Lesson 1: Introduction to Clustering Methods

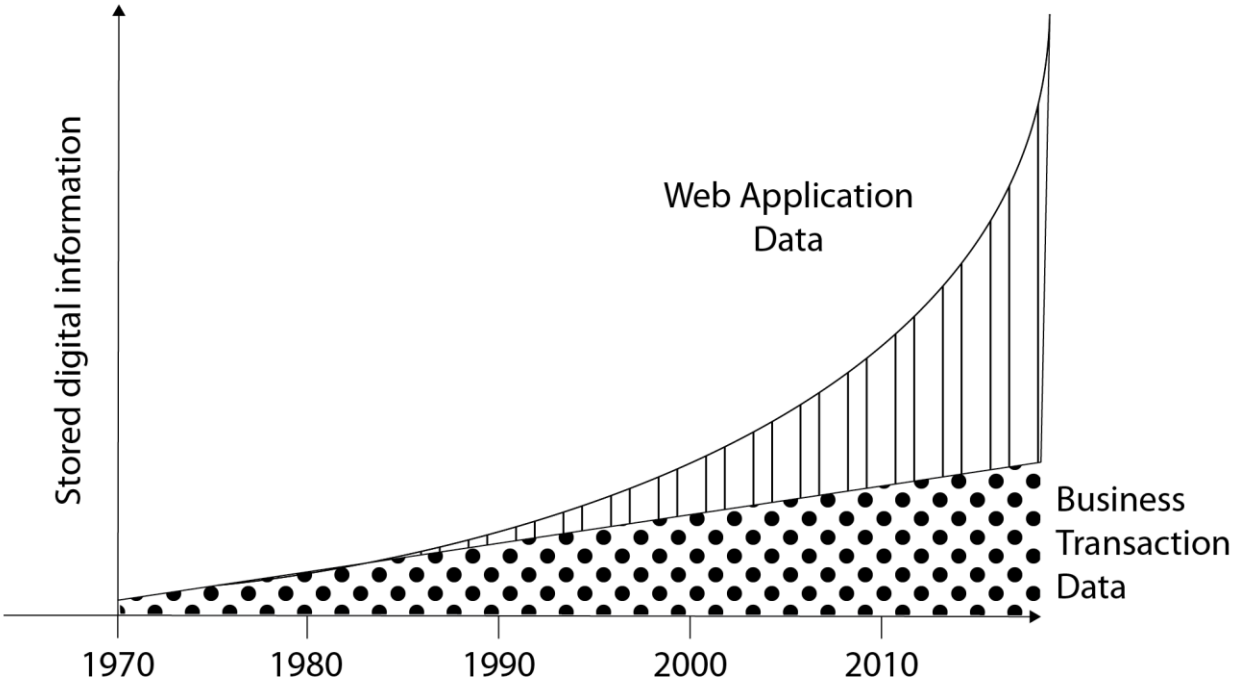


Figure 1.1: The increase in digital data year on year

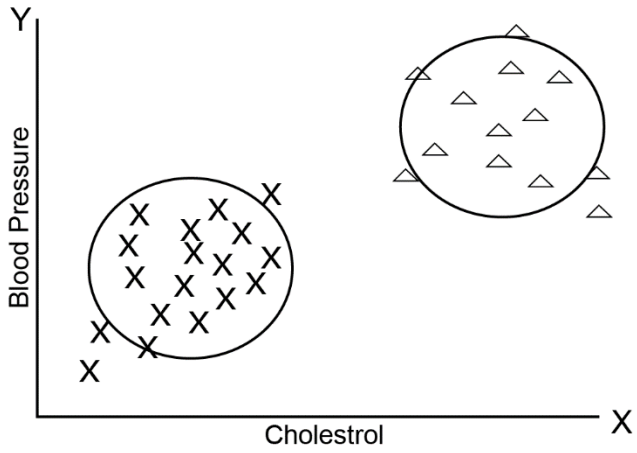


Figure 1.2: A representation of two clusters in a dataset

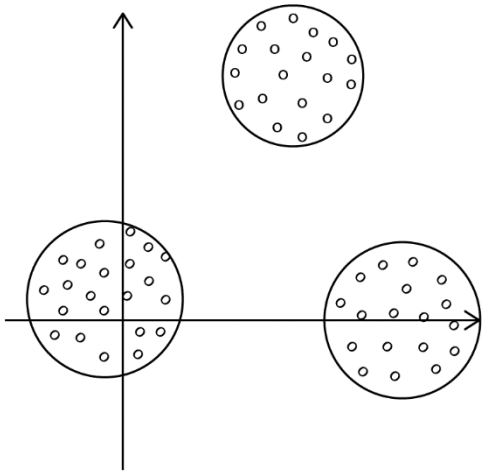


Figure 1.3: A representation of three clusters in a dataset

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

Figure 1.4: The first six rows of the Iris dataset

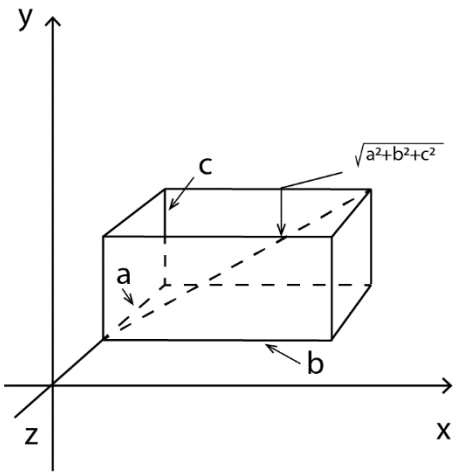


Figure 1.5: Representation of Euclidean distance calculation

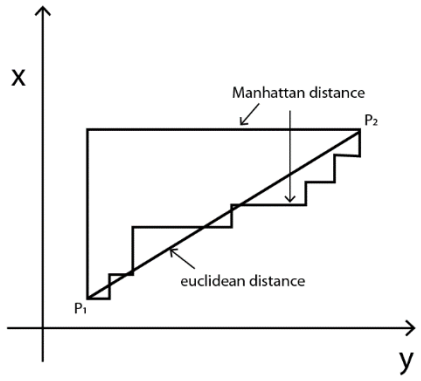


Figure 1.6: Representation of Manhattan distance

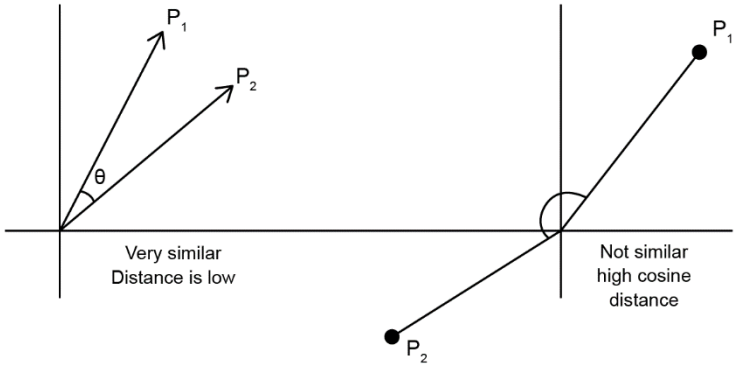


Figure 1.7: Representation of cosine similarity and cosine distance

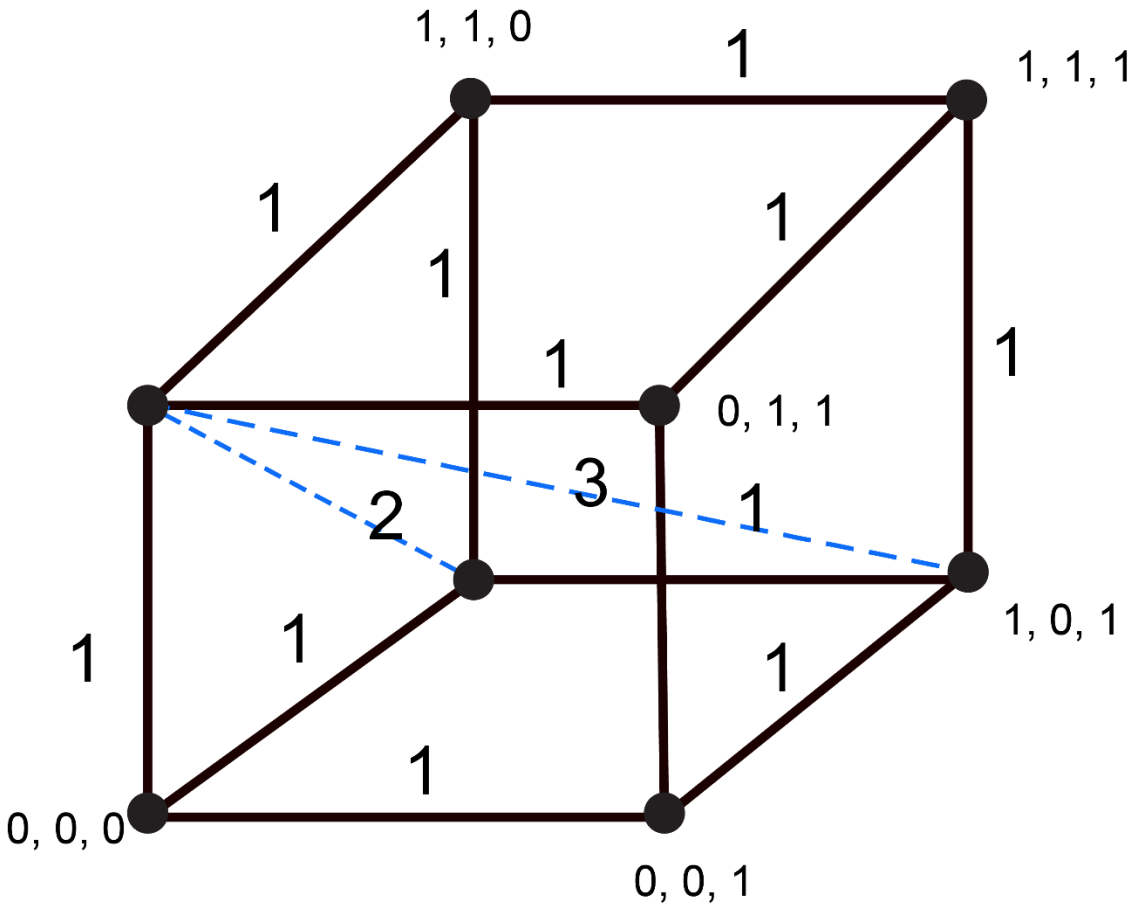


Figure 1.8: Representation of the Hamming distance

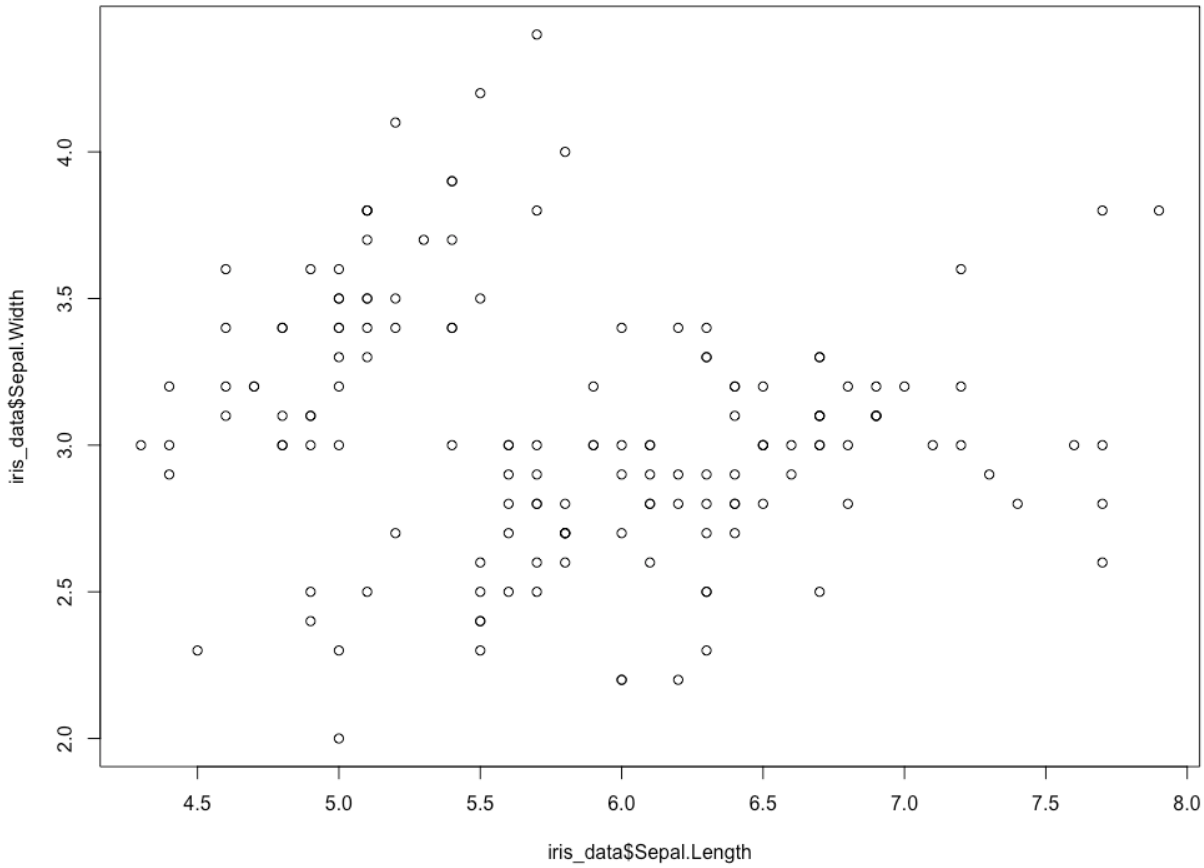


Figure 1.9: A scatter plot of the Iris flowers dataset

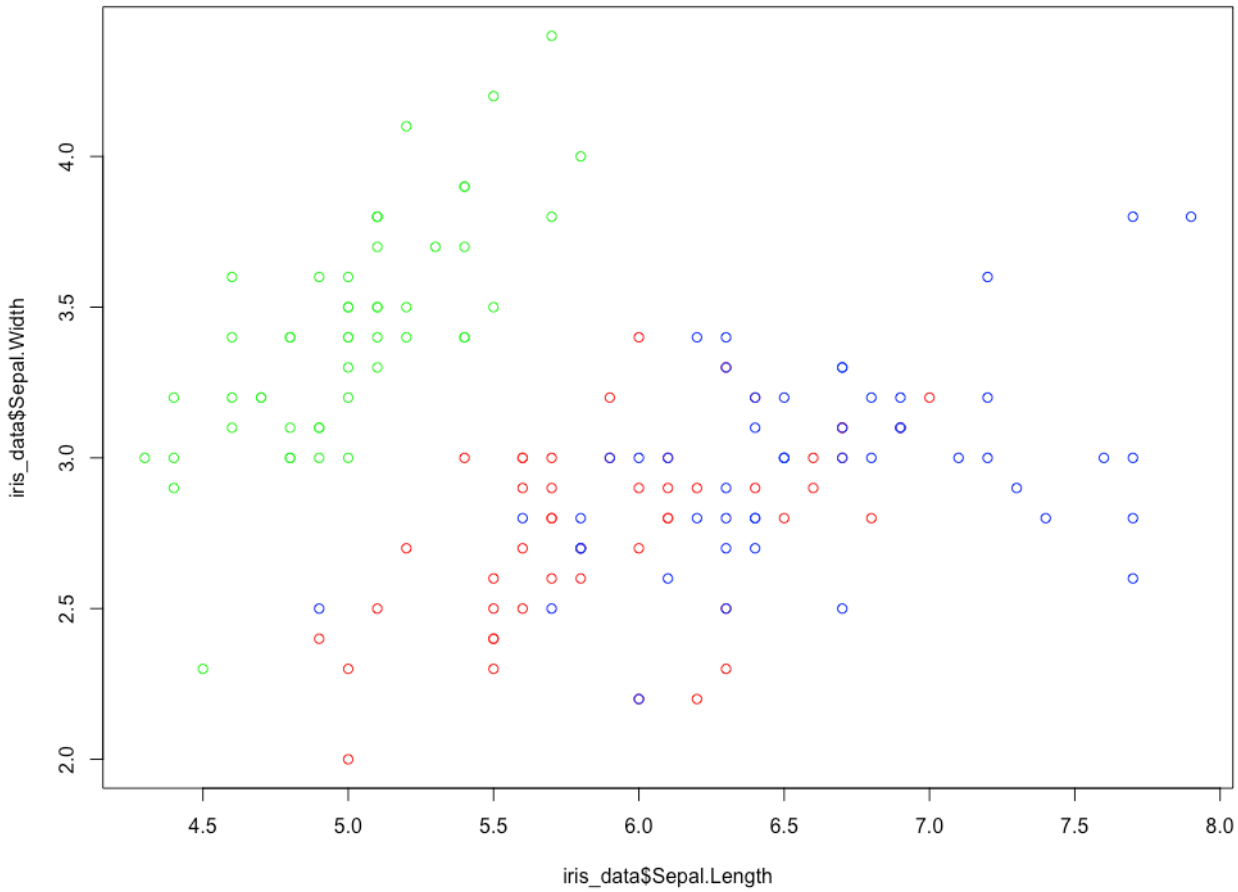


Figure 1.10: A scatter plot showing different species of Iris flowers dataset

Figure 1.11: Iris setosa

Figure 1.12: Iris versicolor

Figure 1.13: Iris virginica

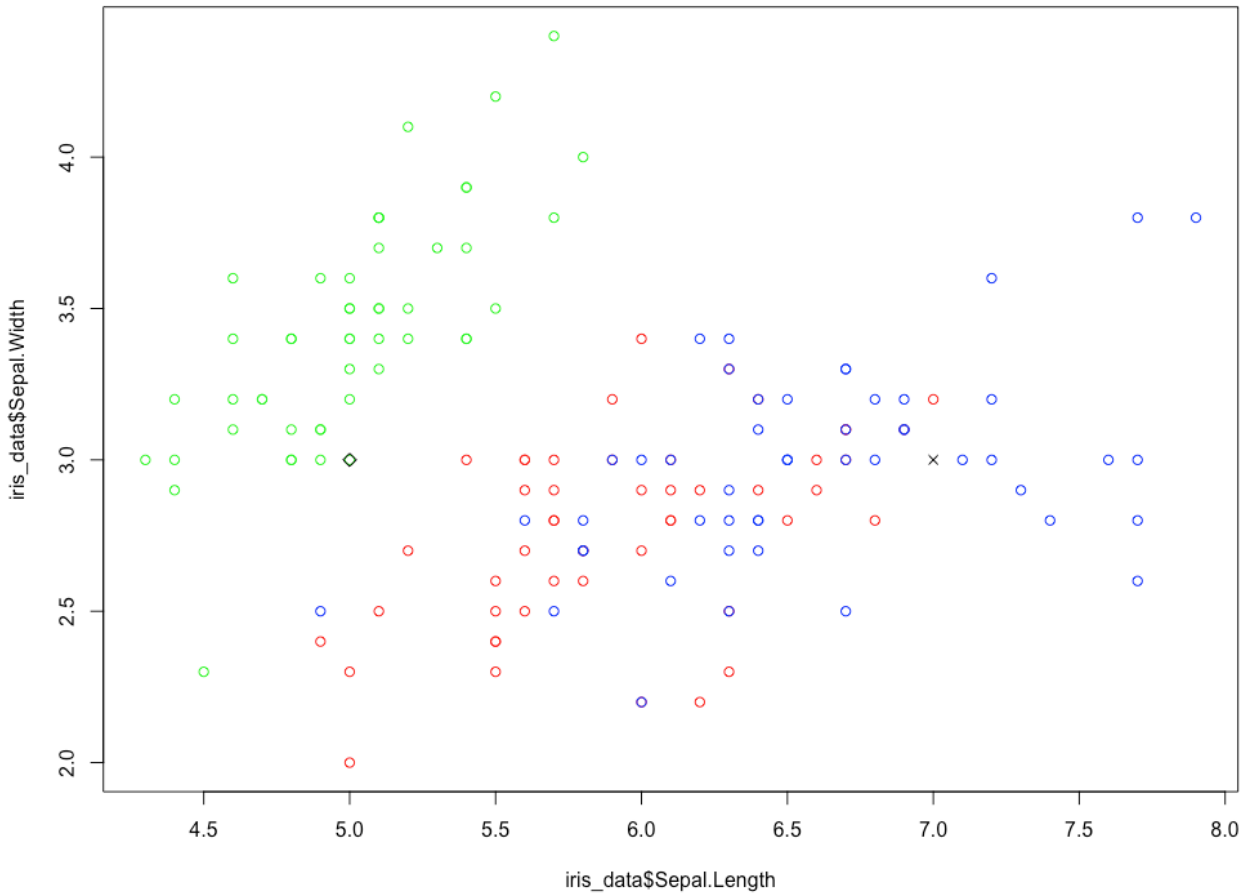


Figure 1.14: A scatter plot of the chosen cluster centers

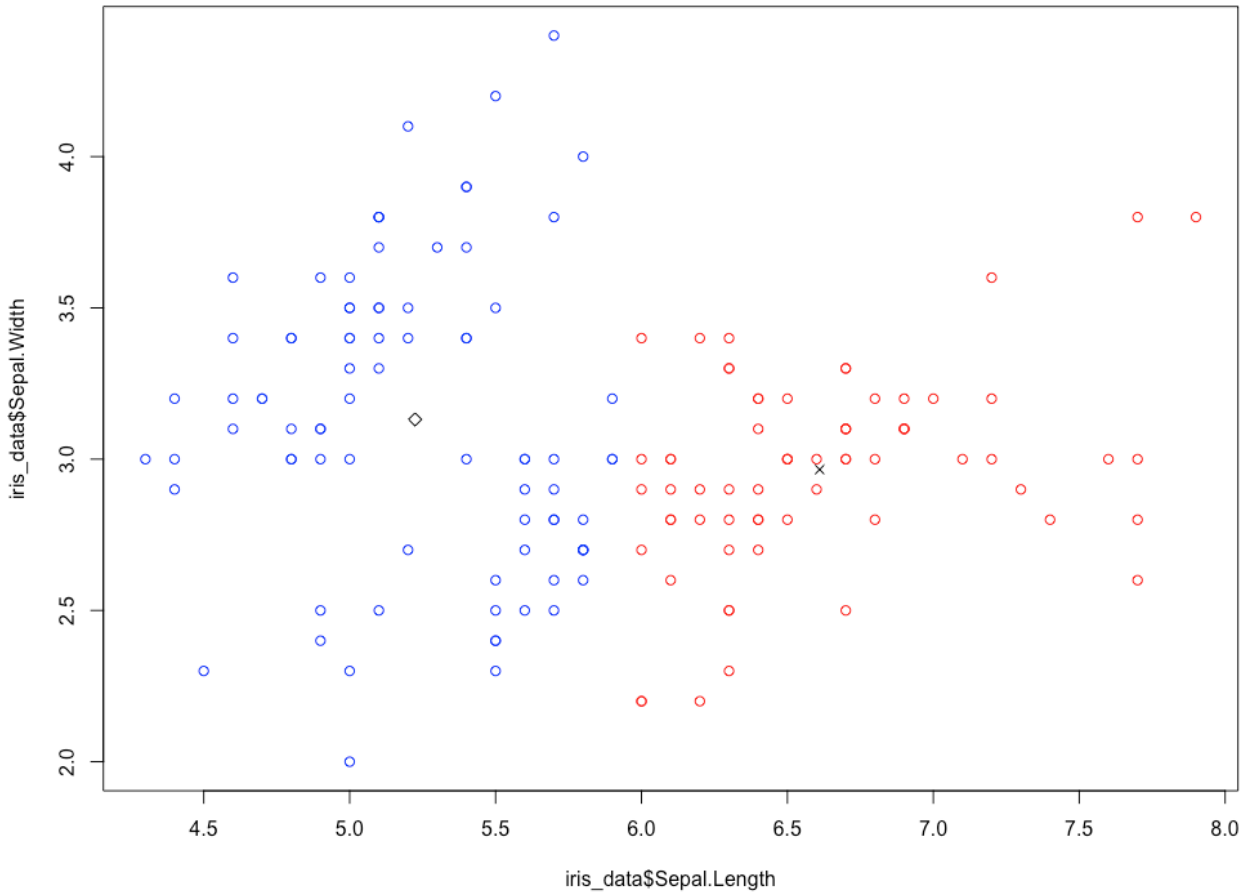


Figure 1.15: A scatter plot representing each cluster with a different color

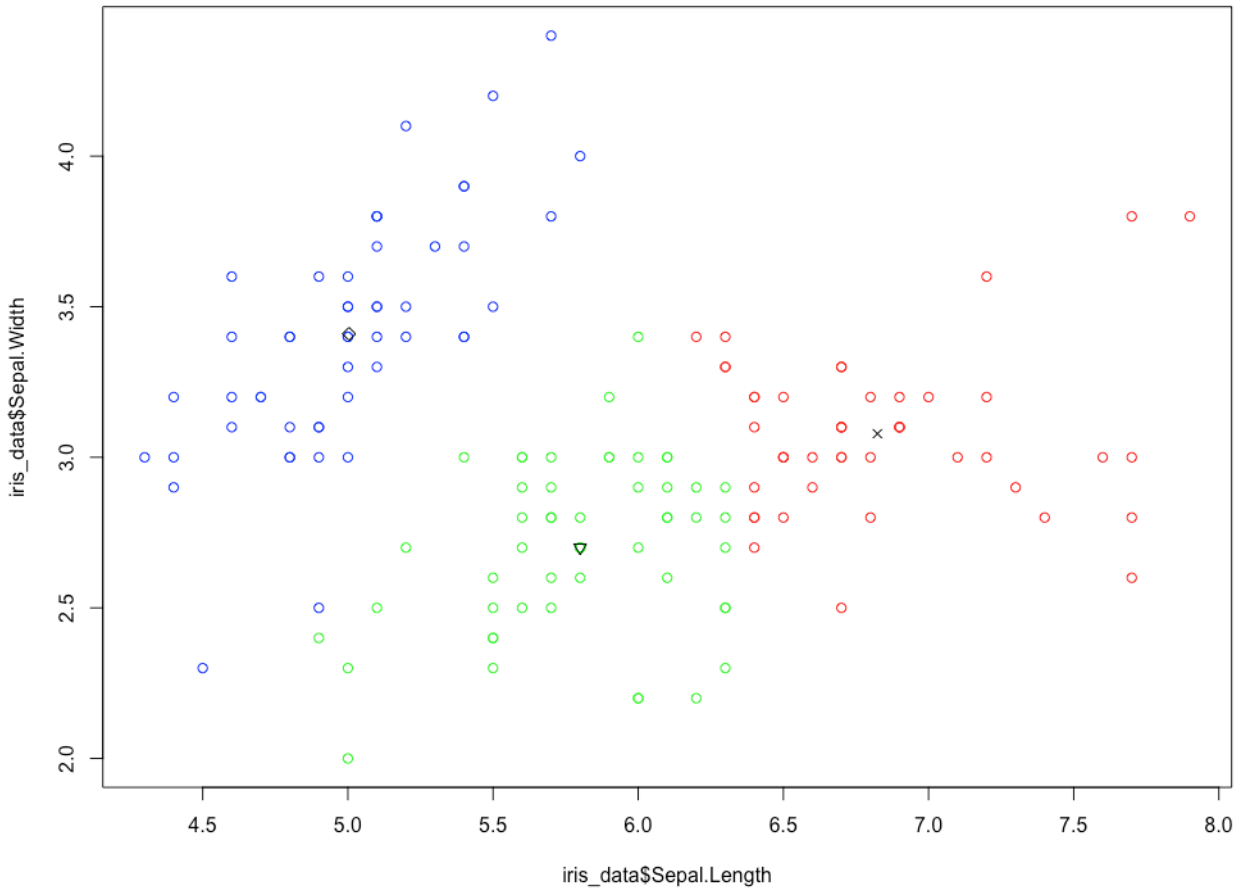


Figure 1.16: The expected scatter plot for the given cluster centers

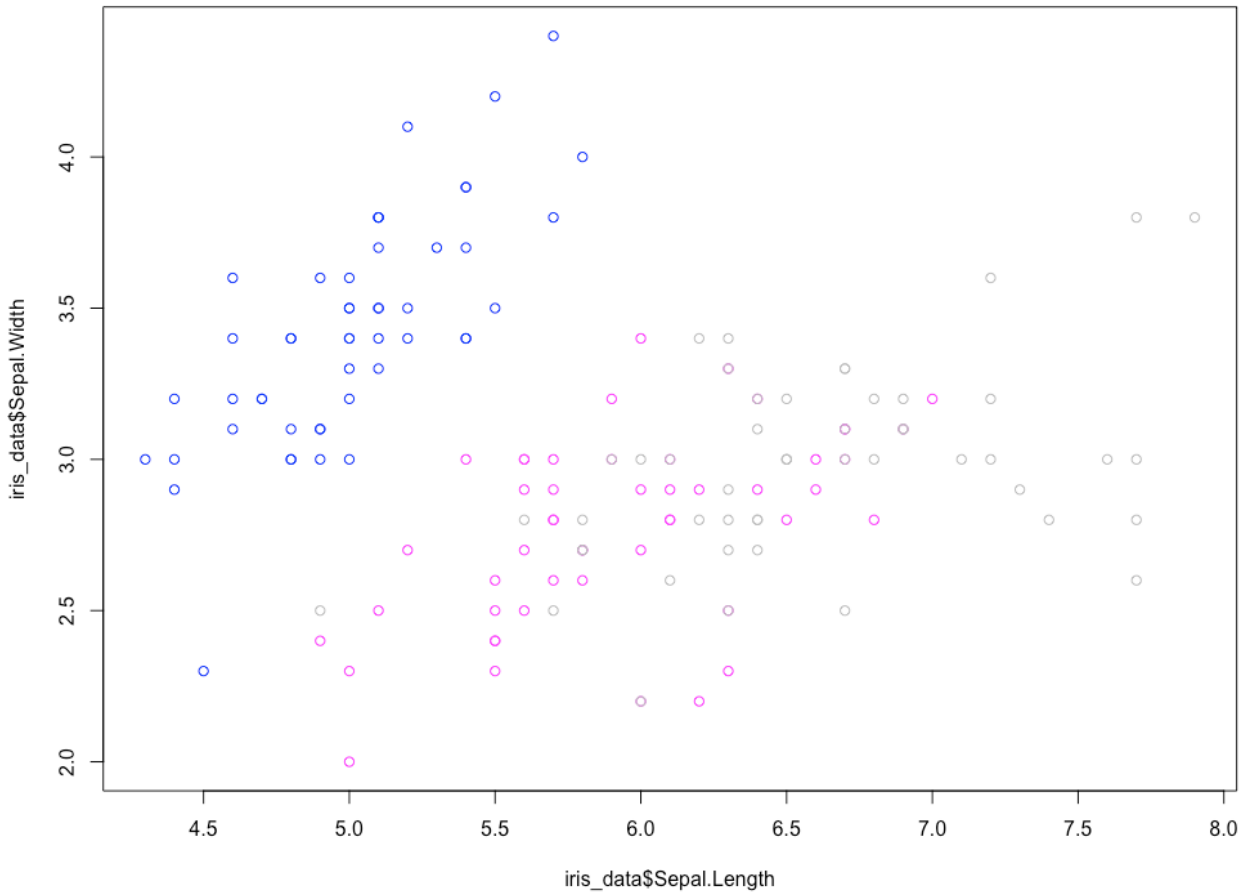


Figure 1.17: A graph representing three species of iris in three colors

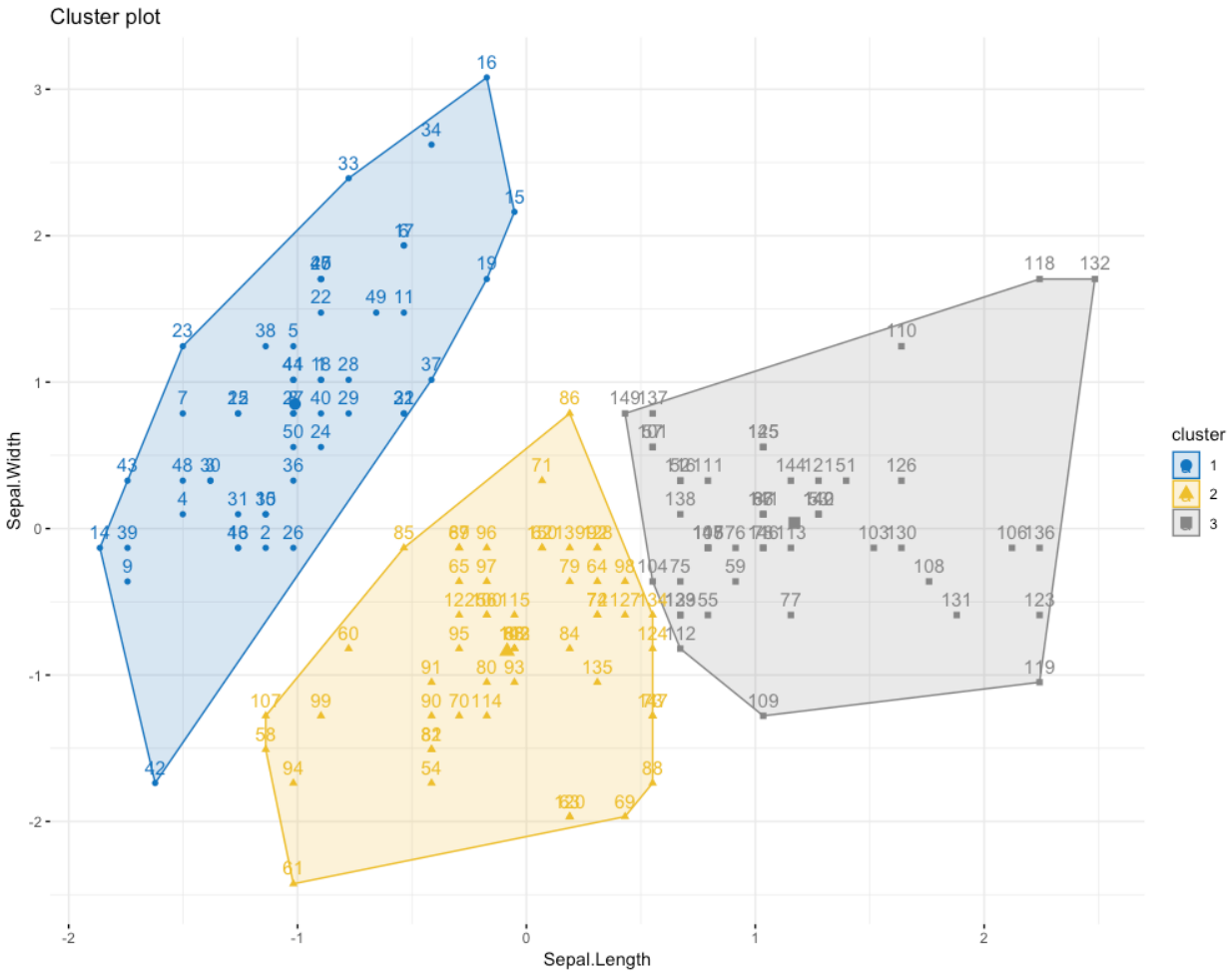


Figure 1.18: Three species of Iris have been clustered into three clusters

	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
1	2	3	12669	9656	7561	214	2674	1338
2	2	3	7057	9810	9568	1762	3293	1776
3	2	3	6353	8808	7684	2405	3516	7844
4	1	3	13265	1196	4221	6404	507	1788
5	2	3	22615	5410	7198	3915	1777	5185
6	2	3	9413	8259	5126	666	1795	1451

Figure 1.19: Columns of the wholesale customer dataset

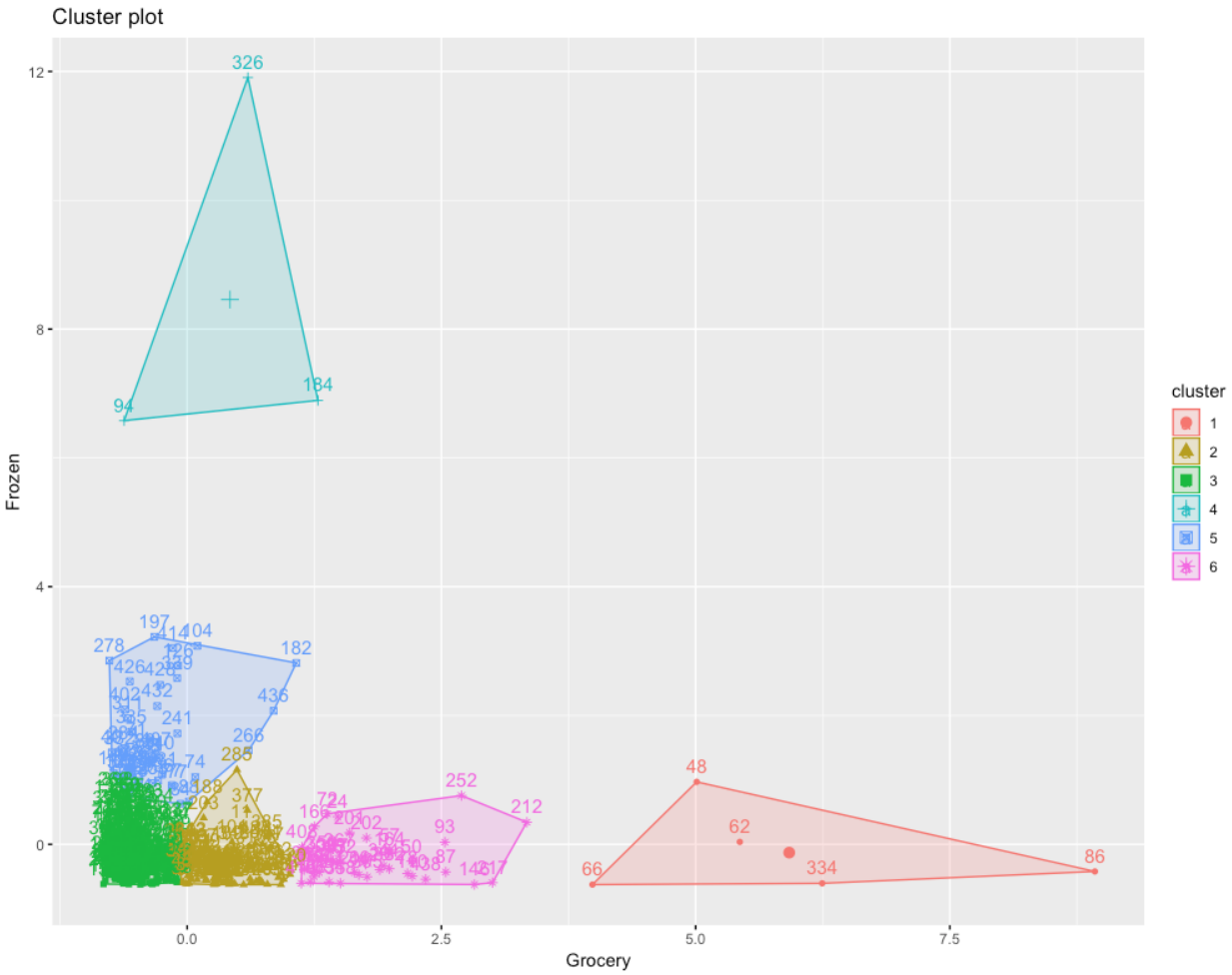


Figure 1.20: The expected chart for six clusters

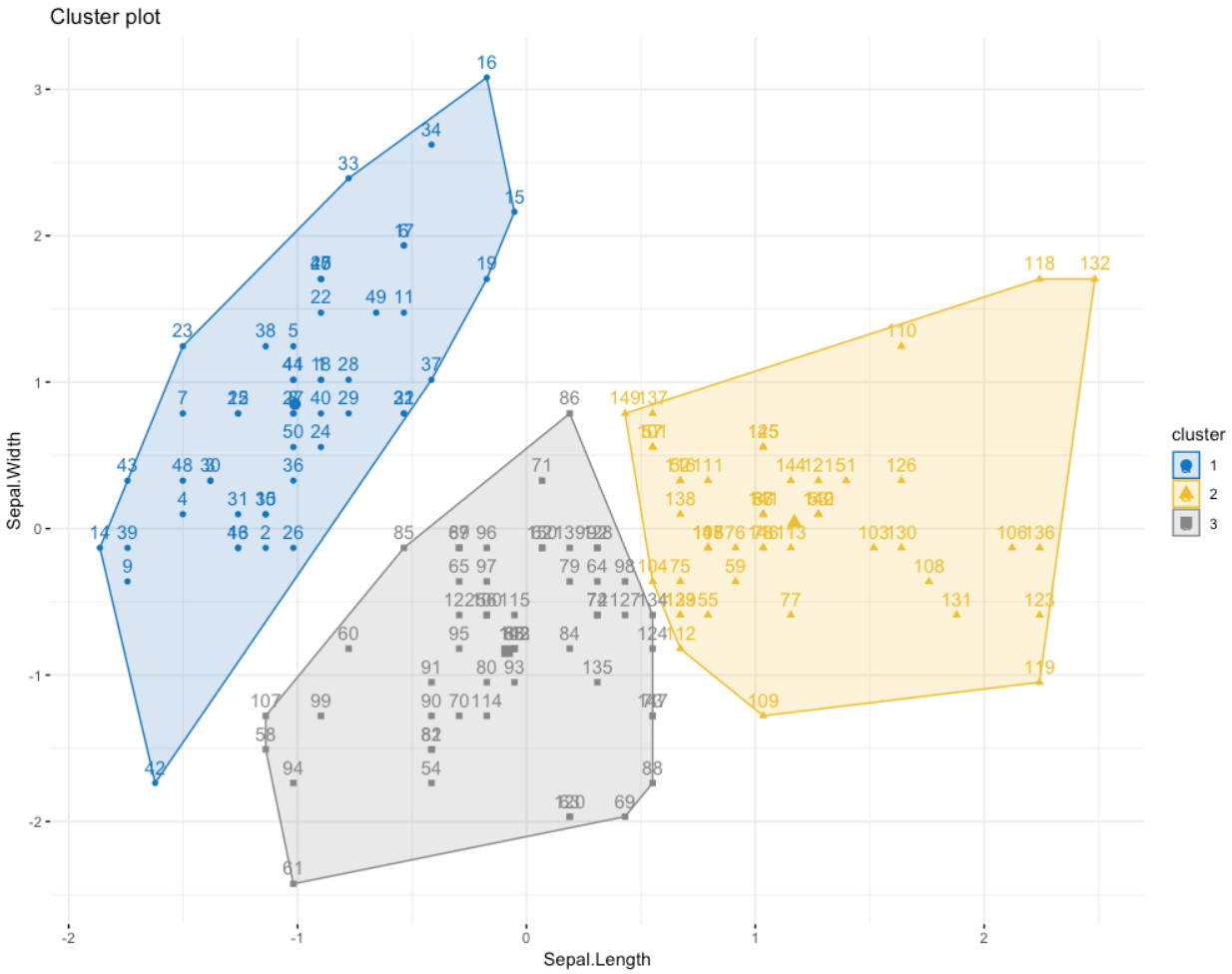


Figure 1.21: Results of k-medoids clustering

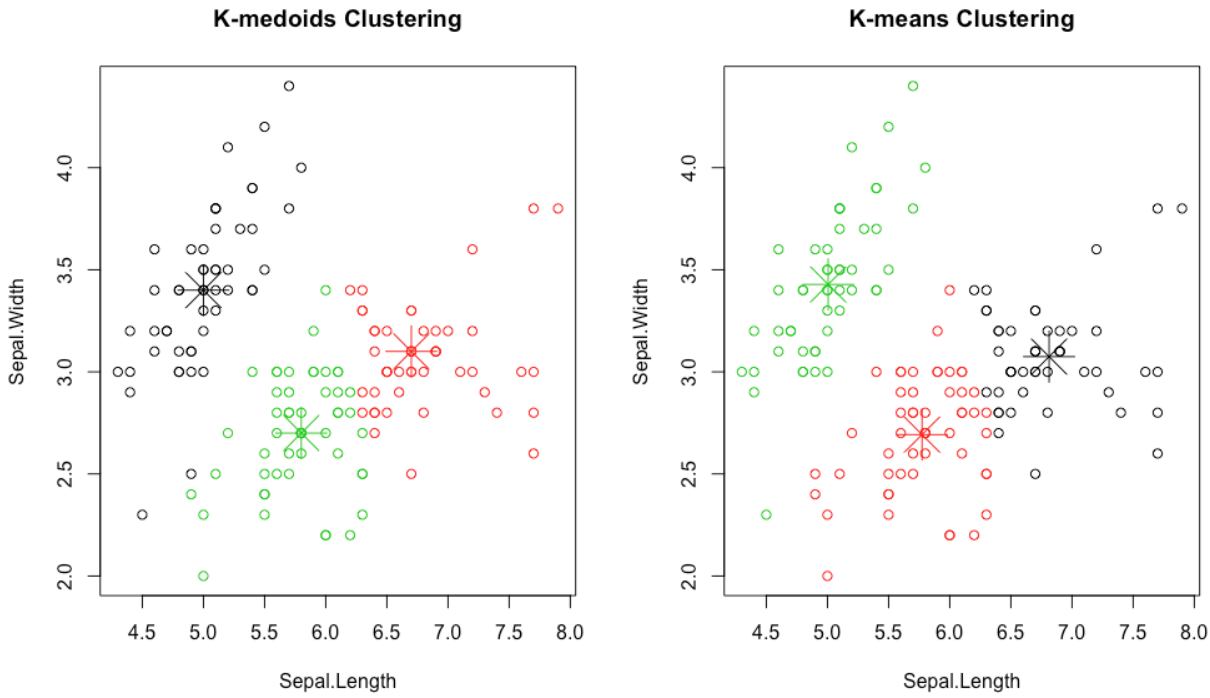


Figure 1.22: The results of k-medoids clustering versus k-means clustering

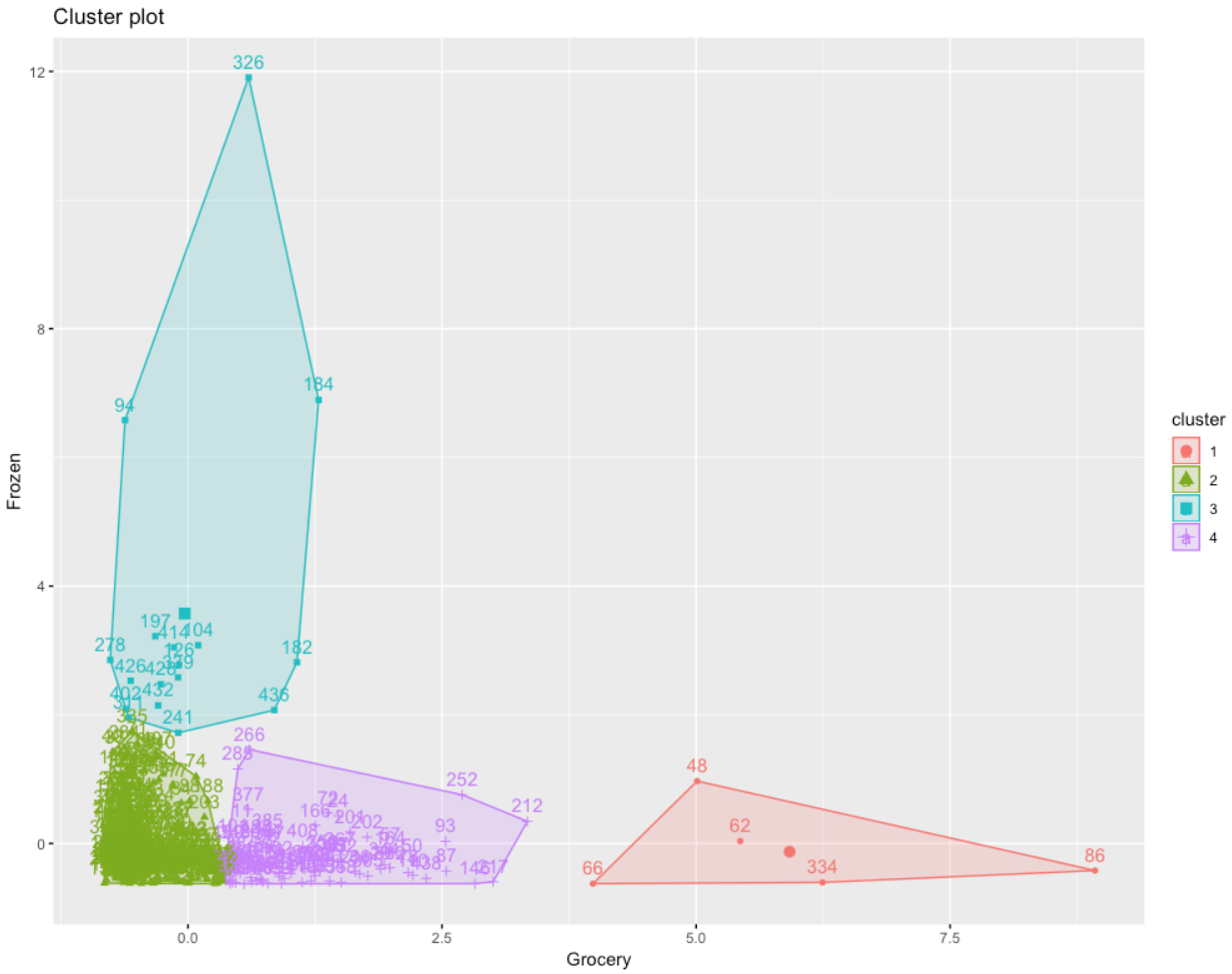


Figure 1.23: The expected k-means plot of the cluster

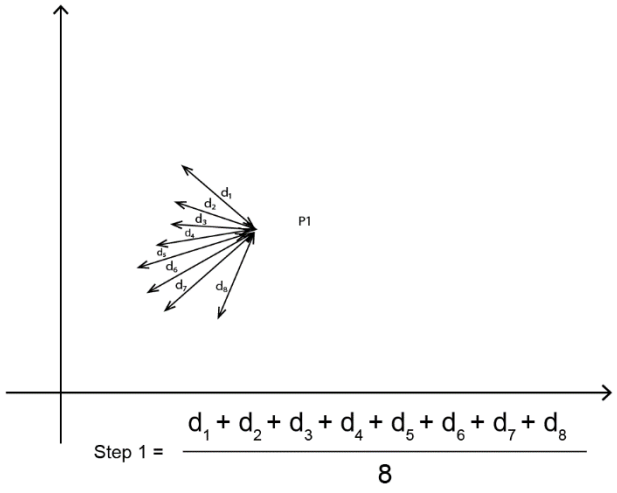


Figure 1.24: Calculating the average distance between point a and all the points of cluster x

