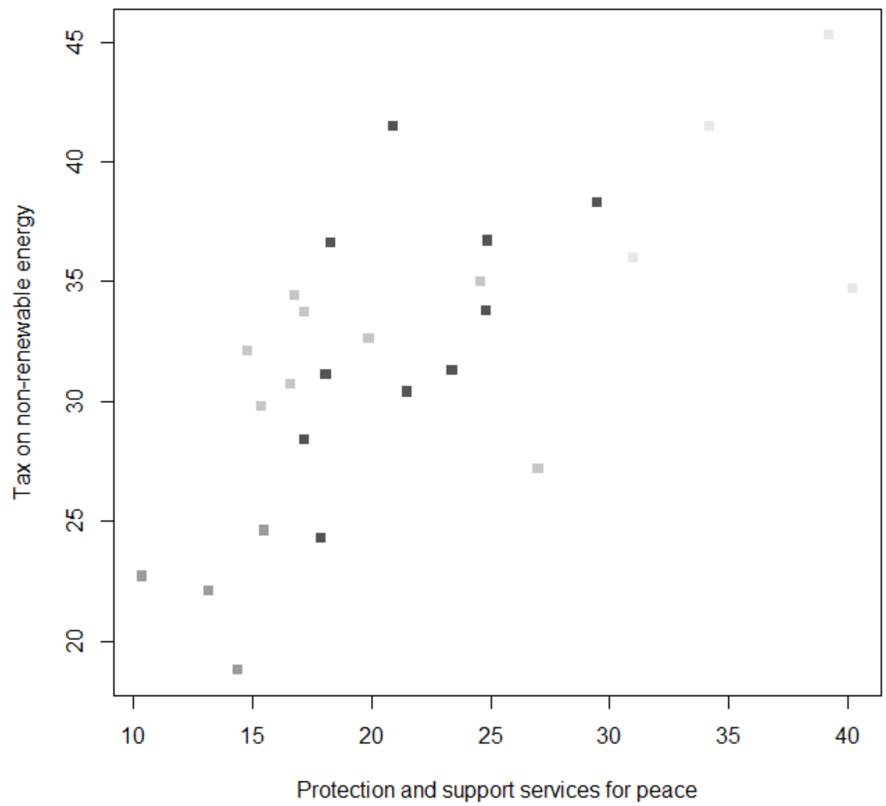


Single linkage

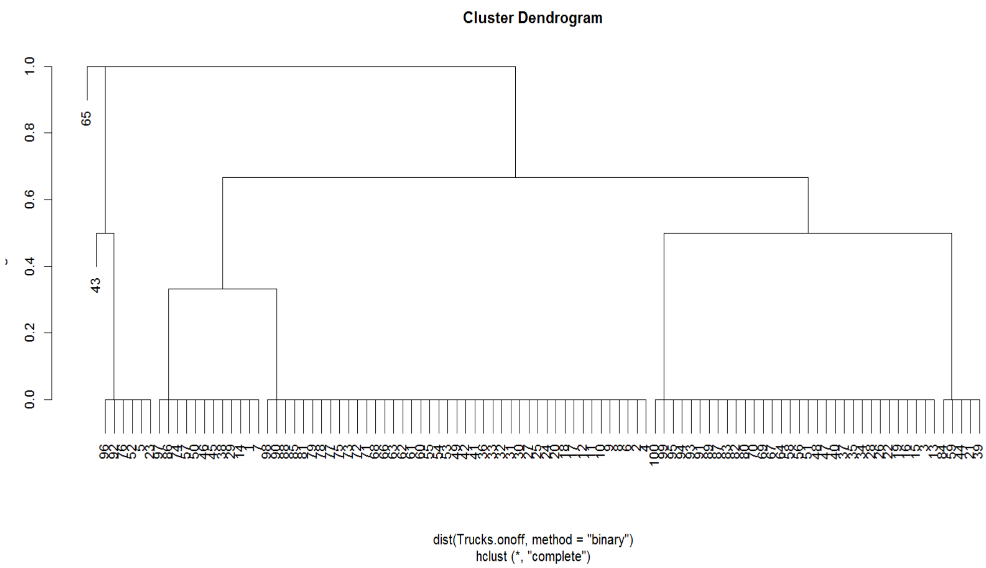


Average linkage

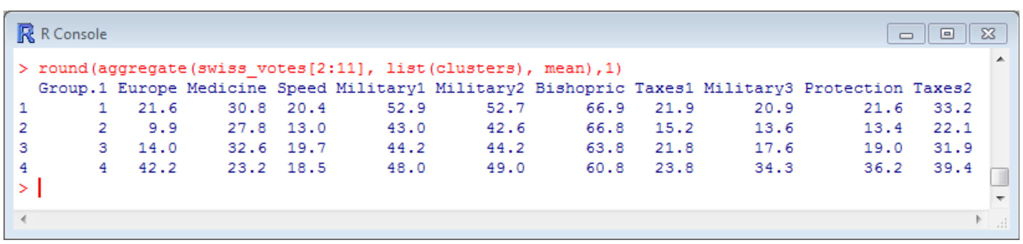




Tax on non-renewable energy
45
40
35
30
25
20
10
15
20
25
30
35
40
Protection and support services for peace



Canton	Europe	Medicine	Speed	Military1	Military2	Bishopric	Taxes1	Military3	Protection	Taxes2
AG	17	34	17	50.9	51.1	64.5	18.8	17.1	17.2	28.4
AI	6.8	24.9	10.6	37.3	37.8	67.2	14.3	11.5	10.4	22.7
AR	13.5	28.4	18.5	45.6	46.1	65.9	20.9	17.6	17.2	33.7
BE	23.4	31.4	22.1	57.7	57.4	60.2	24.2	19.6	20.9	41.5
BL	22.6	30.6	23.1	54.5	53.2	67.3	23.6	22.9	23.4	31.3

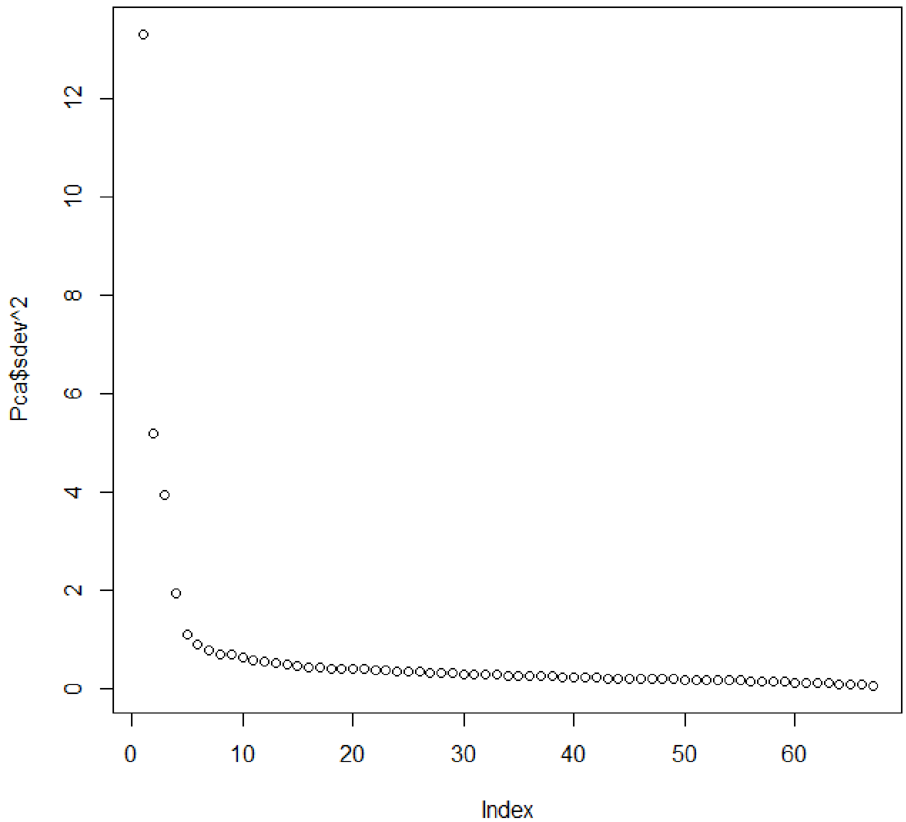



```
> round(aggregate(swiss_votes[2:11], list(clusters), mean),1)
  Group.1 Europe Medicine Speed Military1 Military2 Bishopric Taxes1 Military3 Protection Taxes2
1       1   21.6     30.8  20.4      52.9      52.7      66.9   21.9      20.9       21.6   33.2
2       2    9.9     27.8  13.0      43.0      42.6      66.8   15.2      13.6       13.4   22.1
3       3   14.0     32.6  19.7      44.2      44.2      63.8   21.8      17.6       19.0   31.9
4       4   42.2     23.2  18.5      48.0      49.0      60.8   23.8      34.3       36.2   39.4
>
```

Chapter 6: Dimensionality Reduction with Principal Component Analysis

```
> myPCA(iris[1:4])
[[1]]
[1] 4.22824171 0.24267075 0.07820950 0.02383509

[[2]]
            [,1]        [,2]        [,3]       [,4]
[1,]  0.36138659 -0.65658877 -0.58202985  0.3154872
[2,] -0.08452251 -0.73016143  0.59791083 -0.3197231
[3,]  0.85667061  0.17337266  0.07623608 -0.4798390
[4,]  0.35828920  0.07548102  0.54583143  0.7536574

[[3]]
              [,1]         [,2]        [,3]        [,4]
  [1,] -2.02428352 -0.482693188  0.31226649 -0.95505451
  [2,] -2.01460877  0.513488442 -0.23304573 -0.66448564
  [3,] -2.18920532  0.327211755  0.17756654 -0.86020941
  [4,] -2.11639895  0.593665486  0.11931399 -0.87931870
  [5,] -2.08731760 -0.570920955  0.51973192 -1.06650727
  [6,] -1.73132921 -1.341378475  0.80628723 -1.01796720
  [7,] -2.17609795  0.091188104  0.59813723 -0.97332307
  [8,] -2.00000554 -0.226060611  0.24969535 -0.94698205
  [9,] -2.21342812  1.077467178 -0.01878407 -0.78162853
 [10,] -2.03247716  0.345887445 -0.16315864 -0.86389521
 [11,] -1.88361212 -1.045786708  0.38007671 -1.01464540
 [12,] -2.03876163 -0.057655800  0.39458964 -1.05036235
```

```
> apply(is.na(motiv),2,sum)
     active      afraid       alert       angry     anxious     aroused     ashamed  astonished     at.ease
          6           5          11           9        1849           6          11          13          17
    at.rest   attentive        blue       bored        calm    cheerful  clutched.up   confident     content
         17           6           5           4          82        1850          23           7          22
  delighted   depressed  determined  distressed      drowsy        dull      elated   energetic enthusiastic
          6          17           7           8          12           9          15           6           6
    excited     fearful  frustrated full.of.pep      gloomy    grouchy      guilty       happy     hostile
          6          20          11          12          12           5           5          16          11
       idle    inactive    inspired     intense  interested   irritable     jittery      lively      lonely
       1848        1846           6           7          12          16           6          10           6
    nervous      placid     pleased       proud  quiescent       quiet     relaxed         sad   satisfied
         17          19          13           7         136           5           7          10           7
     scared      serene      sleepy    sluggish    sociable       sorry       still      strong   surprised
         10          12          16           8           6          15          12           7           6
      tense       tired    tranquil     unhappy       upset    vigorous     wakeful warmhearted  wide.awake
         10          10        1843           5           8          10          10           7          12
```