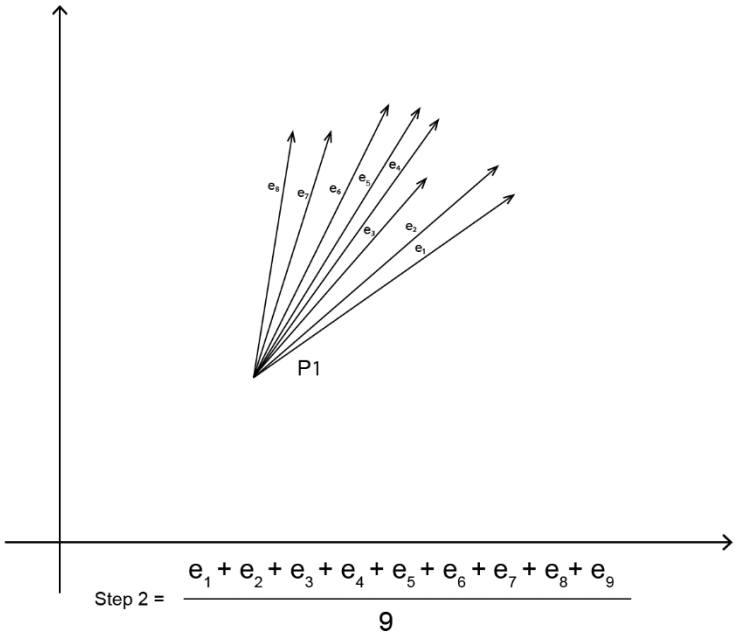


Figure 1.25: Calculating the average distance between point a and all the points near cluster x

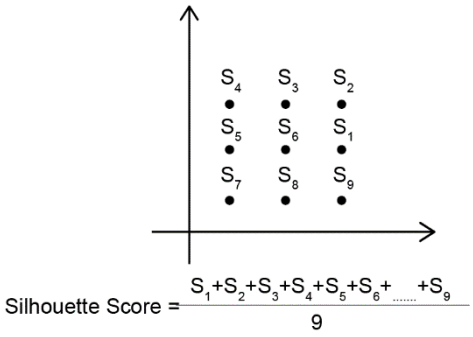


Figure 1.26: Calculating the silhouette score

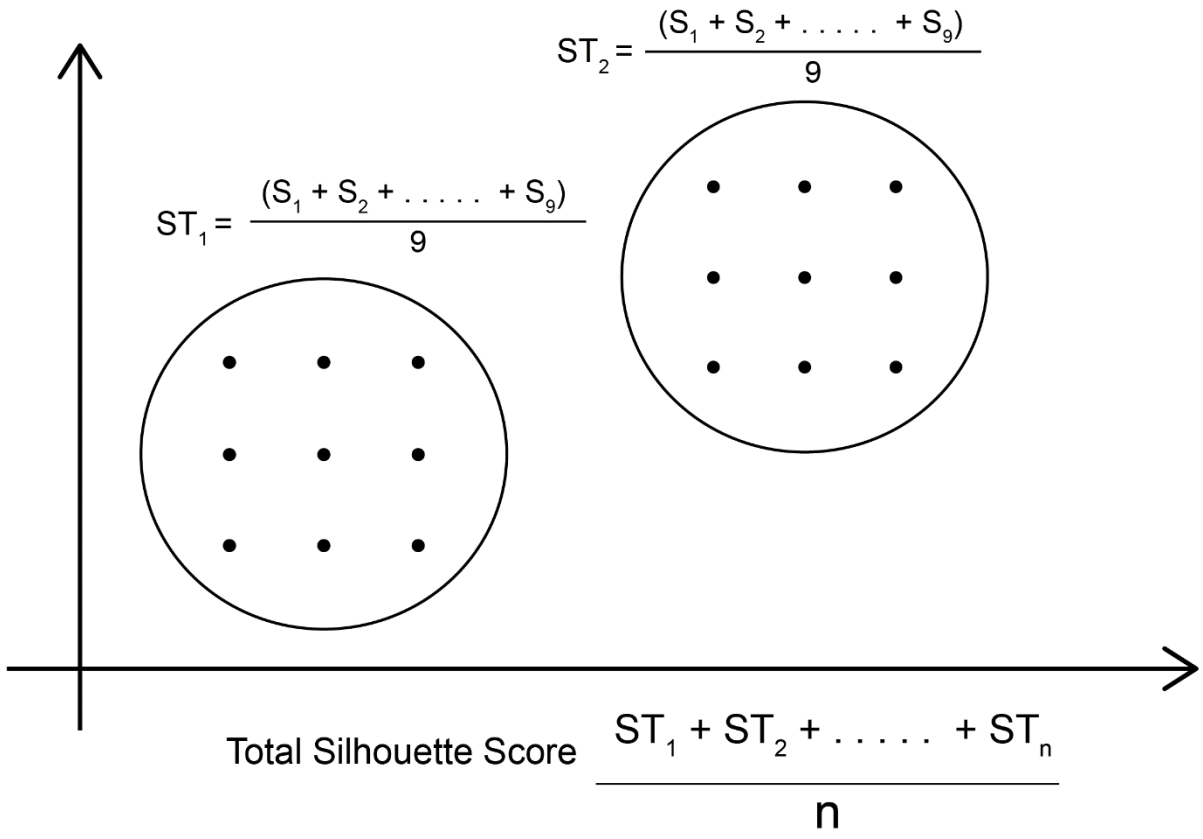


Figure 1.27: Calculating the average silhouette score

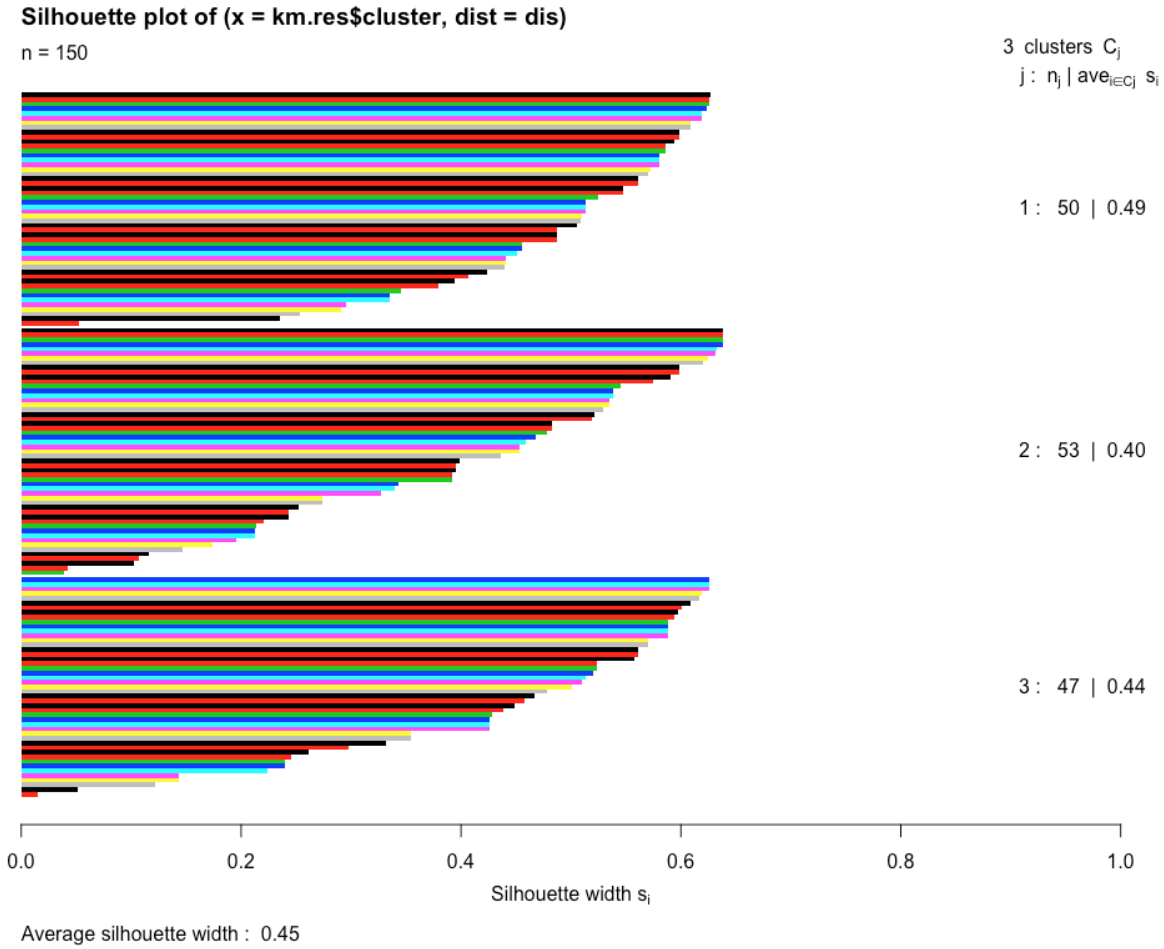


Figure 1.28: The silhouette score for each point in every cluster is represented by a single bar

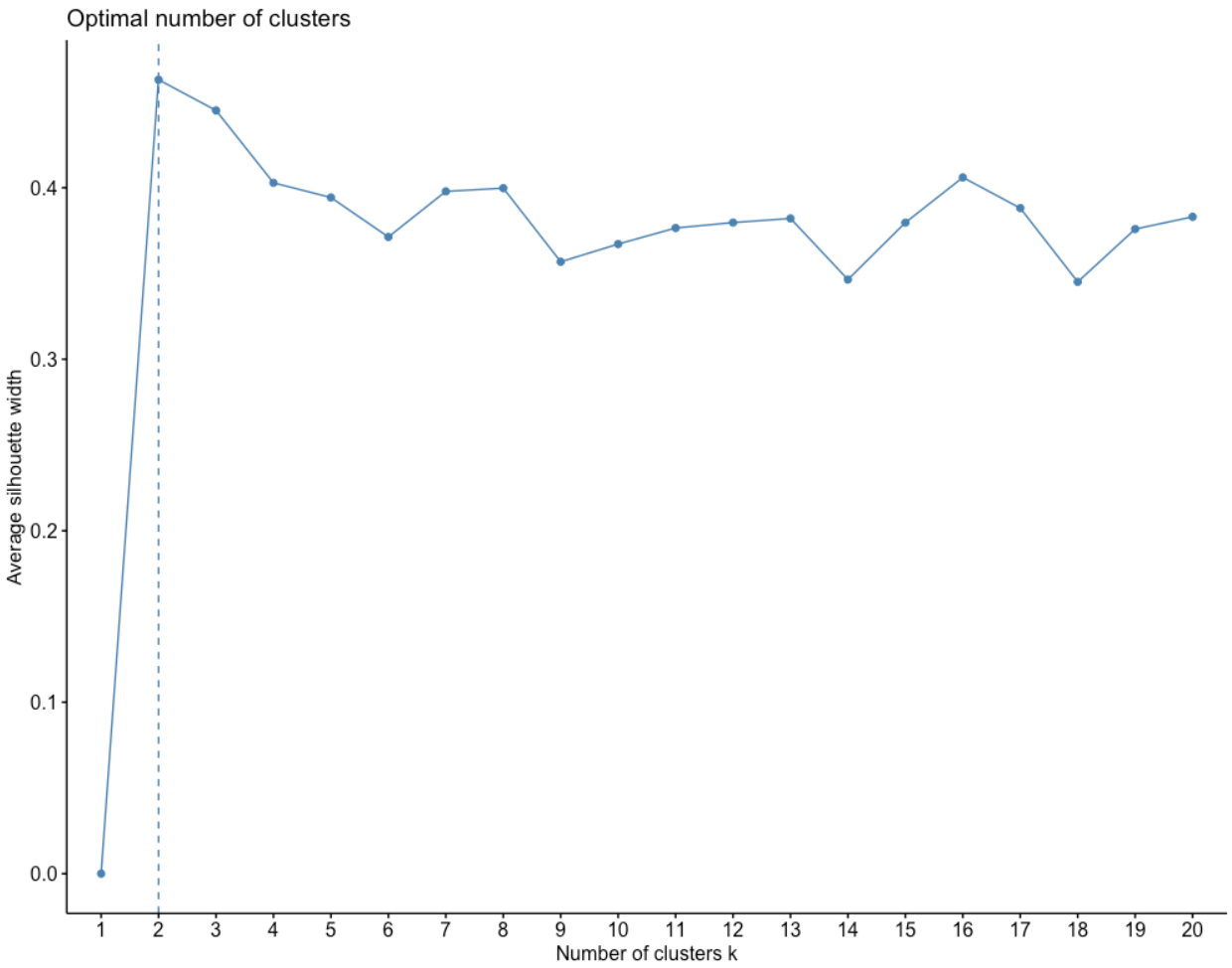


Fig 1.29: The number of clusters versus average silhouette score

$$\sum_{k=1}^{K} \sum_{i \in S_k} \sum_{j=1}^{p} (x_{ij} - \bar{x}_{kj})^2$$

where S_k is the set of observations in the *k*th cluster and $\bar{x}_{kj}$ is the *j*th variable of the cluster center for the *k*th cluster.

Figure 1.30: The formula to calculate WSS where p is the total number of dimensions of the data

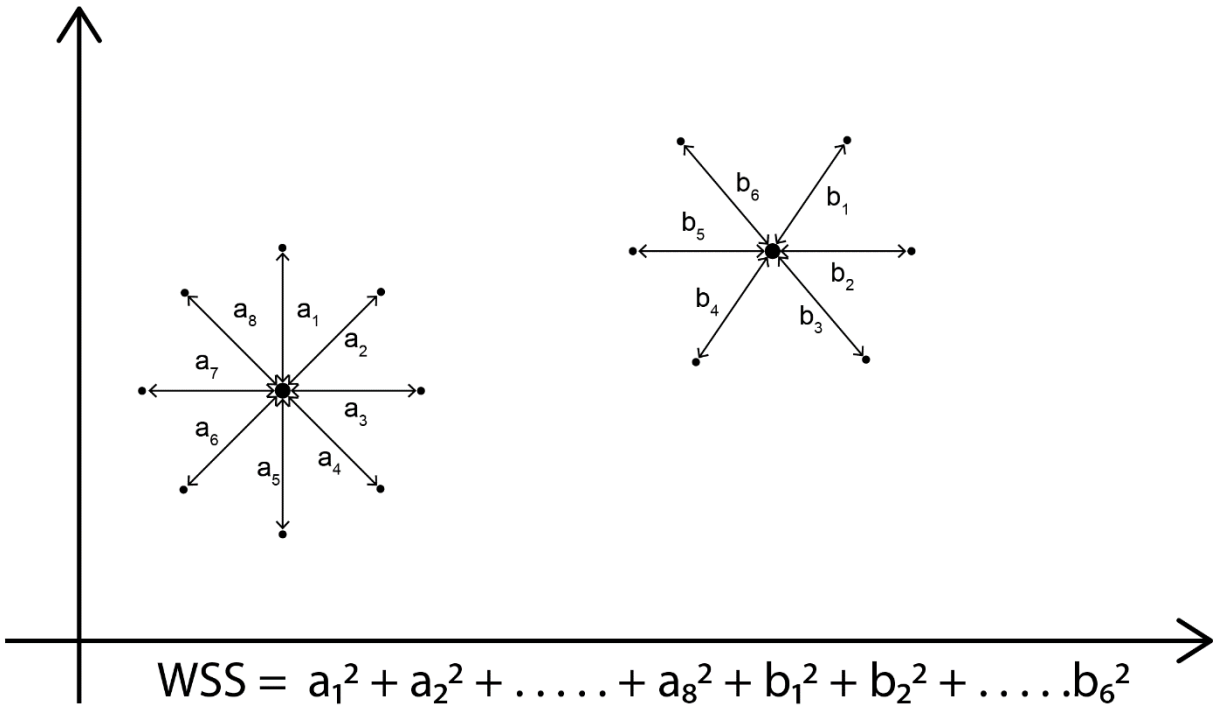


Figure 1.31: Illustration of WSS score

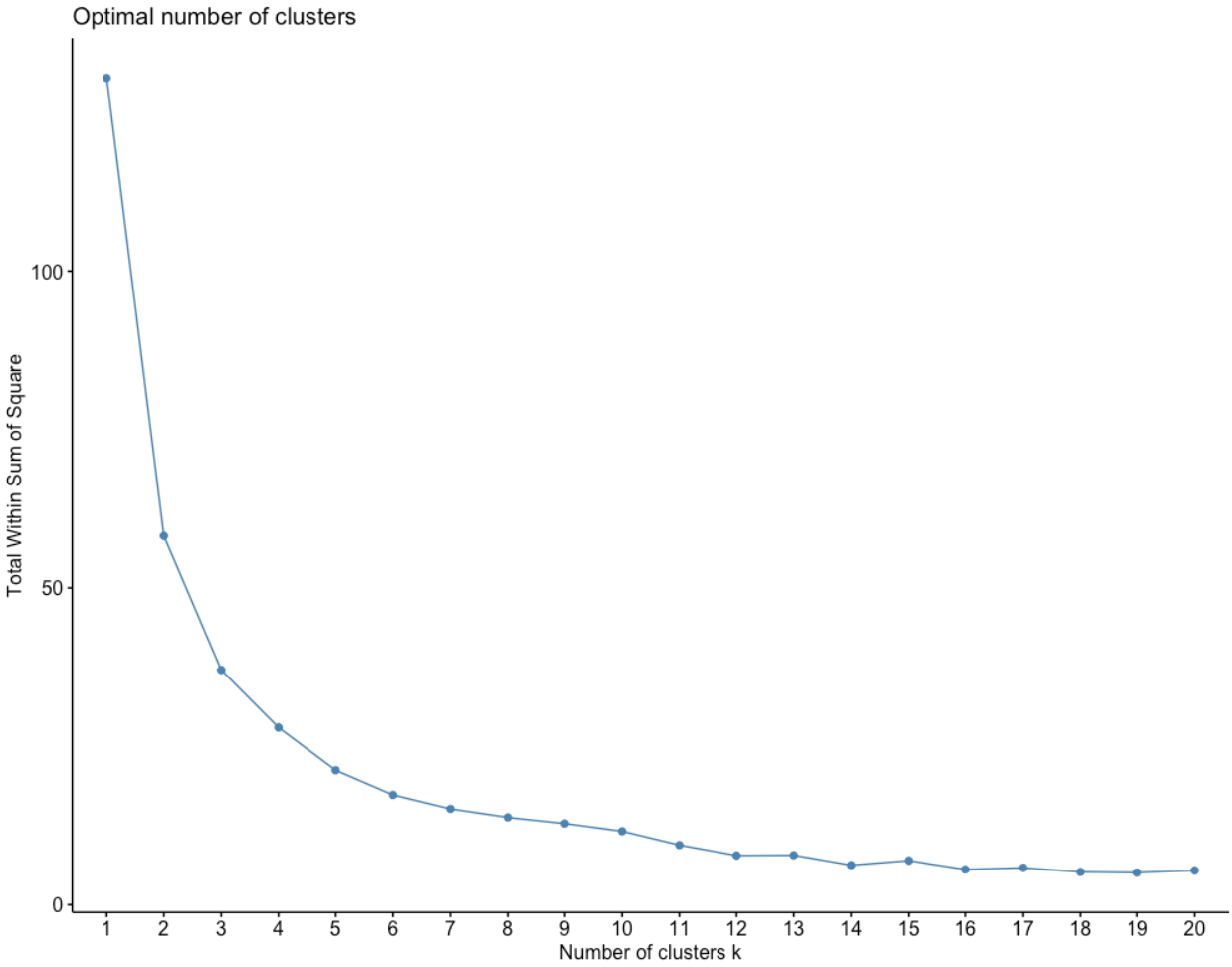


Fig 1.32: WSS versus number of clusters

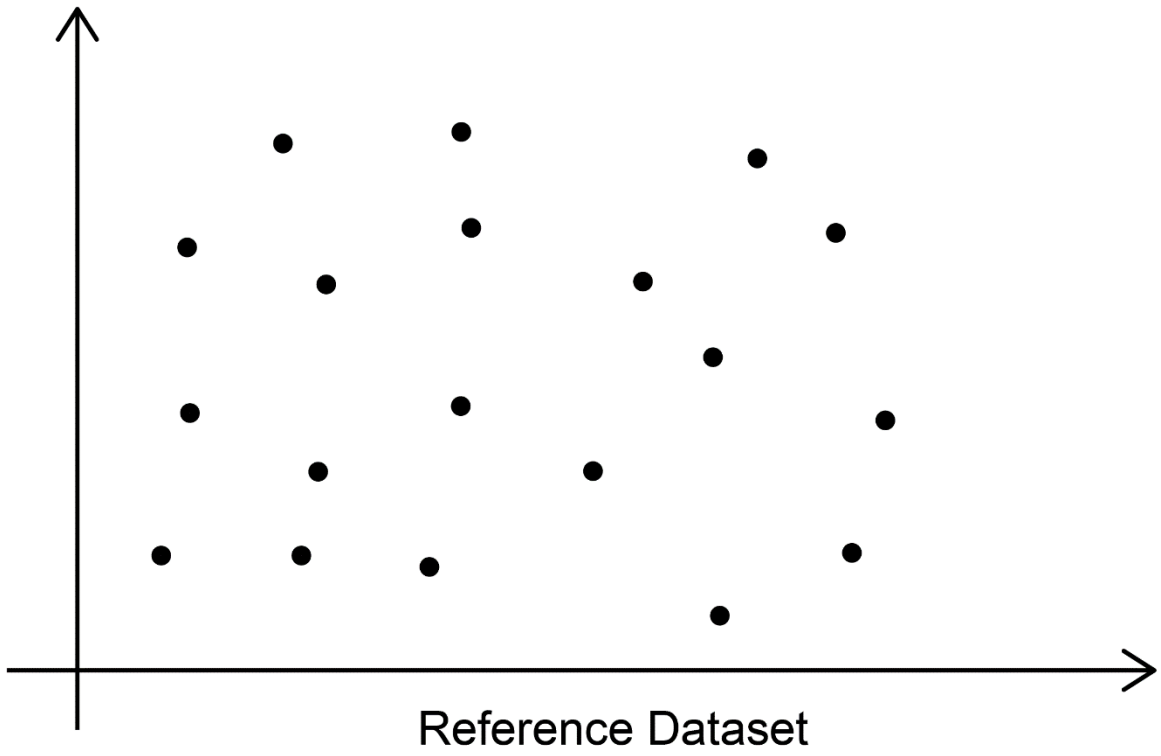


Figure 1.33: The reference dataset

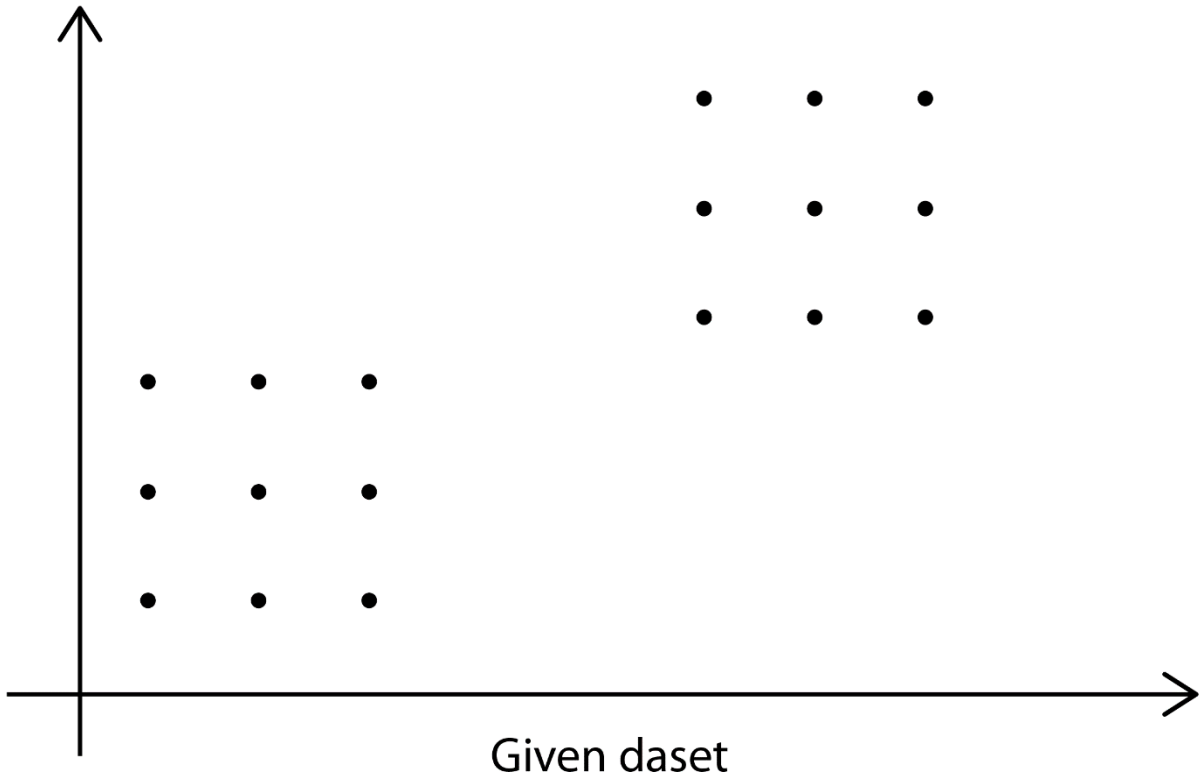


Figure 1.34: The observed dataset

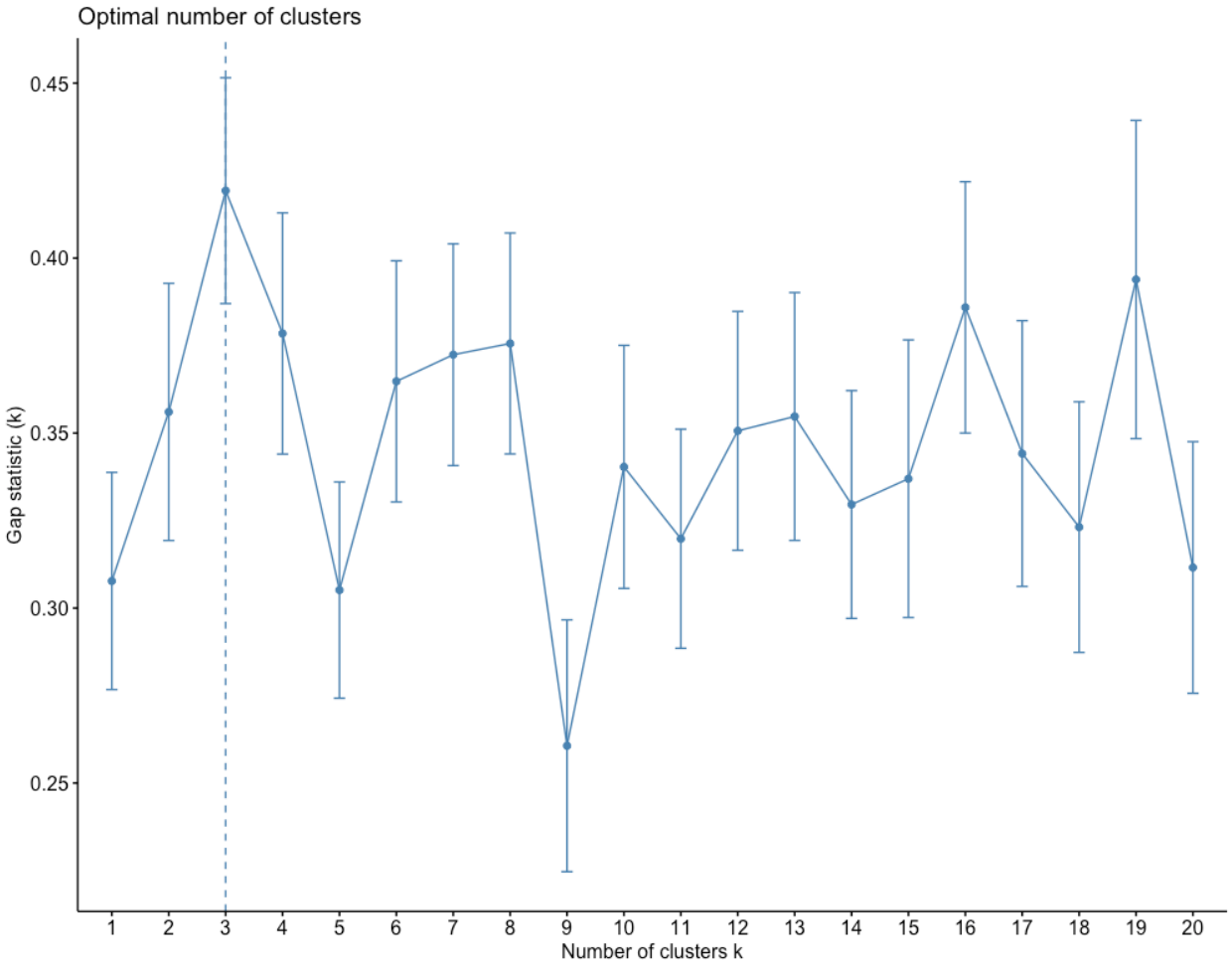


Figure 1.35: Gap statistics versus number of clusters

Lesson 2: Advanced Clustering Methods

```
> k.centers
K-modes clustering with 2 clusters of sizes 185, 101

Cluster modes:
     V2      V3    V4  V5 V6 V7    V8       V9 V10
1 40-49 premeno 25-29 0-2 no  2  left left_low  no
2 50-59    ge40 30-34 0-2 no  1 right  left_up  no

Clustering vector:
  [1] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 2 1 1 2 2 1 1 2 1 1 1 1 1 2 2 2 2 2
 [59] 2 2 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 2 1 2 1 2 1 2 2 2 1 2 2 2 1 1 2 1 1 1 1 2 1 1
[117] 1 1 1 2 1 2 2 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 2 1 1 2 2 2 2 2 2 1 1 1 2 2
[175] 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1 2 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 1
[233] 2 1 2 1 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 1 2 2 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 2 2 1
[291] 1 1 1 2 1 2

within cluster simple-matching distance by cluster:
[1] 671 335

Available components:
[1] "cluster"    "size"       "modes"      "withindiff" "iterations" "weighted"
```

Figure 2.1: Screenshot of the cluster centers

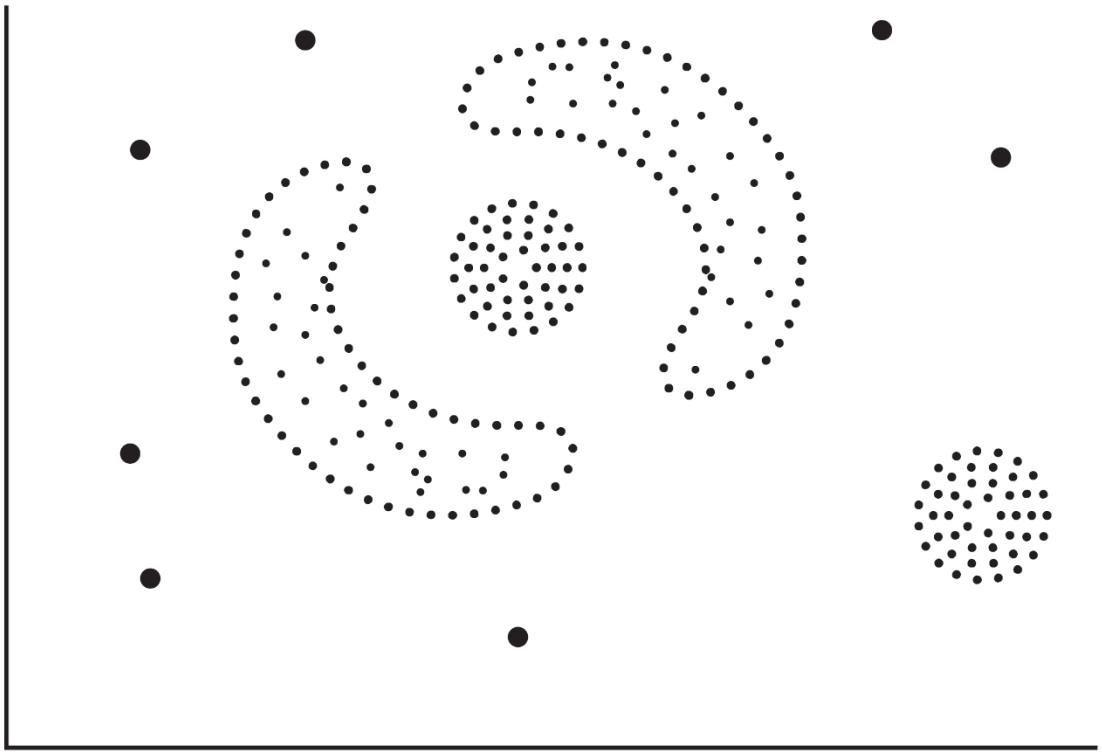


Figure 2.2: A sample scatter plot

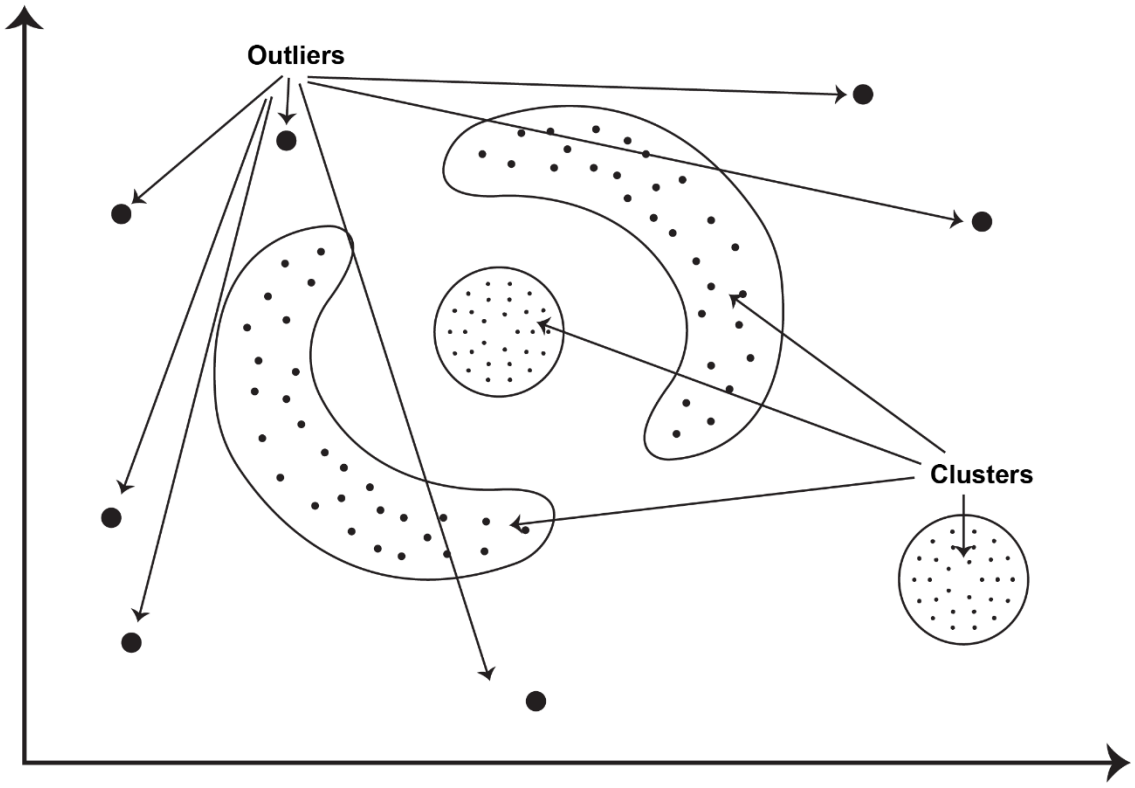


Figure 2.3: Clusters and outliers classified by DBSCAN

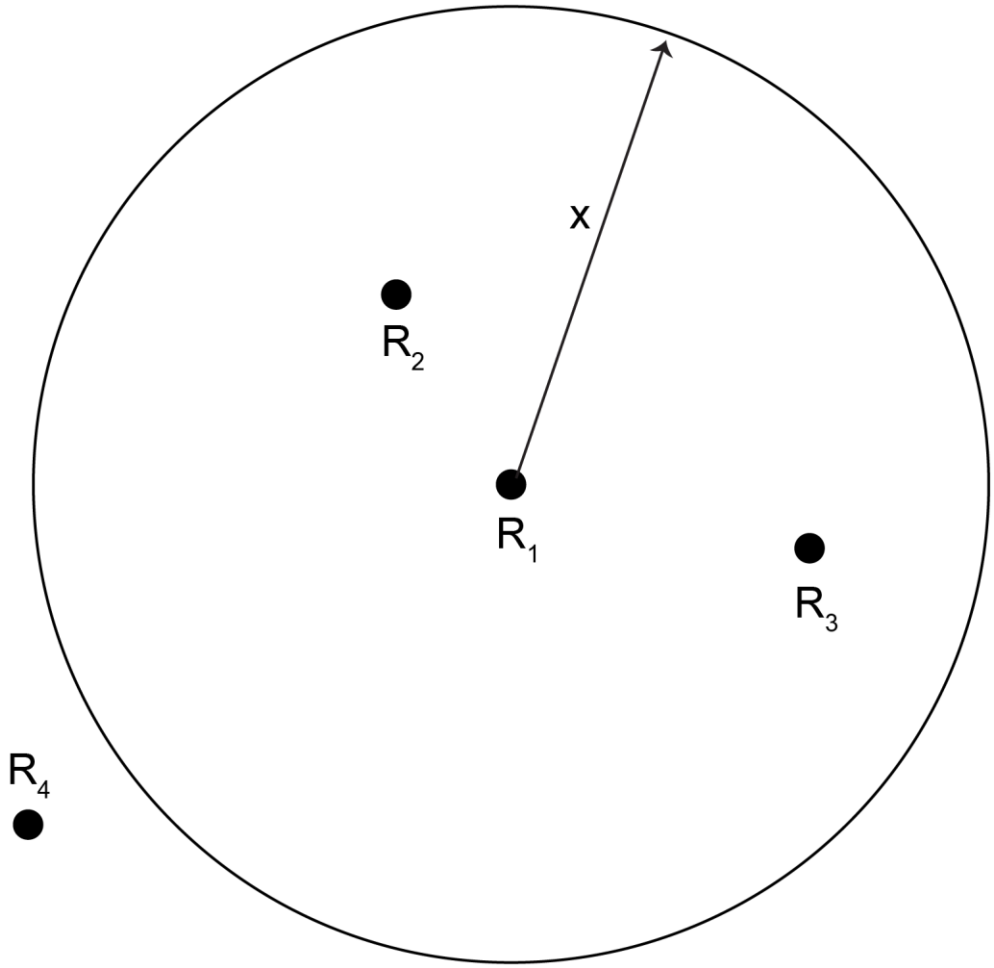


Figure 2.4: Only 2 points lie within x distance of point R1

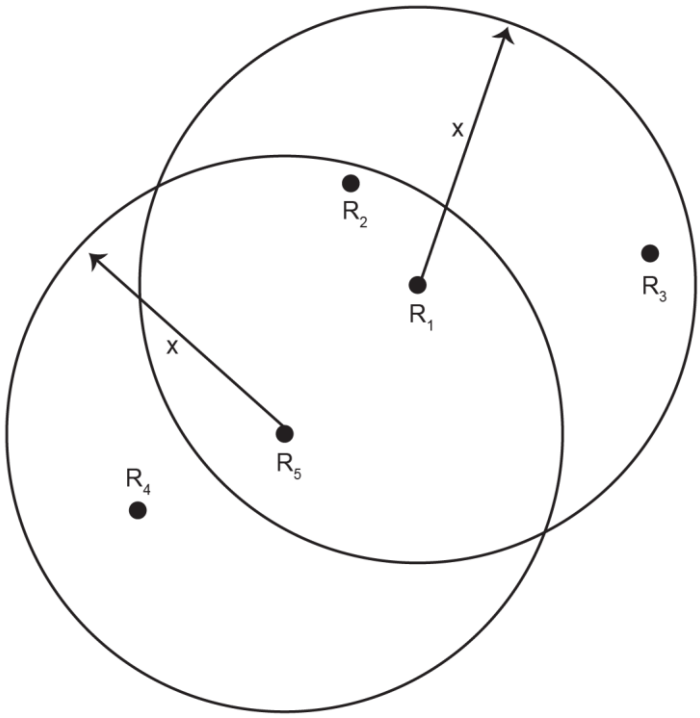


Figure 2.5: All of these four points belong to a cluster

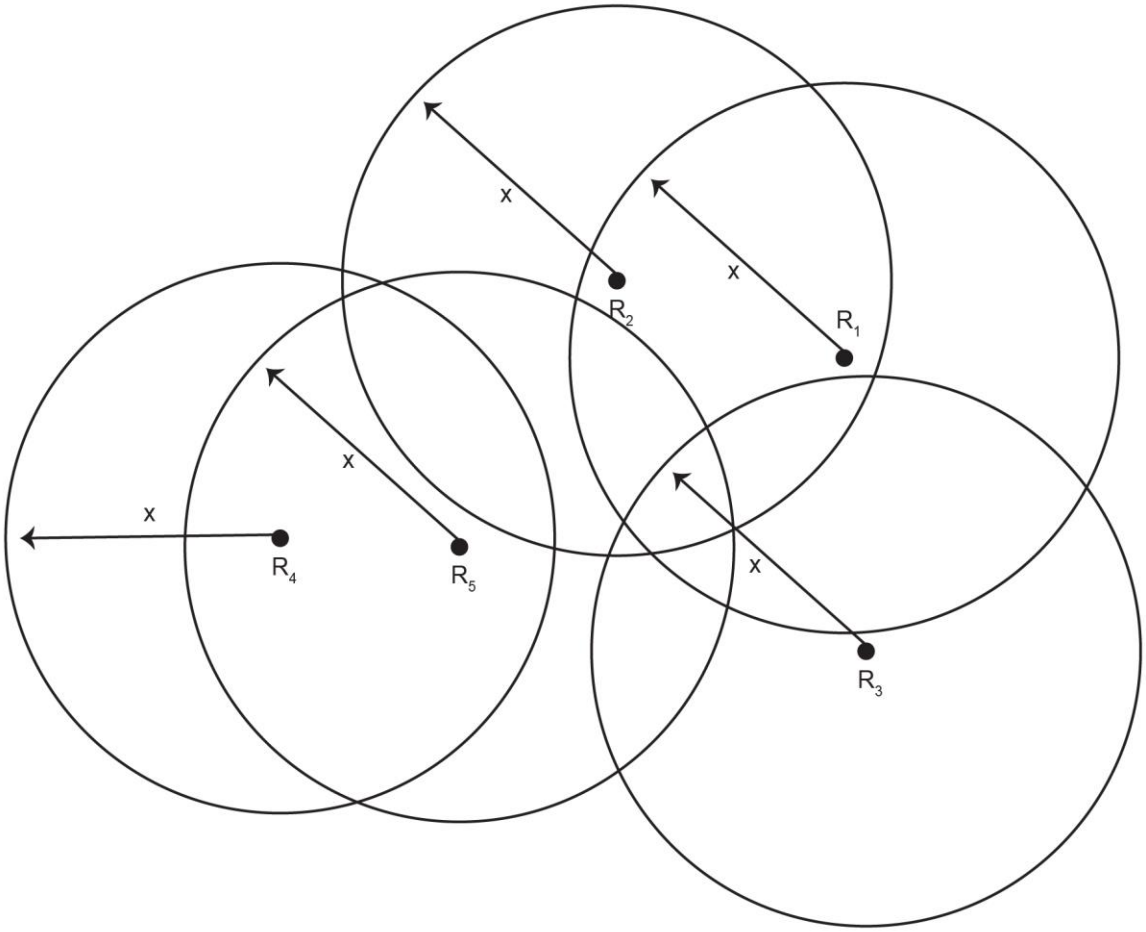


Figure 2.6: Core versus non-core points

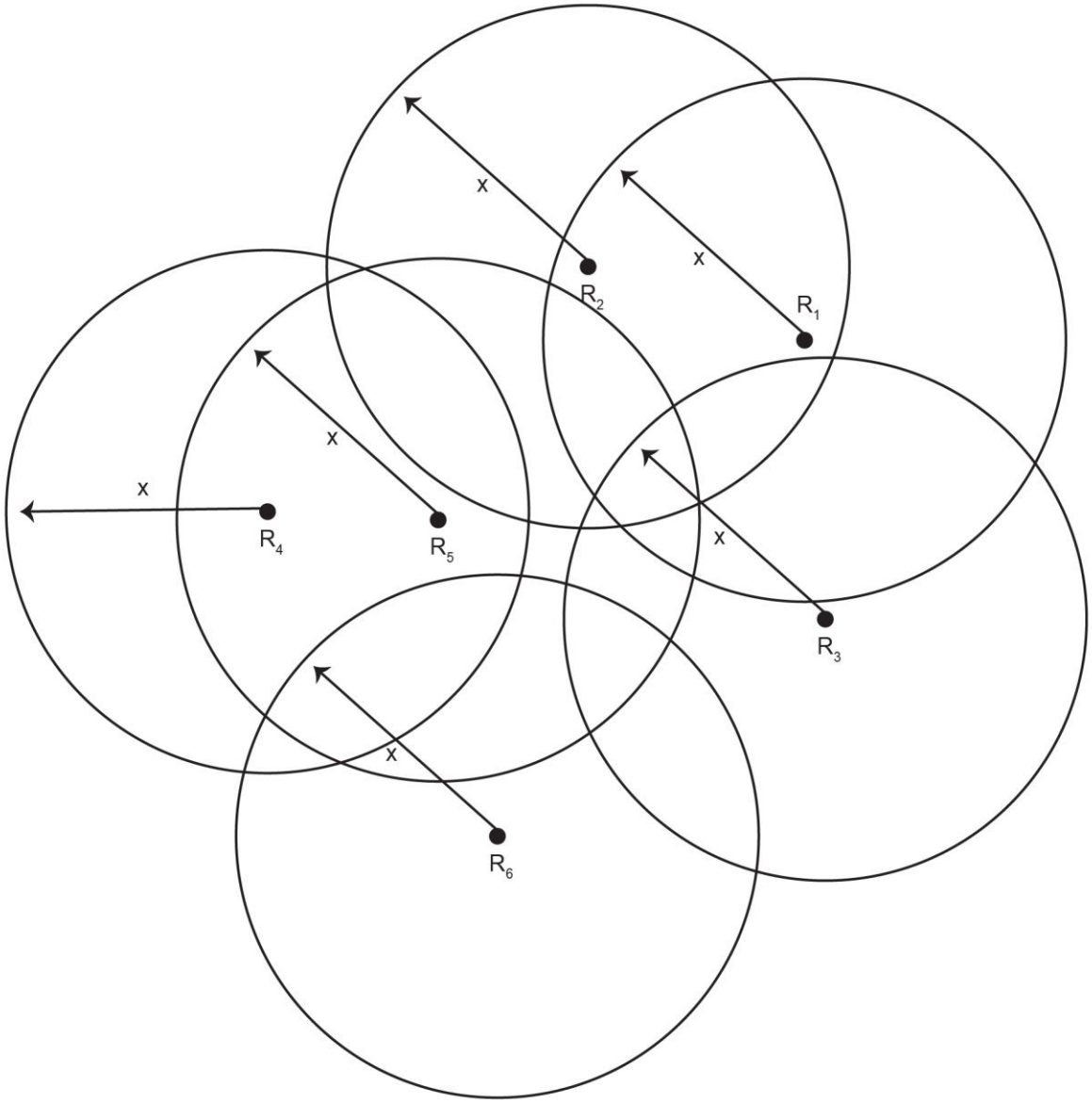


Figure 2.7: Noise point R6

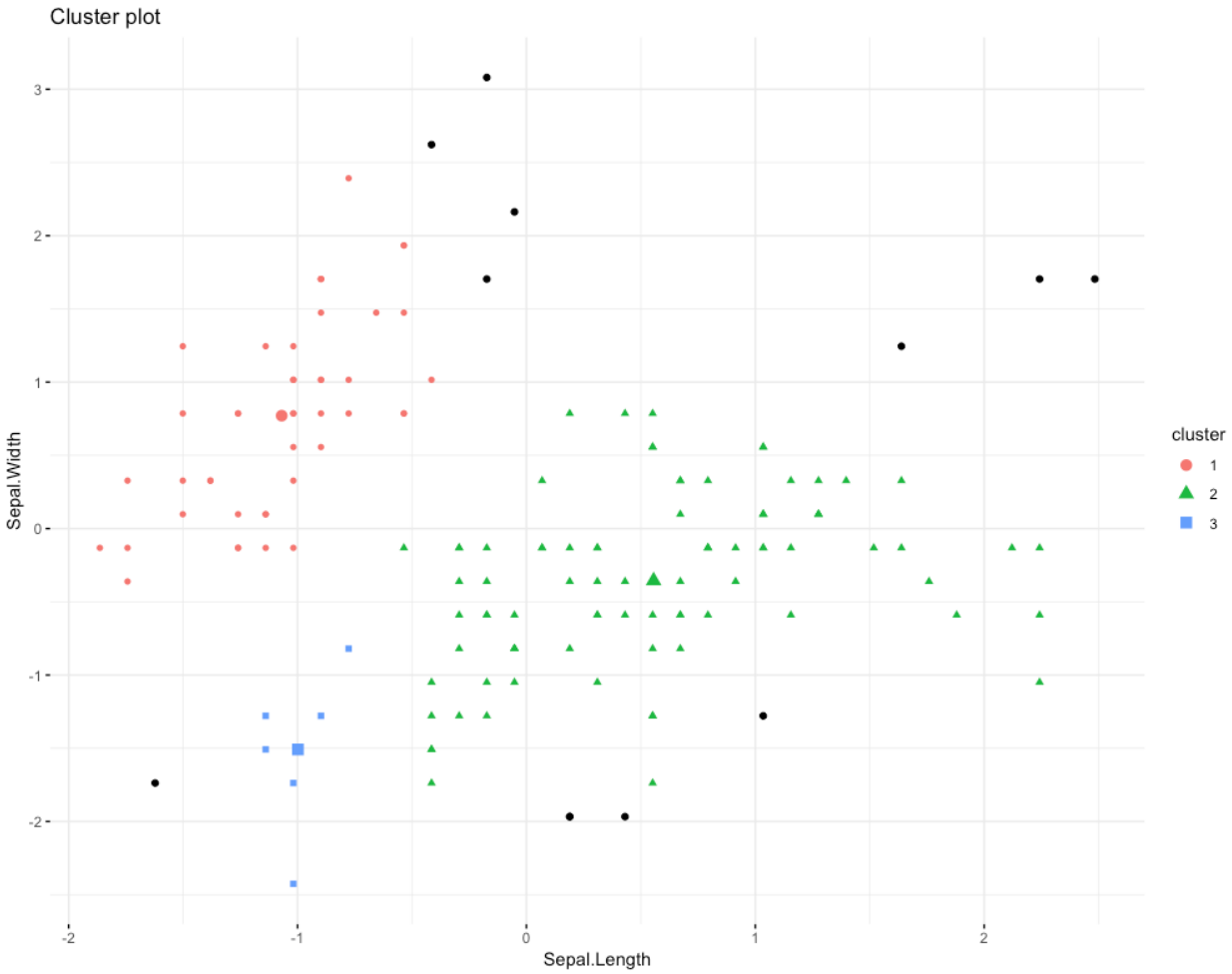


Figure 2.8: DBSCAN clusters in different colors

Figure 2.9: Expected plot of DBCAN on the multishapes dataset

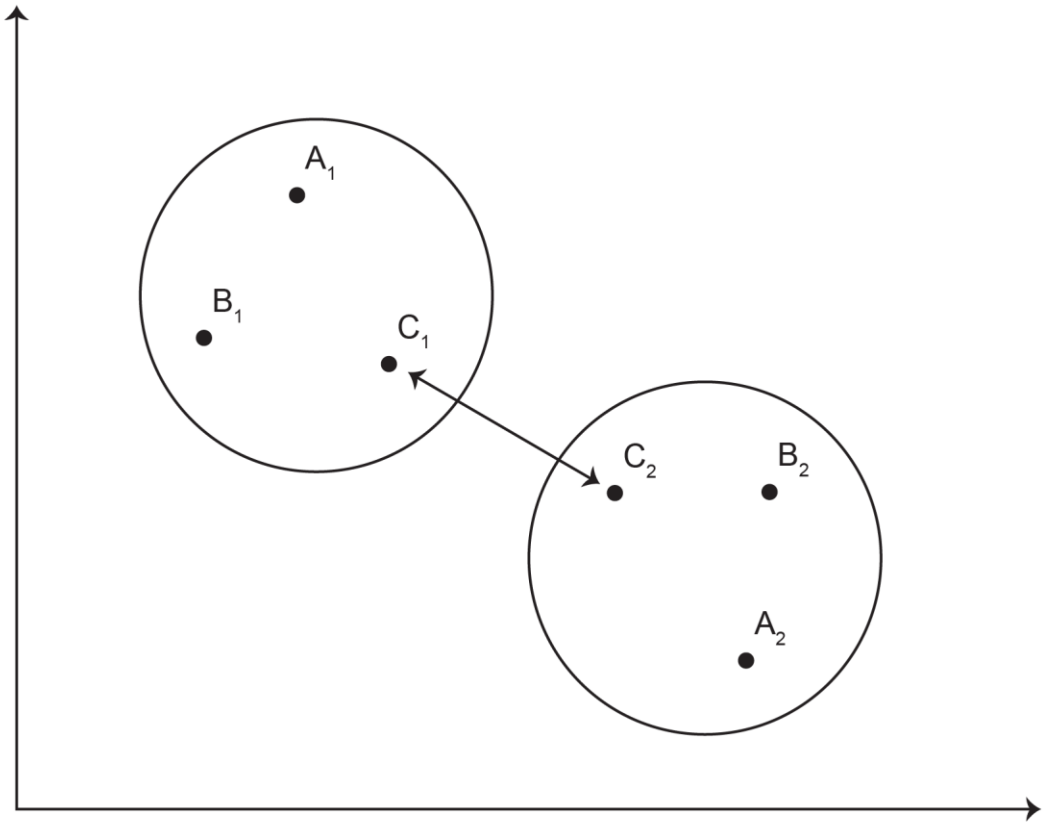


Figure 2.10: Demonstration of the single-link metric

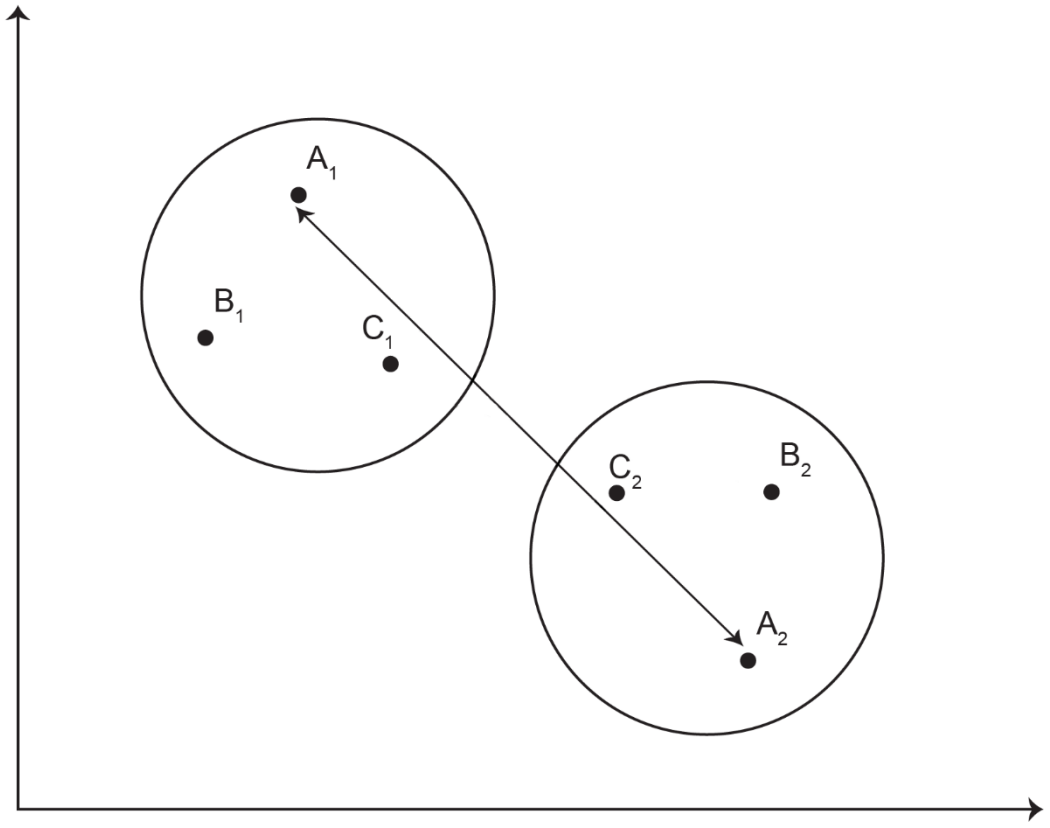


Figure 2.11: Demonstration of the complete-link metric

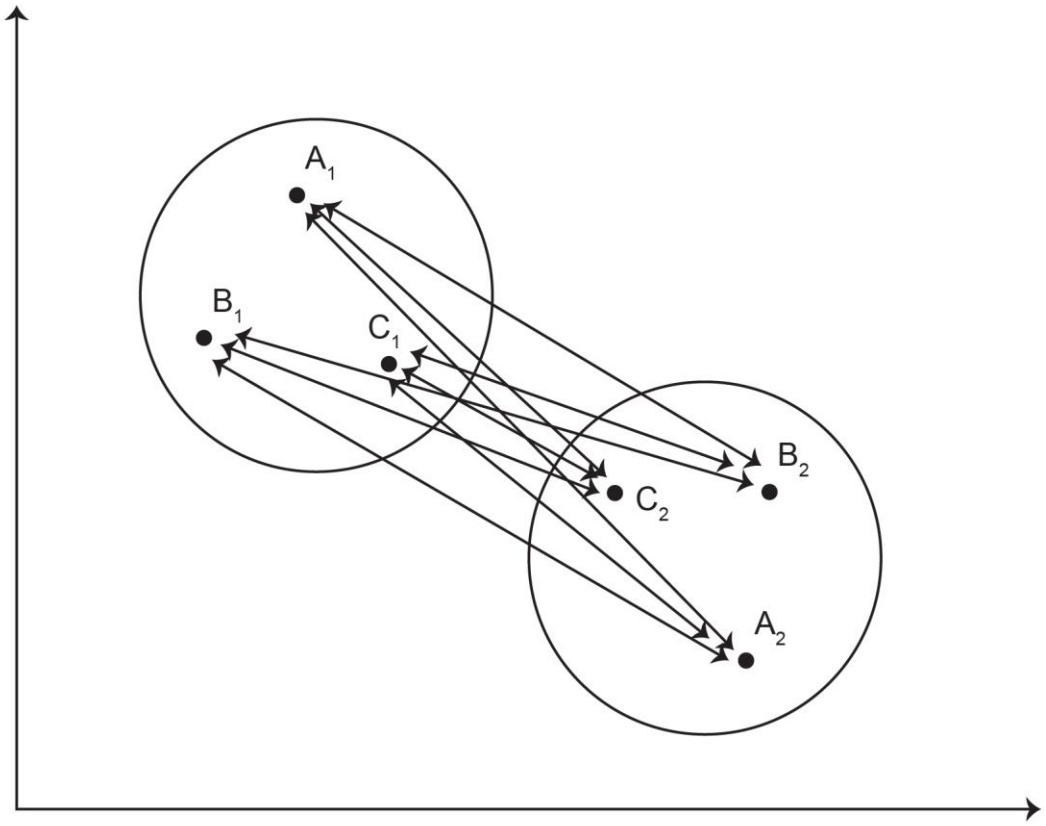


Figure 2.12: Demonstration of the group-average metric

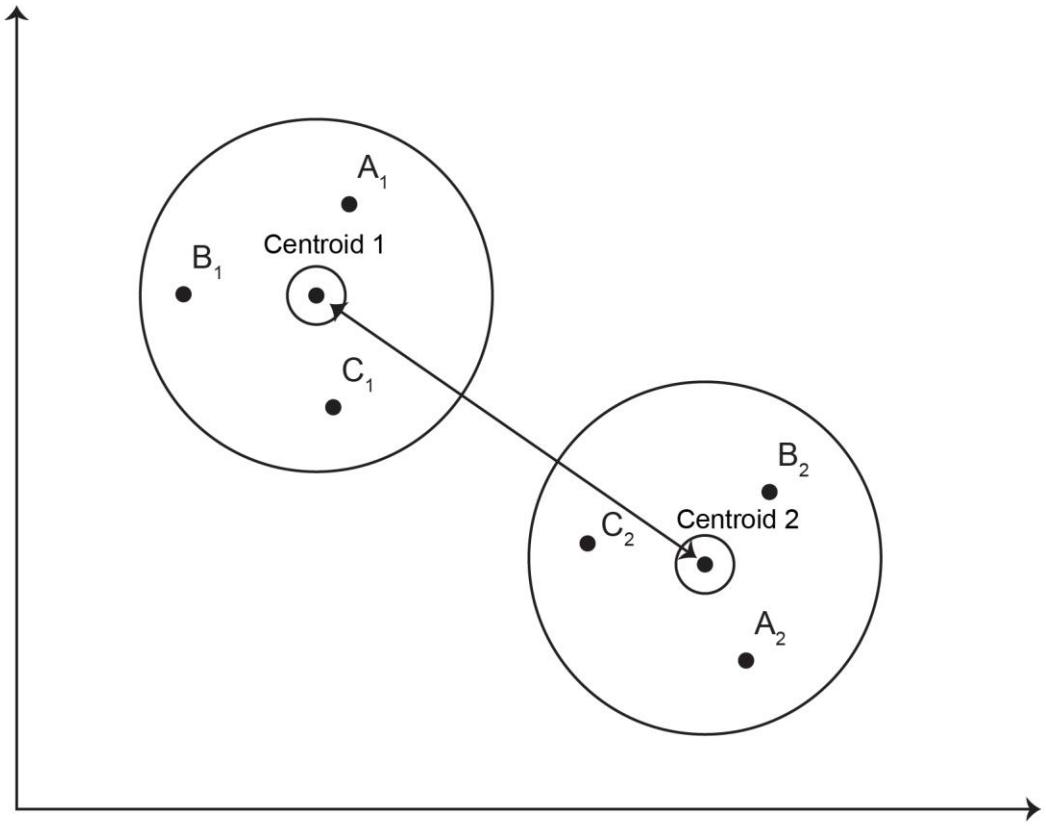


Figure 2.13: Demonstration of centroid similarity

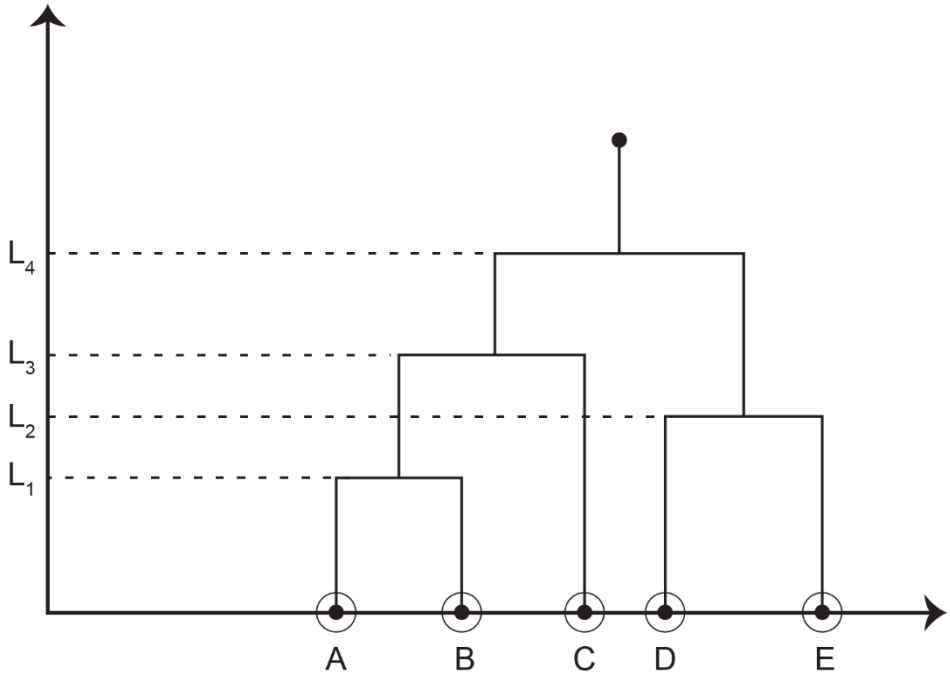


Figure 2.14: A sample dendrogram

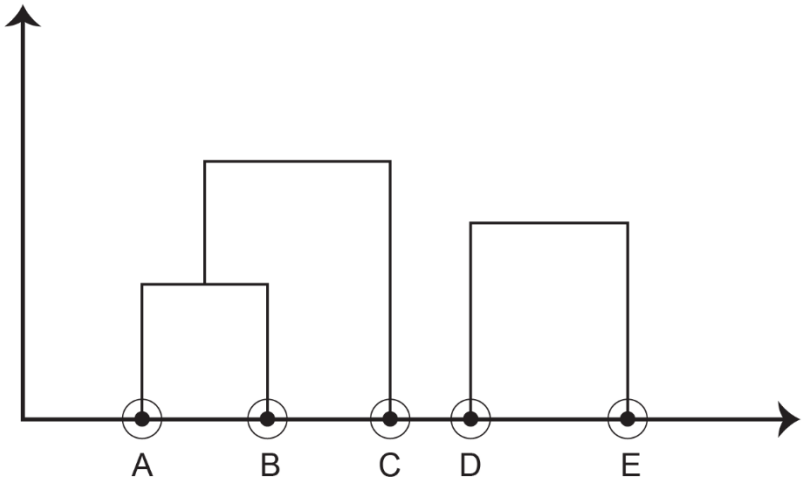


Figure 2.15: Clusters represented in the dendrogram

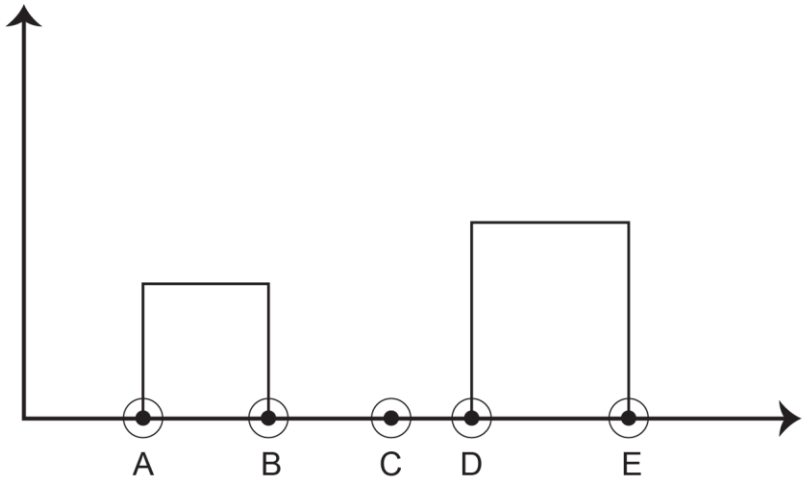


Figure 2.16: Representation of clusters in a dendrogram

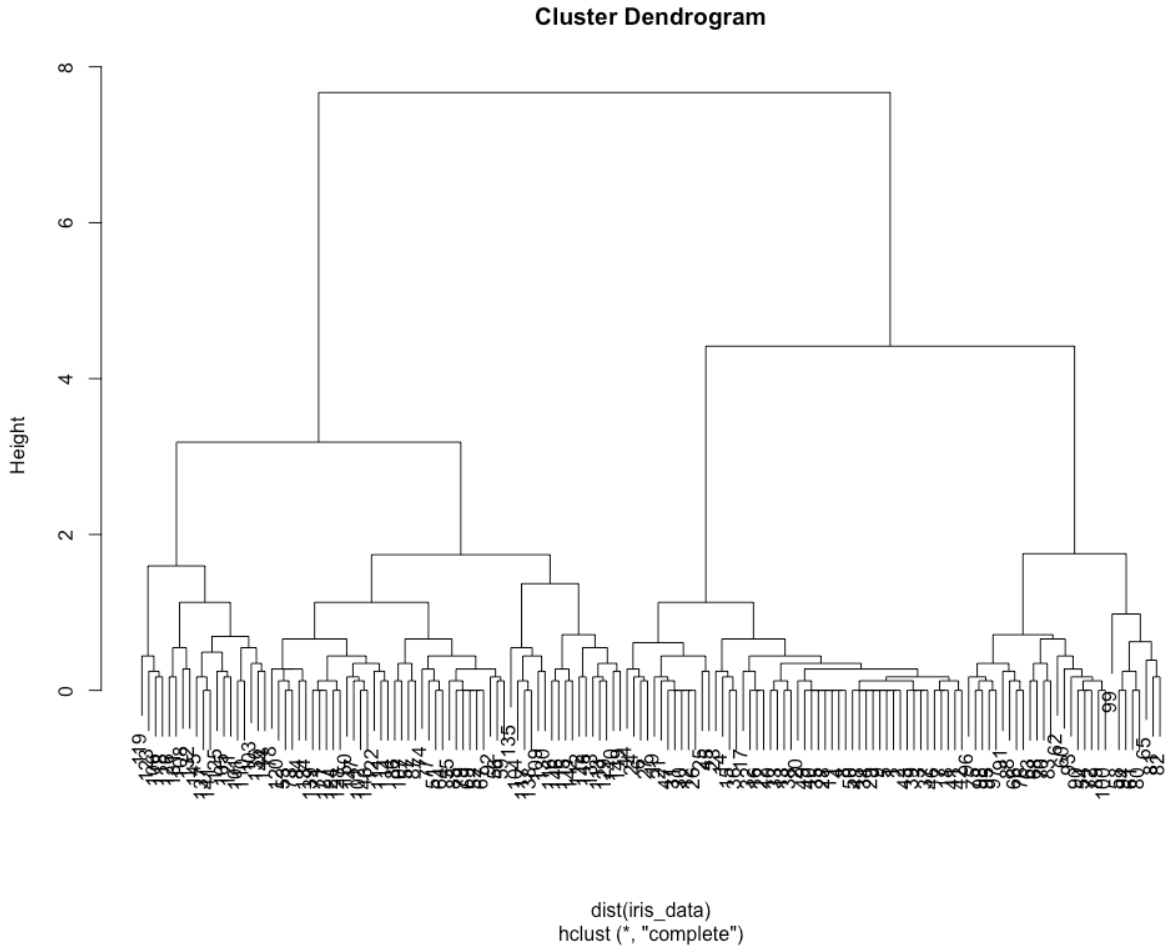


Figure 2.17: Dendrogram derived from the complete similarity metric

clusterCut	setosa	versicolor	virginica
1	50	0	0
2	0	21	50
3	0	29	0

Figure 2.18: Table displaying the distribution of clusters

Figure 2.19: Dendrogram derived from the single similarity metric

clusterCut	setosa	versicolor	virginica
1	50	0	0
2	0	49	50
3	0	1	0

Figure 2.20: Table displaying the distribution of clusters

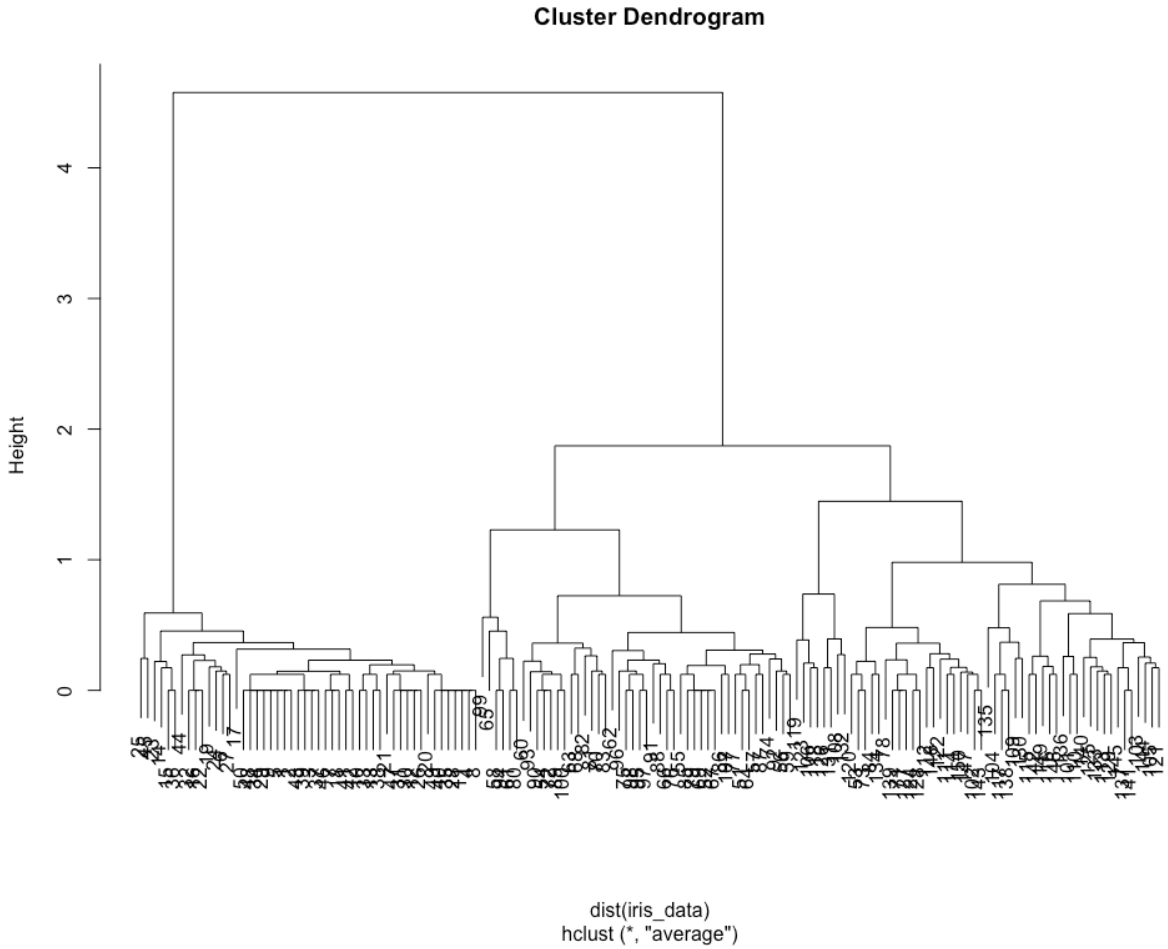


Figure 2.21: Dendrogram derived from the average similarity metric

clusterCut	setosa	versicolor	virginica
1	50	0	0
2	0	45	1
3	0	5	49

Figure 2.22: Table displaying the distribution of clusters

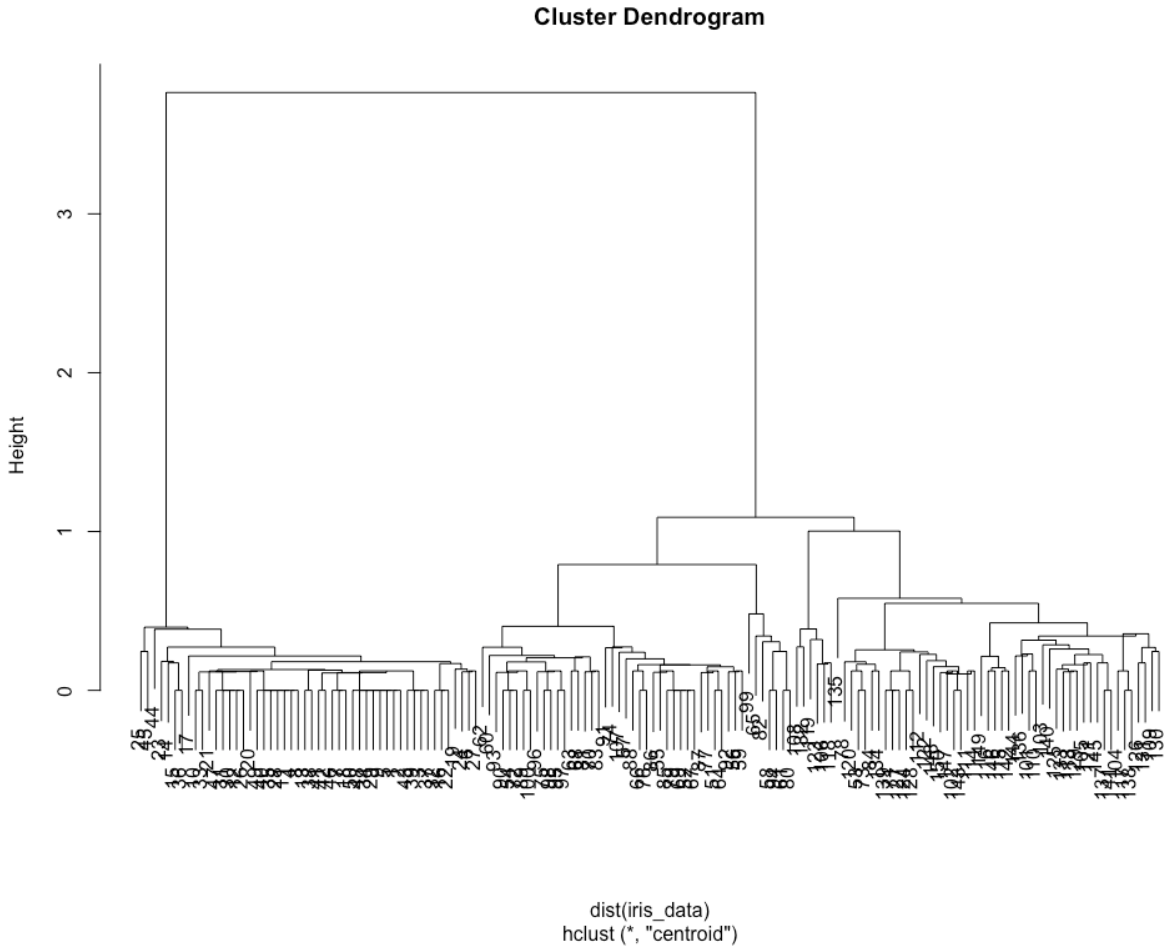


Figure 2.23: Dendrogram derived from the centroid similarity metric

clusterCut	setosa	versicolor	virginica
1	50	0	0
2	0	45	1
3	0	5	49

Figure 2.24: Table displaying the distribution of clusters