Pca$sdev^2
Index
0
2
4
6
8
10
12
0
10
20
30
40
50
60

```
> print.psych(Pca2, sort =T)
Principal Components Analysis
Call: principal(r = motiv[, -ToSuppress], nfactors = 5, rotate = "varimax",
    scores = T, missing = T)
Standardized loadings (pattern matrix) based upon correlation matrix
```

	item	RC1	RC2	RC3	RC4	RC5	h2	u2
lively	40	0.82	-0.06	-0.05	-0.27	-0.03	0.75	0.25
excited	26	0.80	-0.08	-0.11	-0.13	0.10	0.69	0.31
enthusiastic	25	0.80	-0.16	0.03	-0.12	0.07	0.68	0.32
full.of.pep	29	0.79	-0.07	-0.10	-0.31	-0.05	0.73	0.27
energetic	24	0.79	-0.04	-0.06	-0.36	-0.04	0.75	0.25
active	1	0.78	-0.02	-0.07	-0.27	-0.06	0.70	0.30
elated	23	0.77	-0.07	-0.02	-0.05	0.01	0.60	0.40
vigorous	64	0.75	0.05	-0.11	-0.25	-0.06	0.63	0.37
happy	33	0.72	-0.30	0.24	-0.09	0.00	0.68	0.32
pleased	44	0.71	-0.22	0.24	-0.05	0.04	0.62	0.38
aroused	5	0.71	0.03	-0.10	-0.20	0.03	0.56	0.44
inspired	35	0.71	0.02	-0.02	-0.09	0.13	0.52	0.48
proud	45	0.70	-0.05	0.14	0.04	-0.03	0.51	0.49
determined	19	0.69	0.09	0.03	-0.05	0.17	0.51	0.49
strong	58	0.69	0.07	0.07	-0.07	-0.05	0.49	0.51
delighted	17	0.68	-0.21	0.05	-0.02	0.06	0.52	0.48
sociable	55	0.66	-0.23	0.10	-0.11	0.01	0.51	0.49
confident	15	0.63	-0.11	0.29	-0.08	-0.15	0.53	0.47
alert	3	0.63	0.02	0.05	-0.54	-0.06	0.69	0.31
warmhearted	66	0.61	-0.25	0.35	0.01	0.09	0.57	0.43
satisfied	50	0.61	-0.28	0.33	-0.05	-0.03	0.57	0.43
interested	37	0.61	-0.14	0.20	-0.17	0.14	0.48	0.52
attentive	10	0.61	-0.02	0.15	-0.46	0.00	0.60	0.40
wakeful	65	0.56	-0.02	0.07	-0.56	-0.05	0.64	0.36
content	16	0.55	-0.31	0.46	-0.09	-0.07	0.62	0.38
intense	36	0.53	0.34	-0.22	-0.05	0.11	0.46	0.54
surprised	59	0.40	0.12	-0.12	0.01	0.21	0.24	0.76
astonished	7	0.34	0.17	-0.09	0.04	0.30	0.25	0.75
unhappy	62	-0.16	0.74	0.00	0.03	0.25	0.64	0.36
irritable	38	-0.09	0.70	-0.22	0.22	-0.03	0.60	0.40
grouchy	31	-0.13	0.70	-0.11	0.28	-0.03	0.60	0.40
upset	63	-0.08	0.70	-0.06	0.02	0.36	0.62	0.38
gloomy	30	-0.18	0.69	0.04	0.16	0.22	0.59	0.41
blue	11	-0.14	0.69	0.08	0.04	0.28	0.58	0.42
depressed	18	-0.18	0.68	0.03	0.05	0.31	0.60	0.40
frustrated	28	-0.01	0.68	-0.11	0.06	0.32	0.58	0.42
sad	49	-0.10	0.68	0.07	0.01	0.34	0.59	0.41
angry	4	0.00	0.66	-0.13	0.00	0.21	0.50	0.50
hostile	34	0.00	0.64	-0.21	0.11	-0.01	0.47	0.53
distressed	20	0.02	0.59	-0.07	0.04	0.47	0.57	0.43
lonely	41	-0.10	0.57	0.11	0.01	0.26	0.42	0.58
tense	60	0.18	0.52	-0.34	0.03	0.31	0.51	0.49
clutched.up	14	0.17	0.50	-0.27	0.07	0.24	0.41	0.59
bored	12	-0.20	0.34	0.05	0.31	-0.21	0.29	0.71
calm	13	0.11	-0.18	0.73	0.00	-0.10	0.59	0.41
serene	52	0.14	-0.15	0.71	0.05	-0.03	0.55	0.45
at.ease	8	0.30	-0.26	0.67	-0.11	-0.15	0.63	0.37
relaxed	48	0.22	-0.27	0.65	-0.05	-0.11	0.56	0.44
still	57	-0.16	0.06	0.64	0.19	-0.03	0.47	0.53
at.rest	9	0.18	-0.12	0.64	-0.17	-0.08	0.49	0.51
placid	43	-0.04	0.05	0.59	0.18	-0.01	0.38	0.62
quiet	47	-0.22	0.24	0.52	0.17	0.05	0.40	0.60
quiescent	46	0.12	0.14	0.41	0.14	0.05	0.22	0.78
jittery	39	0.34	0.23	-0.39	-0.03	0.19	0.36	0.64
sleepy	53	-0.23	0.15	0.11	0.84	0.06	0.80	0.20
drowsy	21	-0.22	0.16	0.13	0.83	0.04	0.79	0.21
tired	61	-0.27	0.18	0.11	0.80	0.05	0.76	0.24
sluggish	54	-0.32	0.22	0.14	0.68	0.05	0.63	0.37
wide.awake	67	0.59	0.02	0.03	-0.59	-0.08	0.71	0.29
dull	22	-0.34	0.37	0.19	0.39	-0.01	0.44	0.56
afraid	2	0.07	0.26	-0.09	0.05	0.76	0.66	0.34
fearful	27	0.06	0.26	-0.08	0.04	0.74	0.63	0.37
scared	51	0.08	0.27	-0.12	0.05	0.72	0.62	0.38
ashamed	6	-0.03	0.33	0.01	-0.01	0.61	0.49	0.51
guilty	32	0.02	0.31	0.02	0.01	0.60	0.45	0.55
sorry	56	-0.02	0.45	0.06	0.01	0.57	0.53	0.47
nervous	42	0.18	0.30	-0.26	0.03	0.55	0.50	0.50

	RC1	RC2	RC3	RC4	RC5
SS loadings	14.25	8.56	5.09	4.86	4.57
Proportion Var	0.21	0.13	0.08	0.07	0.07
Cumulative Var	0.21	0.34	0.42	0.49	0.56
Proportion Explained	0.38	0.23	0.14	0.13	0.12
Cumulative Proportion	0.38	0.61	0.75	0.88	1.00

```
Test of the hypothesis that 5 components are sufficient.

The degrees of freedom for the null model are  2211  and the objective function was  45.86
The degrees of freedom for the model are 1886  and the objective function was  6.28
The total number of observations was  3896  with MLE Chi Square =  24294.41  with prob <  0

Fit based upon off diagonal values = 0.99>
```

$$(A - \lambda I)^k \mathrm{v} = 0$$

$$partVar_i = \frac{eigen_i}{\sum_{i=i}^{n} eigen_i}$$

Chapter 7: Exploring Association Rules with Apriori

```
> rules = apriori(Groceries)

parameter specification:
 confidence minval smax arem  aval originalSupport support minlen maxlen target   ext
        0.8    0.1    1 none FALSE            TRUE     0.1      1     10  rules FALSE

algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

apriori - find association rules with the apriori algorithm
version 4.21 (2004.05.09)        (c) 1996-2004   Christian Borgelt
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.02s].
sorting and recoding items ... [8 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing ... [0 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

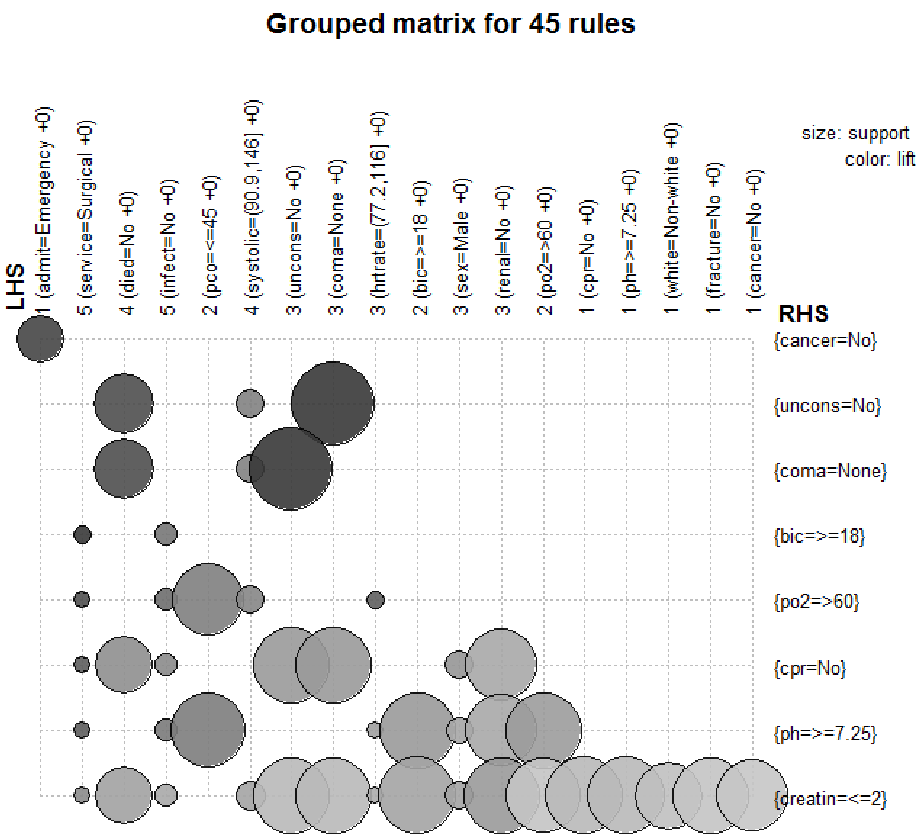
```
> rules = apriori(Groceries, parameter = list(support = 0.05, confidence = .1))

parameter specification:
 confidence minval smax arem  aval originalSupport support minlen maxlen target   ext
        0.1    0.1    1 none FALSE            TRUE    0.05      1     10  rules FALSE

algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

apriori - find association rules with the apriori algorithm
version 4.21 (2004.05.09)        (c) 1996-2004   Christian Borgelt
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [28 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing ... [14 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
> inspect(rules)
   lhs                    rhs                  support confidence     lift
1  {}                  => {bottled water}    0.11052364  0.1105236 1.000000
2  {}                  => {tropical fruit}   0.10493137  0.1049314 1.000000
3  {}                  => {root vegetables}  0.10899847  0.1089985 1.000000
4  {}                  => {soda}             0.17437722  0.1743772 1.000000
5  {}                  => {yogurt}           0.13950178  0.1395018 1.000000
6  {}                  => {rolls/buns}       0.18393493  0.1839349 1.000000
7  {}                  => {other vegetables} 0.19349263  0.1934926 1.000000
8  {}                  => {whole milk}       0.25551601  0.2555160 1.000000
9  {yogurt}            => {whole milk}       0.05602440  0.4016035 1.571735
10 {whole milk}        => {yogurt}           0.05602440  0.2192598 1.571735
11 {rolls/buns}        => {whole milk}       0.05663447  0.3079049 1.205032
12 {whole milk}        => {rolls/buns}       0.05663447  0.2216474 1.205032
13 {other vegetables} => {whole milk}       0.07483477  0.3867578 1.513634
14 {whole milk}        => {other vegetables} 0.07483477  0.2928770 1.513634
```

```
 > summary(ICU)
  died          age            sex          race         service       cancer     renal      infect
 No :160   Min.   :16.00   Female: 76   Black: 15   Medical : 93   No :180   No :181   No :116
 Yes: 40   1st Qu.:46.75   Male  :124   Other: 10   Surgical:107   Yes: 20   Yes: 19   Yes: 84
           Median :63.00                White:175
           Mean   :57.55
           3rd Qu.:72.00
           Max.   :92.00
  cpr          systolic        hrtrate        previcu        admit        fracture    po2
 No :187   Min.   : 36.0   Min.   : 39.00   No :170   Elective : 53   No :185   >60 :184
 Yes: 13   1st Qu.:110.0   1st Qu.: 80.00   Yes: 30   Emergency:147   Yes: 15   <=60: 16
           Median :130.0   Median : 96.00
           Mean   :132.3   Mean   : 98.92
           3rd Qu.:150.0   3rd Qu.:118.25
           Max.   :256.0   Max.   :192.00
      ph          pco        bic       creatin       coma           white       uncons
 >=7.25:187   <=45:180   >=18:185   <=2:190   None  :185   White    : 25   No :185
 <7.25 : 13   >45 : 20   <18 : 15   >2 : 10   Stupor:  5   Non-white:175   Yes: 15
                                              Coma  : 10
```

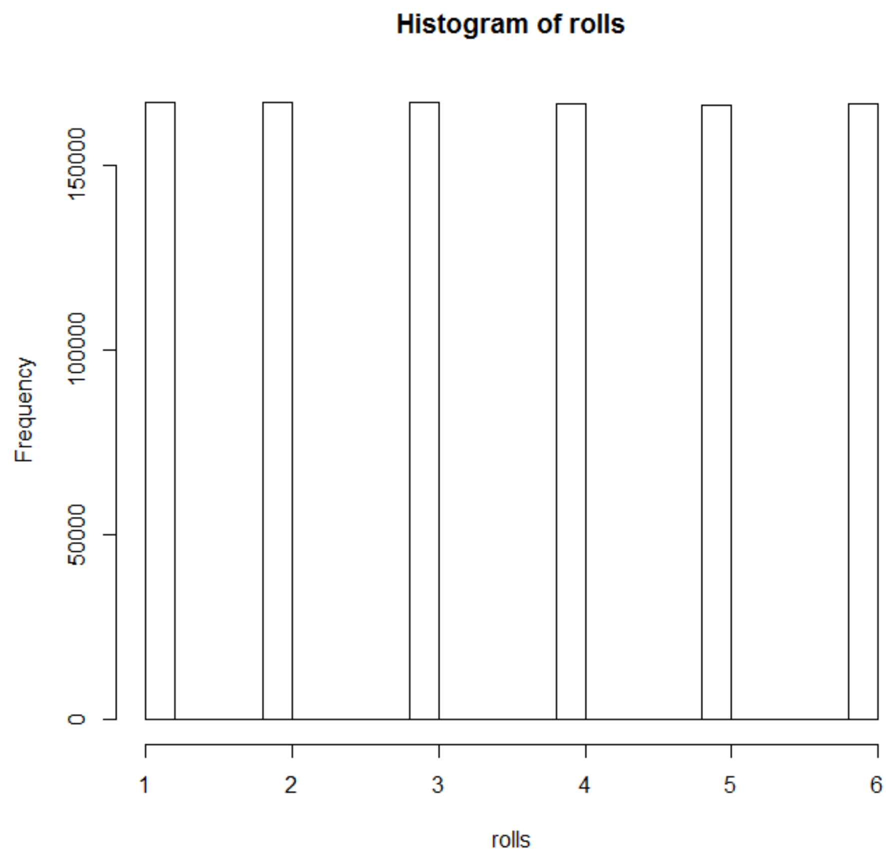
```
> inspect(rules)
   lhs               rhs             support confidence     lift
1  {}             => {creatin=<=2}    0.950  0.9500000 1.000000
2  {cancer=No}    => {creatin=<=2}    0.855  0.9500000 1.000000
3  {pco=<=45}     => {po2=>60}        0.860  0.9555556 1.038647
4  {pco=<=45}     => {ph=>=7.25}      0.875  0.9722222 1.039810
5  {renal=No}     => {cpr=No}         0.860  0.9502762 1.016338
6  {renal=No}     => {ph=>=7.25}      0.860  0.9502762 1.016338
7  {renal=No}     => {creatin=<=2}    0.880  0.9723757 1.023553
8  {po2=>60}      => {ph=>=7.25}      0.880  0.9565217 1.023018
9  {po2=>60}      => {creatin=<=2}    0.875  0.9510870 1.001144
10 {uncons=No}    => {coma=None}      0.925  1.0000000 1.081081
```

```
> inspect(rulesDeath)
   lhs                     rhs          support confidence     lift
1  {infect=Yes,
    admit=Emergency} => {died=Yes}    0.120  0.3478261 1.739130
2  {service=Medical,
    white=Non-white} => {died=Yes}    0.120  0.3037975 1.518987
3  {infect=Yes,
    admit=Emergency,
    white=Non-white} => {died=Yes}    0.115  0.3650794 1.825397
4  {infect=Yes,
    admit=Emergency,
    pco=<=45}        => {died=Yes}    0.100  0.3448276 1.724138
5  {cancer=No,
    infect=Yes,
    admit=Emergency} => {died=Yes}    0.115  0.3432836 1.716418
6  {infect=Yes,
    admit=Emergency,
    po2=>60}         => {died=Yes}    0.105  0.3620690 1.810345
7  {infect=Yes,
    admit=Emergency,
    fracture=No}     => {died=Yes}    0.110  0.3437500 1.718750
8  {infect=Yes,
    admit=Emergency,
    ph=>=7.25}       => {died=Yes}    0.100  0.3333333 1.666667
9  {cancer=No,
    infect=Yes,
    white=Non-white} => {died=Yes}    0.110  0.3098592 1.549296
10 {infect=Yes,
    po2=>60,
    white=Non-white} => {died=Yes}    0.100  0.3076923 1.538462
```