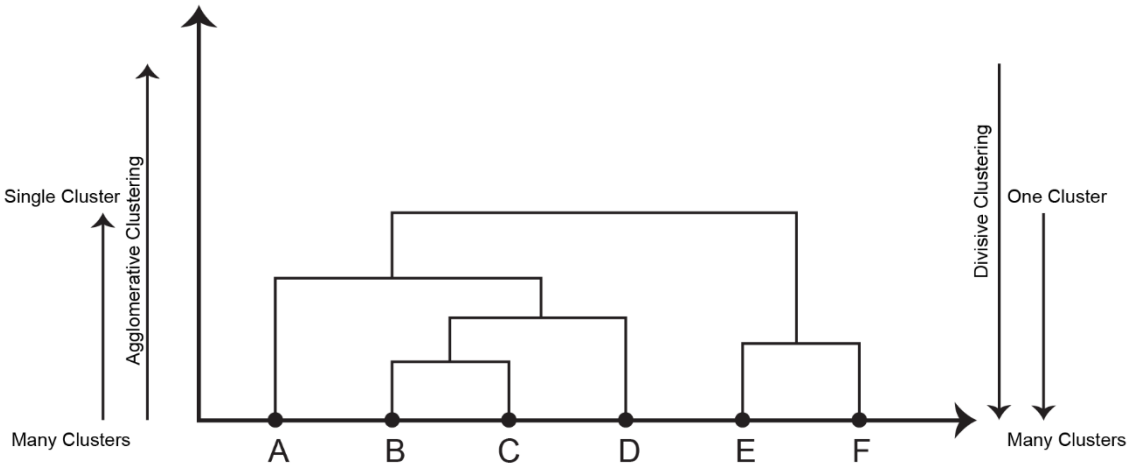


Figure 2.25: Representation of agglomerative and divisive clustering

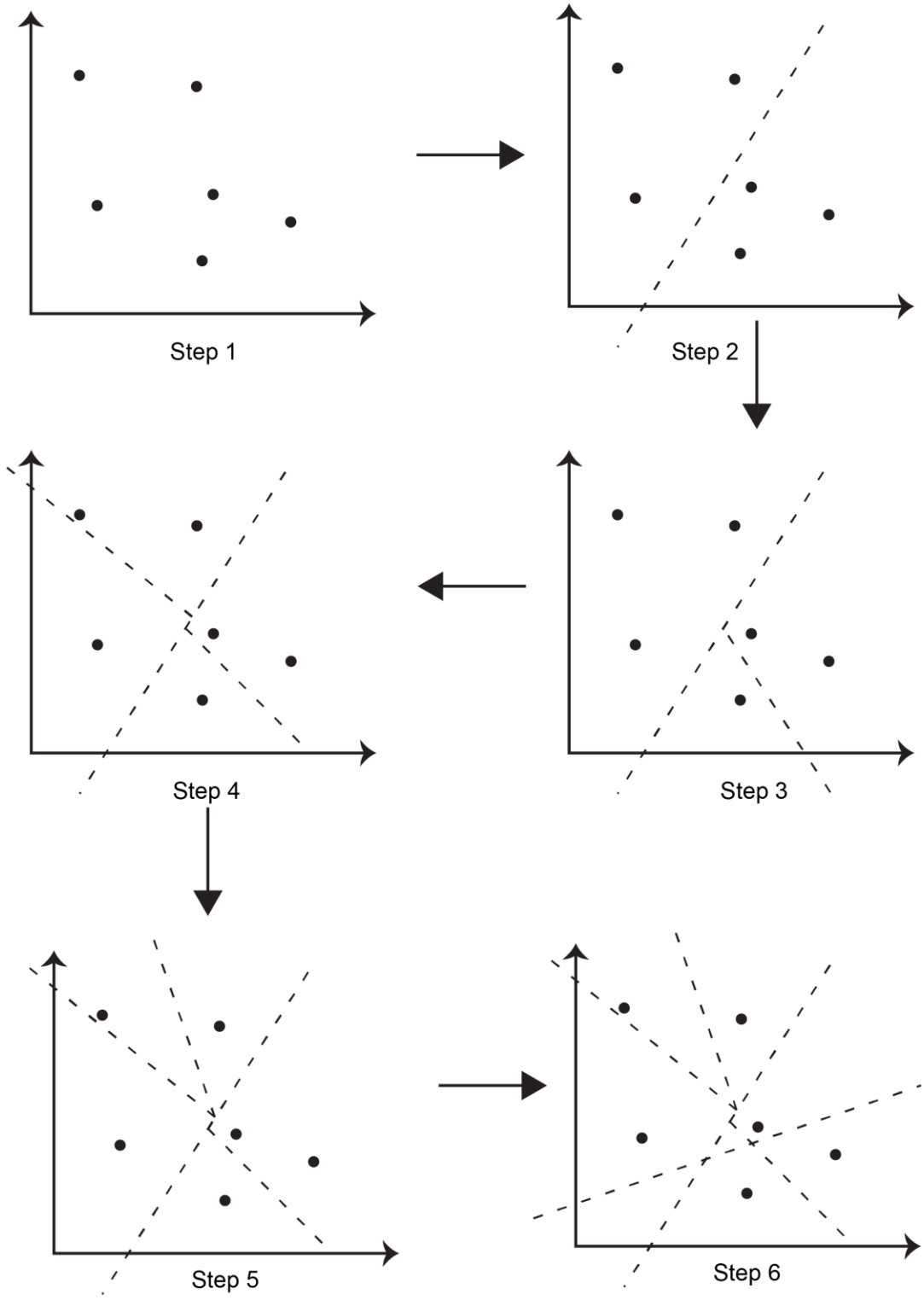


Figure 2.26: Representation of the divisive clustering process

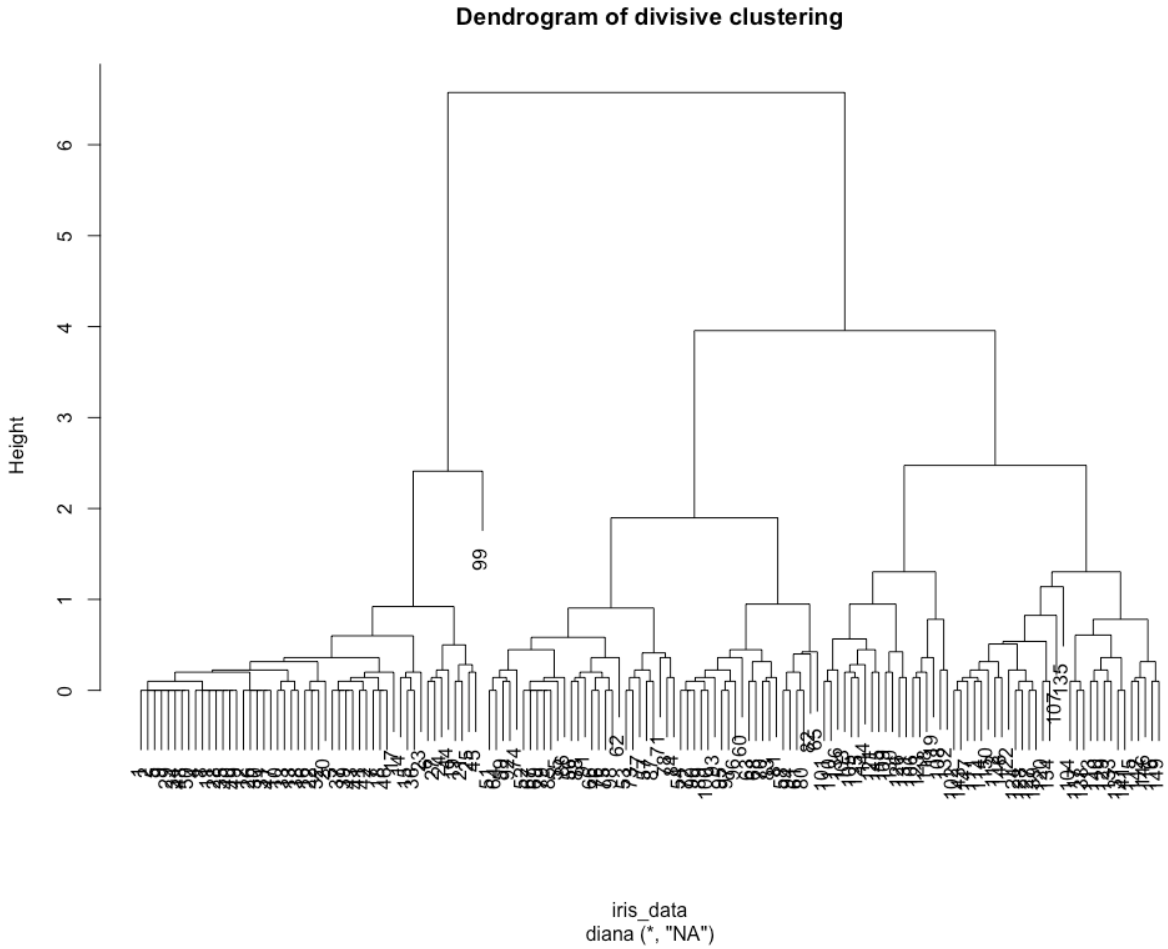


Figure 2.27: Dendrogram of divisive clustering

memb			
	1	2	3
1	65	3	1
2	6	0	64
3	9	61	0

Figure 2.28: Expected table classifying the three types of seeds

Lesson 3: Probability Distributions

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

Figure 3.1: Mathematical formula for a uniform distribution

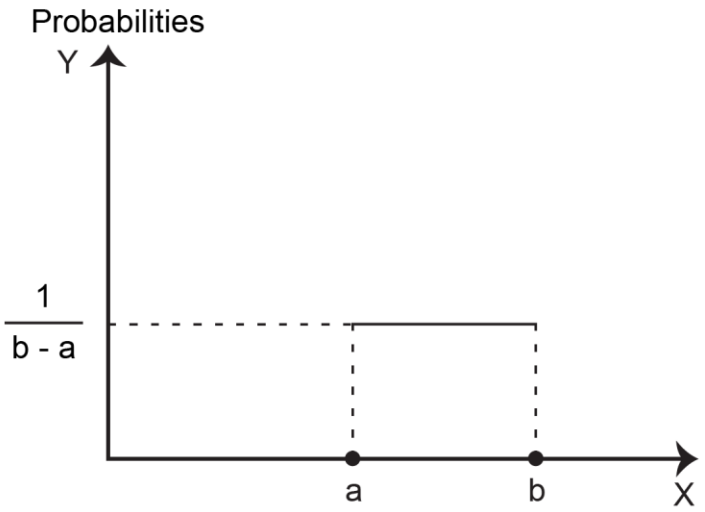


Figure 3.2: Graph of a uniform distribution

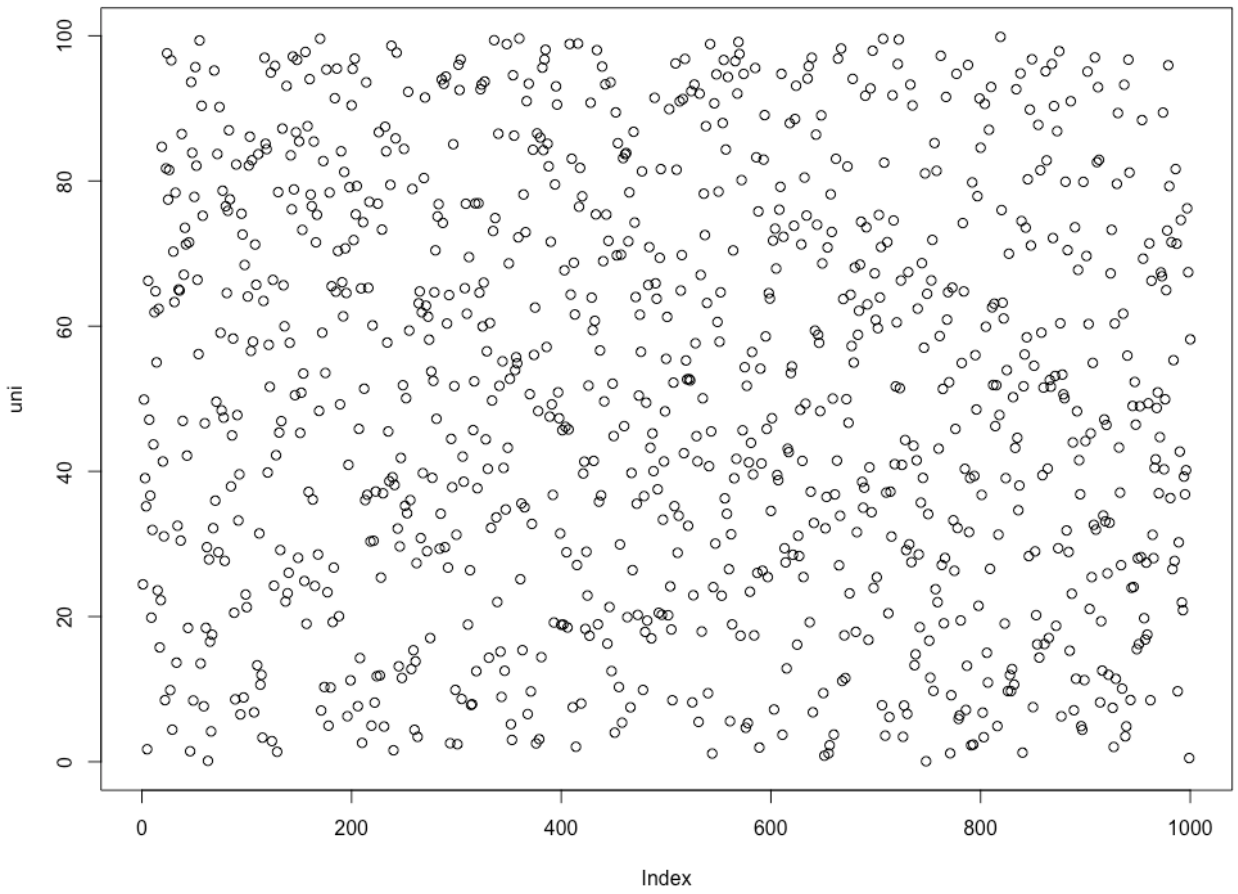


Figure 3.3: Uniform distribution

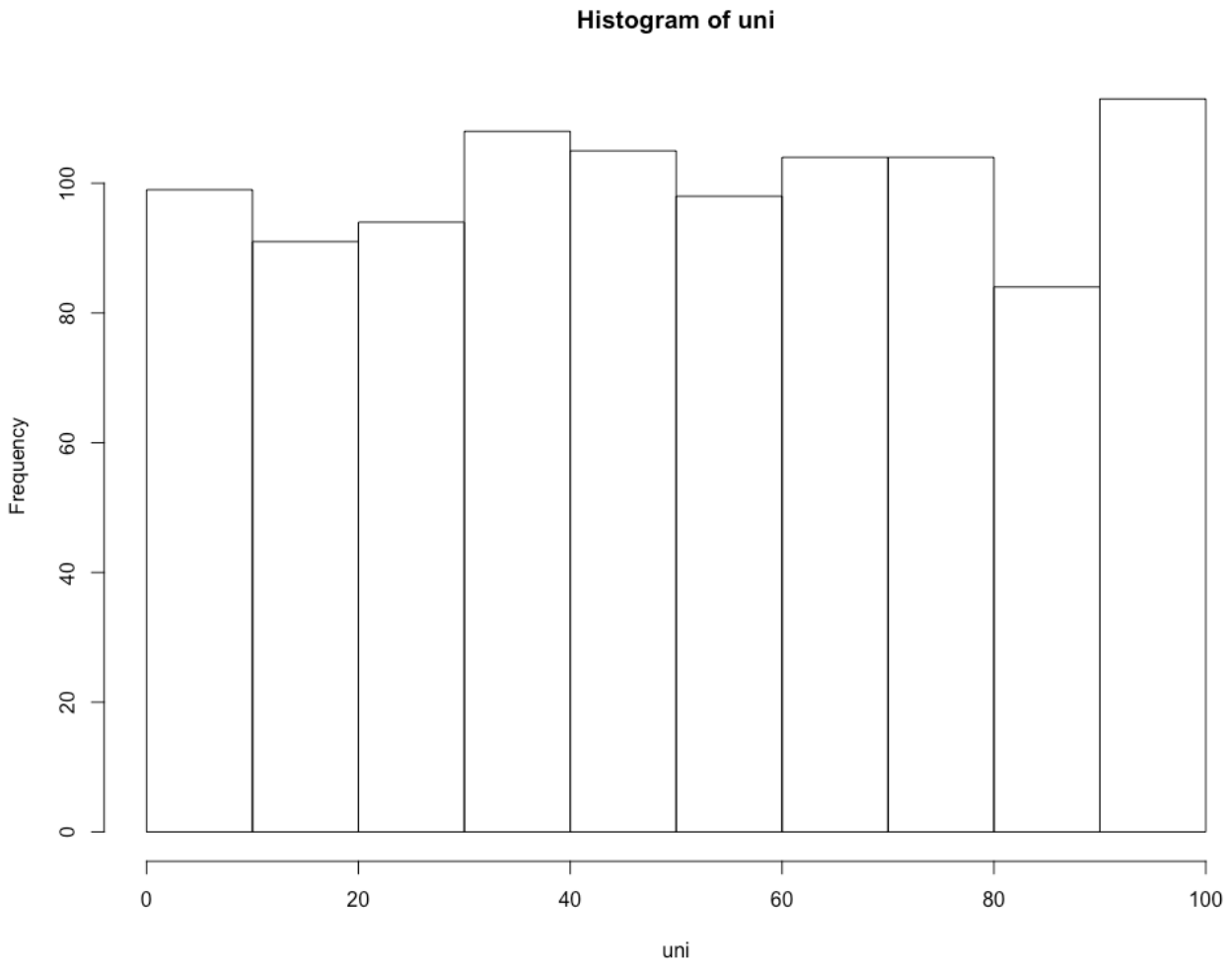


Figure 3.4: Histogram of the distribution

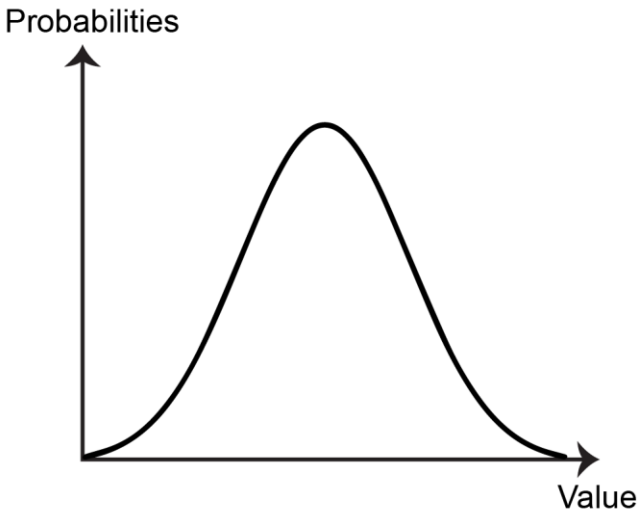


Figure 3.5: Approximate representation of a bell curve, typical with normally distributed data

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

Figure 3.6: Equation for the normal distribution

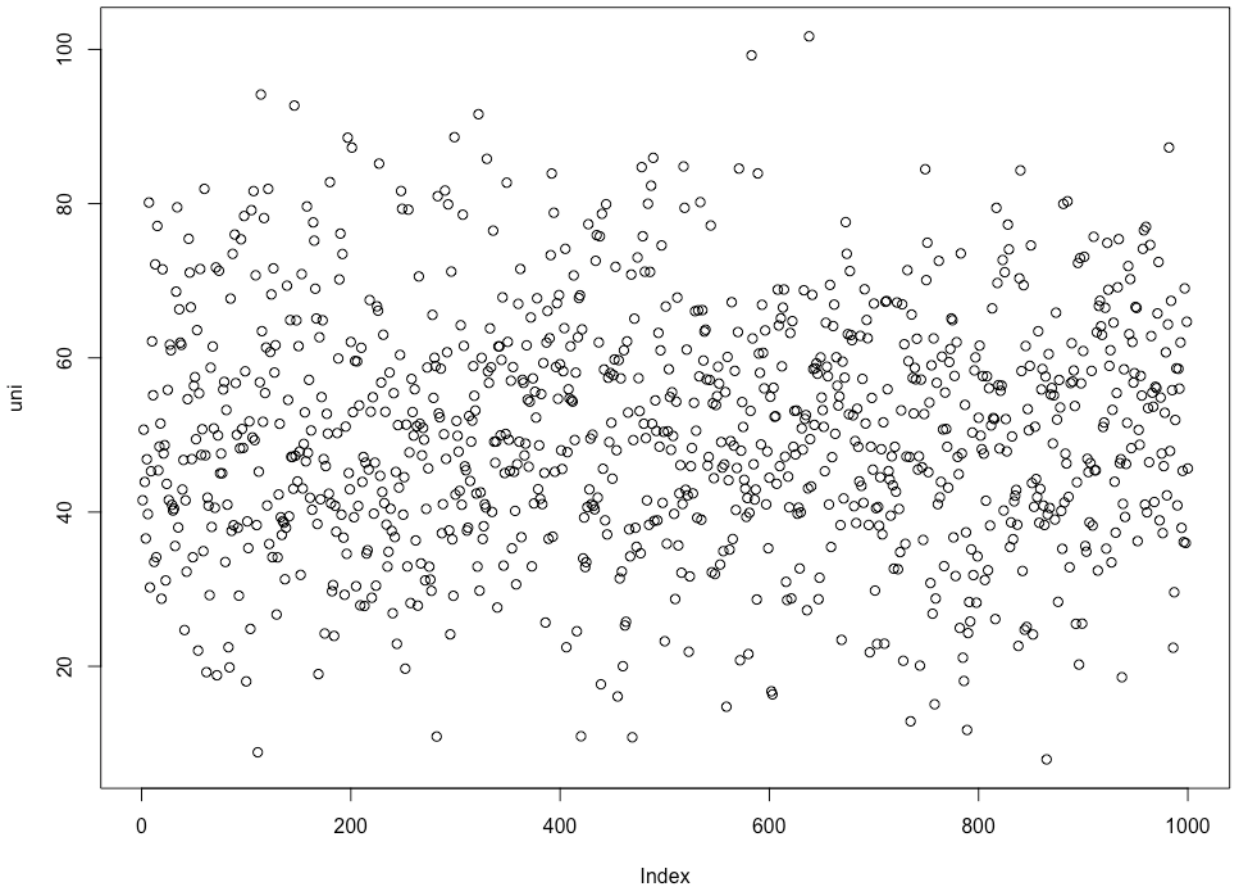


Figure 3.7: Normal distribution

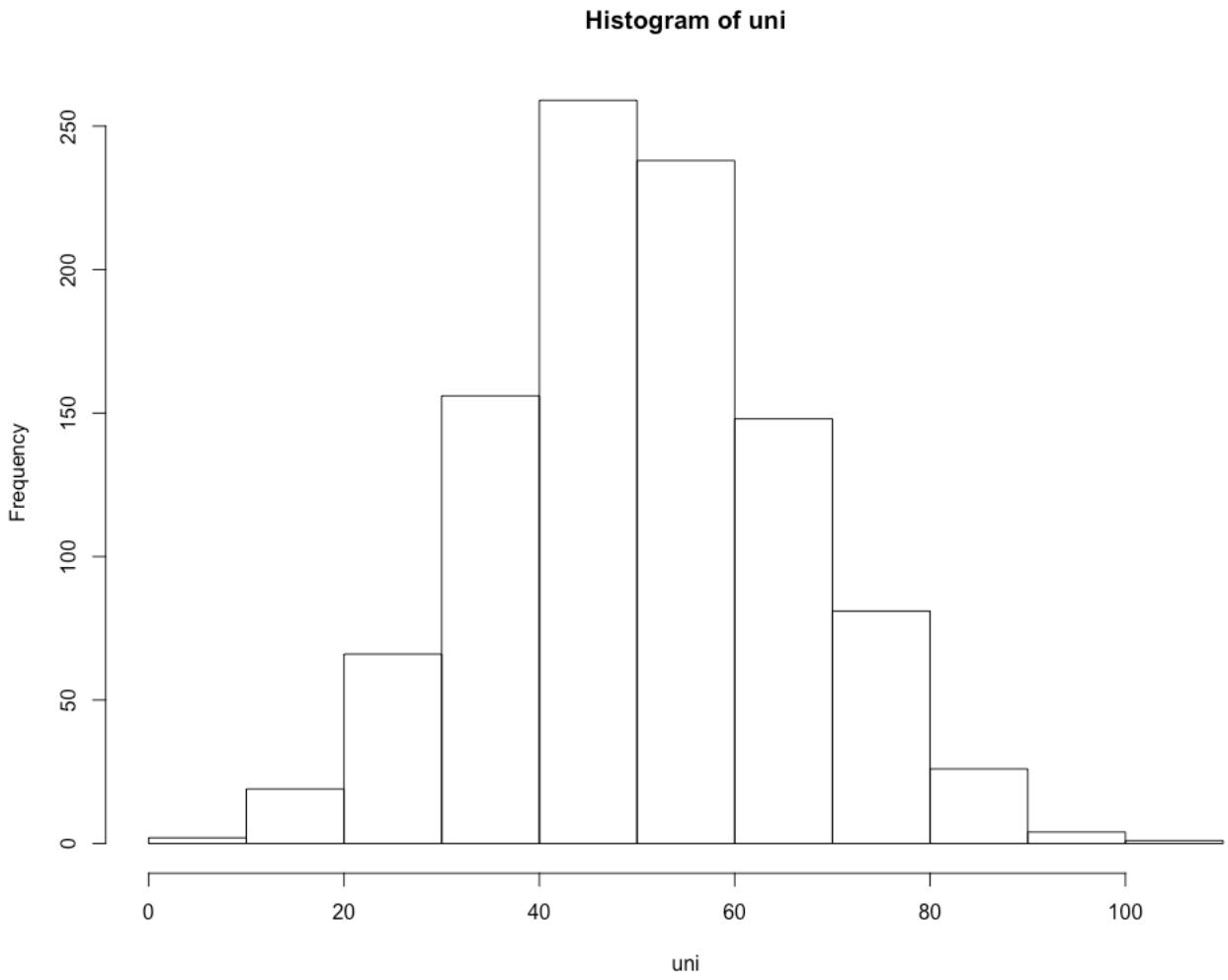


Figure 3.8: Normal distribution histogram

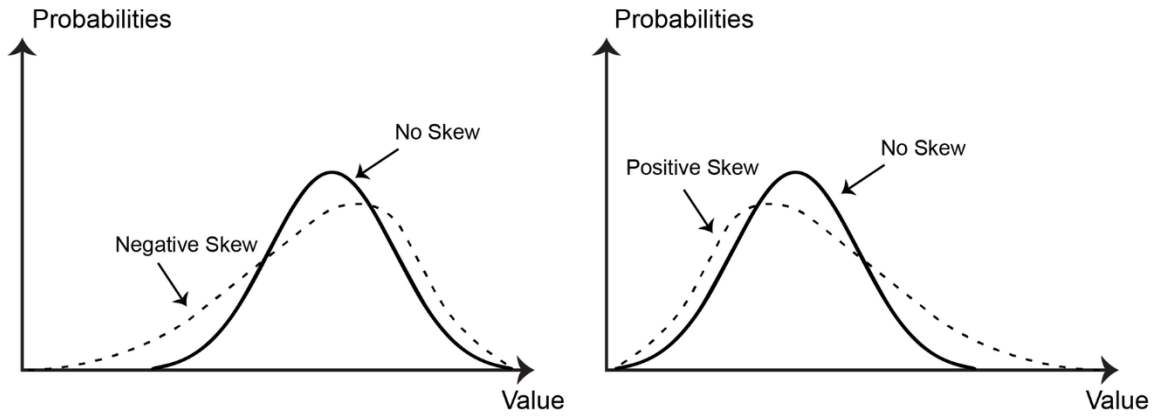


Figure 3.9: Negative skew and positive skew

$$\boldsymbol{Skewness} = \frac{E[(X-\mu)^3]}{\sigma^3}$$

Figure 3.10: Mathematical formula for skewness

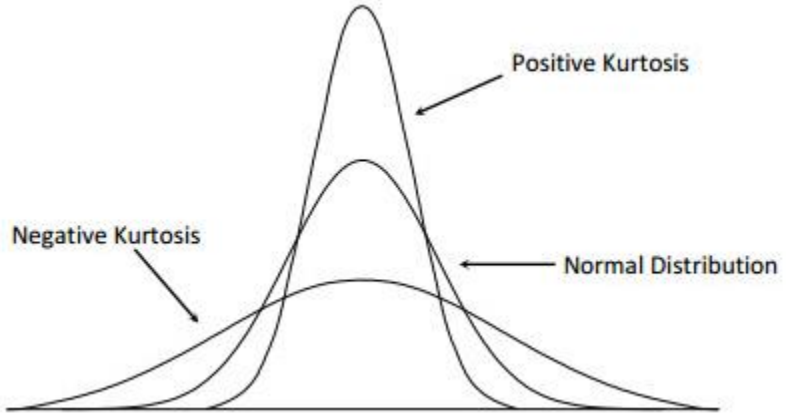


Figure 3.11: Kurtosis demonstration

$$K = \frac{E[(X-\mu)^4]}{\sigma^4}$$

Figure 3.12: Mathematical formula for Kurtosis

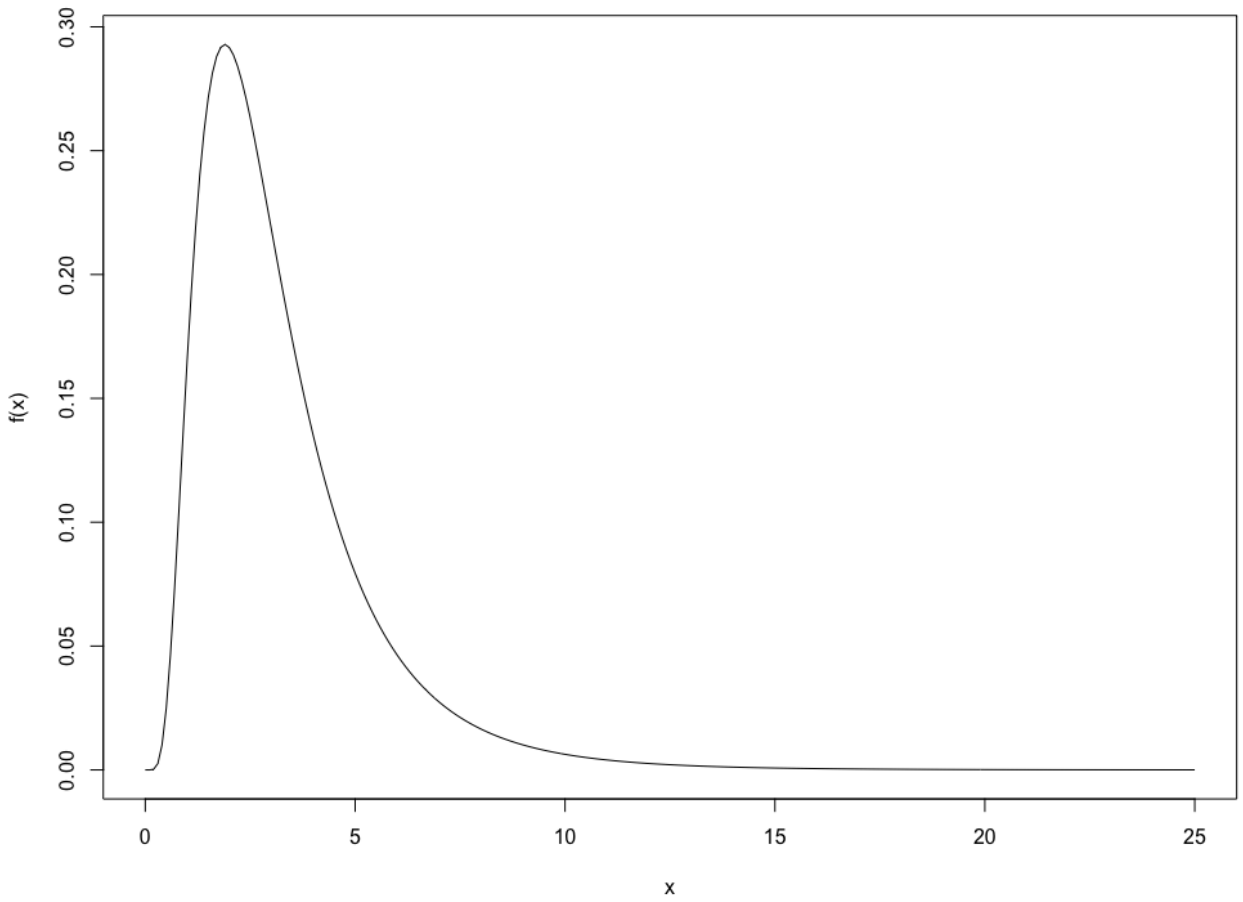


Figure 3.13: Log-normal distribution

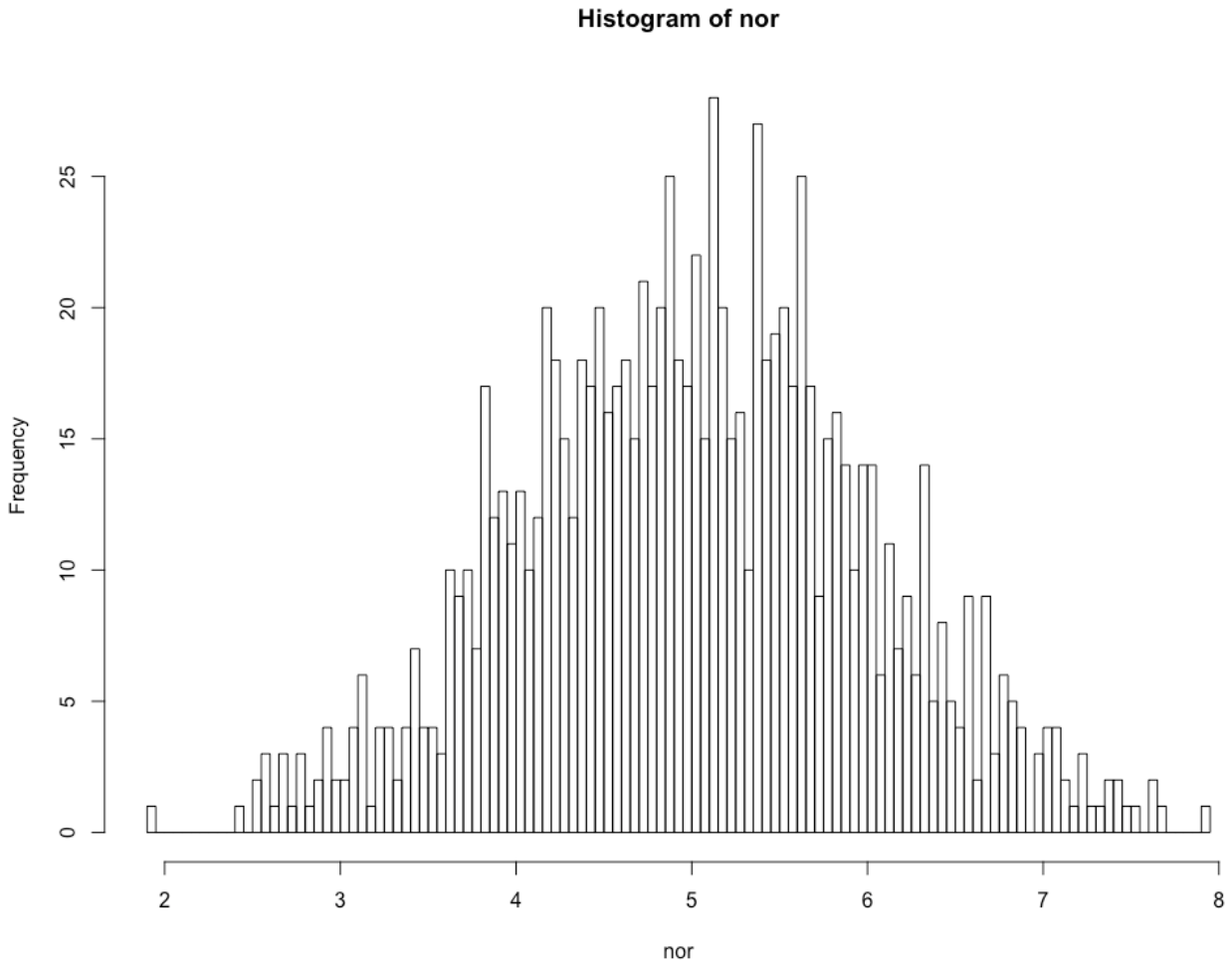


Figure 3.14: Normal distribution with a mean of 5 and a standard deviation of 1

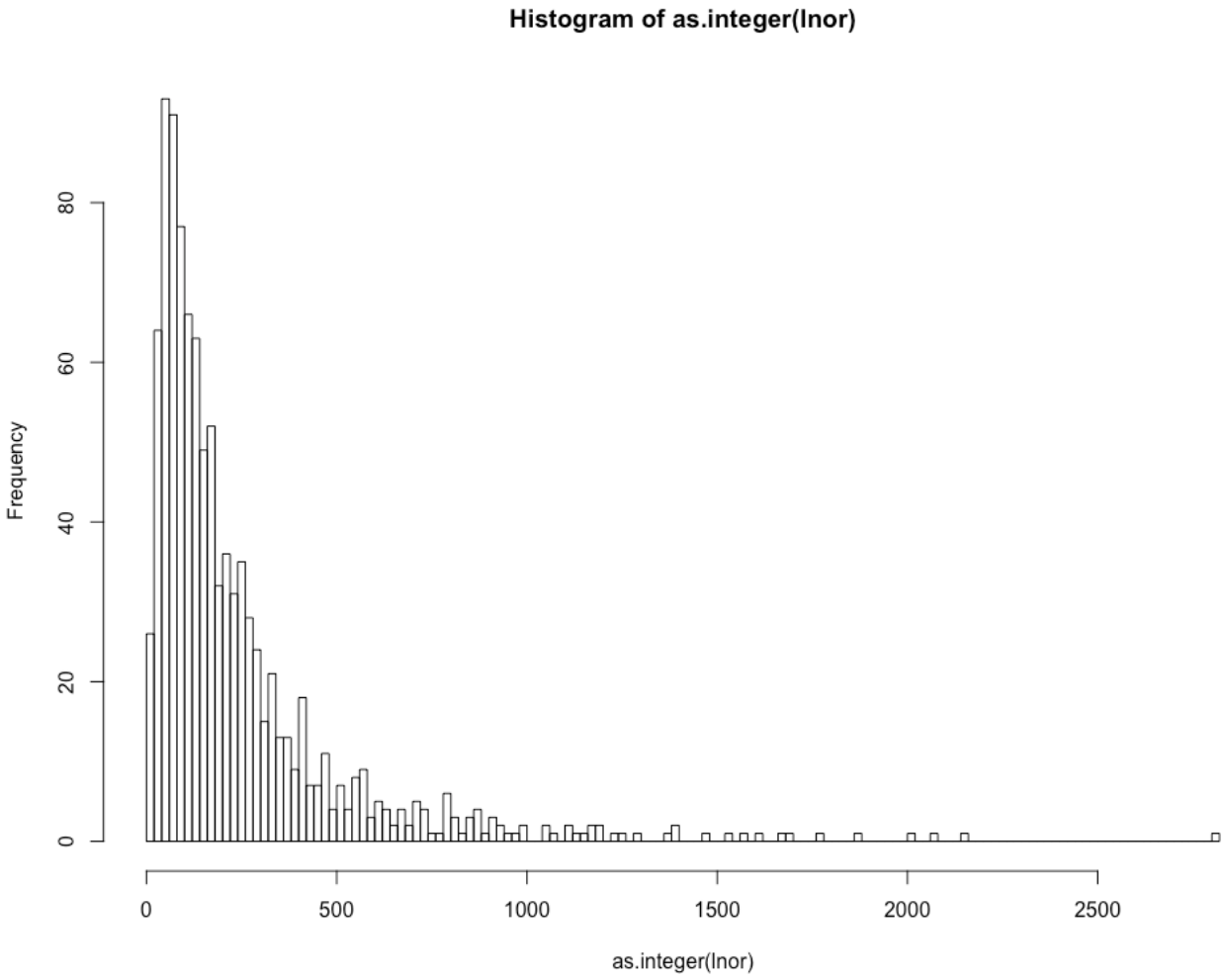


Figure 3.15: Log-normal distribution

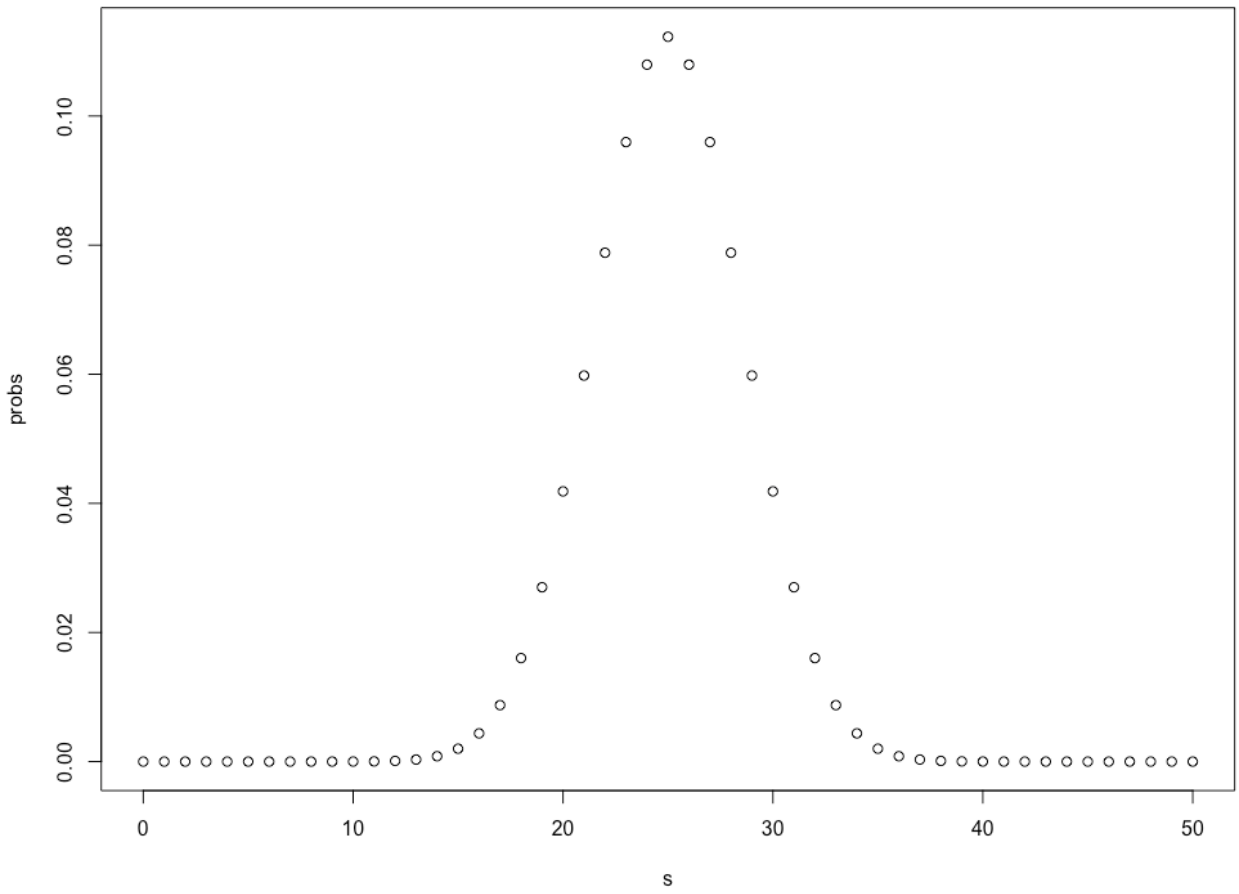


Figure 3.16: Binomial distribution

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Figure 3.17: Formula for poisson distribution

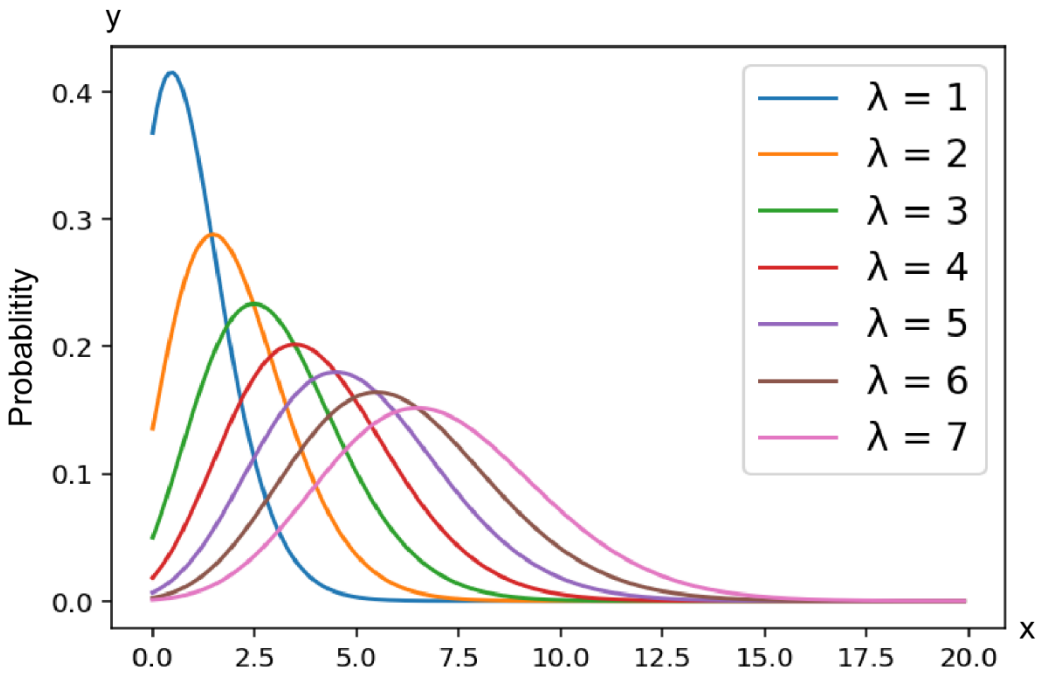


Figure 3.18: Plot for poisson distribution

$$F(x) = \begin{cases} \dfrac{\alpha x_m^{\alpha}}{x_m^{\alpha+1}} & x \geq x_m, \\ 0 & x < x_m. \end{cases}$$