```
> head(rulesDeath.df.sorted)
                                                                rules support confidence     lift
45 {cancer=No,infect=Yes,admit=Emergency,po2=>60,white=Non-white} => {died=Yes}   0.100  0.3921569 1.960784
19           {infect=Yes,admit=Emergency,po2=>60,white=Non-white} => {died=Yes}   0.100  0.3846154 1.923077
47     {cancer=No,infect=Yes,admit=Emergency,fracture=No,po2=>60} => {died=Yes}   0.100  0.3773585 1.886792
23               {infect=Yes,admit=Emergency,fracture=No,po2=>60} => {died=Yes}   0.100  0.3703704 1.851852
21                 {cancer=No,infect=Yes,admit=Emergency,po2=>60} => {died=Yes}   0.105  0.3684211 1.842105
3                   {infect=Yes,admit=Emergency,white=Non-white} => {died=Yes}   0.115  0.3650794 1.825397
```

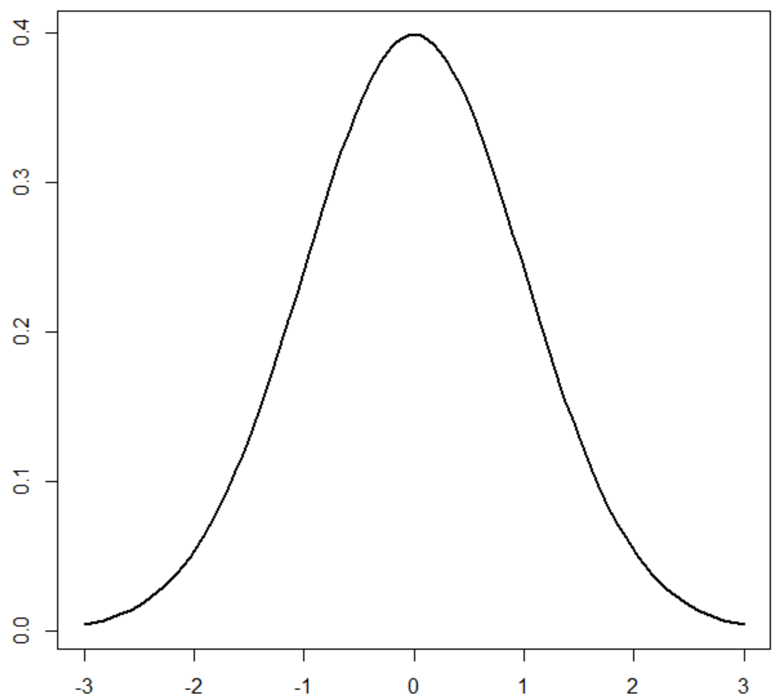


Chapter 8: Probability Distributions, Covariance, and Correlation

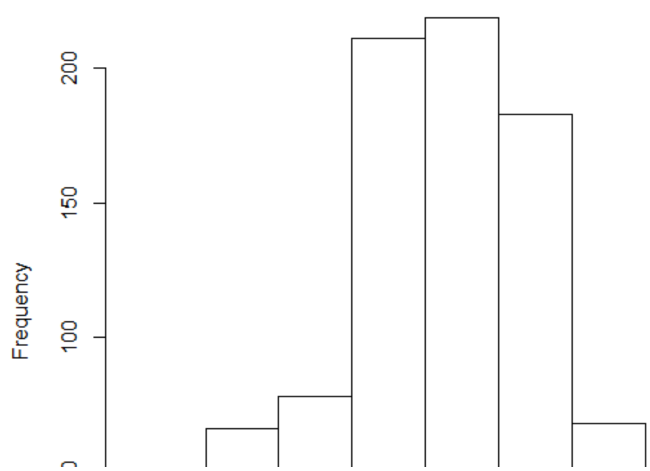


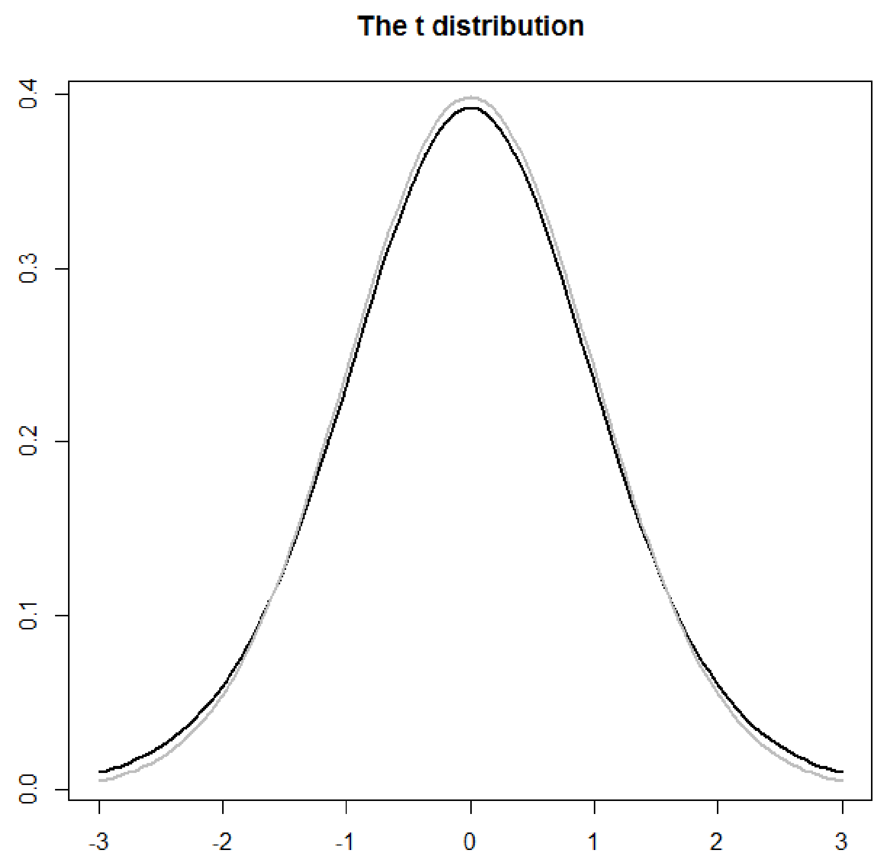
Histogram of rolls
Frequency
0
50000
100000
150000
1
2
3
4
5
6
rolls

The standard normal distribution

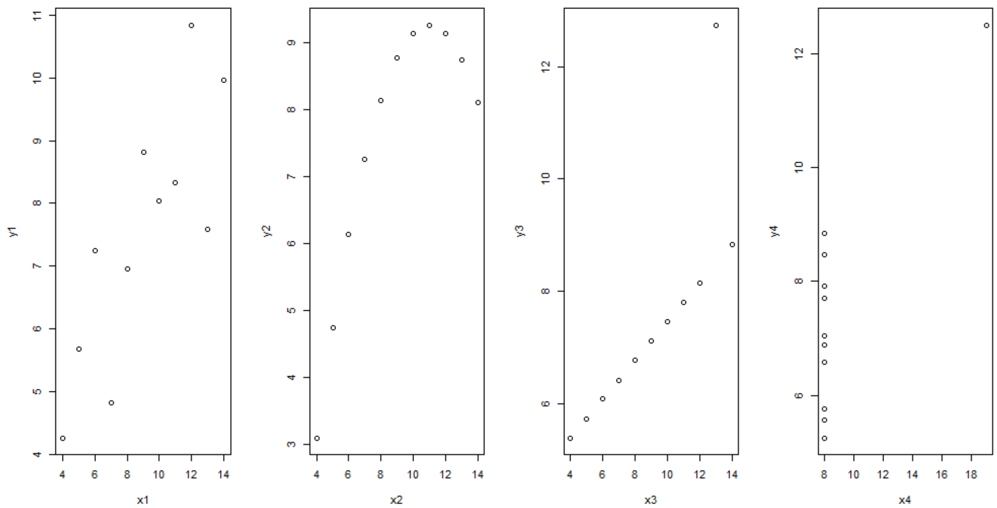


Height of adults in inches

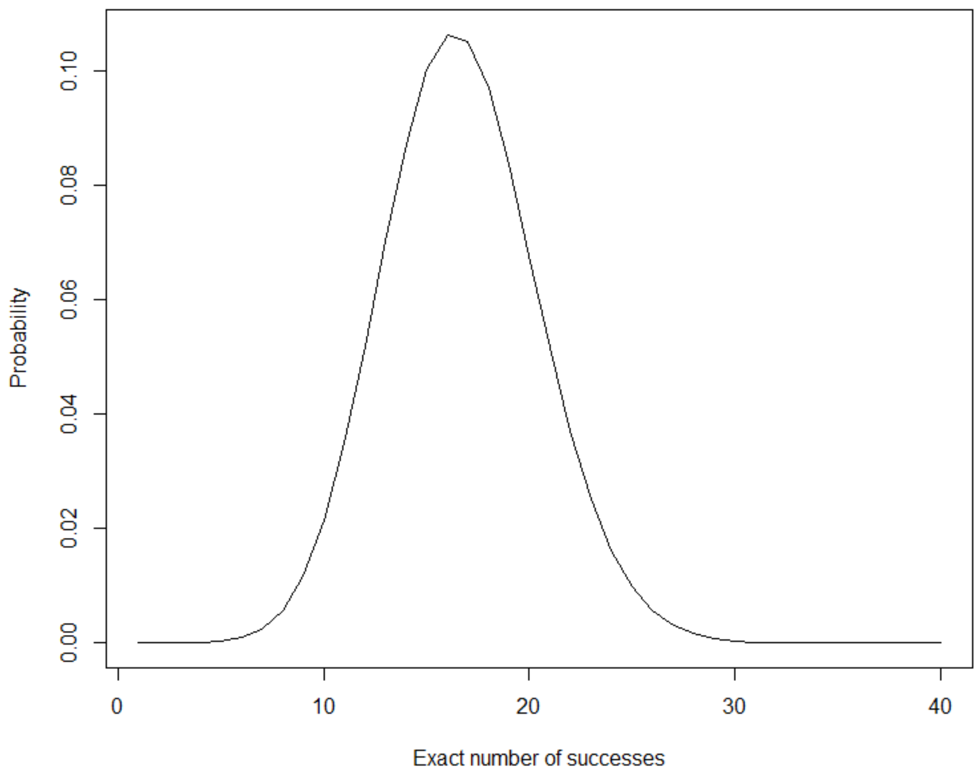




The t distribution
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3



y1
x1
y2
x2
y3
x3
y4
x4



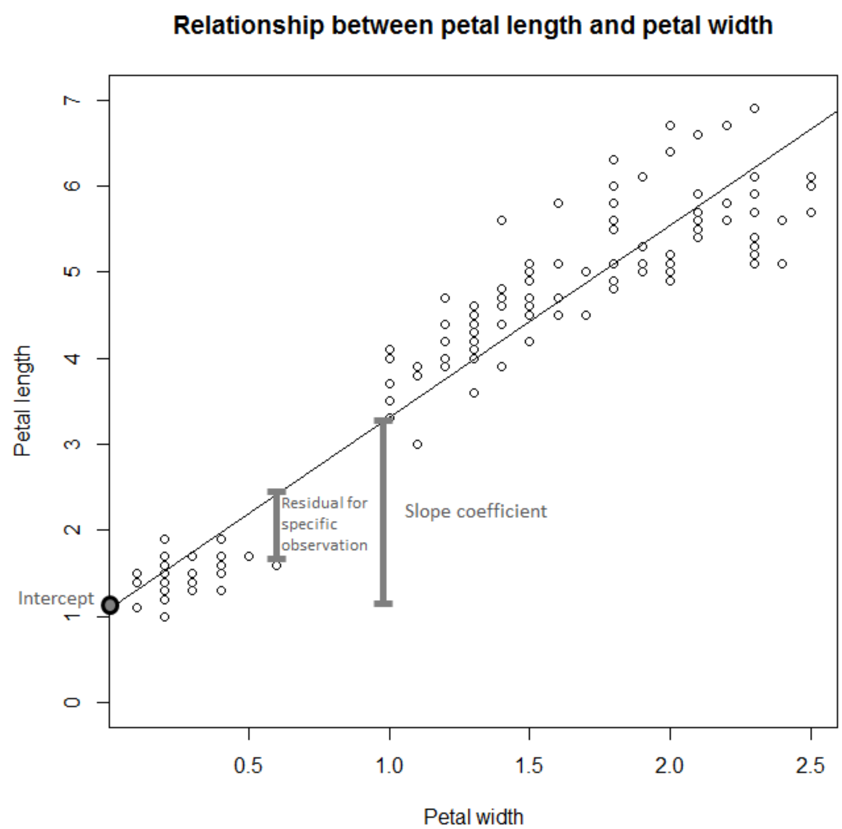
$$\bar{x} = \frac{\sum_{i=1}^{n} x}{n}$$

$$s^2 = \frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n-1}$$

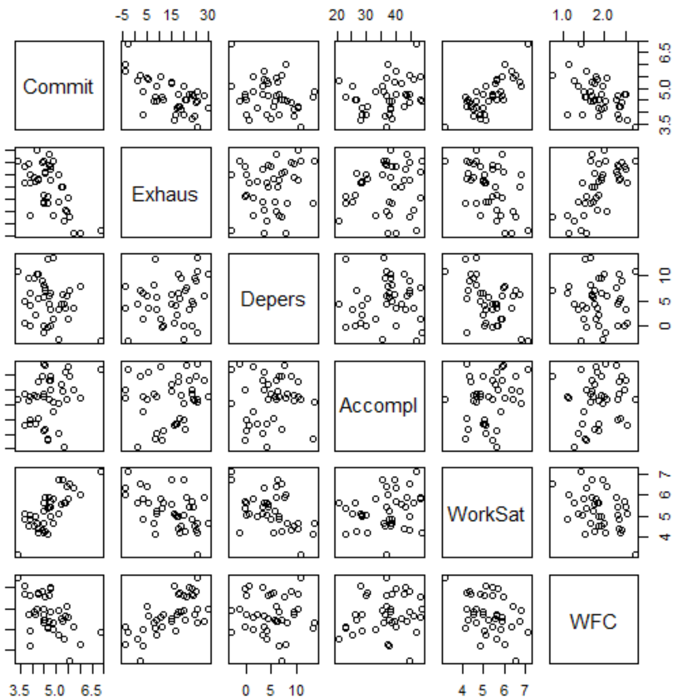
$$s = \sqrt{\frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n-1}}$$

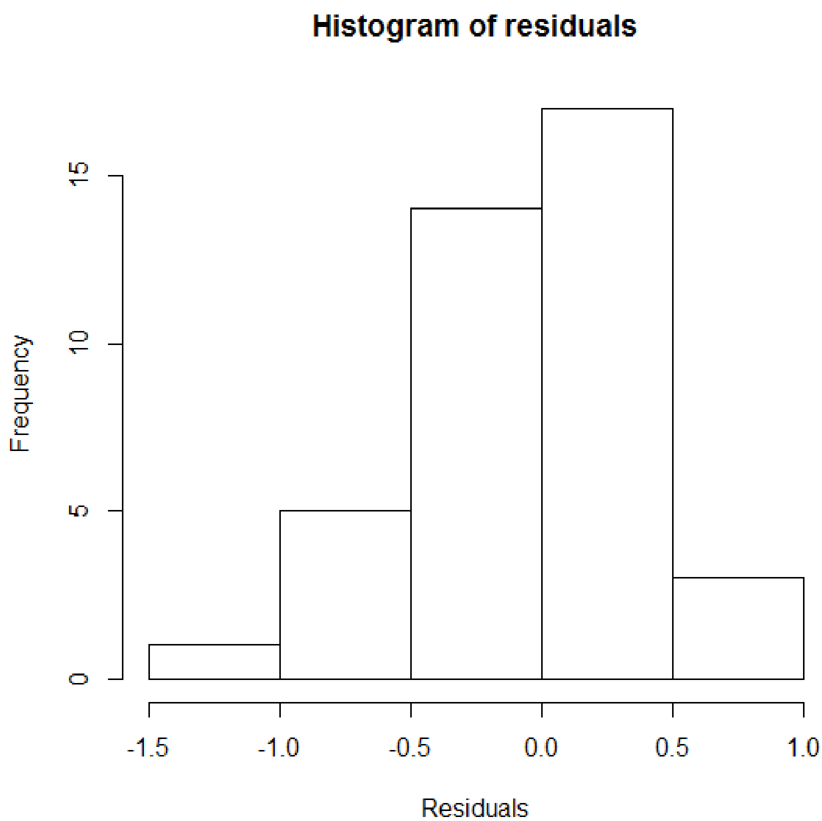
$$cov(x, y) = \frac{\sum_{i=1}^{n}(x - \bar{x})^2 \, (y - \bar{y})^2}{n-1}$$

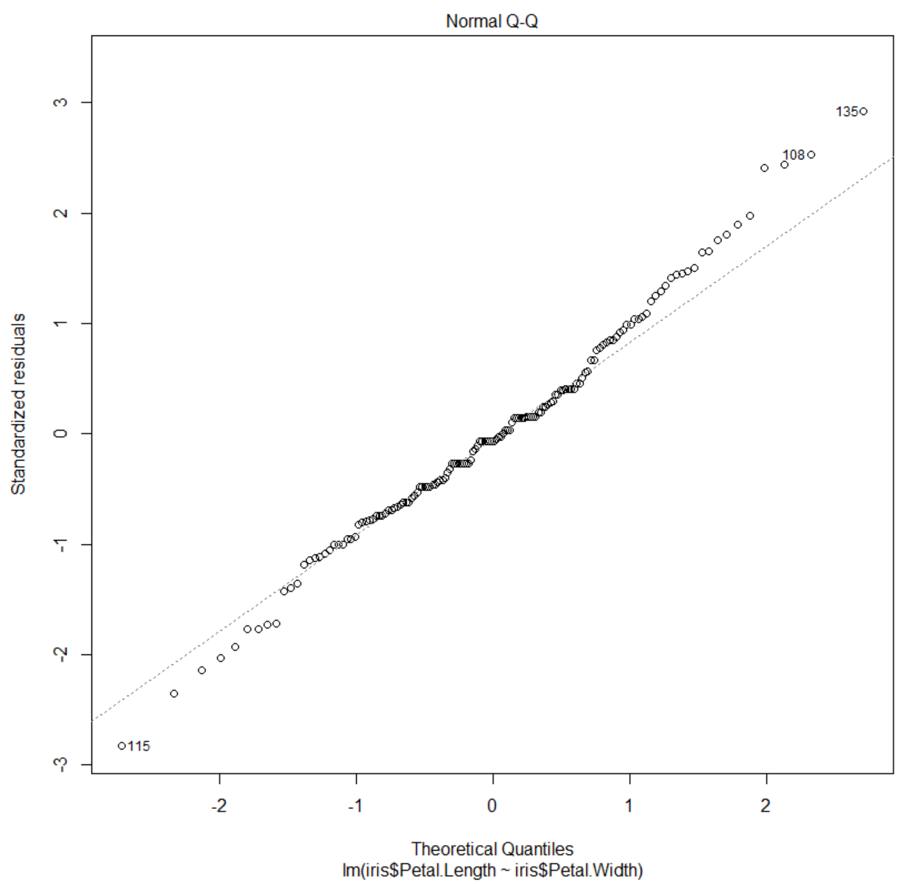
Chapter 9: Linear Regression



Relationship between petal length and petal width
Petal length
Petal width
0
1
2
3
4
5
6
7
0.5
1.0
1.5
2.0
2.5
Intercept
Residual for specific observation
Slope coefficient


Commit
Exhaus
Depers
Accompl
WorkSat
WFC
-5
5
15
30
20
30
40
1.0
2.0
3.5
5.0
6.5
0
10
4
5
6
7

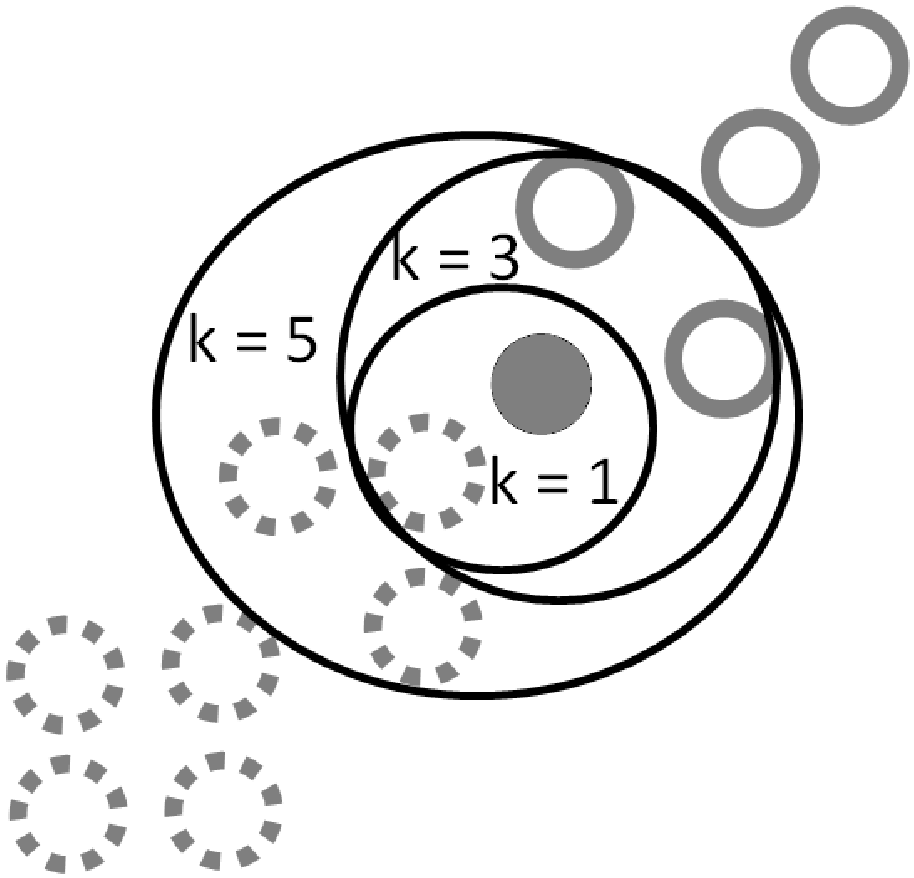

Histogram of residuals
Frequency
0
5
10
15
-1.5
-1.0
-0.5
0.0
0.5
1.0
Residuals



$$z = \frac{a * b}{\sqrt{(b^2 * s_a^2 + a^2 * s_b^2)}}$$

$$\bar{x} \pm z * \frac{s}{\sqrt{n}}$$

Chapter 10: Classification with k-Nearest Neighbors and Naïve Bayes


k = 3
k = 5
k = 1

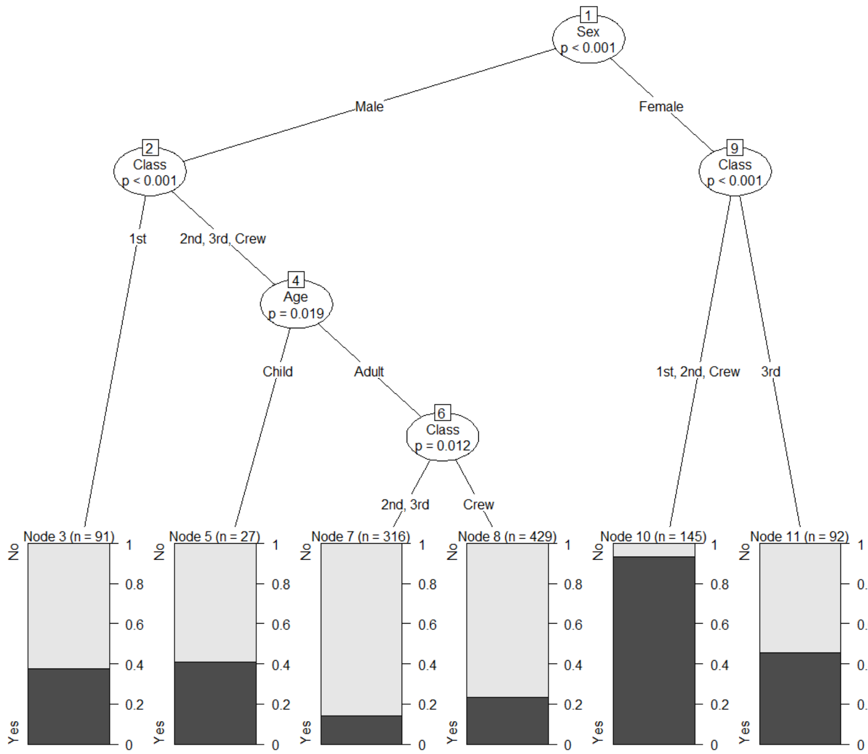
```
Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = as.matrix(DiseaseZ[1:10, 1:6]), y = as.matrix(DiseaseZ[1:10,
    7]))