Figure 3.19: Mathematical formula of the Pareto distribution

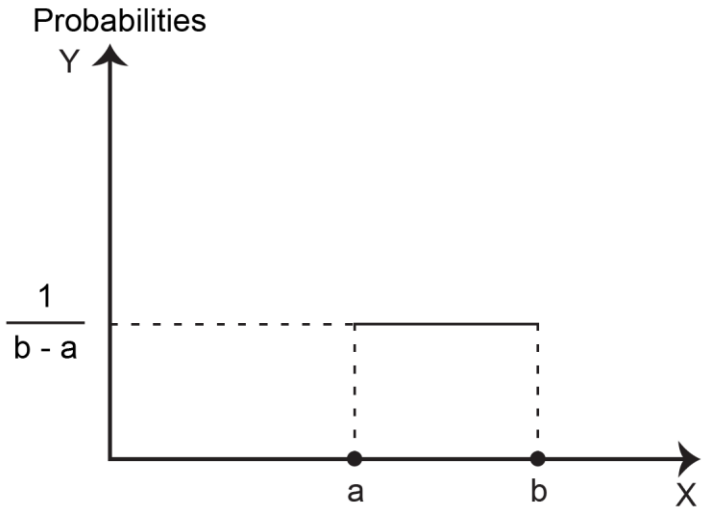


Figure 3.20: Representation of a uniform kernel function

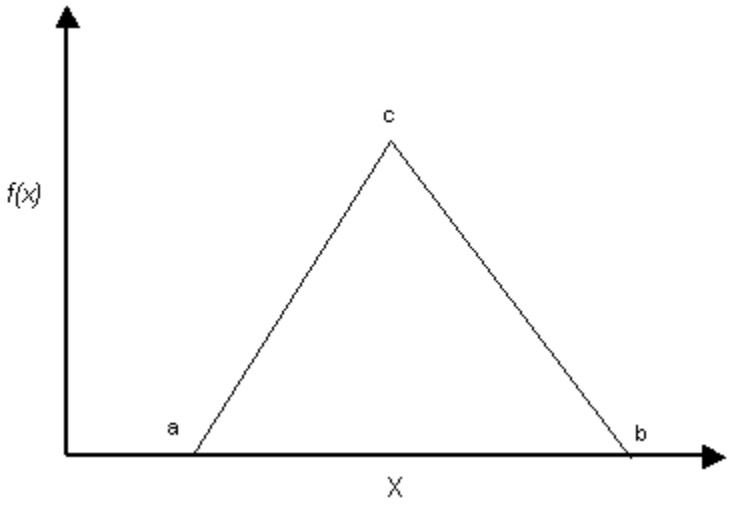


Figure 3.21: Representation of a triangular kernel function

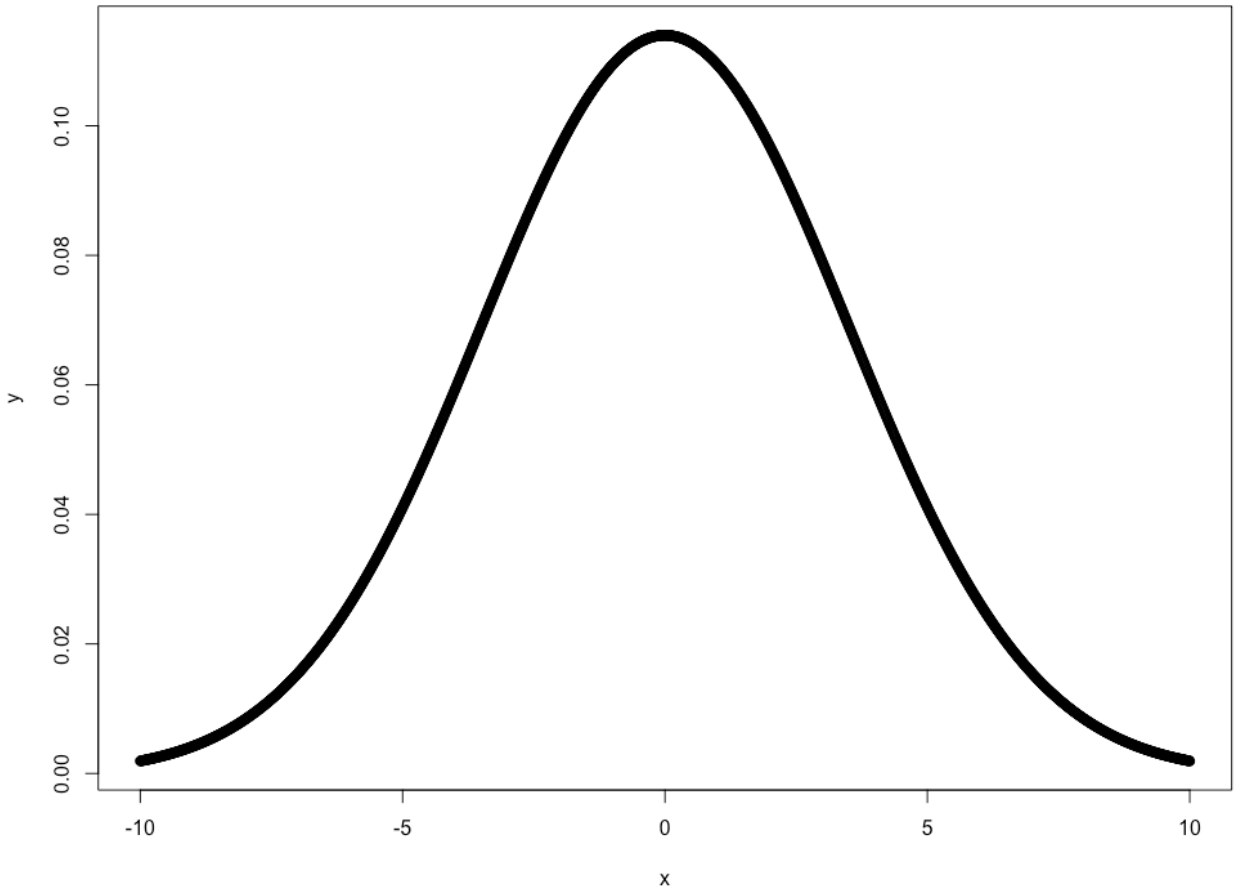


Figure 3.22: Representation of a Gaussian kernel function

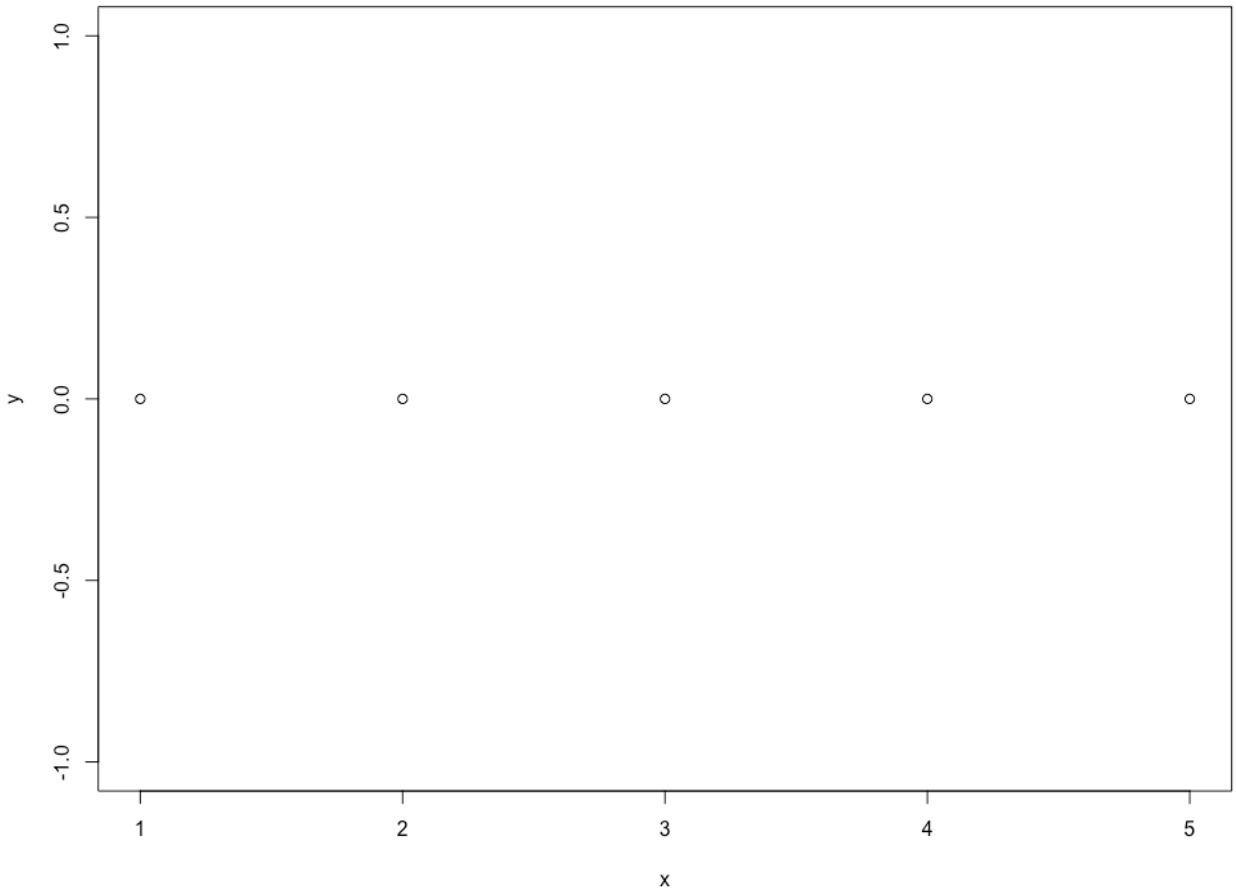


Figure 3.23: Plot of the five points

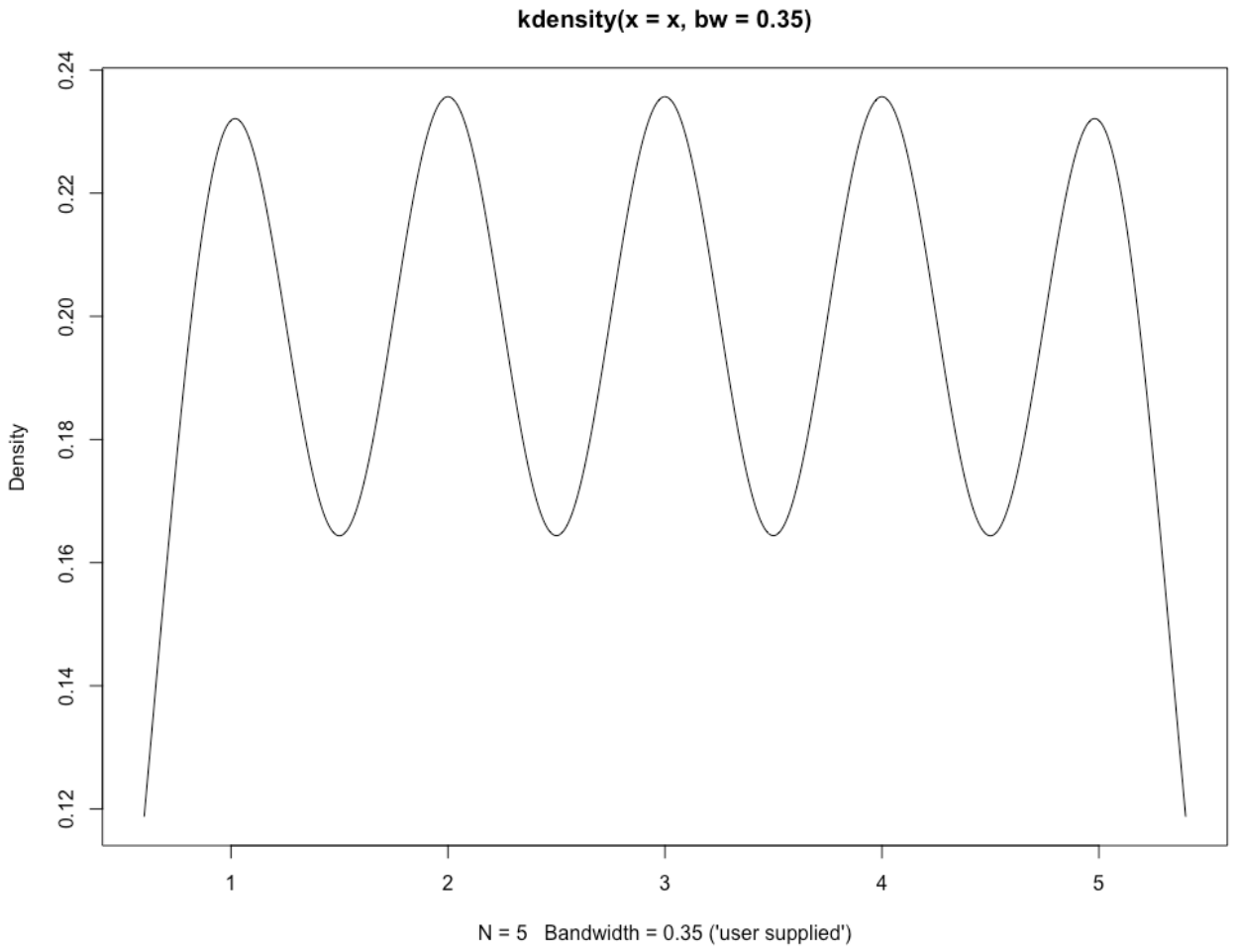


Figure 3.24: Plot of the Gaussian kernel

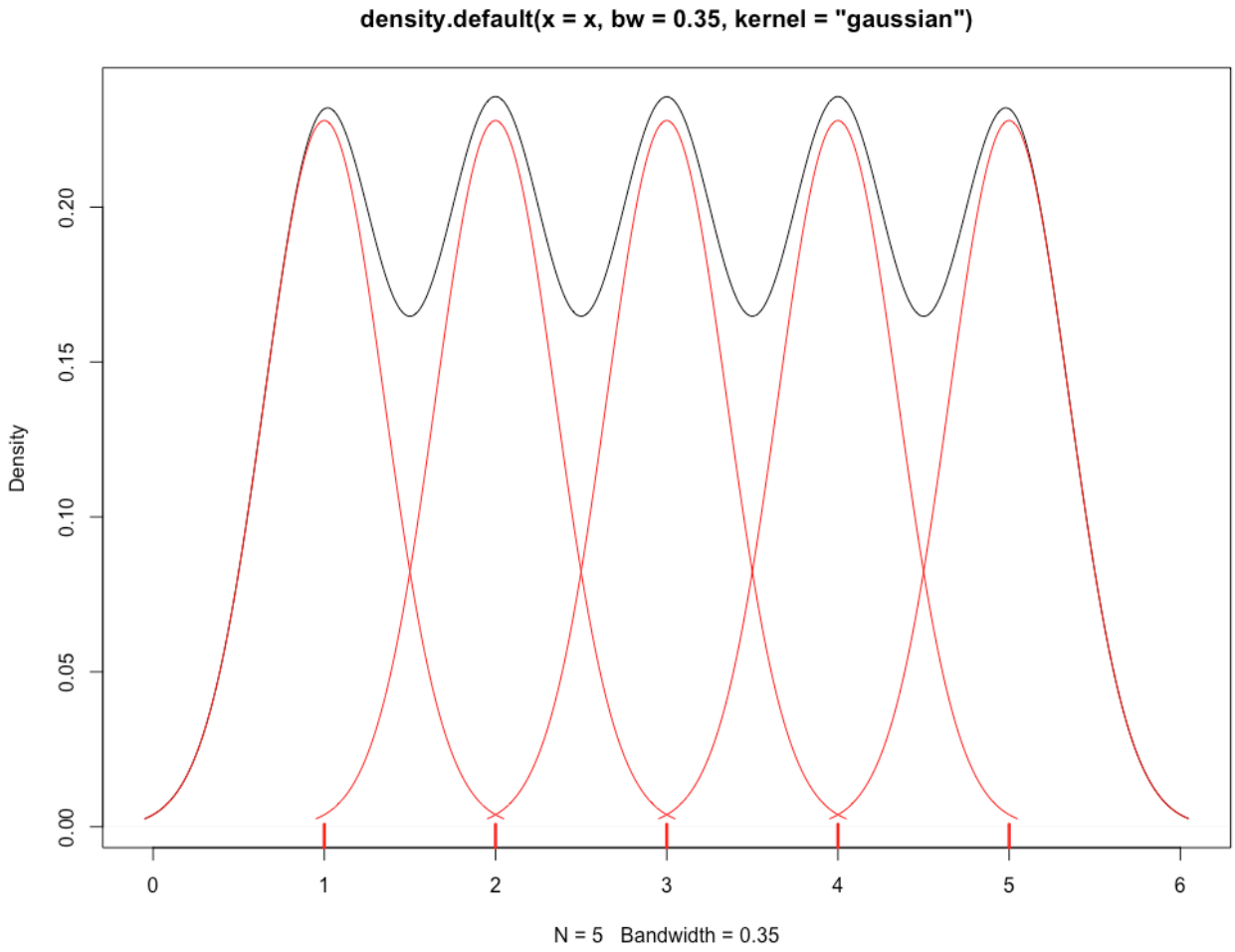


Figure 3.25: Gaussian kernel plotted on each point

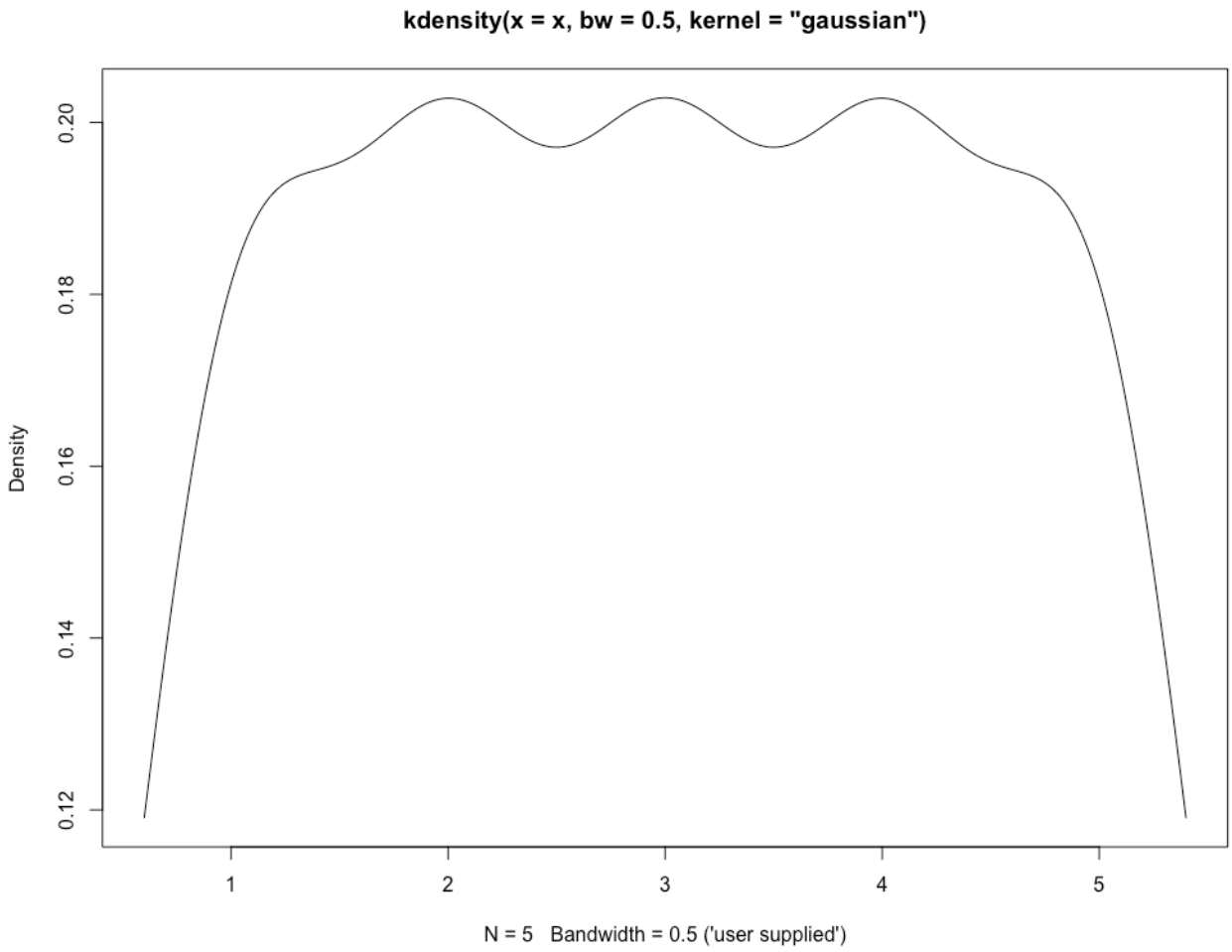


Figure 3.26: Plot of the Gaussian kernel with a bandwidth of 0.5

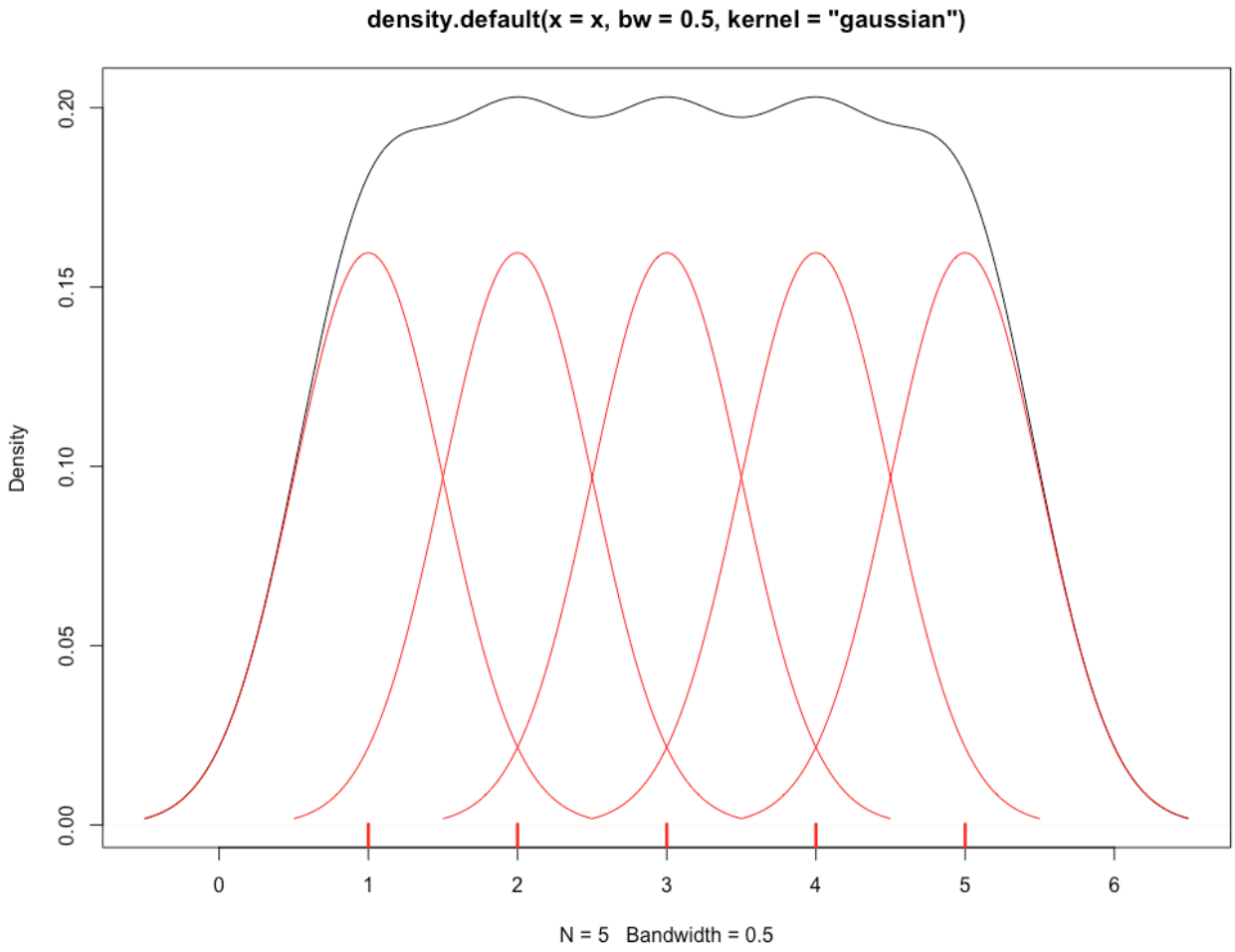


Figure 3.27: Gaussian kernel plotted on each point

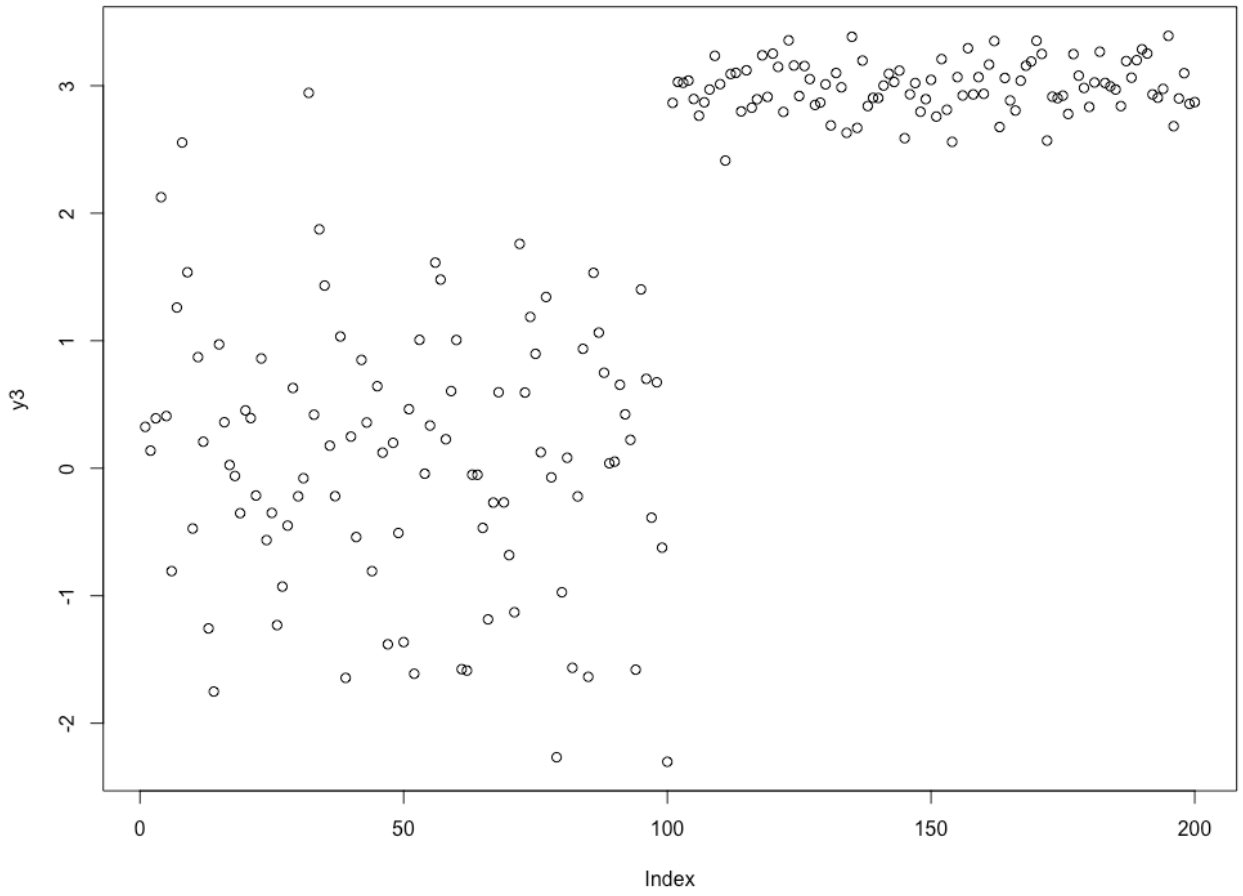


Figure 3.28: Plot of combined distributions

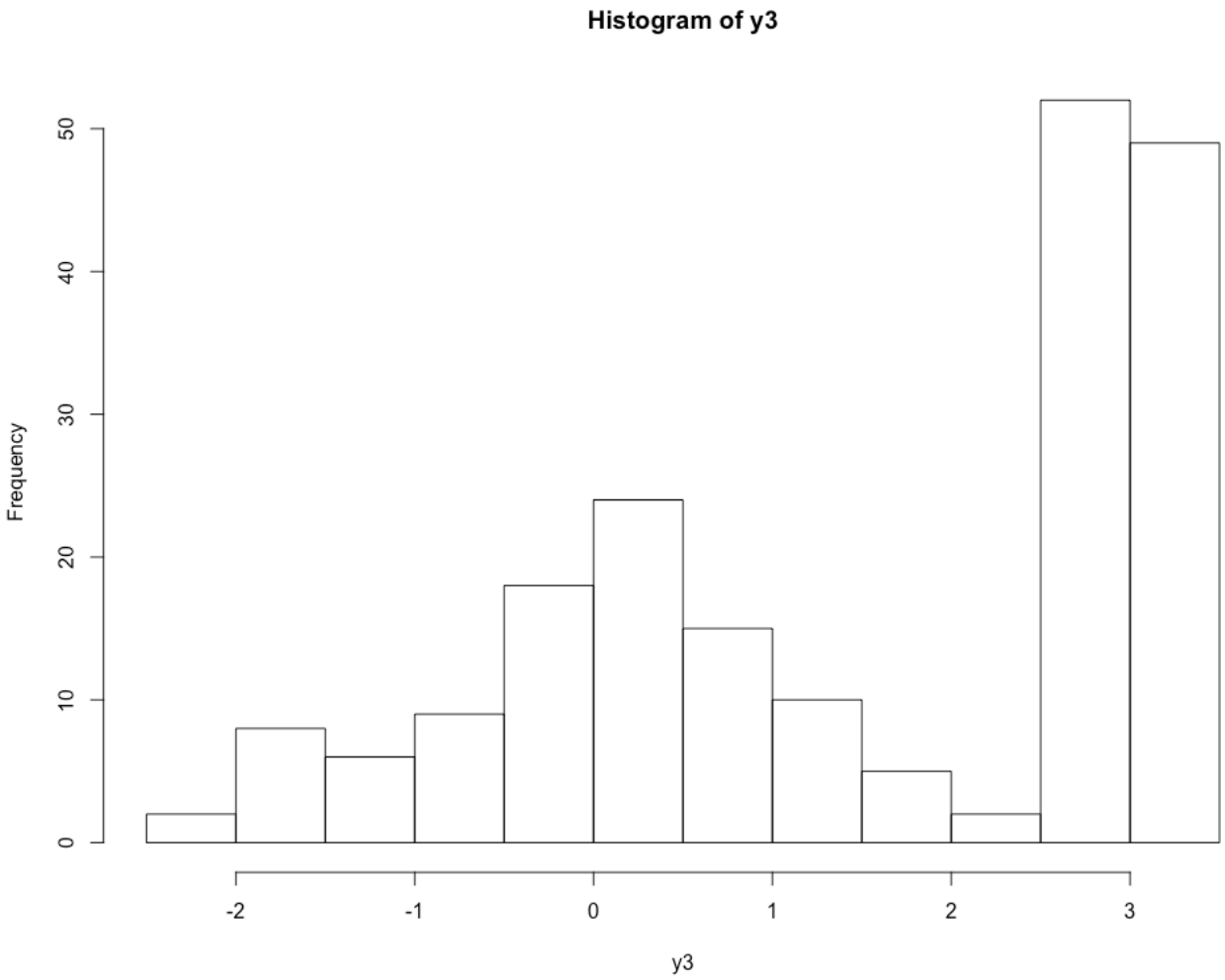


Figure 3.29: Histogram of the resultant distribution

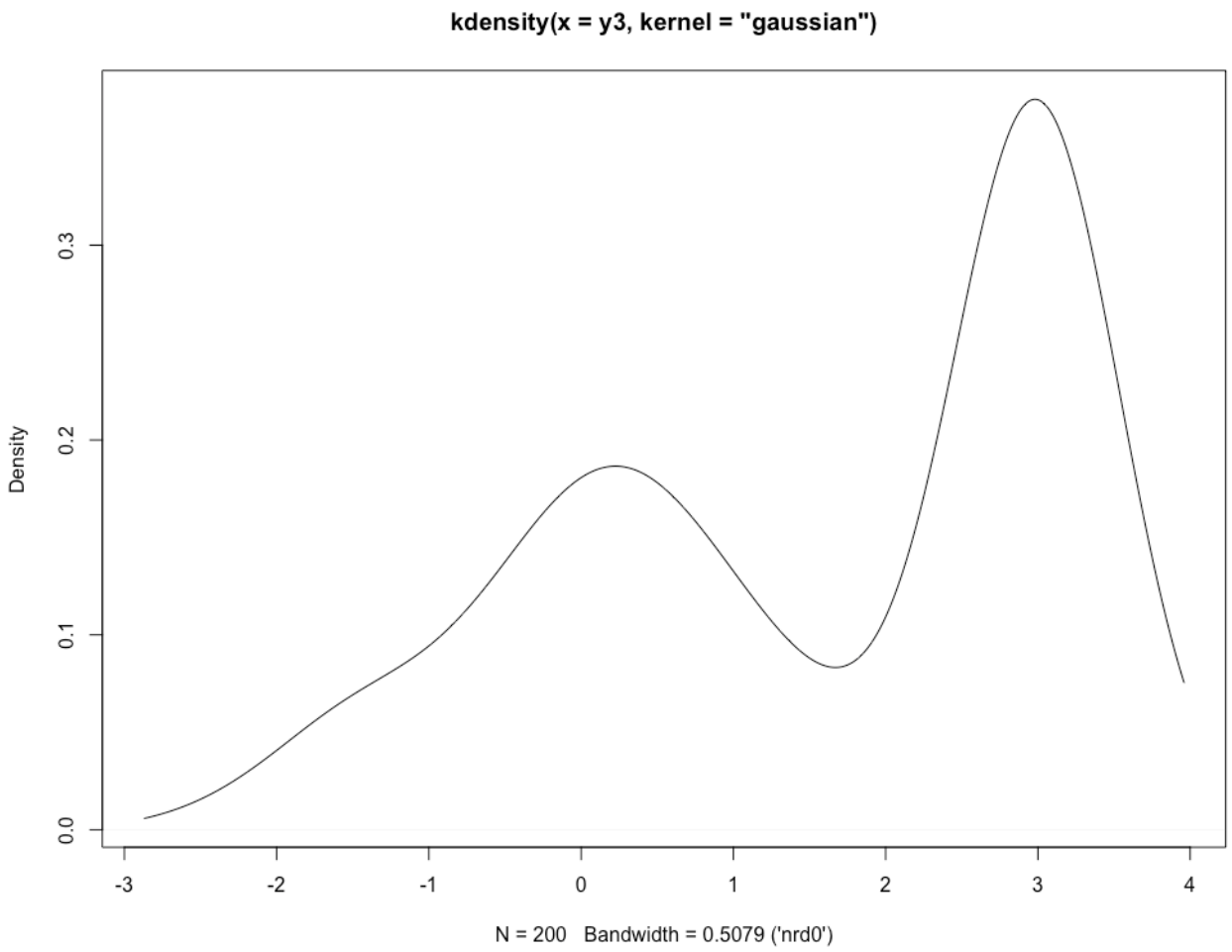


Figure 3.30: Plot of Gaussian kernel density

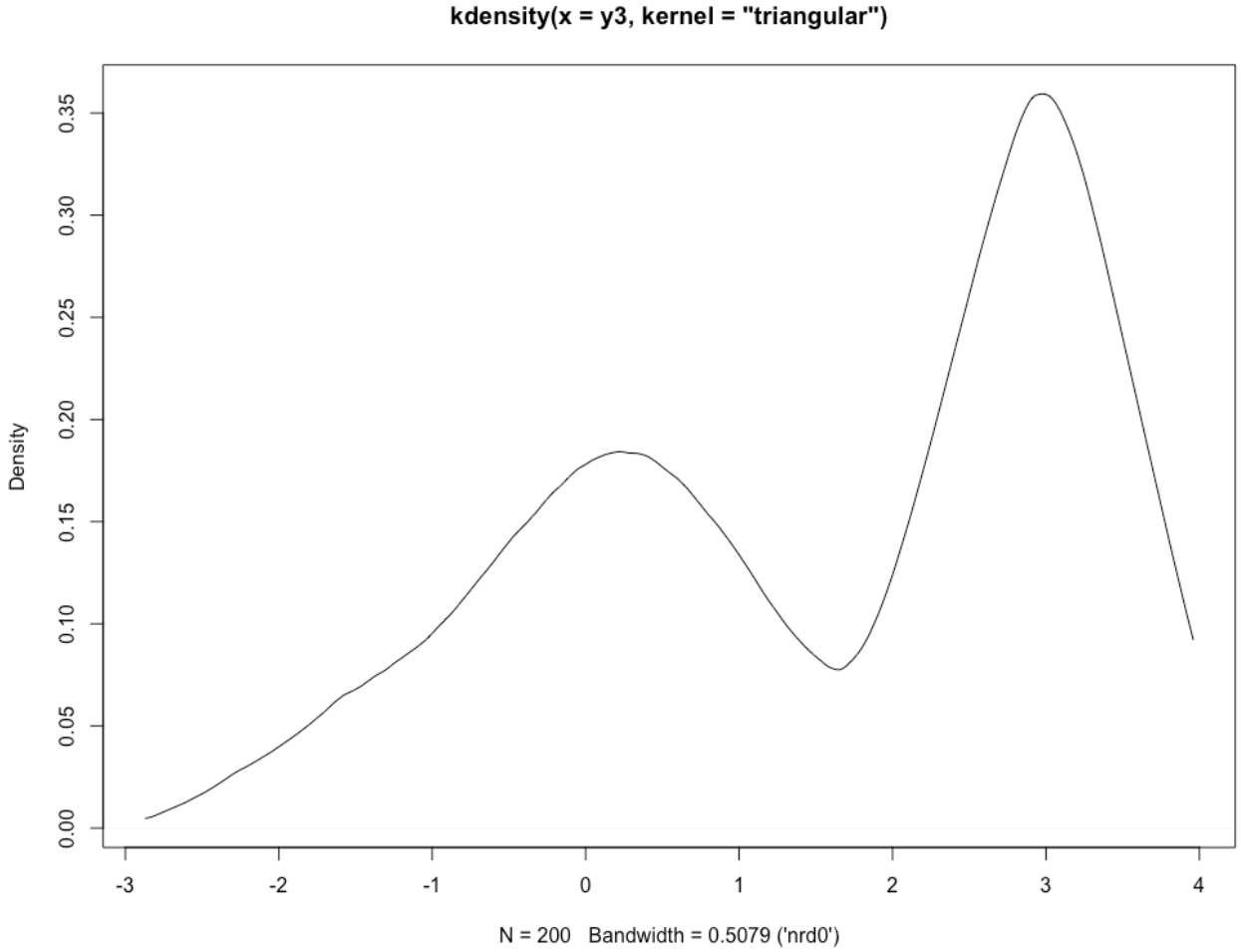


Figure 3.31: KDE with triangular kernel

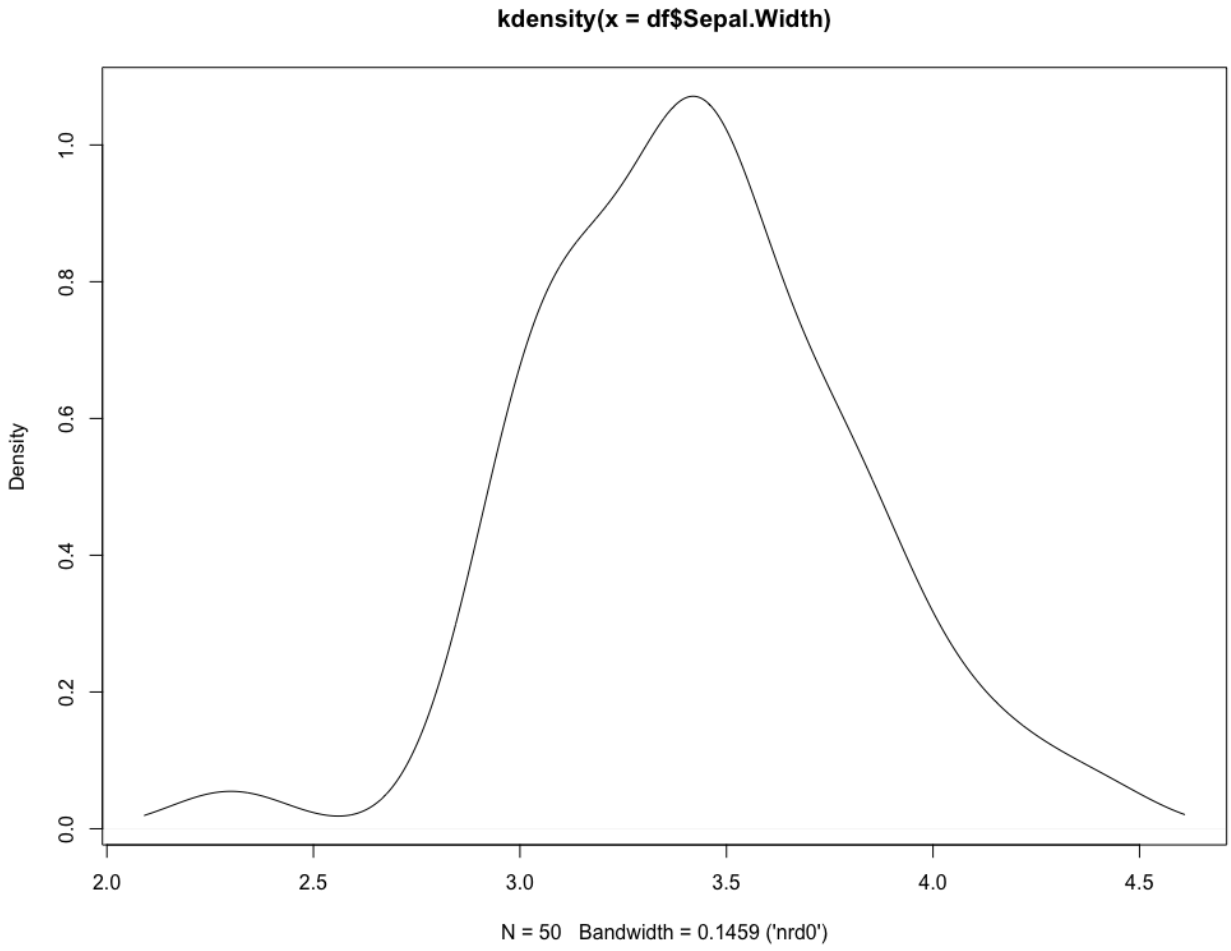


Figure 3.32: Expected plot of the KDE for sepal width

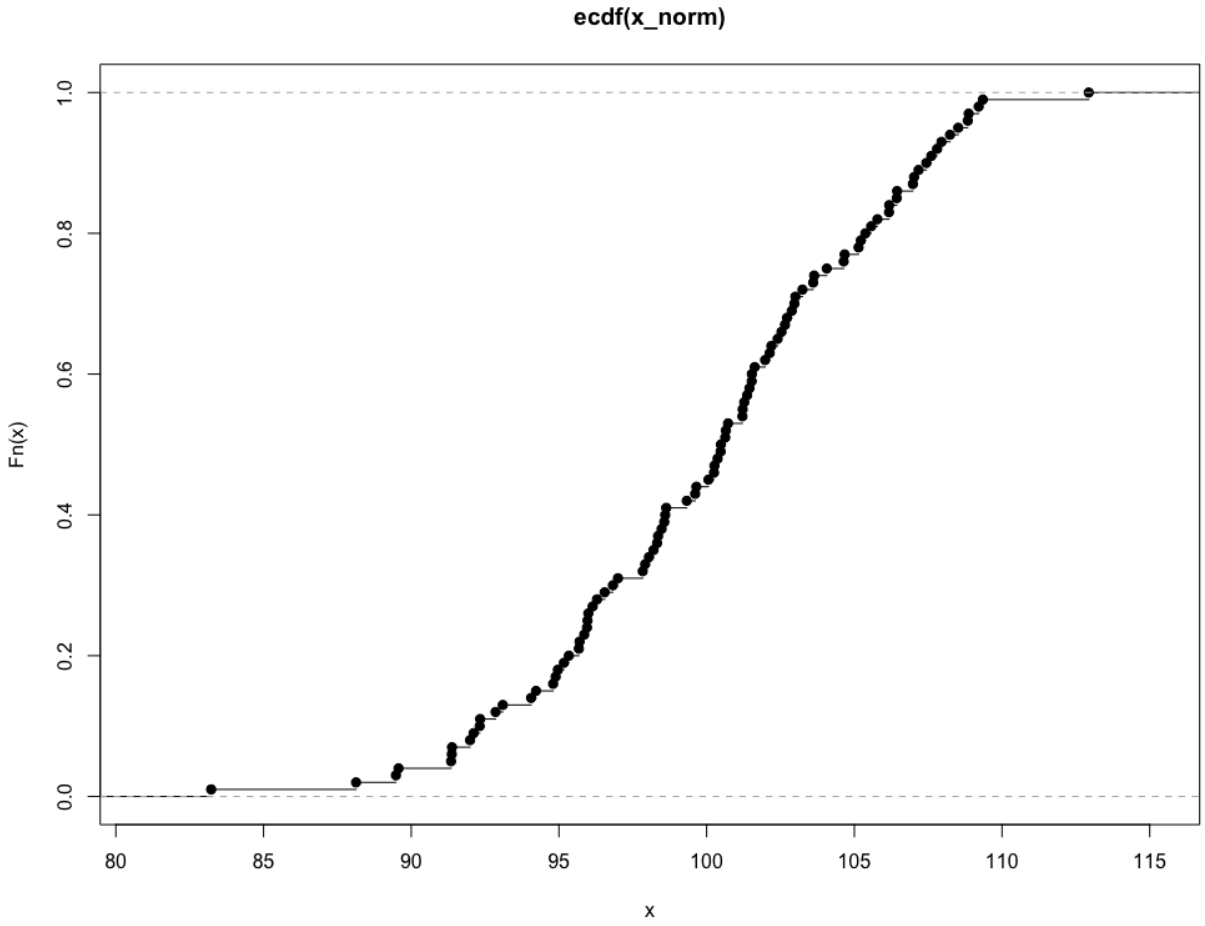


Figure 3.33: Plot of ecdf(x_norm)

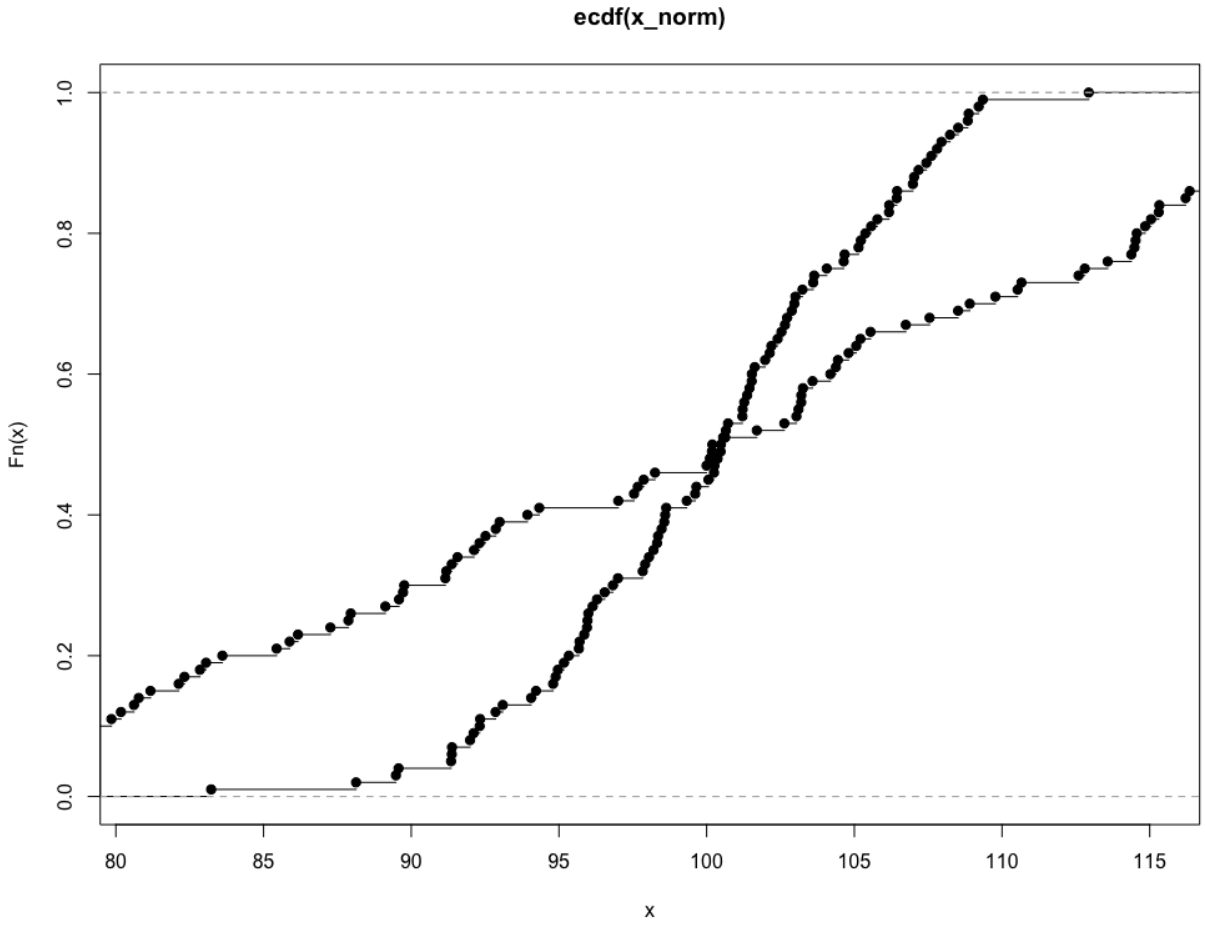


Figure 3.34: Plot of ecdf(y_unif)