A-priori probabilities:
as.matrix(DiseaseZ[1:10, 7])
 NO YES
0.4 0.6

Conditional probabilities:
                            Smoking
as.matrix(DiseaseZ[1:10, 7])        NO       YES
                         NO  0.7500000 0.2500000
                         YES 0.3333333 0.6666667

                            Drinking
as.matrix(DiseaseZ[1:10, 7])        NO       YES
                         NO  0.7500000 0.2500000
                         YES 0.1666667 0.8333333

                            PhysicalActivity
as.matrix(DiseaseZ[1:10, 7])   NO  YES
                         NO  0.25 0.75
                         YES 0.50 0.50

                            Movies
as.matrix(DiseaseZ[1:10, 7])        NO       YES
                         NO  0.5000000 0.5000000
                         YES 0.6666667 0.3333333

                            Music
as.matrix(DiseaseZ[1:10, 7])   NO  YES
                         NO  0.75 0.25
                         YES 0.50 0.50

                            Sunbathing
as.matrix(DiseaseZ[1:10, 7])        NO       YES
                         NO  0.7500000 0.2500000
                         YES 0.3333333 0.6666667
```

```
Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = TRAIN[1:3], y = TRAIN[[4]])

A-priori probabilities:
TRAIN[[4]]
       No       Yes
0.6873905 0.3126095

Conditional probabilities:
          Class
TRAIN[[4]]        1st        2nd        3rd       Crew
       No  0.08535032 0.11719745 0.33503185 0.46242038
       Yes 0.30812325 0.15406162 0.23529412 0.30252101

          Sex
TRAIN[[4]]       Male     Female
       No  0.91592357 0.08407643
       Yes 0.50140056 0.49859944