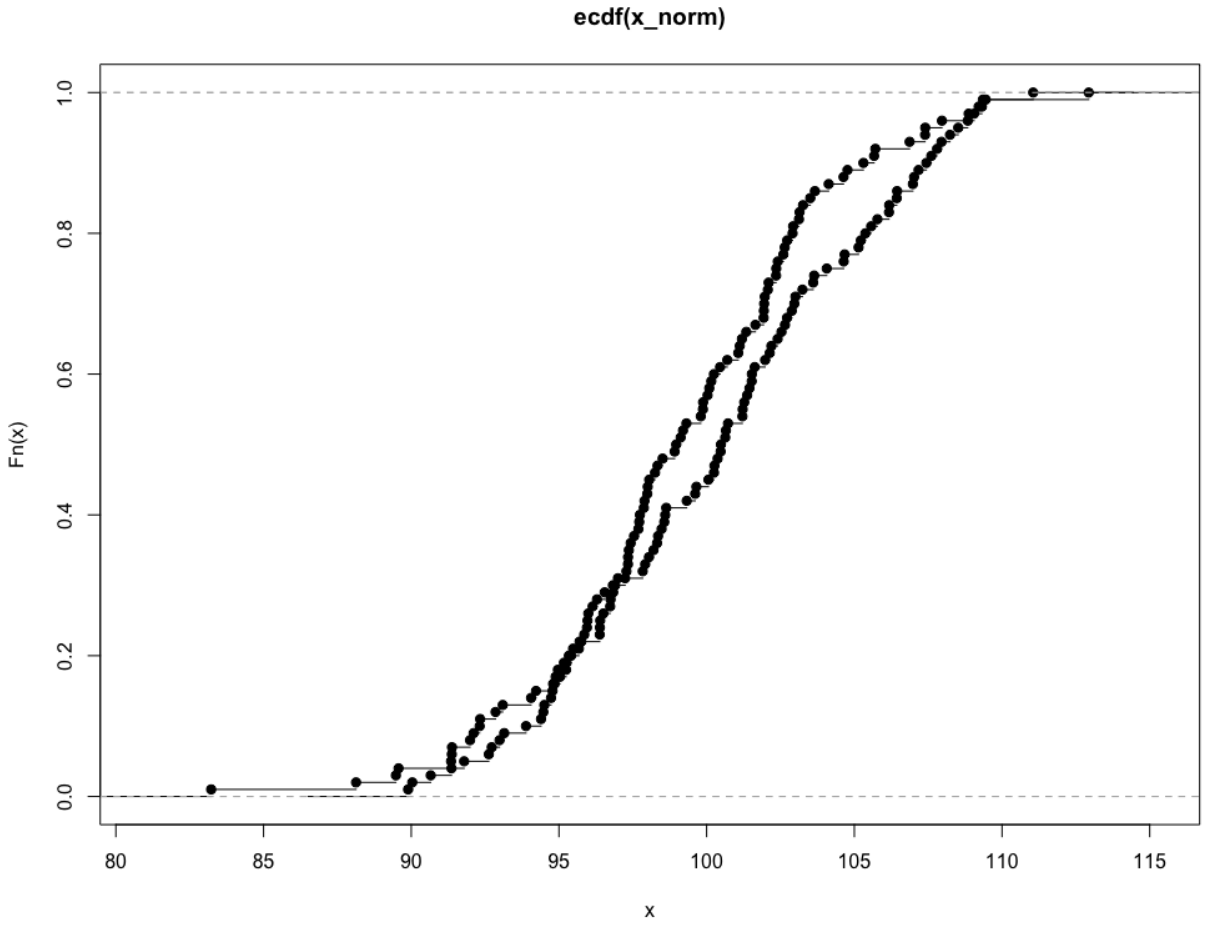


Figure 3.35 Plot of combined cdf

Lesson 4: Dimension Reduction

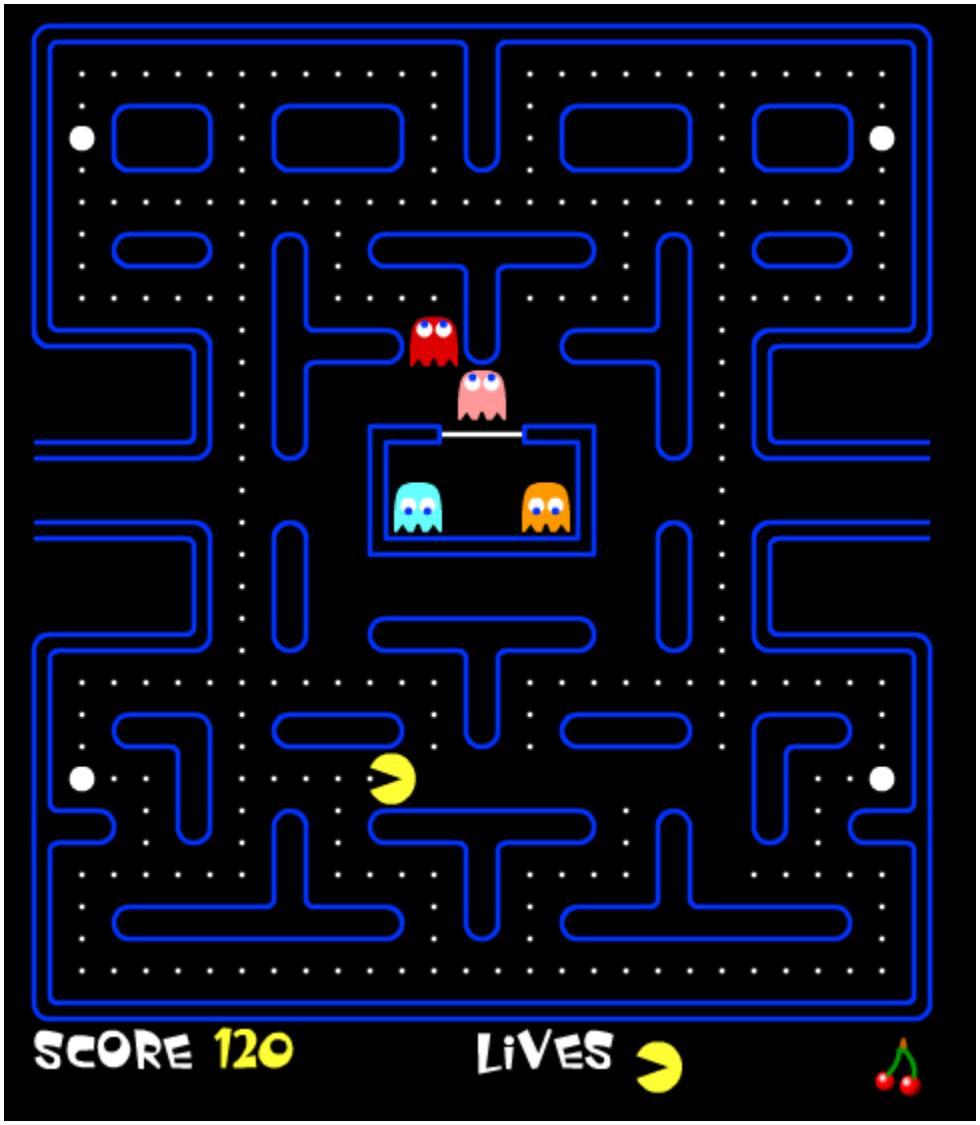


Figure 4.1: Illustration of a Pac-Man-style game

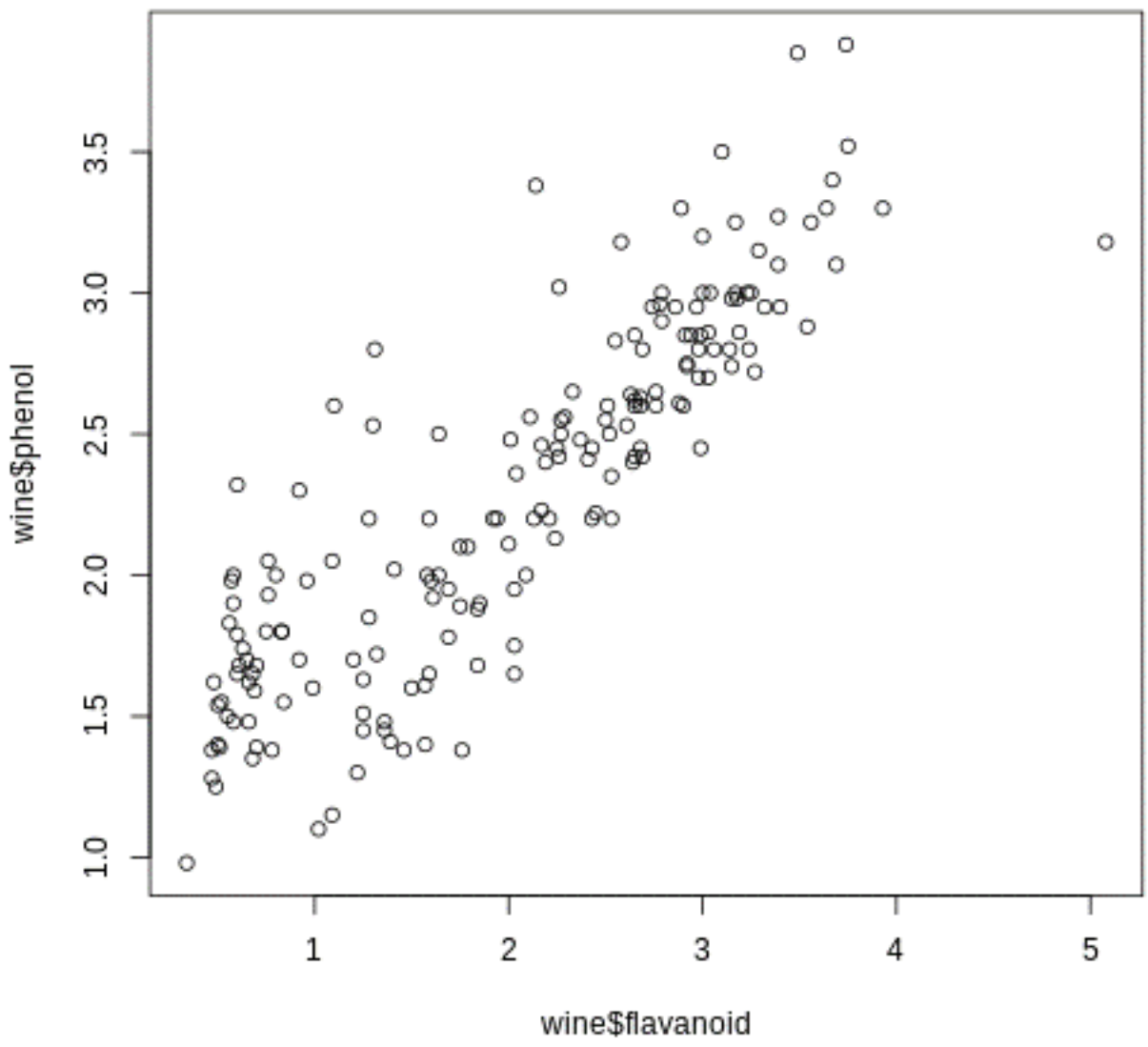


Figure 4.2: Scatterplot of two-dimensional data of flavanoids and phenol

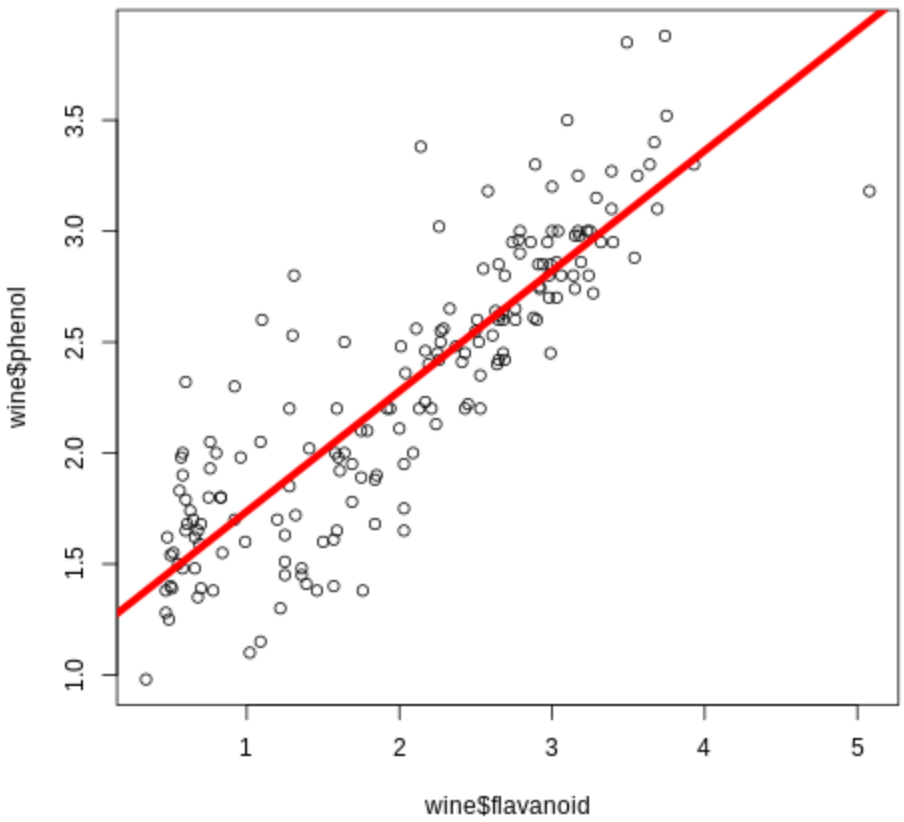


Figure 4.3: Scatterplot with a line representing correlation between flavanoids and phenol

	Peanut Butter	Jelly	Bread	Milk
Transaction 1	0	1	0	0
Transaction 2	1	1	0	0
Transaction 3	0	0	0	1
Transaction 4	1	1	1	1
Transaction 5	1	0	0	0

Figure 4.4: Table demonstrating transactions of the customers

```
40, Private,121772, Assoc-voc,11, Married-civ-spouse, Craft-repair, Husband, Male,0,0,40, ?, >50K
34, Private,245487, 7th-8th,4, Married-civ-spouse, Transport-moving, Husband, Male,0,0,45, Mexico, <=50K
25, Self-emp-not-inc,176756, HS-grad,9, Never-married, Farming-fishing, Own-child, Male,0,0,35, United-States, <=50K
32, Private,186824, HS-grad,9, Never-married, Machine-op-inspct, Unmarried, Male,0,0,40, United-States, <=50K
38, Private,28887, 11th,7, Married-civ-spouse, Sales, Husband, Male,0,0,50, United-States, <=50K
43, Self-emp-not-inc,292175, Masters,14, Divorced, Exec-managerial, Unmarried, Female,0,0,45, United-States, >50K
40, Private,193524, Doctorate,16, Married-civ-spouse, Prof-specialty, Husband, Male,0,0,60, United-States, >50K
54, Private,302146, HS-grad,9, Separated, Other-service, Unmarried, Female,0,0,20, United-States, <=50K
35, Federal-gov,76845, 9th,5, Married-civ-spouse, Farming-fishing, Husband, Male,0,0,40, United-States, <=50K
43, Private,117037, 11th,7, Married-civ-spouse, Transport-moving, Husband, Male,0,2042,40, United-States, <=50K
59, Private,109015, HS-grad,9, Divorced, Tech-support, Unmarried, Female,0,0,40, United-States, <=50K
56, Local-gov,216851, Bachelors,13, Married-civ-spouse, Tech-support, Husband, Male,0,0,40, United-States, >50K
19, Private,168294, HS-grad,9, Never-married, Craft-repair, Own-child, Male,0,0,40, United-States, <=50K
54, ?,180211, Some-college,10, Married-civ-spouse, ?, Husband, Male,0,0,60, South, >50K
39, Private,367260, HS-grad,9, Divorced, Exec-managerial, Not-in-family, Male,0,0,80, United-States, <=50K
49, Private,193366, HS-grad,9, Married-civ-spouse, Craft-repair, Husband, Male,0,0,40, United-States, <=50K
23, Local-gov,190709, Assoc-acdm,12, Never-married, Protective-serv, Not-in-family, Male,0,0,52, United-States, <=50K
20, Private,266015, Some-college,10, Never-married, Sales, Own-child, Male,0,0,44, United-States, <=50K
45, Private,386940, Bachelors,13, Divorced, Exec-managerial, Own-child, Male,0,1408,40, United-States, <=50K
30, Federal-gov,59951, Some-college,10, Married-civ-spouse, Adm-clerical, Own-child, Male,0,0,40, United-States, <=50K
22, State-gov,311512, Some-college,10, Married-civ-spouse, Other-service, Husband, Male,0,0,15, United-States, <=50K
48, Private,242406, 11th,7, Never-married, Machine-op-inspct, Unmarried, Male,0,0,40, Puerto-Rico, <=50K
21, Private,197200, Some-college,10, Never-married, Machine-op-inspct, Own-child, Male,0,0,40, United-States, <=50K
19, Private,544091, HS-grad,9, Married-AF-spouse, Adm-clerical, Wife, Female,0,0,25, United-States, <=50K
31, Private,84154, Some-college,10, Married-civ-spouse, Sales, Husband, Male,0,0,38, ?, >50K
48, Self-emp-not-inc,265477, Assoc-acdm,12, Married-civ-spouse, Prof-specialty, Husband, Male,0,0,40, United-States, <=50K
31, Private,507875, 9th,5, Married-civ-spouse, Machine-op-inspct, Husband, Male,0,0,43, United-States, <=50K
53, Self-emp-not-inc,88506, Bachelors,13, Married-civ-spouse, Prof-specialty, Husband, Male,0,0,40, United-States, <=50K
24, Private,172987, Bachelors,13, Married-civ-spouse, Tech-support, Husband, Male,0,0,50, United-States, <=50K
49, Private,94638, HS-grad,9, Separated, Adm-clerical, Unmarried, Female,0,0,40, United-States, <=50K
25, Private,289980, HS-grad,9, Never-married, Handlers-cleaners, Not-in-family, Male,0,0,35, United-States, <=50K
57, Federal-gov,337895, Bachelors,13, Married-civ-spouse, Prof-specialty, Husband, Male,0,0,40, United-States, >50K
53, Private,144361, HS-grad,9, Married-civ-spouse, Machine-op-inspct, Husband, Male,0,0,38, United-States, <=50K
44, Private,128354, Masters,14, Divorced, Exec-managerial, Unmarried, Female,0,0,40, United-States, <=50K
41, State-gov,101603, Assoc-voc,11, Married-civ-spouse, Craft-repair, Husband, Male,0,0,40, United-States, <=50K
29, Private,271466, Assoc-voc,11, Never-married, Prof-specialty, Not-in-family, Male,0,0,43, United-States, <=50K
```

Figure 4.5: Screenshot of the dataset

```
  V1                 V2     V3           V4 V5                  V6
1 39          State-gov  77516  Bachelors 13       Never-married
2 50  Self-emp-not-inc  83311  Bachelors 13  Married-civ-spouse
3 38            Private 215646    HS-grad  9            Divorced
4 53            Private 234721       11th  7  Married-civ-spouse
5 28            Private 338409  Bachelors 13  Married-civ-spouse
6 37            Private 284582    Masters 14  Married-civ-spouse
                 V7             V8      V9  V10 V11 V12
1      Adm-clerical  Not-in-family    Male 2174   0  40
2   Exec-managerial        Husband    Male    0   0  13
3  Handlers-cleaners  Not-in-family    Male    0   0  40
4  Handlers-cleaners        Husband    Male    0   0  40
5     Prof-specialty           Wife  Female    0   0  40
6   Exec-managerial           Wife  Female    0   0  40
            V13    V14
1  United-States  <=50K
2  United-States  <=50K
3  United-States  <=50K
4  United-States  <=50K
5           Cuba  <=50K
6  United-States  <=50K
```

Figure 4.6: Screenshot of the data

old	young	government_employee	self_employed	never_worked	private_employment
0.4877000092	0.5122999908	0.1041429932	0.0780381438	0.0002149811	0.6970301895
other_employment	high_school_incomplete	high_school_complete	bachelors	post_bachelors	married
0.0568164368	0.1203894229	0.5888639784	0.1644605510	0.0832898253	0.4599367341
never_married	divorced_separated	widowed	clerical	managerial	moving
0.3280918891	0.1679309604	0.0304966064	0.1157826848	0.1248733147	0.0490464052
farming_fishing	craft_repair	sales	tech_support	service	armed_forces
0.0305273180	0.1258867971	0.1120972943	0.0285003532	0.0245078468	0.0002764043
other_occupation	male	female	high_hours	low_hours	usa
0.2873069009	0.6692054912	0.3307945088	0.2942477197	0.7057522803	0.8958570068
not_usa	low_income	high_income			
0.1041429932	0.7591904426	0.2408095574			

Figure 4.7: Section of resulting dataset of dummy variables

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1	33	12	0.1481834	0.8645404	1.879694
[2,]	6	33	12	0.1296029	0.8502922	1.848715
[3,]	26	33	12	0.1823654	0.8913239	1.937927
[4,]	29	33	12	0.1069992	0.8742785	1.900867
[5,]	30	33	12	0.1878628	0.8530191	1.854644

Figure 4.8: Output of thirdpass_high

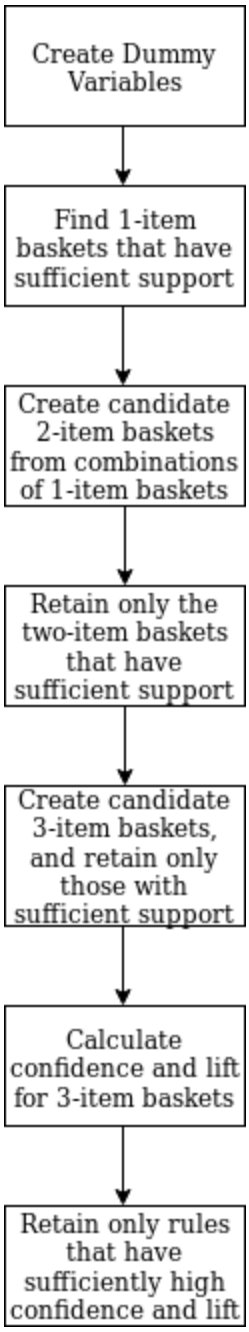


Figure 4.9: Flowchart of steps followed in market basket analysis

```
              [,1]          [,2]         [,3]         [,4]         [,5]         [,6]         [,7]          [,8]          [,9]         [,10]
 [1,] -0.0016592647 -1.203406e-03  0.016873809 -0.141446778  0.020336977 -0.194120104  0.923280337 -2.848207e-01  8.660061e-02 -2.245000e-03
 [2,]  0.0006810156 -2.154982e-03  0.122003373 -0.160389543 -0.612883454 -0.742472963 -0.150109941  6.467447e-02  1.566214e-02 -1.850935e-02
 [3,] -0.0001949057 -4.593693e-03  0.051987430  0.009772810  0.020175575 -0.041752912  0.045009549  1.493395e-01  7.364985e-02 -8.679965e-02
 [4,]  0.0046713006 -2.645039e-02  0.938593003  0.330965260  0.064352340  0.024065303  0.031526583 -1.515391e-02  2.044578e-03  3.554028e-03
 [5,] -0.0178680075 -9.993442e-01 -0.029780248  0.005393756 -0.006149345  0.001923782  0.001797363  3.552212e-03 -1.963668e-03 -4.051542e-05
 [6,] -0.0009898297 -8.779622e-04 -0.040484644  0.074584656  0.315245063 -0.278716809 -0.020185710  1.772379e-01  2.556729e-01  8.471951e-01
 [7,] -0.0015672883  5.185073e-05 -0.085443339  0.169086724  0.524761088 -0.433597955 -0.038868518  2.481166e-01  3.783067e-01 -5.201384e-01
 [8,]  0.0001230867  1.354479e-03  0.013510780 -0.010805561 -0.029647512  0.021952834 -0.004665483 -6.497968e-03  3.675204e-02  3.771319e-02
 [9,] -0.0006006078 -5.004400e-03 -0.024659382  0.050120952  0.251182529 -0.241884488 -0.309799487 -8.704332e-01 -5.152017e-02  9.722752e-03
[10,] -0.0023271432 -1.510035e-02  0.291398464 -0.878893693  0.331747051 -0.002739609 -0.112836514  8.128692e-02 -9.902908e-02 -2.314712e-02
[11,] -0.0001713800  7.626731e-04 -0.025977662  0.060034945  0.051524077  0.023776167  0.030819813  2.951904e-03  3.306512e-02 -3.846983e-02
[12,] -0.0007049316  3.495364e-03 -0.070323969  0.178200254  0.260639176 -0.288912753  0.101973518  1.867145e-01 -8.737465e-01  1.701708e-02
[13,] -0.9998229365  1.777381e-02  0.004528682  0.003112916 -0.002298569  0.001212255 -0.001076189 -1.034095e-05 -7.255852e-05  4.926638e-05
             [,11]         [,12]         [,13]
 [1,] -0.0149715080  1.565141e-02  8.029245e-03
 [2,] -0.0231876506 -6.729555e-02 -1.109039e-02
 [3,]  0.9540106426  1.320630e-01 -1.736857e-01
 [4,] -0.0528216953 -5.393806e-03  1.939563e-03
 [5,] -0.0030248882 -6.208885e-04  2.284536e-03
 [6,]  0.0088016070 -3.882903e-03 -2.669144e-02
 [7,] -0.1332046120  3.748803e-02  6.959853e-02
 [8,]  0.1991789841 -1.475524e-01  9.664662e-01
 [9,]  0.1356214601  1.311883e-02 -1.760357e-02
[10,] -0.0098196717 -5.035557e-02 -4.632943e-03
[11,]  0.0975106606 -9.755619e-01 -1.665508e-01
[12,]  0.0284851062 -1.163025e-02  4.419224e-02
[13,] -0.0002404522  9.999951e-05  3.626701e-05
```

Figure 4.10: Eigenvectors of wine

```
 [1] 9.920179e+04 1.725353e+02 9.438114e+00 4.991179e+00 1.228845e+00 8.410639e-01 2.789735e-01 1.513813e-01 1.120968e-01 7.170260e-02 3.757598e-02
[12] 2.107237e-02 8.203703e-03
```

Figure 4.11: Eigenvalues of wine

```
 [1] -0.0016592647  0.0006810156 -0.0001949057  0.0046713006 -0.0178680075 -0.0009898297 -0.0015672883  0.0001230867 -0.0006006078 -0.0023271432
[11] -0.0001713800 -0.0007049316 -0.9998229365
```

Figure 4.12: This first eigenvector expresses a linear combination of our original dimensions

```
              [,1]
[1,] -1067.0557
[2,] -1051.5901
[3,] -1186.5538
[4,] -1481.7328
[5,]  -736.9213
[6,] -1451.7239
```

Figure 4.13: Transformed dataset

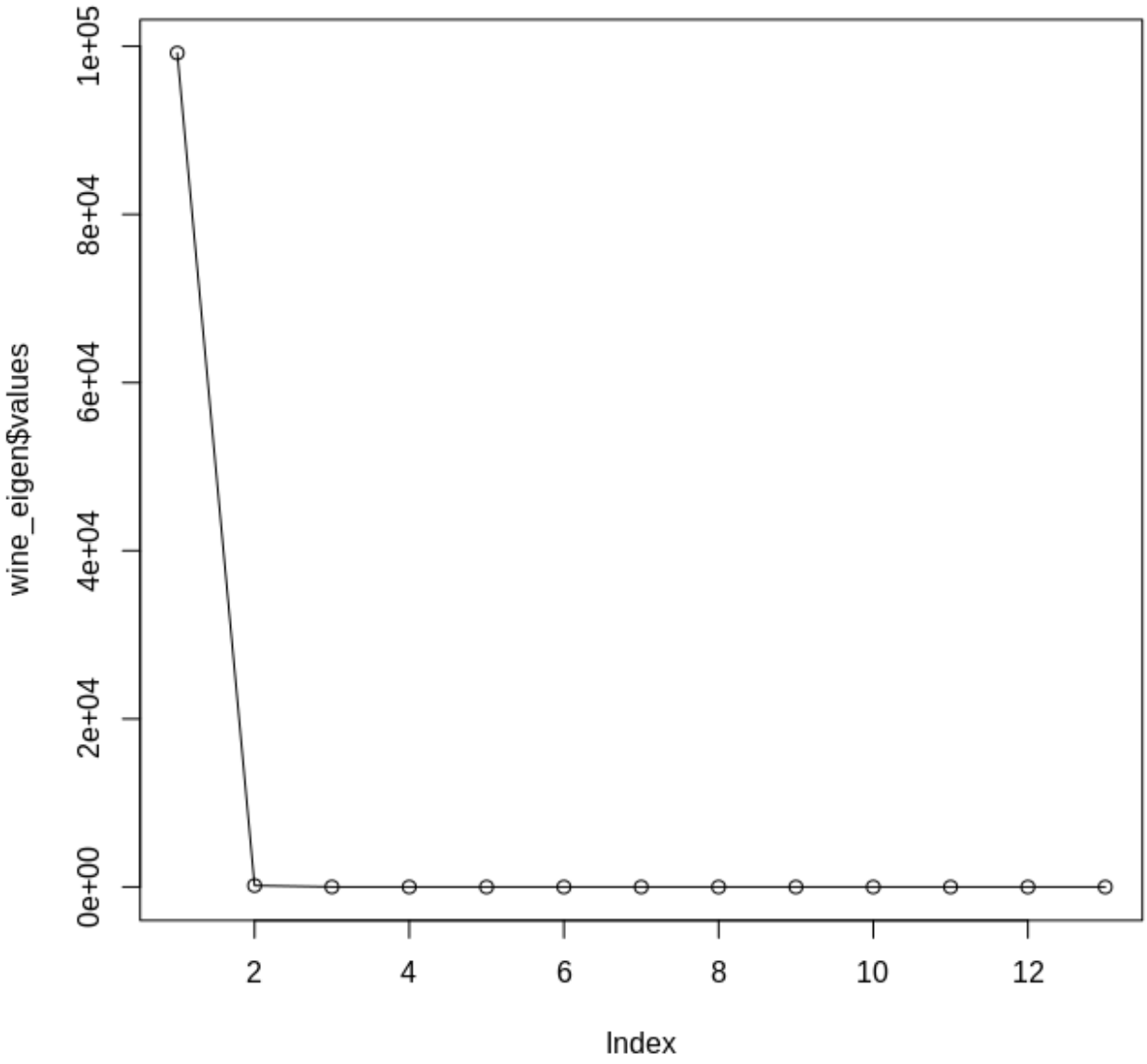


Figure 4.14: Scree plot showing the eigenvalues of a covariance matrix

```
               [,1]          [,2]          [,3]          [,4]          [,5]
 [1,]  2.964674e-02 -0.015003651 -0.0268303390  0.3105083975  0.9159233539
 [2,] -4.492354e-02  0.631037941 -0.7639084529  0.0914984764 -0.0419675589
 [3,]  2.939532e-02 -0.088515154  0.0130376306  0.0541523272 -0.1258473131
 [4,] -5.070965e-05 -0.000906056 -0.0009292491 -0.0055661084  0.0009662470
 [5,]  4.615941e-04 -0.001817055 -0.0006921813  0.0002758662 -0.0003315697
 [6,] -1.225450e-03  0.005008013 -0.0063668357 -0.0463339423  0.0208901648
 [7,]  8.630861e-02 -0.752355714 -0.6397040816 -0.0812146848 -0.0035253723
 [8,] -6.760251e-03  0.044759063 -0.0017451705  0.0325596672 -0.0278668129
 [9,]  4.670538e-02  0.002571526  0.0181608586 -0.0234532588  0.2379011114
[10,]  9.926372e-01  0.101541990  0.0199846897 -0.0305981550 -0.0272793198
[11,]  5.910888e-03 -0.011370960  0.0329866611  0.0589679625 -0.0184675768
[12,]  2.321636e-02 -0.096883535 -0.0409221274  0.4586710733 -0.0858251068
[13,] -2.565800e-02  0.076330017 -0.0528757677 -0.8167640810  0.2778687767
               [,6]          [,7]          [,8]          [,9]          [,10]
 [1,] -0.155644295 -0.158326612  0.111224831  0.025486508 -0.0227333012
 [2,] -0.038206577 -0.006260425 -0.057018732  0.017656920  0.0318490448
 [3,] -0.860395716 -0.075885802 -0.465215293  0.009900762 -0.0977504102
 [4,] -0.006737114  0.005445449 -0.006327226 -0.009680569 -0.0052830397
 [5,] -0.004868223  0.001270763 -0.003446759 -0.015765167  0.0154002725
 [6,]  0.011854493 -0.009388874 -0.006015218 -0.009814883 -0.0007255794
 [7,]  0.078477626 -0.061457820 -0.007848599  0.009174481 -0.0238738299
 [8,]  0.110677220 -0.041812471  0.013583151  0.116127372 -0.9839320643
 [9,]  0.358117027  0.407606104 -0.793550182 -0.126111107 -0.0106763230
[10,]  0.003158506 -0.016277550  0.044752857  0.001579001 -0.0006268944
[11,]  0.088474600 -0.046285998 -0.146719744  0.973113016  0.1268199497
[12,] -0.170432847  0.807481699  0.287200268  0.071157157 -0.0279765721
[13,] -0.224240115  0.378038741  0.178558244  0.130435940 -0.0545477337
               [,11]         [,12]         [,13]
 [1,] -3.579451e-03 -1.386274e-03  7.168469e-04
 [2,] -2.722109e-03 -1.079790e-04  1.021309e-05
 [3,]  1.260538e-02  8.190011e-03 -4.138848e-03
 [4,] -2.085826e-02 -9.995483e-01 -1.406496e-02
 [5,] -3.828993e-04 -1.392499e-02  9.996394e-01
 [6,]  9.982814e-01 -2.054663e-02  2.645536e-05
 [7,] -5.357718e-03  1.068461e-03 -8.705608e-04
 [8,]  7.023865e-04  2.505075e-03  1.772891e-02
 [9,] -1.221601e-02  6.746324e-03 -3.164339e-03
[10,]  7.268209e-05 -4.231324e-04 -4.147717e-05
[11,]  1.046531e-02 -1.078571e-02  1.320768e-02
[12,]  3.536485e-02 -6.651794e-05  3.293986e-04
[13,] -3.579491e-02  6.976225e-03  2.455507e-03
```

Figure 4.15: Principal components of the original data

```
     [,1] [,2] [,3]      [,4]      [,5]     [,6]
[1,]    1    3    6 0.2588933 0.9776119 1.917332
[2,]    1    3    8 0.2509881 0.9477612 1.018189
[3,]    1    3   10 0.2411067 0.9104478 1.693701
[4,]    1    3   14 0.2332016 0.8805970 1.761194
[5,]    1    3   15 0.2371542 0.8955224 1.791045
[6,]    1    3   18 0.2015810 0.7611940 1.254606
```

Figure 4.16: Three-item rules for market basket analysis

Lesson 5: Data Comparison Methods

Figure 5.1: Alamo image

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.8013922	0.8222856	0.8685930	0.8860392	0.8790931	0.8543929
[2,]	0.7894936	0.8038676	0.8190953	0.8486792	0.8580144	0.8787474
[3,]	0.8364491	0.8344346	0.8684280	0.8793931	0.8654631	0.8381220
[4,]	0.8285232	0.8608995	0.8577416	0.8791815	0.8901511	0.8678896
[5,]	0.7722914	0.7996539	0.7830312	0.7346689	0.7528258	0.8230921
[6,]	0.5818810	0.6800845	0.7049547	0.6165779	0.5959824	0.6451469
[7,]	0.3869945	0.5499423	0.5784381	0.5099081	0.5700767	0.5818399
[8,]	0.1772511	0.4792003	0.4952705	0.3892928	0.4537441	0.5686928
[9,]	0.1416325	0.2946302	0.3490931	0.2626835	0.3579248	0.4588876
[10,]	0.2487561	0.2265129	0.2497341	0.2668862	0.2668478	0.2782614

	[,7]	[,8]	[,9]	[,10]
[1,]	0.8131941	0.8397566	0.7535993	0.5631498
[2,]	0.8876371	0.8710443	0.6074711	0.3195716
[3,]	0.8250339	0.7549411	0.5641089	0.3053245
[4,]	0.8860902	0.7095673	0.3035896	0.2377038
[5,]	0.8191305	0.7778448	0.4690702	0.1963085
[6,]	0.6660920	0.6378310	0.5663493	0.2437251
[7,]	0.4905826	0.4436965	0.2017358	0.1559014
[8,]	0.5182057	0.4634460	0.1914863	0.1269013
[9,]	0.4728699	0.4246325	0.1942542	0.1460421
[10,]	0.2671916	0.2420319	0.1878105	0.1782178

Figure 5.2: Screenshot of the output matrix

```
  [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 [55]  0  0  0  0  0  0  0 -1 -1  0  0  0  1  1 -1 -1 -1  0  0  0  1  0 -1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
[109]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 -1  0  0  0  0  0  0 -1 -1 -1 -1  0  1  1  1  0 -1 -1  0  1  1  1  1  0  0  0  1  0  0
[163]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 -1  0  0  0
[217]  1  0  1  0 -1 -1 -1  1  1  1  1 -1 -1 -1 -1  0  1  1  0 -1  0  0  0  0  0  0  0  0  0  0  0  0  0 -1  0  0  0  0  0  0  0  0 -1  0  0  0  0  0 -1 -1 -1  0
[271]  0  0  0  0 -1 -1  0  0  0  0  0  0 -1  0  0  0  0  0  0  0  0  0 -1 -1 -1  0  1  1  1  0 -1 -1 -1  0  1  1  1  1 -1 -1  0  1  1  1  1 -1  0  0  0  1  0  0  0 -1
[325] -1  0  1  0  0  0 -1 -1 -1  0  0  0  0  0 -1  0 -1  0  1  0  1  0  0 -1 -1  0  1  1  1  0 -1 -1 -1  0  1  1  0  0 -1 -1 -1  0  0  0  0  0 -1 -1  1  0 -1  0  1  0
[379] -1 -1 -1 -1 -1  0  1  1  1  1 -1 -1 -1 -1  1  1  1  1  0  0  0  1  0  0  0 -1  0  1  1  1  1 -1 -1 -1 -1  0  1  1  1  0 -1 -1 -1 -1  0  1  1  1  0 -1  0  0  1  1
[433]  1  0 -1  0  0  0  1  1  1  0 -1  0 -1  0  1  1  1  0 -1 -1 -1  0  1  1  1  0 -1 -1 -1 -1 -1  1  1  1 -1 -1 -1 -1 -1  1  1 -1 -1  0  0  0 -1  0  0  0 -1  1  1
[487]  1  1  0 -1 -1 -1  0  0  1  1  1  0 -1 -1 -1 -1  0  1  1  1  0 -1 -1  1  1  1  0 -1 -1 -1  0  0  1  0  0  0 -1  0 -1 -1  1  1  1  1  1  0  0  0  1  1  1  1  1  0
[541] -1 -1  1  1  1  1  1  0 -1 -1  1  1  1  1  1 -1  0  0  0  1  0  0  0 -1  1  1  1  1 -1 -1 -1 -1 -1  0  1  1  1  0 -1 -1 -1 -1  0  1  1  1 -1 -1  0  1  1  1  1 -1
[595] -1 -1  0  1  1  1  0 -1 -1 -1 -1  0 -1  0  1  0 -1  0 -1 -1 -1  0  0  1  0  0 -1 -1 -1  0  1  1  1  0 -1 -1 -1  0  1  1  1  1  0  0  0  1  0  0  0  1  1  1  1
[649] -1 -1 -1 -1 -1  1  1  1  1 -1 -1 -1 -1 -1  1  1  1  1  0  0  0  1  1  1  0 -1 -1 -1 -1  1  1  1  0 -1 -1 -1 -1  0  1  1  1  0 -1 -1 -1 -1  0  0  1  1 -1 -1 -1 -1
[703] -1  0  1  1  1  0  0 -1 -1  0  1  1  1  1  0  0  0 -1  0  0  0  0  0  0  0  1 -1 -1  0  0  0  0  1  1  1  1  0  0  0  0  0  0  0  1  1  0  0  0  0  0  1  0  0  0  0  0
[757]  0  0  1  1  1  0  0  0  0  0  0  1  1  1  0  0  0  0  0  0  1  1  1  1  0  0  0  0 -1 -1  1  1  1  0  0  0  0 -1  0  0  0  0  0  0  0  0  0 -1
```

Figure: 5.3: Matrix of building_signature

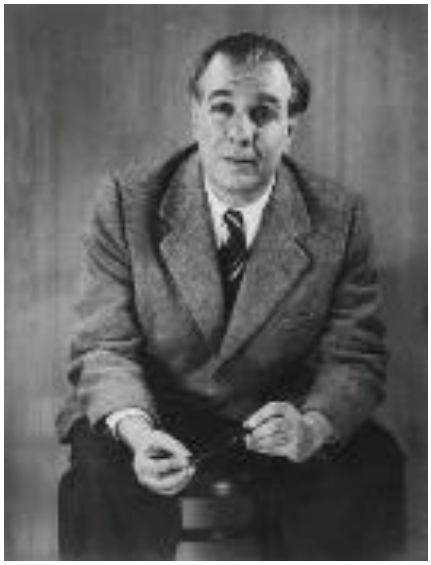
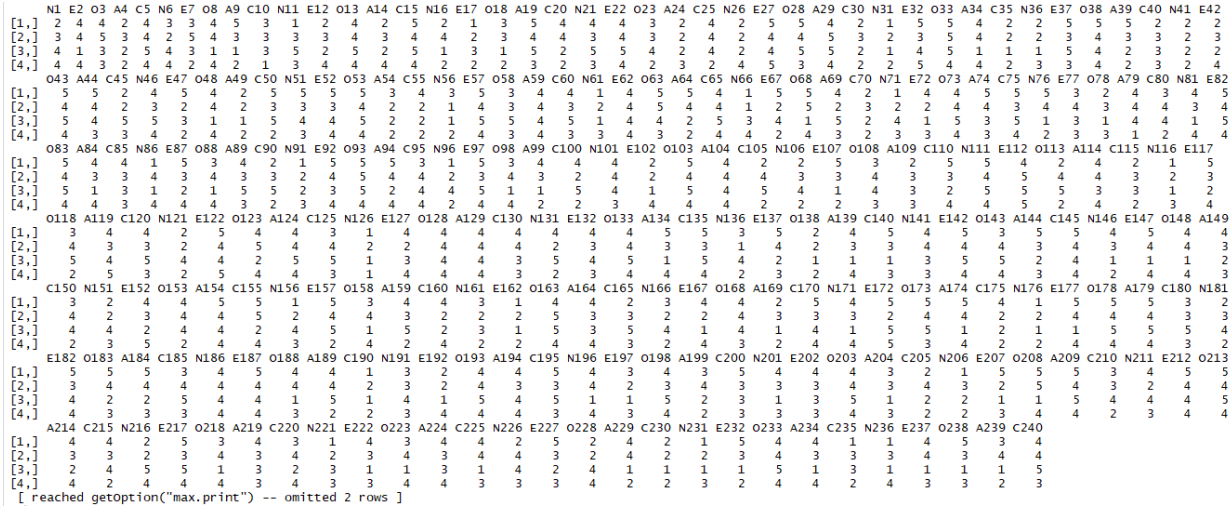


Figure 5.4: Jorge Luis Borges image

Figure 5.5: Alamo_marked image

```
  [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0
 [55]  0  0  0  0  0  0  0 -1 -1  0  0  0  1  1 -1 -1 -1  0  0  0  1  0 -1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0
1  0  0  0  0  0
[109]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 -1  0  0  0  0  0  0  0 -1 -1 -1 -1  0  1  1  1  0 -1 -1  0  1  1  1  1  0
0  0  0  1  0  0
[163]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 -1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0 -1  0  0  0
[217]  1  0  0  1  0 -1 -1 -1  1  1  1  1 -1 -1 -1 -1  0  1  1  0 -1  0  0  0  0  0  0  0  0  0  0  0  0  0 -1  0  0  0  0  0  0  0  0  0 -1  0  0  0
0  0 -1 -1 -1  0
[271]  0  0  0  0  0 -1 -1  0  0  0  0  0  0  0 -1  0  0  0  0  0  0  0  0  0  0  0 -1 -1 -1  0  1  1  1  0 -1 -1 -1  0  1  1  1  1 -1 -1  0  1  1  1  1 -1  0  0
0  1  0  0  0 -1
[325] -1  0  1  0  0  0 -1 -1 -1  0  0  0  0  0 -1  0 -1  0  1  0  1  0  0 -1 -1  0  1  1  1  0 -1 -1 -1  0  1  1  0  0 -1 -1 -1  0  0  0  0  0 -1 -1 -
1  0 -1  0  1  0
[379] -1 -1 -1 -1 -1  0  1  1  1  1 -1 -1 -1 -1  1  1  1  1  0  0  0  1  0  0  0 -1  0  1  1  1  1 -1 -1 -1 -1  0  1  1  1  0 -1 -1 -1 -1  0  1  1  1
0 -1  0  0  1  1
[433]  1  0 -1  0  0  0  1  1  1  0 -1 -1 -1  0  1  1  1  0 -1 -1 -1  0  1  1  1  0 -1 -1 -1 -1 -1  1  1  1 -1 -1 -1 -1 -1 -1  1  1 -1 -1  0  0  0 -1
0  0  0 -1  1  1
[487]  1  1  0 -1 -1 -1 -1  0  1  1  1  0 -1 -1 -1 -1  0  1  1  1  0 -1 -1  1  1  1  0 -1 -1 -1  0  0  1  0  0  0 -1  0 -1 -1  1  1  1  1  1  0  0  0
1  1  1  1  1  0
[541] -1 -1  1  1  1  1  1  0 -1 -1  1  1  1  1  1 -1  0  0  0  1  0  0  0 -1  1  1  1  1 -1 -1 -1 -1 -1  0  1  1  1  0 -1 -1 -1 -1  0  1  1  1 -1 -1
0  1  1  1  1 -1
[595] -1 -1  0  1  1  1  0 -1 -1 -1 -1 -1 -1  0  1  1  0  0 -1  0 -1  0  0  0  0  0 -1 -1 -1  0  1  1  1  0 -1 -1 -1  0  1  1  1  1  0  0  0  1  0  0
0  1  1  1  1  1
[649] -1 -1 -1 -1 -1  1  1  1  1 -1 -1 -1 -1 -1  1  1  1  1  0  0  0  1  1  1  0 -1 -1 -1 -1  1  1  1  0 -1 -1 -1 -1  0  0  1  1  0 -1 -1 -1 -1  0  0
1  1 -1 -1 -1 -1
[703] -1  0  1  1  1  0  0 -1 -1  0  1  1  1  1  0  0  0 -1  0  0  0  0  0  0  1 -1 -1  0  0  0  0  0  1  1  1  0  0  0  0  0  0  1  1  0  0  0  0  0
1  0  0  0  0  0
[757]  0  0  1  1  1  0  0  0  0  0  1  1  1  0  0  0  0  0  1  1  1  1  0  0  0  0 -1 -1  1  1  1  0  0  0  0 -1  0  0  0  0  0  0  0  0 -1
```

Figure 5.6: Expected signature of watermarked image



```
     N1 E2 O3 A4 C5 N6 E7 O8 A9 C10 N11 E12 O13 A14 C15 N16 E17 O18 A19 C20 N21 E22 O23 A24 C25 N26 E27 O28 A29 C30 N31 E32 O33 A34 C35 N36 E37 O38 A39 C40 N41 E42
[1,]  2  4  4  2  4  3  3  4  5   3   1   2   4   2   5   2   1   3   5   4   4   4   3   2   4   2   5   5   4   2   1   5   5   4   2   2   5   5   5   2   2   2
[2,]  3  4  5  3  4  2  5  4  3   3   3   3   4   3   4   4   2   3   4   4   3   4   3   2   4   2   4   4   5   3   2   3   5   4   2   2   3   4   3   3   2   3
[3,]  4  1  3  2  5  4  3  1  1   3   5   2   5   2   5   1   3   1   5   2   5   5   4   2   4   2   4   5   5   2   1   4   5   1   1   1   5   4   2   3   2   2
[4,]  4  4  3  2  4  4  2  4  2   1   3   4   4   4   4   2   2   2   3   2   3   4   2   3   2   4   5   3   4   2   2   5   4   4   2   3   3   4   2   3   2   4
     O43 A44 C45 N46 E47 O48 A49 C50 N51 E52 O53 A54 C55 N56 E57 O58 A59 C60 N61 E62 O63 A64 C65 N66 E67 O68 A69 C70 N71 E72 O73 A74 C75 N76 E77 O78 A79 C80 N81 E82
[1,]   5   5   2   4   5   4   2   5   5   5   5   3   4   3   5   3   4   4   1   4   5   5   4   1   5   5   4   2   1   4   4   5   5   5   3   2   4   3   4   5
[2,]   4   4   2   3   2   4   2   3   3   3   4   2   2   1   4   3   4   3   2   4   5   4   4   1   2   5   2   3   2   2   4   4   3   4   4   3   4   4   3   4
[3,]   5   4   5   5   3   1   1   5   4   4   5   2   2   1   5   5   4   5   1   4   4   2   5   3   4   1   5   2   4   1   5   3   5   1   3   1   4   4   1   5
[4,]   4   3   3   4   2   4   2   2   3   4   4   2   2   2   4   3   4   3   3   4   3   2   4   4   2   4   3   2   3   3   4   3   4   2   3   3   1   2   4   4
     O83 A84 C85 N86 E87 O88 A89 C90 N91 E92 O93 A94 C95 N96 E97 O98 A99 C100 N101 E102 O103 A104 C105 N106 E107 O108 A109 C110 N111 E112 O113 A114 C115 N116 E117
[1,]   5   4   4   1   5   3   4   2   1   5   5   5   3   1   5   3   4   4   4   2   5   4   2   2   5   3   2   5   5   4   2   4   2   1   5
[2,]   4   3   3   4   3   4   3   3   2   4   5   4   4   2   3   4   3   2   4   2   4   4   4   3   3   4   3   3   4   5   4   4   3   2   3
[3,]   5   1   3   1   2   1   5   5   2   3   5   2   4   4   5   1   1   5   4   1   5   4   5   4   1   4   3   2   5   5   5   3   3   1   2
[4,]   4   4   3   4   4   4   3   2   3   4   4   4   4   2   4   4   2   2   3   4   4   4   2   2   2   3   3   4   4   5   2   4   2   3   4
     O118 A119 C120 N121 E122 O123 A124 C125 N126 E127 O128 A129 C130 N131 E132 O133 A134 C135 N136 E137 O138 A139 C140 N141 E142 O143 A144 C145 N146 E147 O148 A149
[1,]    3    4    4    2    5    4    4    3    1    4    4    4    4    4    4    4    5    5    3    5    2    4    5    4    5    3    5    5    4    5    4    4
[2,]    4    3    3    2    4    5    4    4    2    2    4    4    4    2    3    4    3    3    1    4    2    3    3    4    4    4    3    4    3    4    4    3
[3,]    5    4    5    4    4    2    5    5    1    3    4    4    3    5    4    5    1    5    4    2    1    1    1    3    5    5    2    4    1    1    1    2
[4,]    2    5    3    2    5    4    4    3    1    4    4    4    3    2    3    4    4    4    2    3    2    4    3    3    4    4    3    4    2    4    4    3
     C150 N151 E152 O153 A154 C155 N156 E157 O158 A159 C160 N161 E162 O163 A164 C165 N166 E167 O168 A169 C170 N171 E172 O173 A174 C175 N176 E177 O178 A179 C180 N181
[1,]    3    2    4    4    5    5    1    5    3    4    4    3    1    4    4    2    3    4    4    2    5    4    5    5    5    4    1    5    5    5    3    2
[2,]    4    2    3    4    4    5    2    4    4    3    2    2    2    5    3    3    2    2    4    3    3    3    2    4    4    2    2    4    4    4    3    3
[3,]    4    4    2    4    4    2    4    5    1    5    2    3    1    5    3    5    4    1    4    1    4    1    5    5    1    2    1    1    5    5    5    4
[4,]    2    3    5    2    4    4    3    2    4    2    4    2    2    4    4    3    2    4    3    2    4    4    5    3    4    2    2    4    4    4    3    2
     E182 O183 A184 C185 N186 E187 O188 A189 C190 N191 E192 O193 A194 C195 N196 E197 O198 A199 C200 N201 E202 O203 A204 C205 N206 E207 O208 A209 C210 N211 E212 O213
[1,]    5    5    5    3    4    5    4    4    1    3    2    4    4    5    4    3    4    3    5    4    4    4    3    2    1    5    5    5    3    4    5    5
[2,]    3    4    4    4    4    4    4    4    2    3    2    4    3    3    4    2    3    4    3    3    3    4    3    4    3    2    5    4    3    2    4    4
[3,]    4    2    2    5    4    4    1    5    1    4    1    5    4    5    1    1    5    2    3    1    3    5    1    2    2    1    1    5    4    4    4    5
[4,]    4    3    3    3    4    4    3    2    2    3    4    4    4    3    4    3    4    2    3    3    3    4    3    2    2    3    4    4    2    3    4    4
     A214 C215 N216 E217 O218 A219 C220 N221 E222 O223 A224 C225 N226 E227 O228 A229 C230 N231 E232 O233 A234 C235 N236 E237 O238 A239 C240
[1,]    4    4    2    5    3    4    3    1    4    3    4    4    2    5    2    4    2    1    5    4    4    1    1    4    5    3    4
[2,]    3    3    2    3    4    3    4    2    3    4    3    4    4    3    2    4    2    2    3    4    3    3    3    4    3    4    4
[3,]    2    4    5    5    1    3    2    3    1    1    3    1    4    2    4    1    1    1    1    5    1    3    1    1    1    1    5
[4,]    4    2    4    4    4    3    4    3    3    4    4    3    3    3    4    2    2    3    2    4    4    2    4    3    3    2    3
 [ reached getOption("max.print") -- omitted 2 rows ]
```

Figure 5.7: Top section of the data

```
Factor Analysis using method =  pa
Call: fa(r = big_cor, nfactors = 5, rotate = "oblimin", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
       PA1   PA4   PA2   PA3   PA5    h2   u2 com
N1    0.54 -0.03  0.07  0.04  0.09 0.304 0.70 1.1
E2   -0.15 -0.01  0.20  0.31 -0.04 0.173 0.83 2.3
O3    0.11 -0.10 -0.10  0.06  0.47 0.259 0.74 1.3
A4   -0.31  0.02  0.38  0.21 -0.05 0.312 0.69 2.6
C5   -0.15  0.39  0.09 -0.22  0.19 0.269 0.73 2.6
N6    0.45  0.00 -0.27 -0.08  0.01 0.296 0.70 1.7
E7   -0.24 -0.06  0.00  0.38 -0.12 0.225 0.78 2.0
O8    0.08  0.08  0.09  0.11  0.43 0.232 0.77 1.4
A9    0.04  0.10  0.45  0.16 -0.06 0.247 0.75 1.4
C10   0.26  0.31 -0.06 -0.18 -0.13 0.202 0.80 3.0
N11   0.62  0.04  0.00 -0.11  0.18 0.430 0.57 1.2
E12  -0.38  0.15 -0.44  0.21 -0.02 0.444 0.56 2.7
O13   0.12  0.02 -0.05  0.21  0.34 0.193 0.81 2.0
A14  -0.06  0.15  0.56  0.05  0.06 0.367 0.63 1.2
C15   0.03  0.33  0.22  0.05  0.05 0.178 0.82 1.9
N16   0.47 -0.01  0.00 -0.21 -0.03 0.307 0.69 1.4
E17   0.42 -0.01 -0.03 -0.10 -0.01 0.203 0.80 1.1
O18  -0.14 -0.23 -0.03 -0.01  0.16 0.087 0.91 2.5
A19  -0.05  0.01  0.39  0.26  0.05 0.237 0.76 1.8
C20  -0.23  0.37  0.00  0.19 -0.01 0.279 0.72 2.2
N21   0.02 -0.17 -0.15  0.26  0.08 0.133 0.87 2.6
E22   0.12 -0.16 -0.22  0.22  0.18 0.178 0.82 4.3
O23  -0.17 -0.13 -0.12 -0.23  0.52 0.323 0.68 1.9
A24   0.07  0.05  0.50 -0.01 -0.12 0.282 0.72 1.2
C25  -0.20  0.47 -0.06  0.06  0.10 0.326 0.67 1.5
N26   0.63 -0.03 -0.05  0.00 -0.04 0.418 0.58 1.0
E27  -0.06 -0.01  0.01  0.45  0.06 0.228 0.77 1.1
O28  -0.11  0.03  0.13  0.04  0.29 0.124 0.88 1.8
A29   0.12  0.07  0.16 -0.06  0.11 0.057 0.94 3.6
C30  -0.22  0.23  0.21 -0.03 -0.14 0.187 0.81 3.7
N31   0.60  0.04  0.00  0.00 -0.11 0.367 0.63 1.1
E32  -0.12 -0.06  0.10  0.42  0.02 0.224 0.78 1.3
O33   0.05 -0.28  0.06  0.08  0.36 0.218 0.78 2.1
A34  -0.08  0.03  0.24  0.33  0.08 0.206 0.79 2.1
C35   0.01 -0.13 -0.12 -0.09 -0.15 0.075 0.93 3.6
```

Figure 5.8: Section of the output

	PA1	PA4	PA2	PA3	PA5
N1	2.881238e-02	8.670579e-04	-3.040204e-03	7.314594e-03	0.0145400299
E2	-4.617340e-03	-1.835249e-03	1.498291e-02	2.122914e-02	-0.0116023314
O3	2.426148e-03	-6.258903e-03	-1.242467e-02	1.407778e-03	0.0497856169
A4	-8.959417e-03	6.497288e-04	4.158095e-02	1.868088e-02	-0.0149398935
C5	-7.711196e-03	2.157890e-02	8.675828e-03	-2.041749e-02	0.0277668768
N6	2.190748e-02	5.824713e-03	-2.491382e-02	-2.649121e-03	0.0047750369
E7	-1.556929e-02	-1.322104e-02	-4.031547e-03	3.387525e-02	-0.0202769952
O8	1.753012e-02	1.921573e-02	3.786774e-03	9.157273e-03	0.0376477185
A9	2.978894e-03	3.322950e-03	3.830723e-02	1.717055e-02	-0.0053211261
C10	1.254014e-02	2.655267e-02	-1.503292e-02	-8.659995e-03	-0.0133281548
N11	3.662707e-02	1.035695e-02	-1.240881e-03	-1.425144e-02	0.0312693793
E12	-2.599546e-02	2.540984e-02	-6.392001e-02	3.070693e-02	-0.0057103637
O13	1.414654e-02	8.537017e-03	-6.818238e-03	1.599520e-02	0.0317023984
A14	-1.672476e-02	9.509720e-03	6.121387e-02	-4.965671e-04	0.0132508173
C15	4.935832e-04	1.934411e-02	2.081504e-02	3.863660e-03	0.0096609759
N16	1.879551e-02	1.311568e-02	6.006537e-03	-1.621107e-02	0.0043428587
E17	1.630781e-02	4.126107e-03	-7.359349e-03	-1.977539e-03	0.0014600032
O18	-9.600614e-03	-1.067392e-02	-4.602948e-03	-1.305665e-02	0.0123377536
A19	-3.769541e-03	-1.995543e-03	3.311415e-02	1.799977e-02	0.0134641454
C20	4.235881e-04	2.573445e-02	-5.186867e-03	1.800068e-02	-0.0121188467
N21	4.341721e-03	-8.527928e-03	-1.035717e-02	2.709653e-02	-0.0003692319
E22	-7.689687e-04	-1.582864e-02	-1.073823e-02	1.291507e-02	0.0176404242
O23	-1.753918e-02	-1.571162e-02	-1.023660e-02	-4.684526e-02	0.0803286275
A24	3.211651e-04	-7.093895e-03	5.069256e-02	-7.695827e-03	-0.0154206718
C25	-6.921068e-03	3.859124e-02	-1.877422e-03	6.328299e-03	0.0069902781
N26	5.554152e-02	-3.868361e-03	-1.021338e-02	1.222294e-02	-0.0130603015
E27	2.385429e-03	1.852003e-04	1.473539e-03	4.243108e-02	0.0098817470
O28	-3.739258e-03	-2.084281e-03	7.751884e-03	-2.879907e-04	0.0292258995
A29	2.487800e-03	-3.350833e-03	9.095477e-03	-1.020894e-02	0.0193551576
C30	-1.431393e-02	5.489632e-03	1.537105e-02	-8.009814e-03	-0.0142113588
N31	4.830967e-02	1.467822e-02	1.234340e-03	1.500907e-02	-0.0152494634
E32	-4.251237e-03	-6.093333e-03	6.060609e-03	2.044216e-02	-0.0055899979
O33	-1.386843e-03	-1.678646e-02	2.053727e-03	2.469049e-03	0.0435236032

Figure 5.9: Section of the output

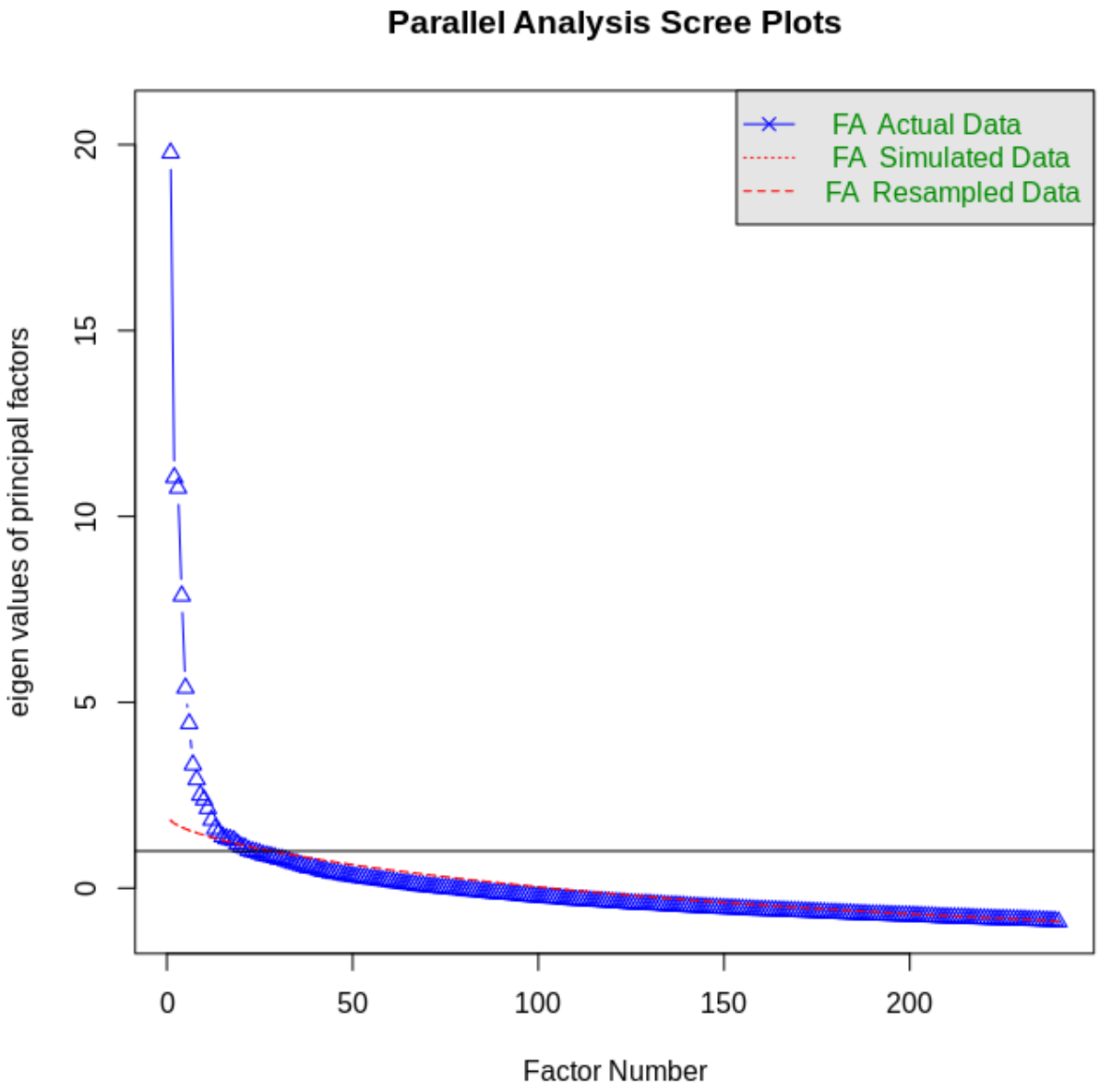


Figure 5.10: Parallel analysis scree plot

```
Factor Analysis using method =  minres
Call: fa(r = ratings_cor, nfactors = 1)
Standardized loadings (pattern matrix) based upon correlation matrix
              MR1      h2   u2 com
Category.1  -0.02 0.00027 1.00   1
Category.2   0.16 0.02454 0.98   1
Category.3   0.68 0.46025 0.54   1
Category.4   0.30 0.08942 0.91   1
Category.5   0.43 0.18654 0.81   1
Category.6   0.61 0.37424 0.63   1
Category.7   0.88 0.77276 0.23   1
Category.8  -0.13 0.01718 0.98   1
Category.9   0.05 0.00257 1.00   1
Category.10 -0.74 0.55225 0.45   1

                MR1
SS loadings    2.48
Proportion Var 0.25
```

Figure 5.11: Expected outcome of factor analysis

Lesson 6: Anomaly Detection

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Figure 6.1: Top six rows of the mtcars dataset

R: Motor Trend Car Road Tests ▾ Find in Topic

`mtcars`

Format

A data frame with 32 observations on 11 (numeric) variables.

[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp Displacement (cu.in.)
[, 4] hp Gross horsepower
[, 5] drat Rear axle ratio
[, 6] wt Weight (1000 lbs)
[, 7] qsec 1/4 mile time
[, 8] vs Engine (0 = V-shaped, 1 = straight)
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

Source

Henderson and Velleman (1981), Building multiple regression models interactively. *Biometrics*, **37**, 391–411.

Examples

```
require(graphics)
pairs(mtcars, main = "mtcars data", gap = 1/4)
coplot(mpg ~ disp | as.factor(cyl), data = mtcars,
       panel = panel.smooth, rows = 1)
## possibly more meaningful, e.g., for summary() or bivariate plots:
```

Figure 6.2: Section of output

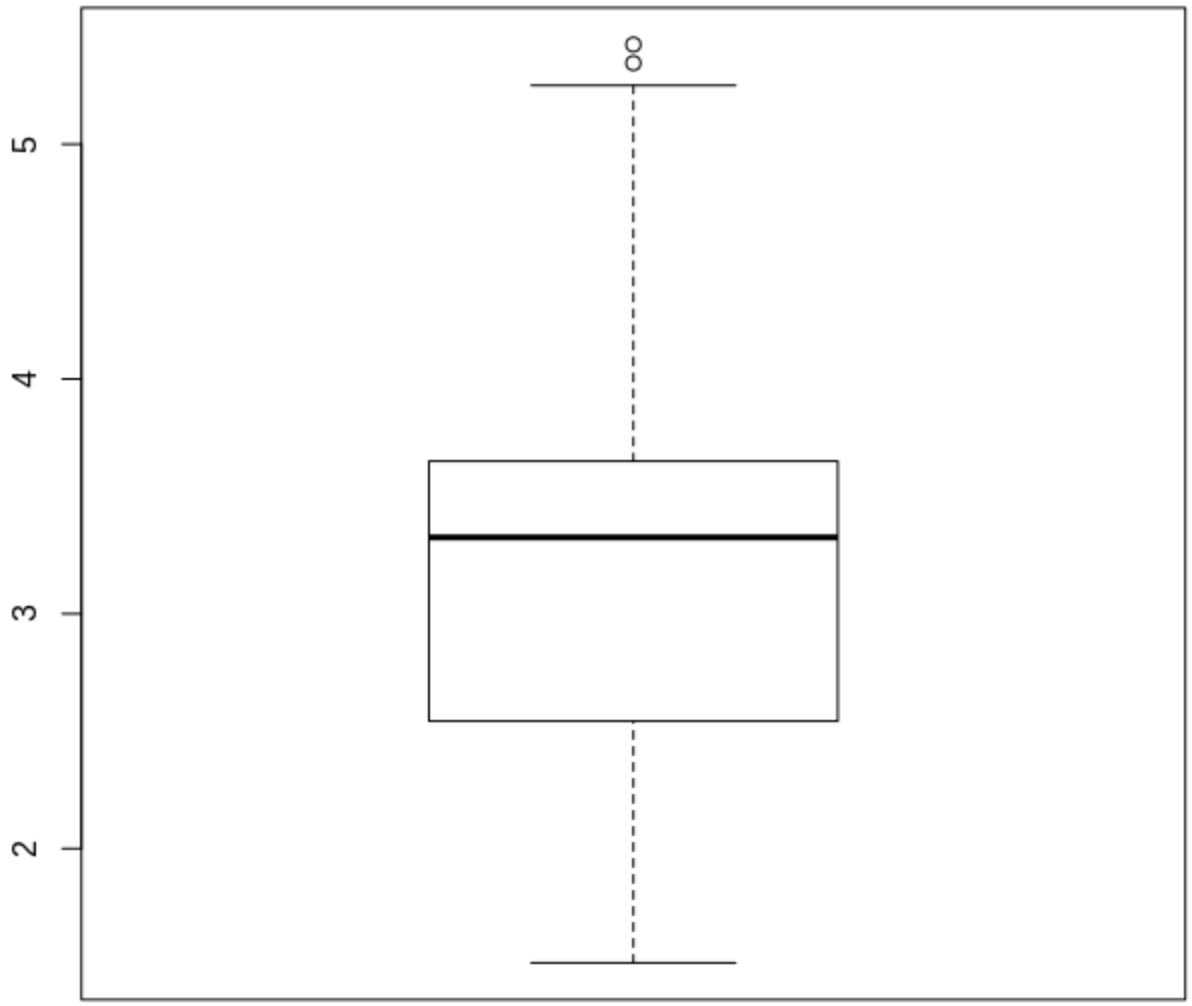


Figure 6.3: Boxplot representing weight of cars

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Cadillac Fleetwood	10.4	8	472	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440	230	3.23	5.345	17.42	0	0	3	4

Figure 6.4: Cars with weights greater than 5,000 pounds

rivers {datasets} R Documentation

Lengths of Major North American Rivers

Description

This data set gives the lengths (in miles) of 141 “major” rivers in North America, as compiled by the US Geological Survey.

Usage

```
rivers
```

Format

A vector containing 141 observations.

Source

World Almanac and Book of Facts, 1975, page 406.

References

McNeil, D. R. (1977) *Interactive Data Analysis*. New York: Wiley.

[Package *datasets* version 3.5.1 Index]