          Age
TRAIN[[4]]      Child      Adult
       No  0.03057325 0.96942675
       Yes 0.07002801 0.92997199
```

$$\sum_{i=1}^{n} |p_i - q_i|$$

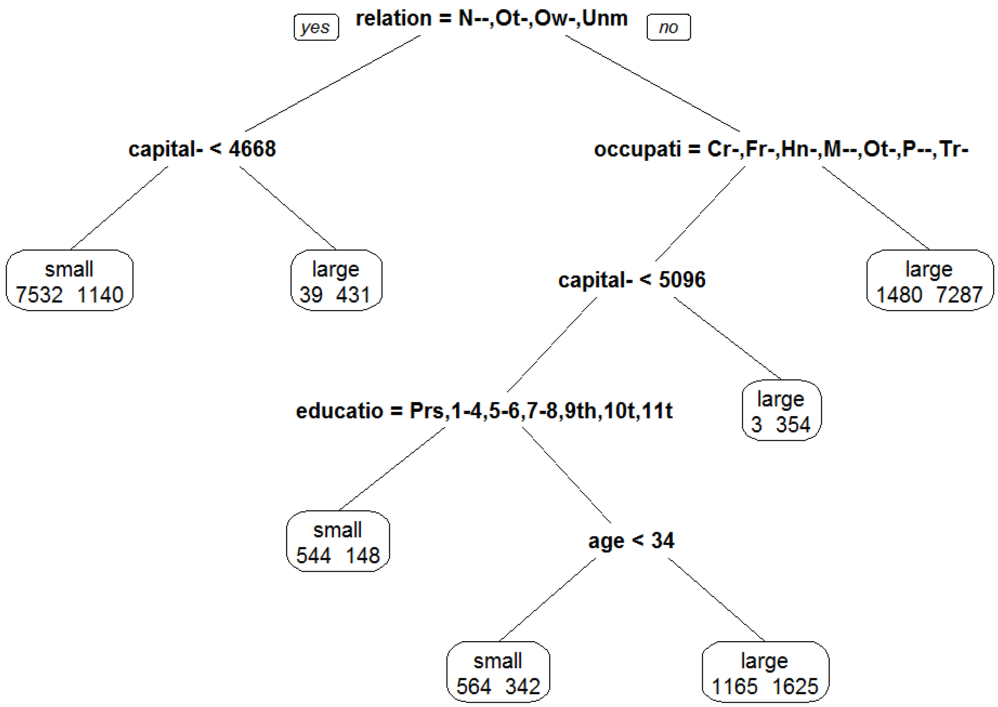
$$\sqrt{\left(\sum_{i=1}^{n} p_i - q_i\right)^2}$$

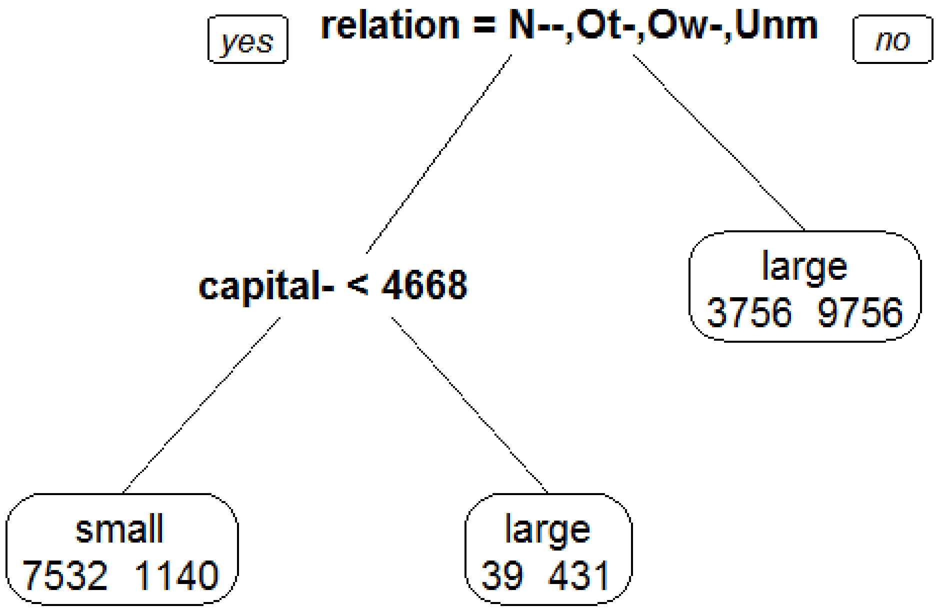
Chapter 11: Classification Trees




```
      age                   workclass          fnlwgt              education      education-num                      marital-status             occupation
 Min.   :17.00   Private            :33906   Min.   :  12285   HS-grad     :15784   Min.   : 1.00   Divorced             : 6633   Prof-specialty : 6172
 1st Qu.:28.00   Self-emp-not-inc: 3862   1st Qu.: 117551   Some-college:10878   1st Qu.: 9.00   Married-AF-spouse    :   37   Craft-repair   : 6112
 Median :37.00   Local-gov          : 3136   Median : 178145   Bachelors   : 8025   Median :10.00   Married-civ-spouse   :22379   Exec-managerial: 6086
 Mean   :38.64   State-gov          : 1981   Mean   : 189664   Masters     : 2657   Mean   :10.08   Married-spouse-absent:  628   Adm-clerical   : 5611
 3rd Qu.:48.00   Self-emp-inc       : 1695   3rd Qu.: 237642   Assoc-voc   : 2061   3rd Qu.:12.00   Never-married        :16117   Sales          : 5504
 Max.   :90.00   (Other)            : 1463   Max.   :1490400   11th        : 1812   Max.   :16.00   Separated            : 1530   (Other)        :16548
                 NA's               : 2799                     (Other)     : 7625                   Widowed              : 1518   NA's           : 2809
         relationship                  race              sex          capital-gain     capital-loss     hours-per-week         native-country    income
 Husband       :19716   Amer-Indian-Eskimo:  470   Female:16192   Min.   :    0   Min.   :   0.0   Min.   : 1.00   United-States:43832   small:24720
 Not-in-family :12583   Asian-Pac-Islander: 1519   Male  :32650   1st Qu.:    0   1st Qu.:   0.0   1st Qu.:40.00   Mexico       :  951   large: 7841
 Other-relative: 1506   Black             : 4685                  Median :    0   Median :   0.0   Median :40.00   Philippines  :  295   NA's :16281
 Own-child     : 7581   Other             :  406                  Mean   : 1079   Mean   :  87.5   Mean   :40.42   Germany      :  206
 Unmarried     : 5125   White             :41762                  3rd Qu.:    0   3rd Qu.:   0.0   3rd Qu.:45.00   Puerto-Rico  :  184
 Wife          : 2331                                             Max.   :99999   Max.   :4356.0   Max.   :99.00   (Other)      : 2517
                                                                                                                   NA's         :  857
```

```
Correctly Classified Instances       20206               89.194  %
Incorrectly Classified Instances      2448               10.806  %
Kappa statistic                          0.7839
Mean absolute error                      0.1448
Root mean squared error                  0.2836
Relative absolute error                 28.9603 %
Root relative squared error             56.7242 %
Coverage of cases (0.95 level)          98.2829 %
Mean rel. region size (0.95 level)      69.6654 %
Total Number of Instances            22654

=== Confusion Matrix ===

     a     b   <-- classified as
  9699  1628 |     a = small
   820 10507 |     b = large
```



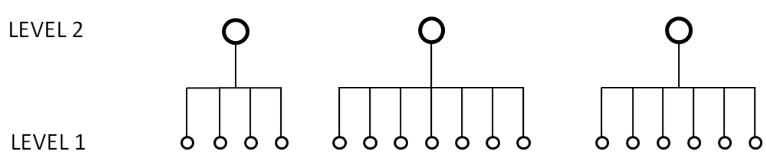


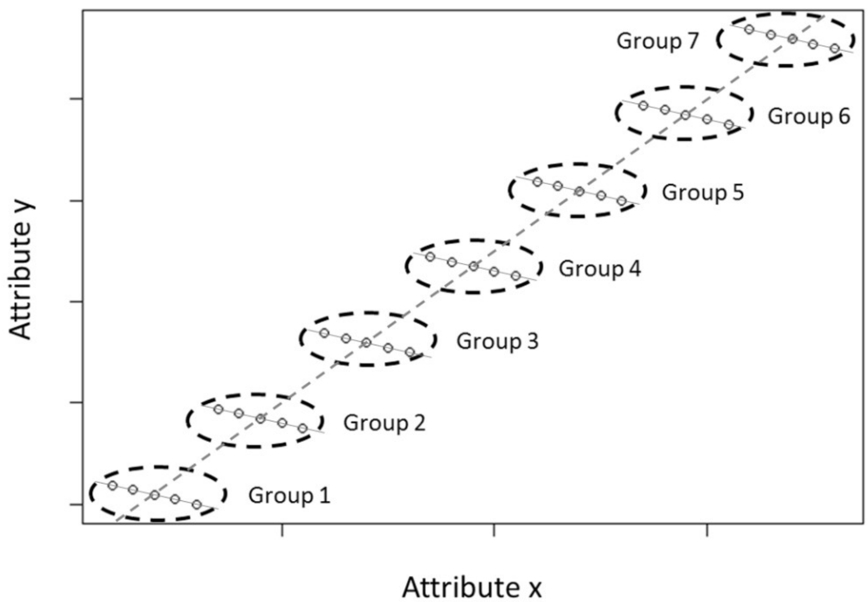
$$-\sum_{i=1}^{c} p_i (log_2\, p_i)$$

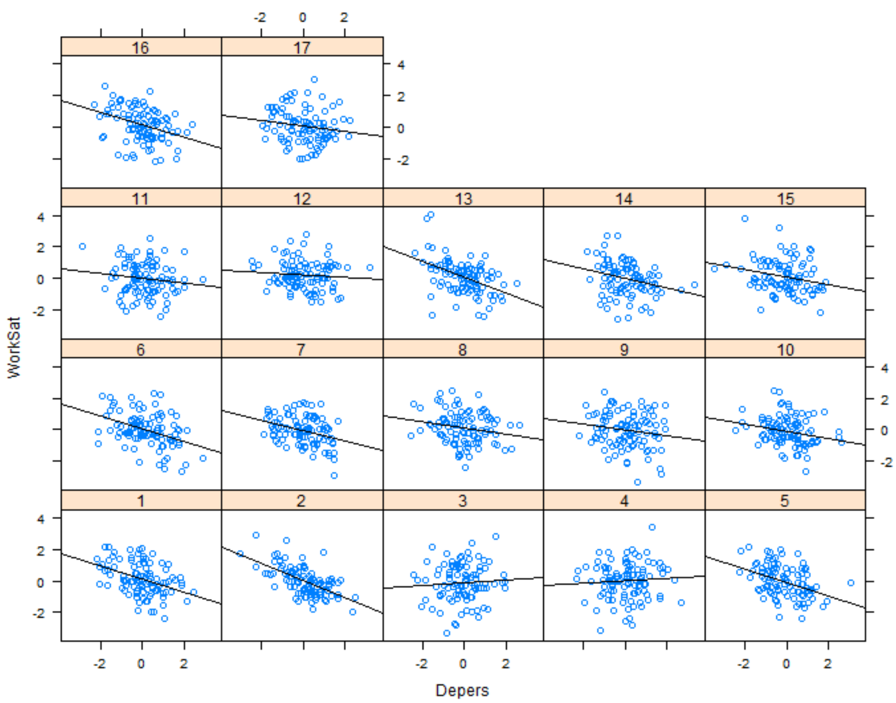
$$-\sum_{i=1}^{c} \frac{|Ti|}{|T|}\, log_2 \frac{|Ti|}{|T|}$$

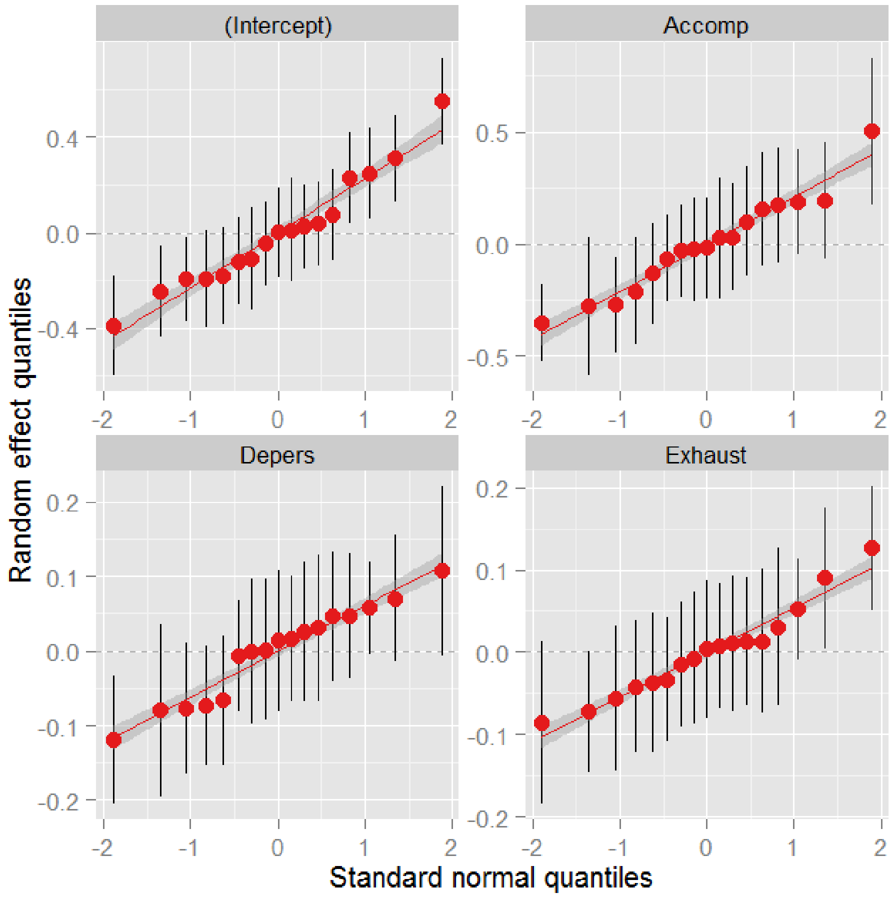
$$1 - \sum_{j=1}^{c} p_j^2$$

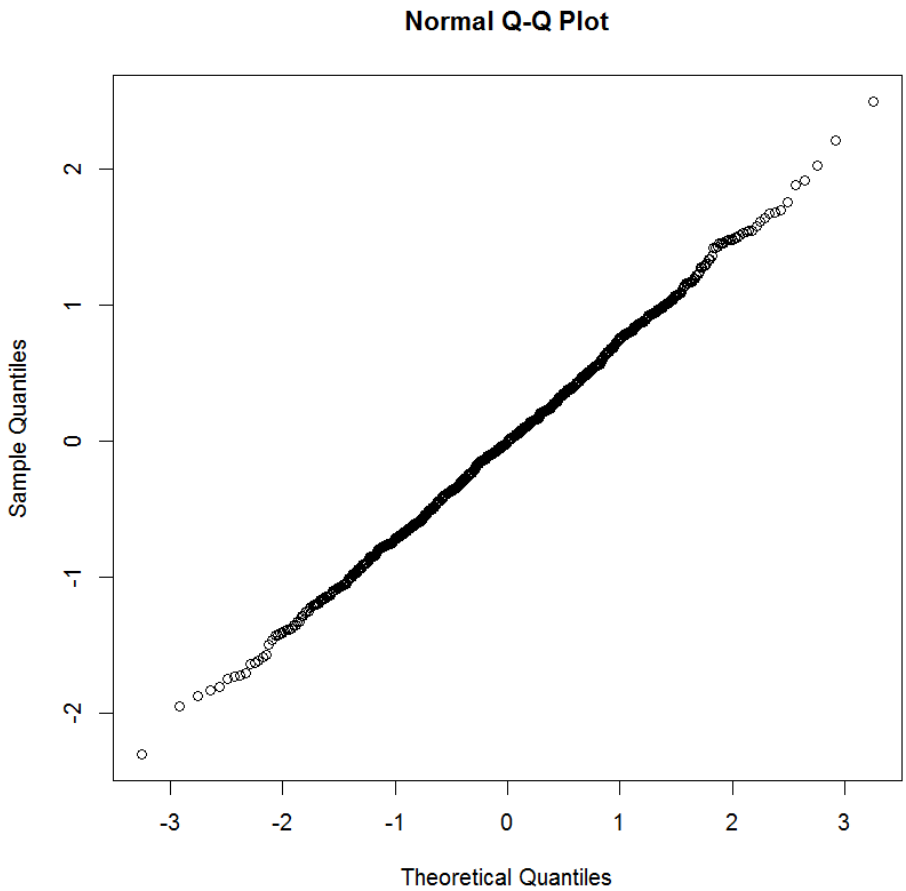
Chapter 12: Multilevel Analyses



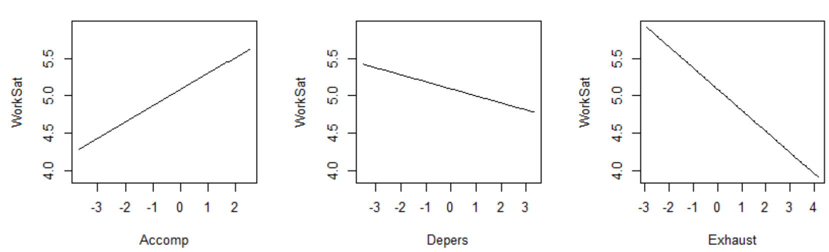

Group 7
Group 6
Group 5
Group 4
Group 3
Group 2
Group 1
Attribute y
Attribute x


16
17
11
12
13
14
15
6
7
8
9
10
1
2
3
4
5
-2
0
2
4
WorkSat
Depers


(Intercept)
Accomp
Depers
Exhaust
Random effect quantiles
Standard normal quantiles
0.4
0.0
-0.4
0.5
0.0
-0.5
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2



Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-3
-2
-1
0
1
2
3



WorkSat
Accomp
Depers
Exhaust
4.0
4.5
5.0
5.5

```
Linear mixed model fit by REML ['lmerMoc
Formula: WorkSat ~ 1 + (1 | hosp)
   Data: NursesML

REML criterion at convergence: 4321.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8527 -0.6556 -0.0038  0.6823  4.0239

Random effects:
 Groups   Name        Variance Std.Dev.
 hosp     (Intercept) 0.06988  0.2643
 Residual             0.72564  0.8518
Number of obs: 1700, groups:  hosp, 17

Fixed effects:
            Estimate Std. Error t value
(Intercept)  5.10679    0.06736   75.81
```

```
Data: NursesMLtrain
Models:
null: WorkSat ~ 1 + (1 | hosp)
model: WorkSat ~ Accomp + Depers + Exhaust + (1 | hosp)
      Df    AIC    BIC   logLik deviance  Chisq Chi Df Pr(>Chisq)
null   3 2156.3 2170.5 -1075.15   2150.3
model  6 1984.2 2012.6  -986.08   1972.2 178.15      3  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

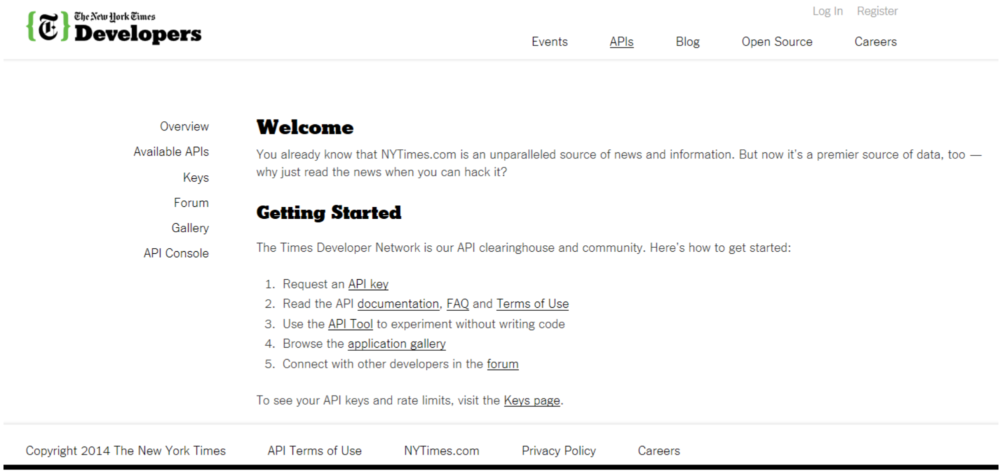
```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: WorkSat ~ Accomp + Depers + Exhaust + (1 | hosp)
   Data: NursesMLtrain

     AIC      BIC   logLik deviance df.resid
  1984.2   2012.6   -986.1   1972.2      844

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0252 -0.6756  0.0225  0.6616  3.4649

Random effects:
 Groups   Name        Variance Std.Dev.
 hosp     (Intercept) 0.06674  0.2583
 Residual             0.57345  0.7573
Number of obs: 850, groups:  hosp, 17

Fixed effects:
            Estimate Std. Error t value
(Intercept)  5.11854    0.06783   75.46
Accomp       0.17611    0.03749    4.70
Depers      -0.07335    0.03135   -2.34
Exhaust     -0.29215    0.02935   -9.95

Correlation of Fixed Effects:
        (Intr) Accomp Depers
Accomp   0.000
Depers   0.000  0.041
Exhaust  0.000  0.044 -0.490
```

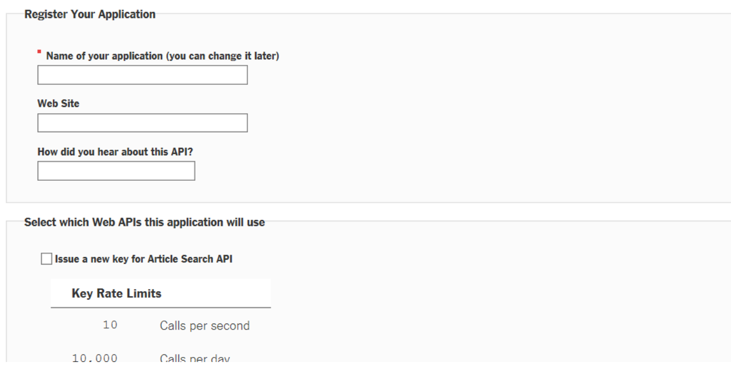
```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: WorkSat ~ Accomp + Depers + Exhaust + (1 + Accomp + Depers +
    Exhaust | hosp)
   Data: NursesMLtrain

     AIC      BIC   logLik deviance df.resid
  1970.5   2041.7   -970.2   1940.5      835

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1422 -0.6826  0.0028  0.6510  3.4045

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 hosp     (Intercept) 0.060492 0.24595
          Accomp      0.058726 0.24233  -0.56
          Depers      0.005805 0.07619  -0.17  0.83
          Exhaust     0.004594 0.06778   0.74 -0.24  0.34
 Residual             0.536849 0.73270
Number of obs: 850, groups:  hosp, 17

Fixed effects:
            Estimate Std. Error t value
(Intercept)  5.06994    0.06531   77.63
Accomp       0.21722    0.07044    3.08
Depers      -0.09408    0.03628   -2.59
Exhaust     -0.28444    0.03371   -8.44

Correlation of Fixed Effects:
        (Intr) Accomp Depers
Accomp  -0.448
Depers  -0.074  0.384
Exhaust  0.336 -0.073 -0.251
```

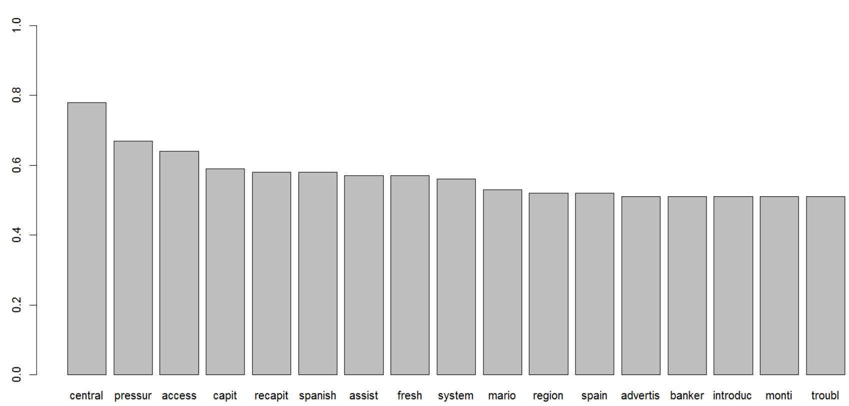
Chapter 13: Text Analytics with R

```
> Terms.seasonal
      Topic 1     Topic 2
 [1,] "vaccin"    "year"
 [2,] "state"     "peopl"
 [3,] "get"       "influenza"
 [4,] "com"       "center"
 [5,] "dai"       "time"
 [6,] "month"     "diseas"
 [7,] "million"   "strain"
 [8,] "viru"      "doctor"
 [9,] "yesterdai" "case"
[10,] "report"    "nytim"
[11,] "week"      "winter"
[12,] "shot"      "week"
[13,] "feder"     "protect"
[14,] "drug"      "url"
[15,] "nytim"     "control"
[16,] "death"     "sai"
[17,] "problem"   "nation"
[18,] "countri"   "recommend"
[19,] "season"    "ag"
[20,] "expect"    "risk"
```

```
> Terms.non.seasonal
      Topic 1   Topic 2
 [1,] "nytim"   "infect"
 [2,] "offici"  "diseas"
 [3,] "viru"    "year"
 [4,] "world"   "week"
 [5,] "million" "peopl"
 [6,] "prevent" "outbreak"
 [7,] "nation"  "pandem"
 [8,] "work"    "influenza"
 [9,] "come"    "sai"
[10,] "confirm" "countri"
[11,] "unit"    "case"
[12,] "peopl"   "com"
[13,] "found"   "human"
[14,] "strain"  "url"
[15,] "month"   "di"
[16,] "case"    "test"
[17,] "start"   "spread"
[18,] "get"     "govern"
[19,] "report"  "includ"
[20,] "effect"  "viru"
```


The New York Times Developers
Log In Register
Events APIs Blog Open Source Careers
Overview
Available APIs
Keys
Forum
Gallery
API Console
Welcome
You already know that NYTimes.com is an unparalleled source of news and information. But now it's a premier source of data, too — why just read the news when you can hack it?
Getting Started
The Times Developer Network is our API clearinghouse and community. Here's how to get started:
1. Request an API key
2. Read the API documentation, FAQ and Terms of Use
3. Use the API Tool to experiment without writing code
4. Browse the application gallery
5. Connect with other developers in the forum
To see your API keys and rate limits, visit the Keys page.
Copyright 2014 The New York Times API Terms of Use NYTimes.com Privacy Policy Careers

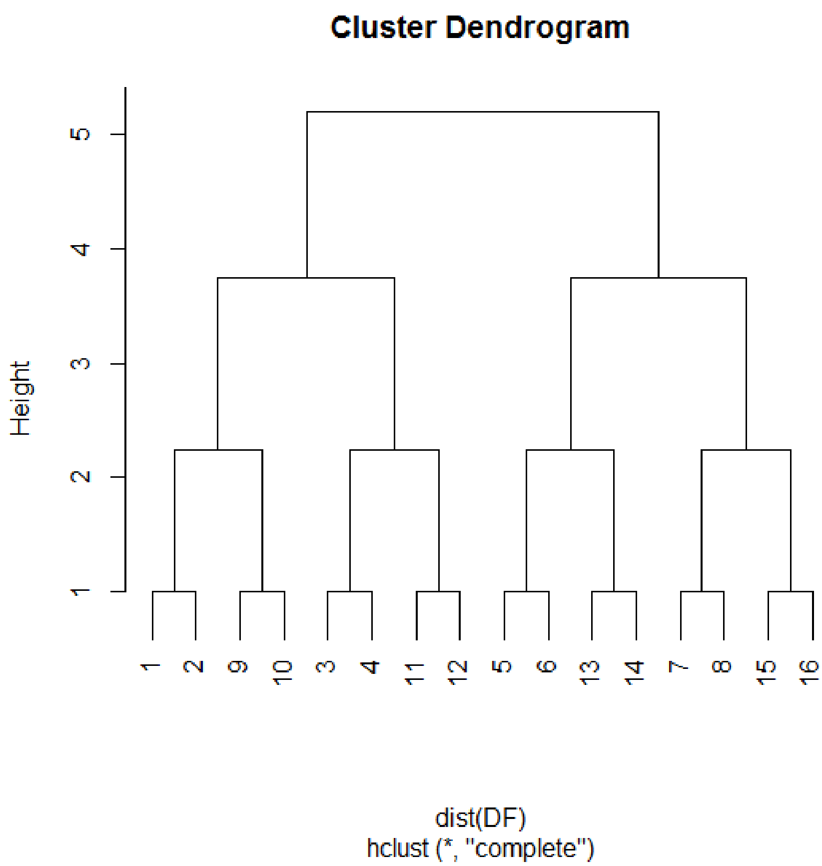

Register Your Application
* Name of your application (you can change it later)
Web Site
How did you hear about this API?
Select which Web APIs this application will use
Issue a new key for Article Search API
Key Rate Limits
10 Calls per second


0.0
0.2
0.4
0.6
0.8
1.0
central
pressur
access
capit
recapit
spanish
assist
fresh
system
mario
region
spain
advertis
banker
introduc
monti
troubl

Chapter 14: Cross-validation and Bootstrapping using Caret and Exporting Predictive Models Using PMML

Fold	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Iteration 6	Iteration 7	Iteration 8	Iteration 9	Iteration 10
Fold 1	Train	Train	Train	Train	Train	Train	Train	Train	Train	Test
Fold 2	Train	Train	Train	Train	Train	Train	Train	Train	Test	Train
Fold 3	Train	Train	Train	Train	Train	Train	Train	Test	Train	Train
Fold 4	Train	Train	Train	Train	Train	Train	Test	Train	Train	Train
Fold 5	Train	Train	Train	Train	Train	Test	Train	Train	Train	Train
Fold 6	Train	Train	Train	Train	Test	Train	Train	Train	Train	Train
Fold 7	Train	Train	Train	Test	Train	Train	Train	Train	Train	Train
Fold 8	Train	Train	Test	Train	Train	Train	Train	Train	Train	Train
Fold 9	Train	Test	Train	Train	Train	Train	Train	Train	Train	Train
Fold 10	Test	Train	Train	Train	Train	Train	Train	Train	Train	Train


```xml
<?xml version="1.0"?>
<PMML version="4.2" xmlns="http://www.dmg.org/PMML-4_2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.dmg.org/PMML-4_2 http://www.dmg.org/v4-2/pmml-4-2.xsd">
 <Header copyright="Copyright (c) 2015 mayore" description="KMeans cluster model">
  <Extension name="user" value="mayore" extender="Rattle/PMML"/>
  <Application name="Rattle/PMML" version="1.4"/>
  <Timestamp>2015-02-18 14:39:28</Timestamp>
 </Header>
 <DataDictionary numberOfFields="4">
  <DataField name="Sepal.Length" optype="continuous" dataType="double"/>
  <DataField name="Sepal.Width" optype="continuous" dataType="double"/>
  <DataField name="Petal.Length" optype="continuous" dataType="double"/>
  <DataField name="Petal.Width" optype="continuous" dataType="double"/>
 </DataDictionary>
 <ClusteringModel modelName="KMeans_Model" functionName="clustering" algorithmName="KMeans: Hartigan and Wong"
modelClass="centerBased" numberOfClusters="3">
  <MiningSchema>
   <MiningField name="Sepal.Length"/>
   <MiningField name="Sepal.Width"/>
   <MiningField name="Petal.Length"/>
   <MiningField name="Petal.Width"/>
  </MiningSchema>
  <Output>
   <OutputField name="predictedValue" feature="predictedValue"/>
   <OutputField name="clusterAffinity_1" feature="clusterAffinity" value="1"/>
   <OutputField name="clusterAffinity_2" feature="clusterAffinity" value="2"/>
   <OutputField name="clusterAffinity_3" feature="clusterAffinity" value="3"/>
  </Output>
  <ComparisonMeasure kind="distance">
   <squaredEuclidean/>
  </ComparisonMeasure>
  <ClusteringField field="Sepal.Length" compareFunction="absDiff"/>
  <ClusteringField field="Sepal.Width" compareFunction="absDiff"/>
  <ClusteringField field="Petal.Length" compareFunction="absDiff"/>
  <ClusteringField field="Petal.Width" compareFunction="absDiff"/>
  <Cluster name="1" size="38" id="1">
   <Array n="4" type="real">6.85 3.07368421052632 5.74210526315789 2.07105263157895</Array>
  </Cluster>
  <Cluster name="2" size="50" id="2">
   <Array n="4" type="real">5.006 3.428 1.462 0.246</Array>
  </Cluster>
  <Cluster name="3" size="62" id="3">
   <Array n="4" type="real">5.90161290322581 2.74838709677419 4.39354838709678 1.43387096774194</Array>
  </Cluster>
 </ClusteringModel>
</PMML>
```


Cluster Dendrogram
Height
1
2
3
4
5
1 2 9 10 3 4 11 12 5 6 13 14 7 8 15 16
dist(DF)
hclust (*, "complete")