Figure 6.5: Information of the rivers dataset

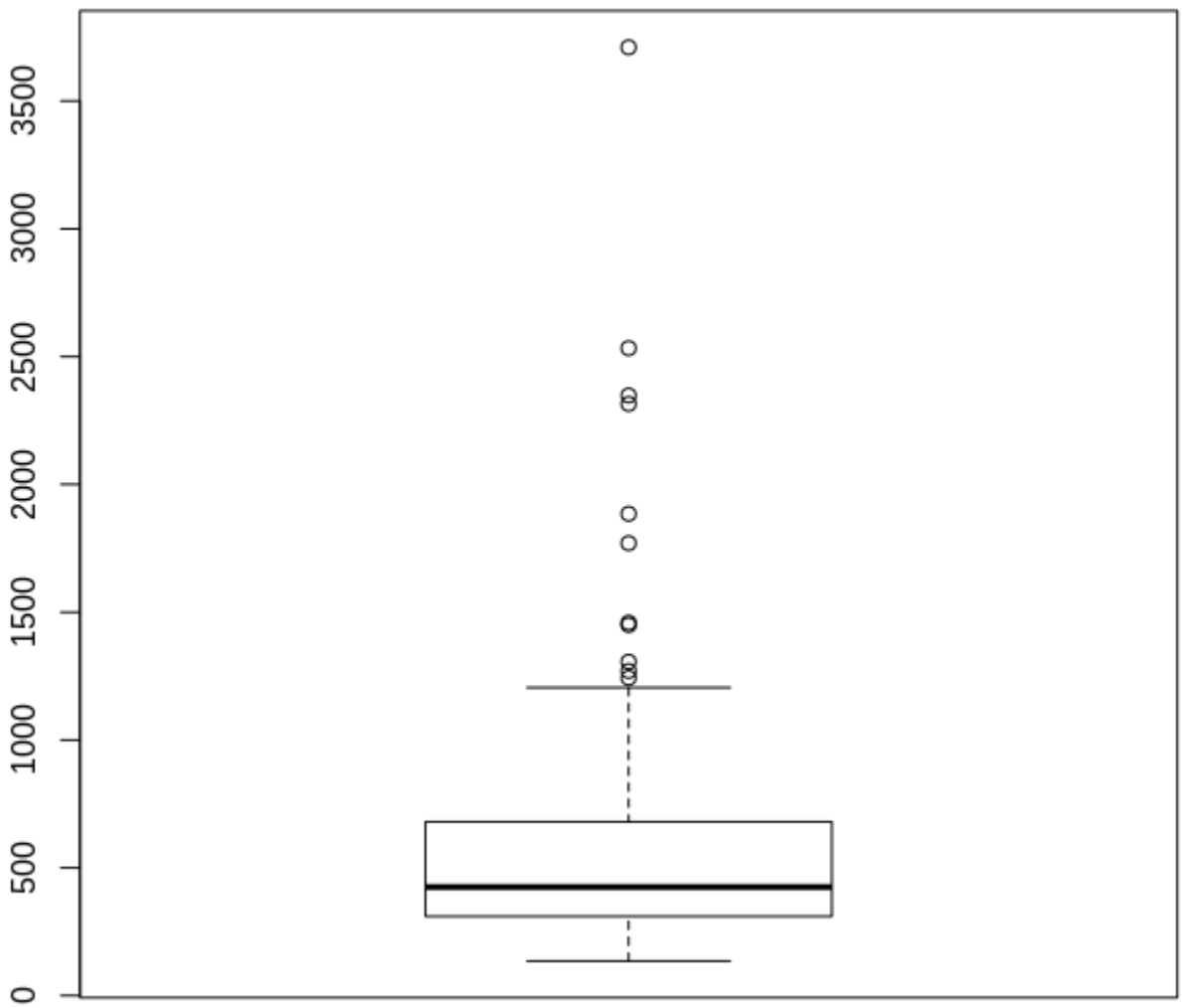


Figure 6.6: Boxplot of the rivers dataset

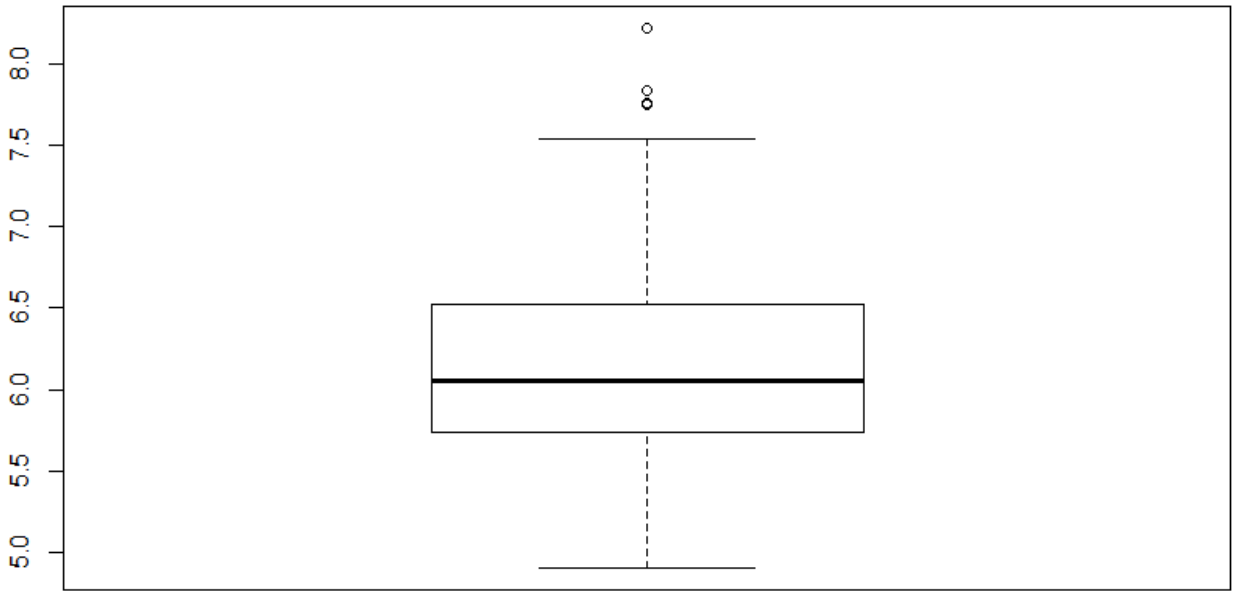


Figure 6.7: Boxplot of transformed dataset

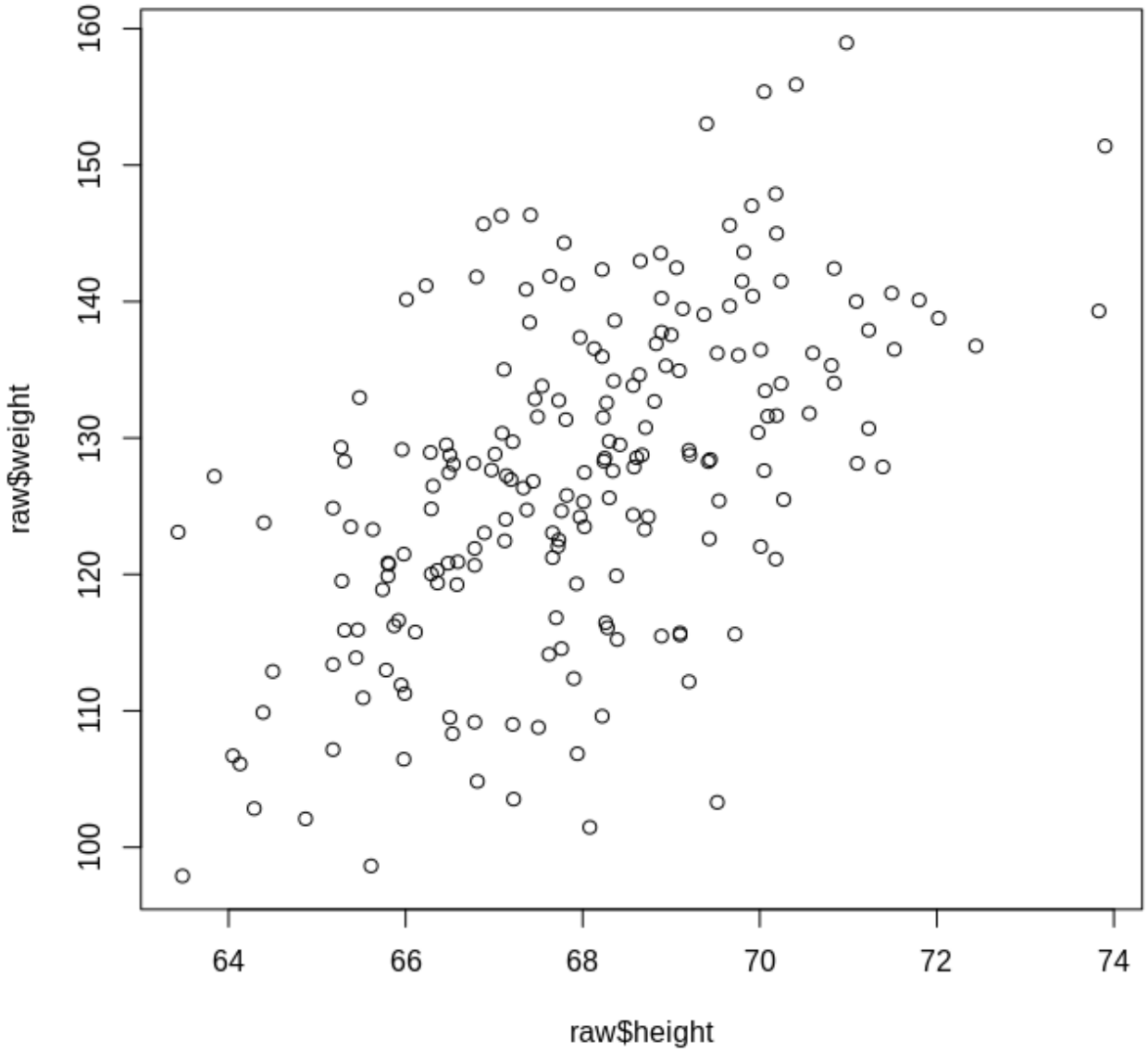


Figure 6.8: Plot of height and weight

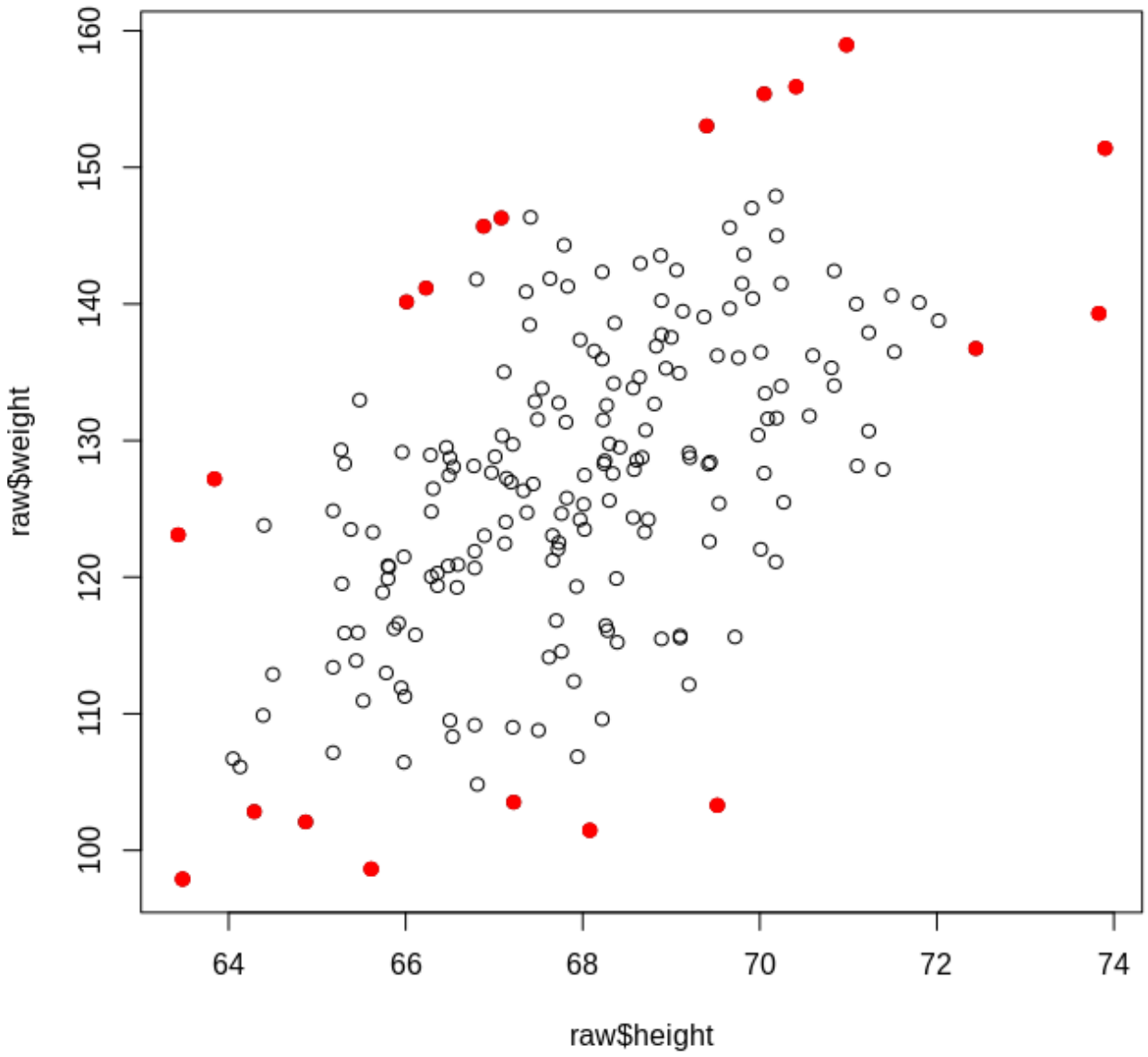


Figure 6.9: Observations with high Mahalanobis distances

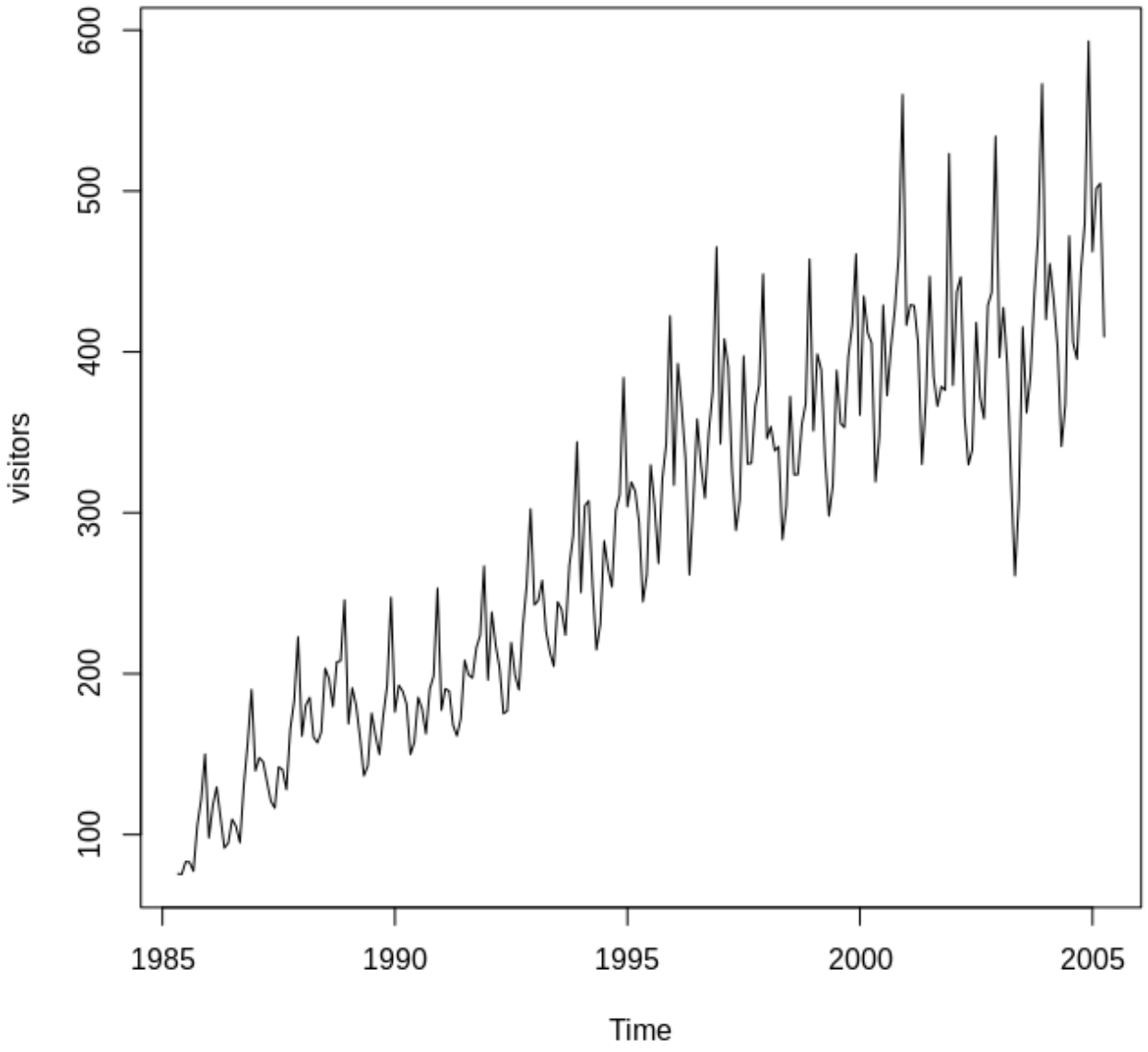


Figure 6.10: Plot of monthly visitors to Australia between 1985 and 2005

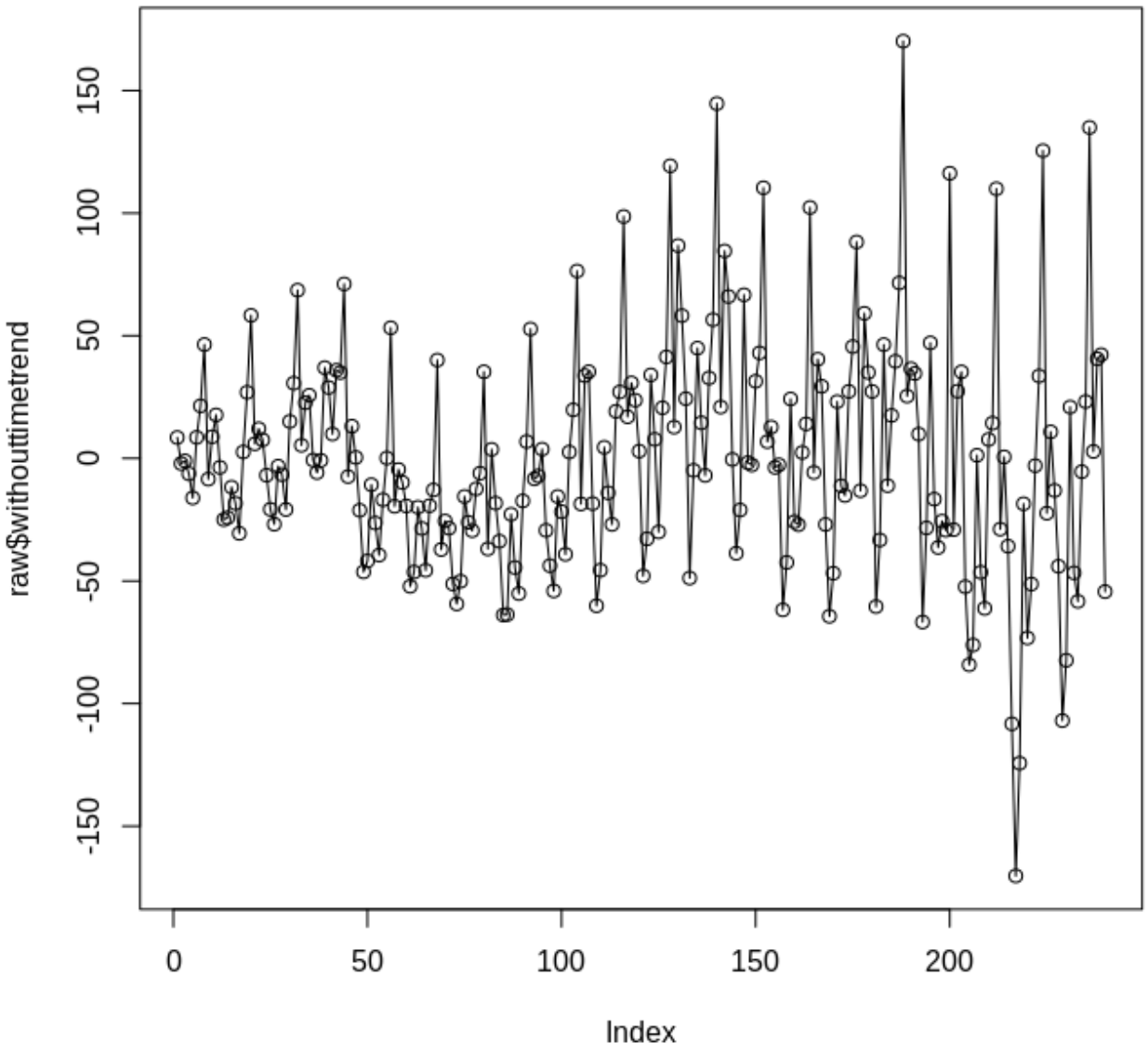


Figure 6.11: Plot of the de-trended data

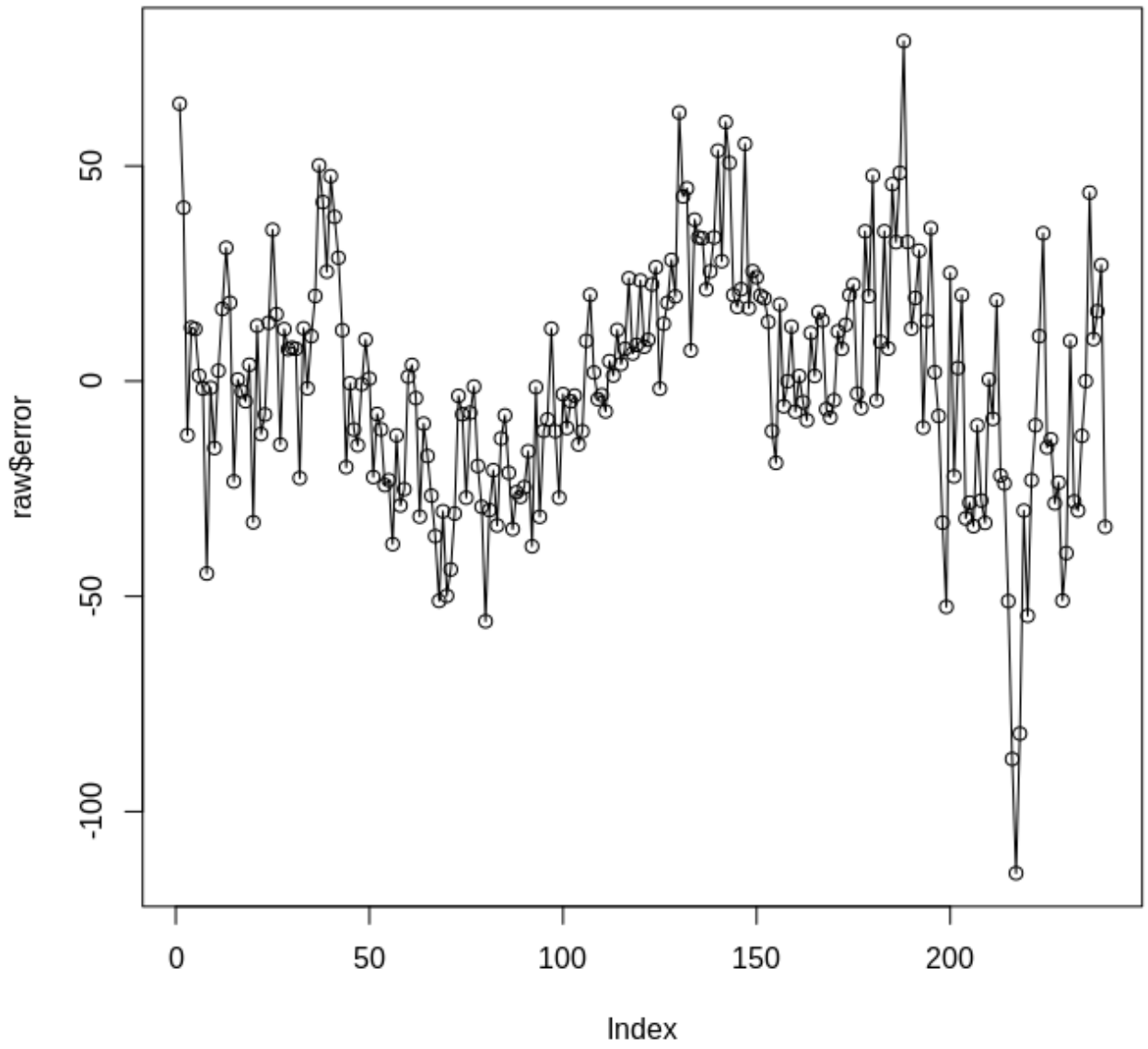


Figure 6.12: Plot of the de-trended data

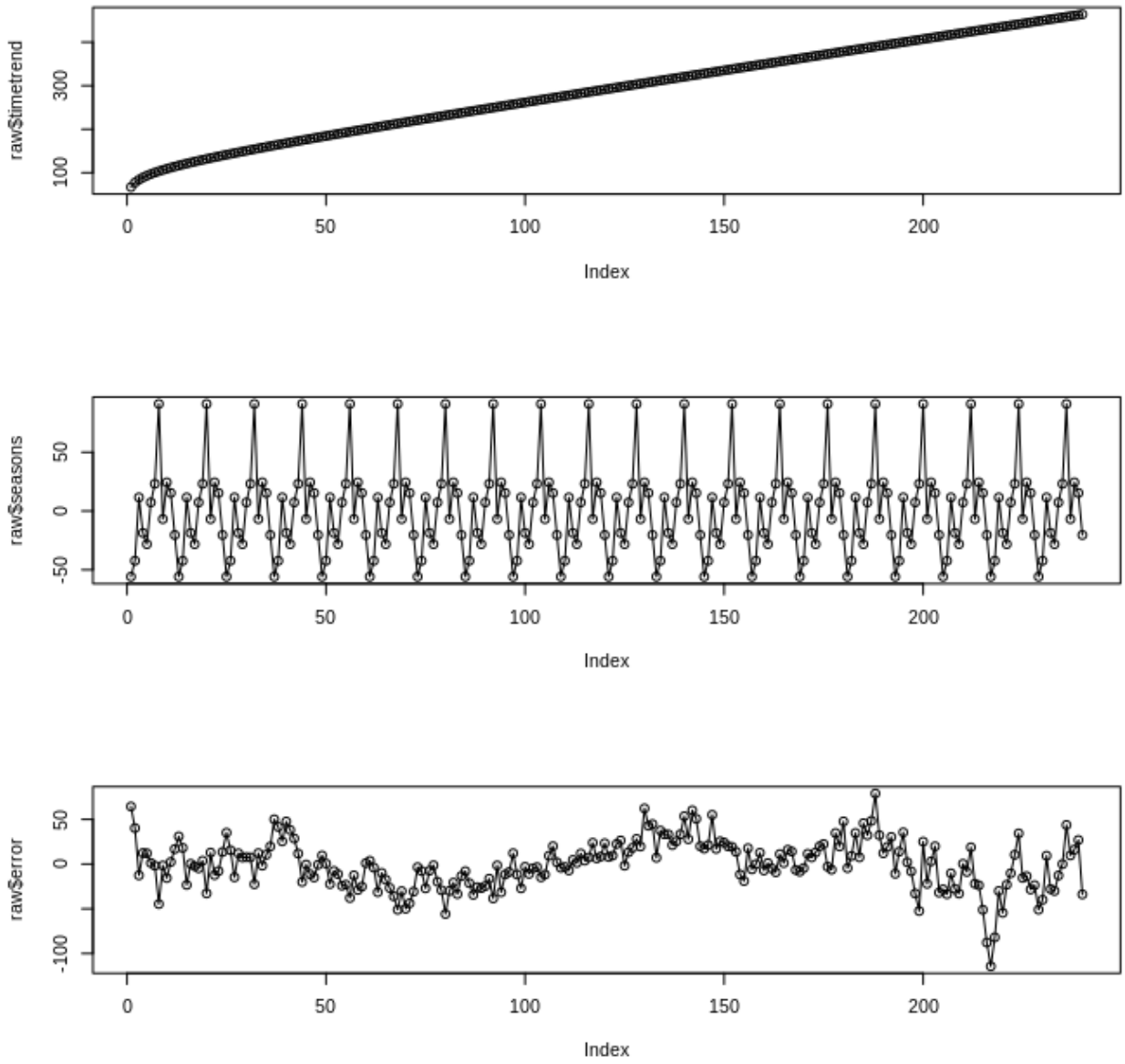


Figure 6.13: Plot of elements of seasonality modelling

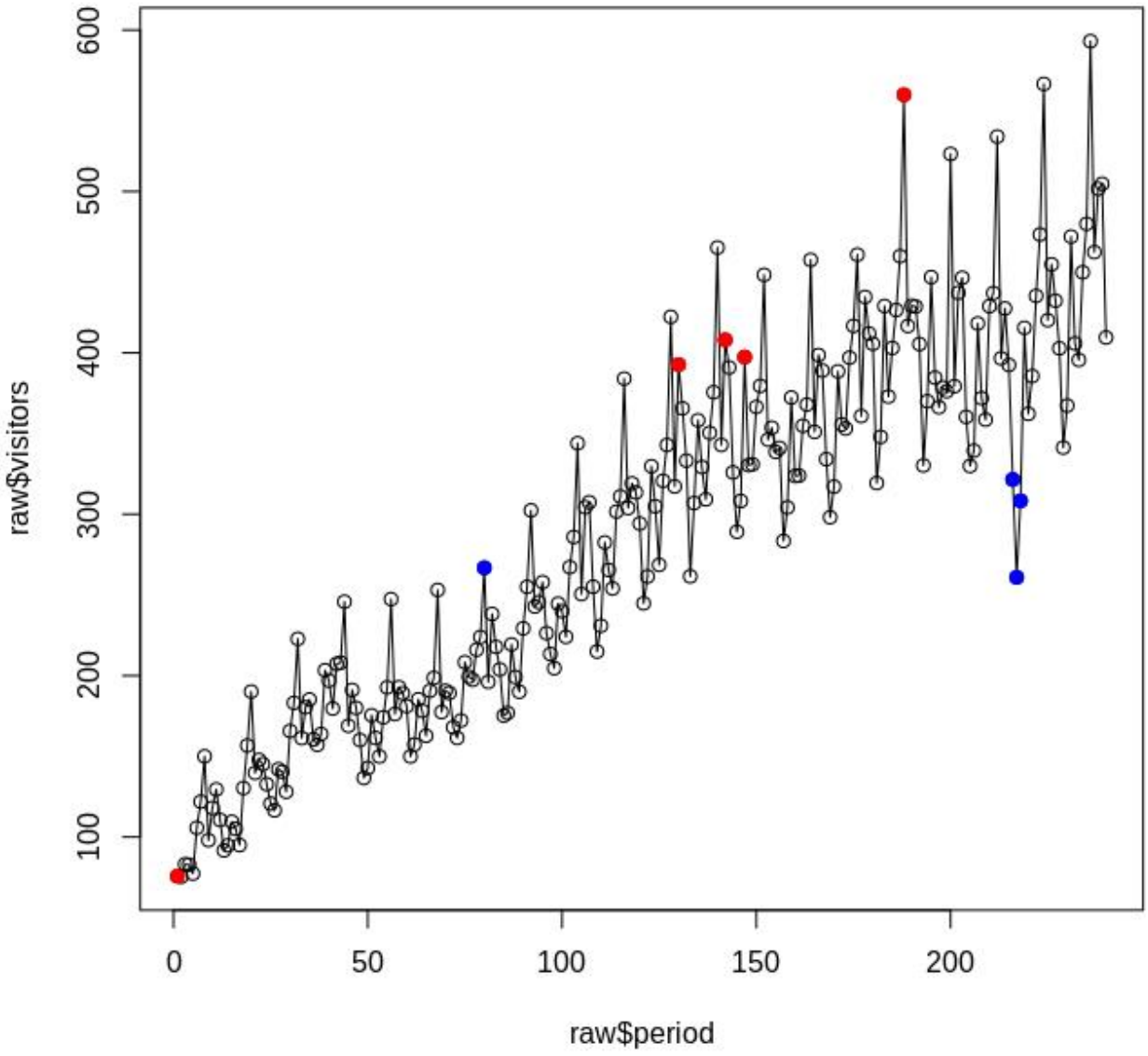


Figure 6.14: Plot of data classified as anomalies

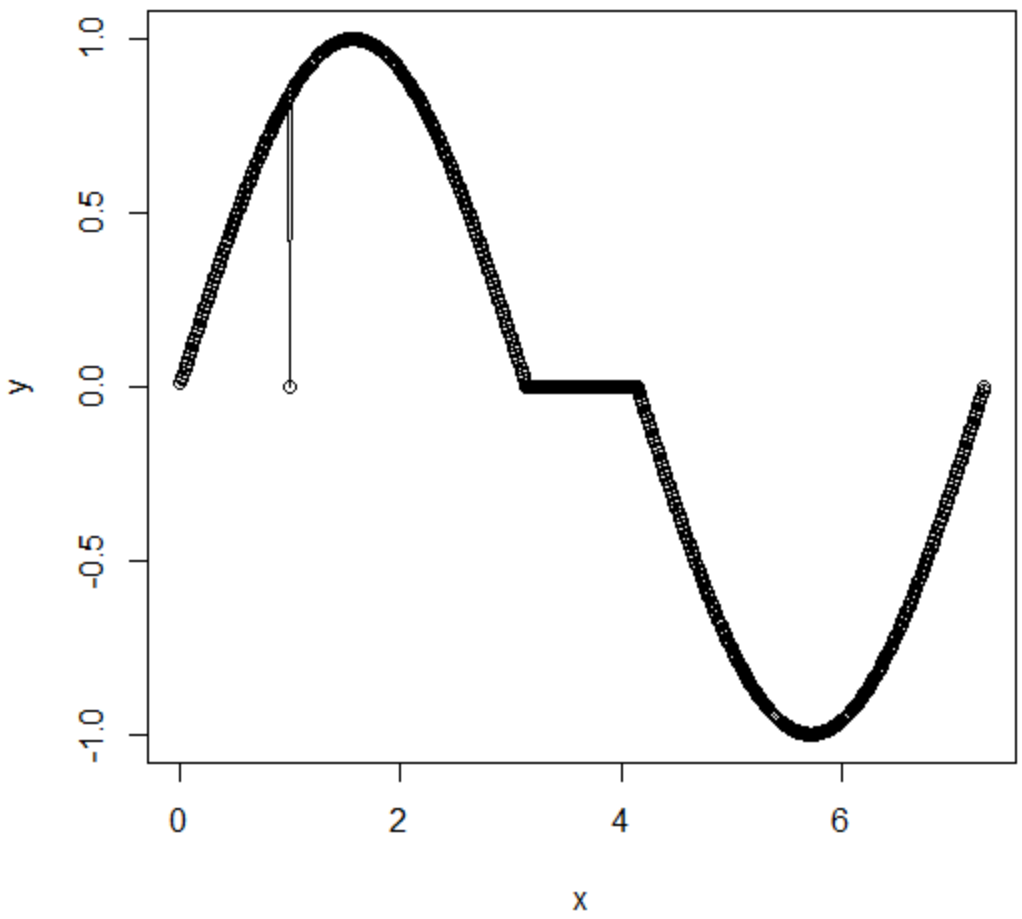


Figure 6.15: Plot of generated dataset

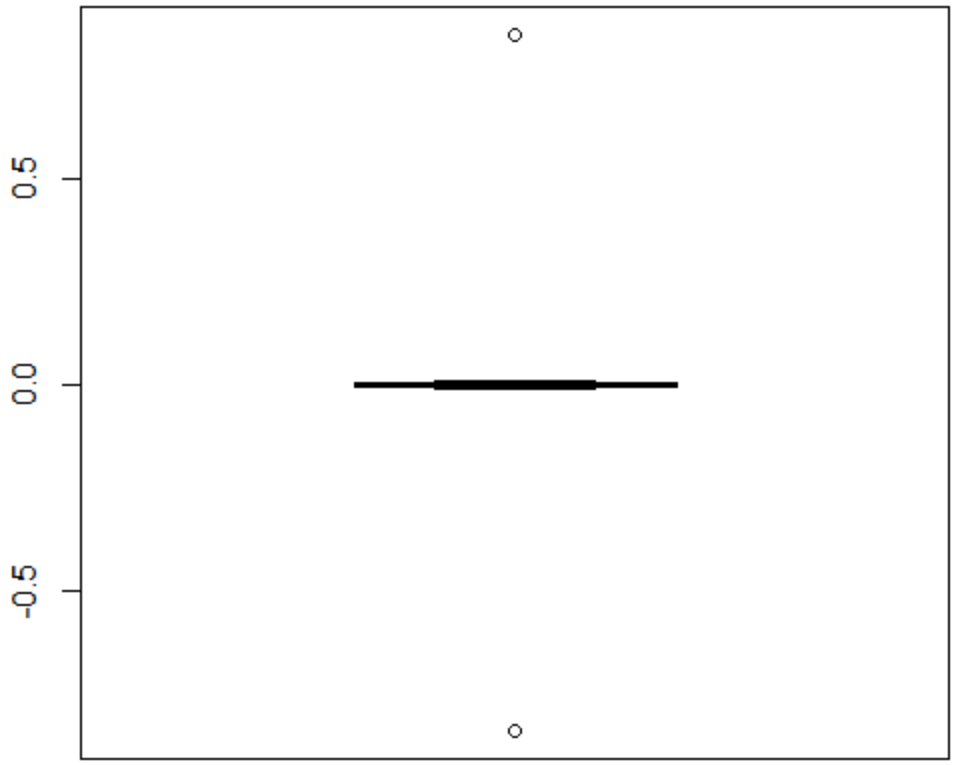


Figure 6.16: Boxplot of the first difference data

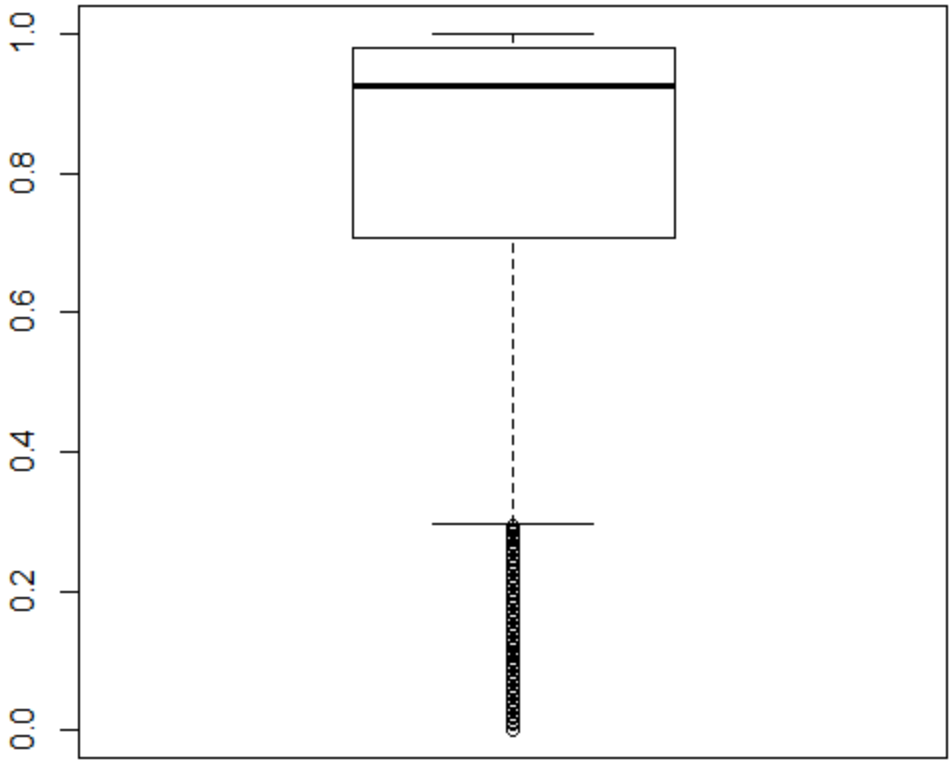


Figure 6.17: Boxplot of neighborhood changes

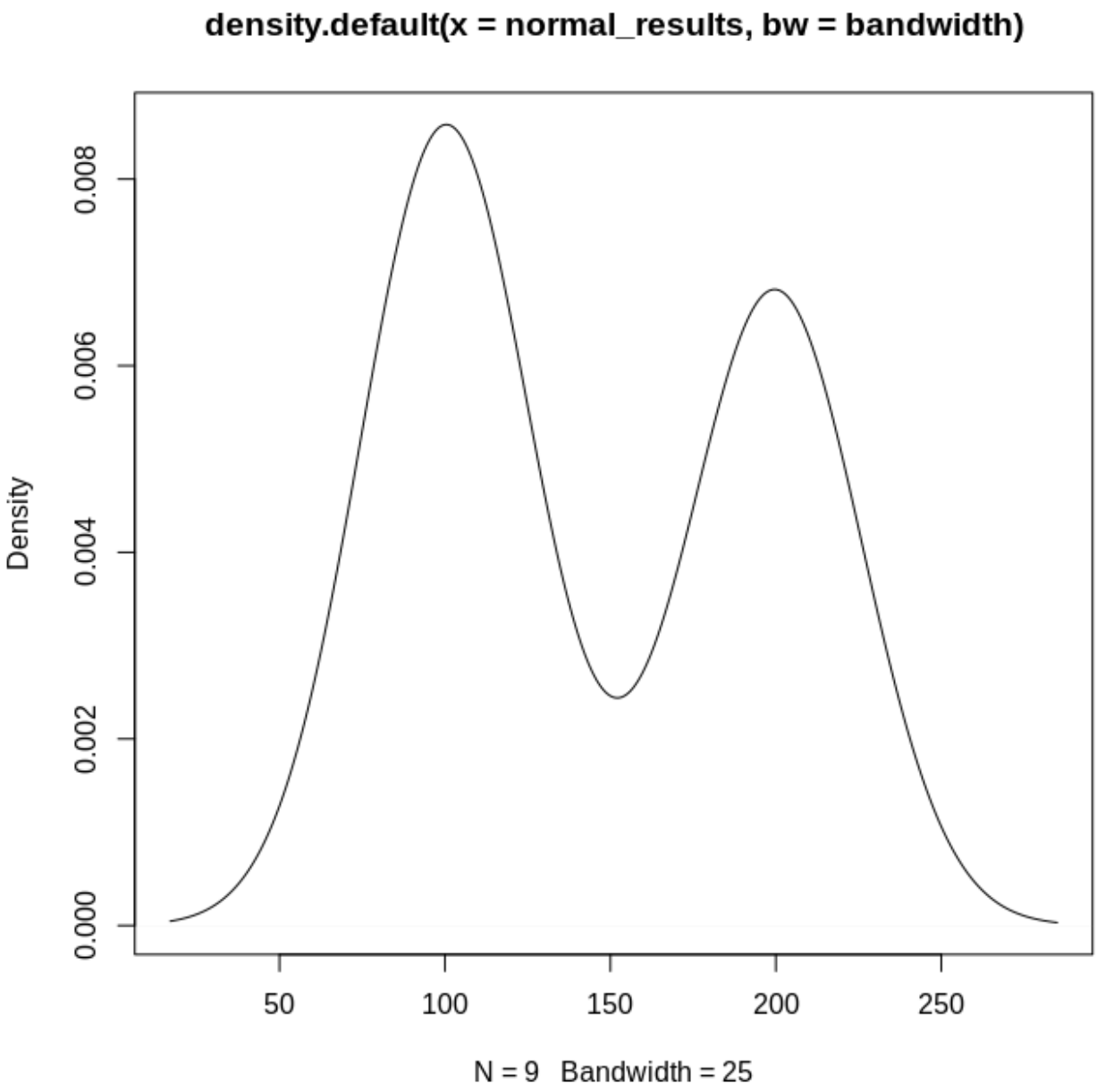


Figure 6.18: Plot of density estimate

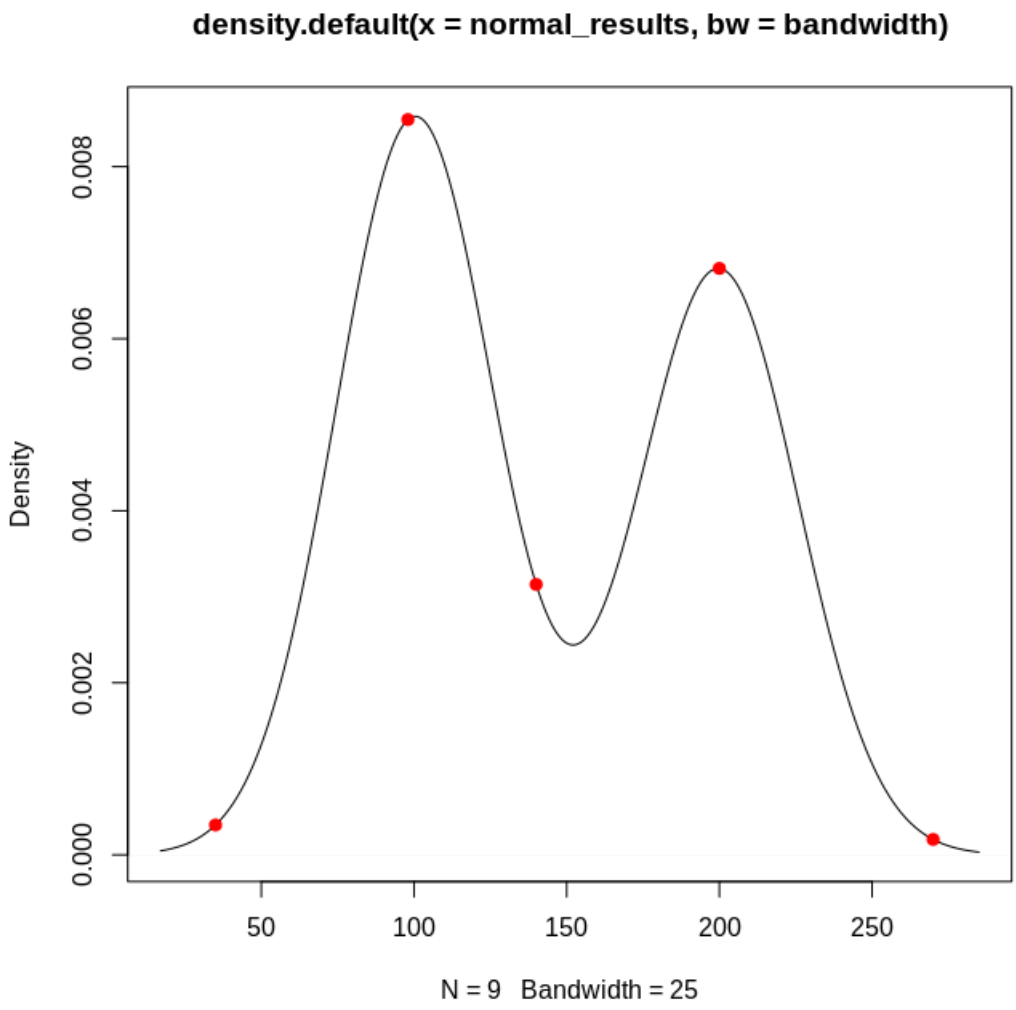


Figure 6.19: Points mapped on density plot

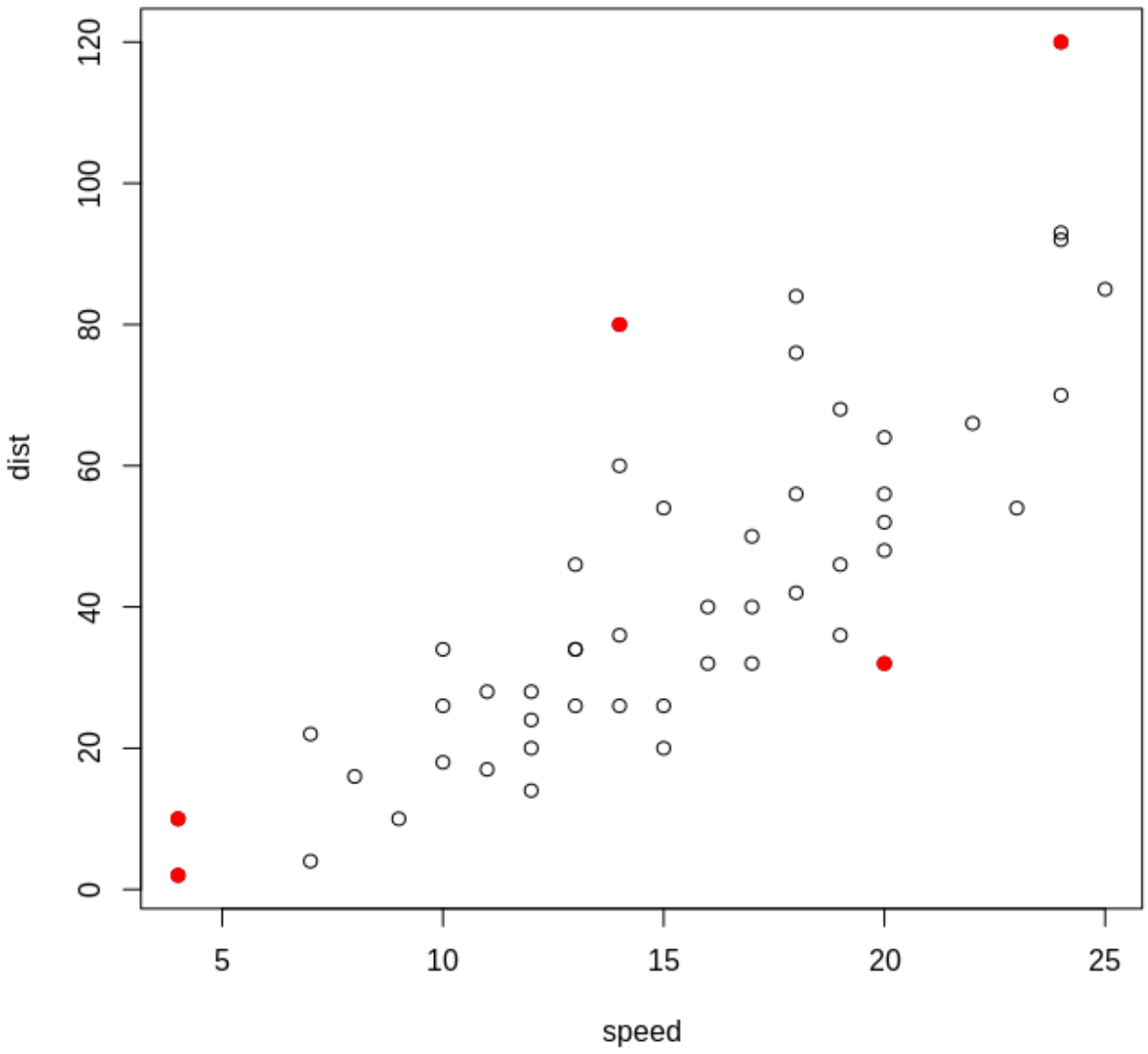


Figure 6.20: Plot with outliers marked

Solutions

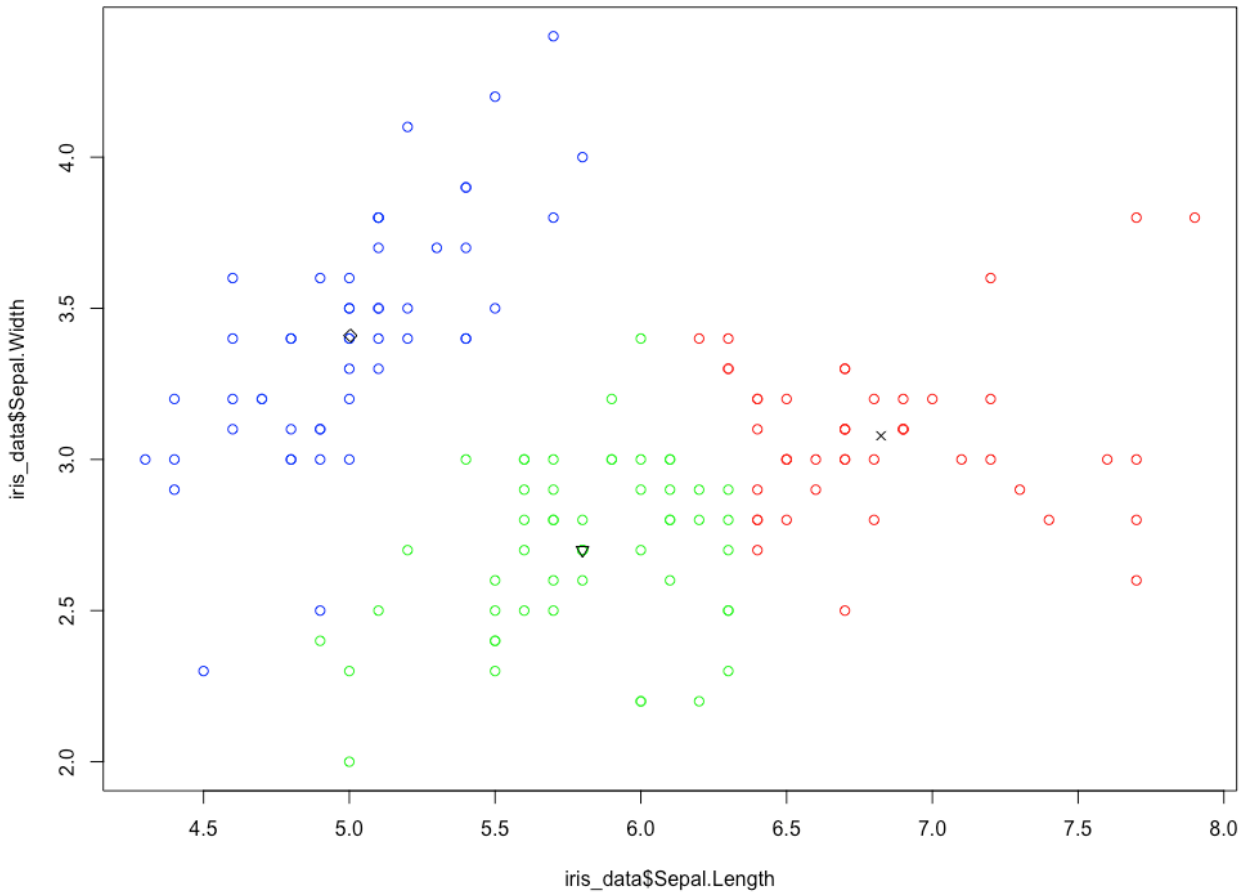


Figure 1.36: Scatter plot for the given cluster centers

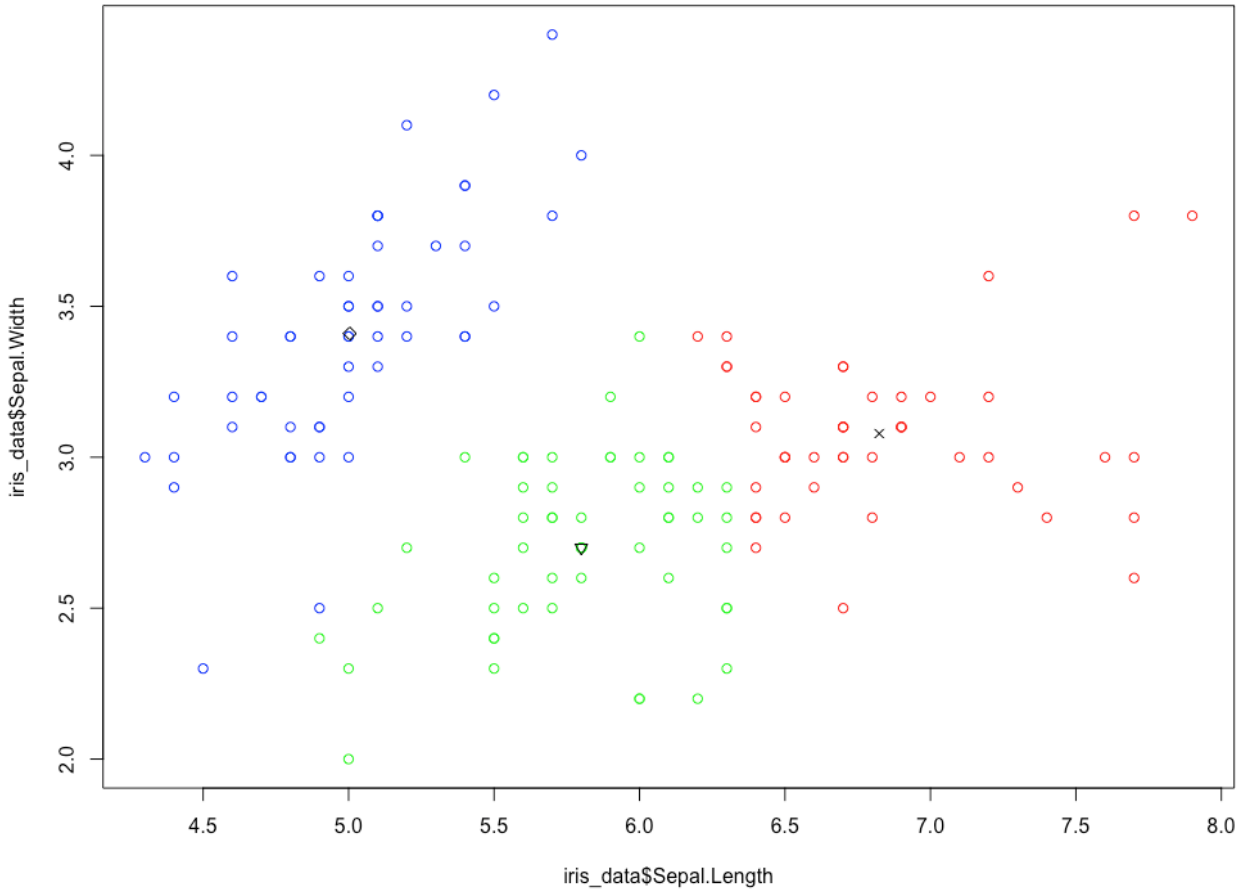


Figure 1.37: Scatter plot representing different species in different colors

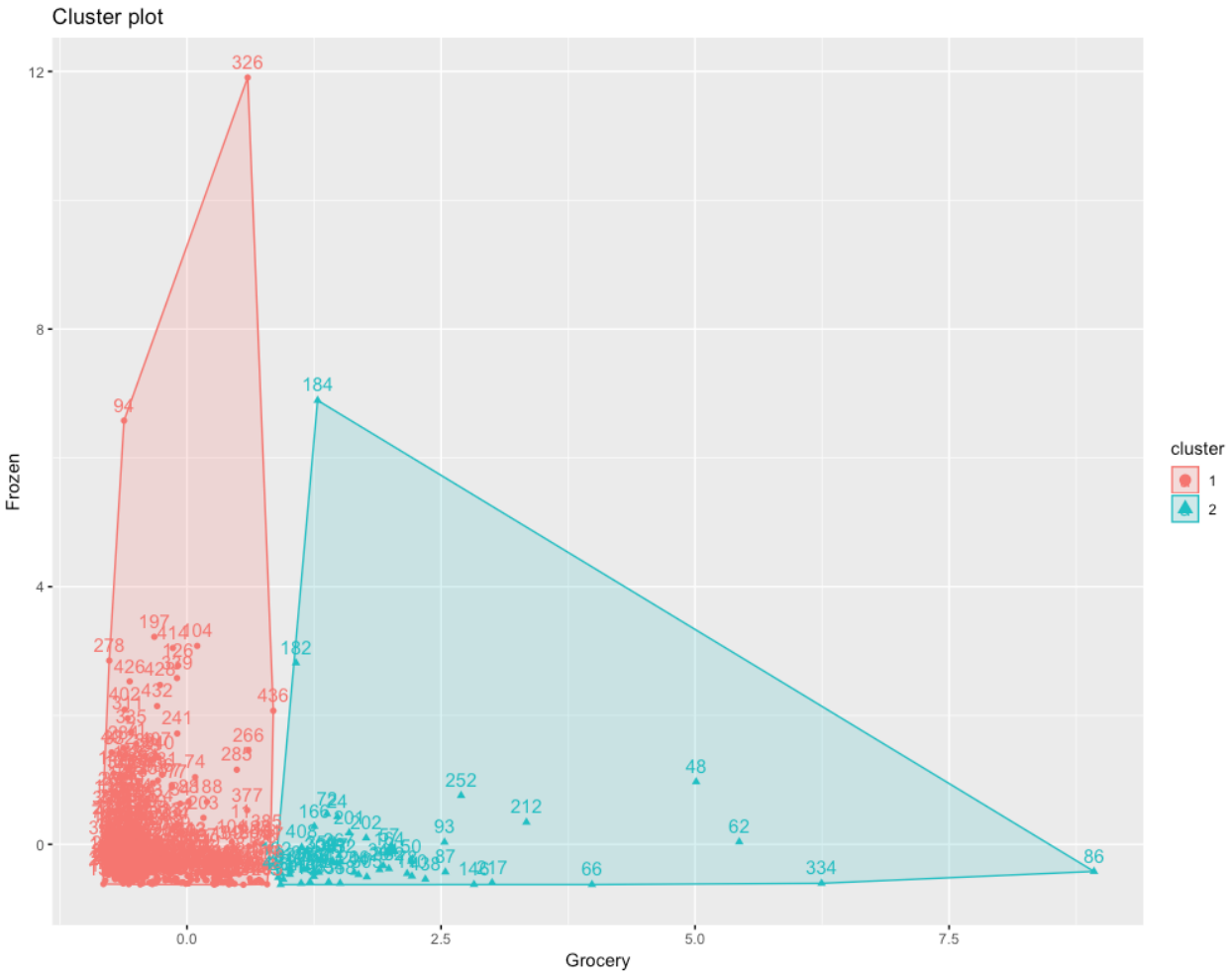


Figure 1.38: Chart for two clusters

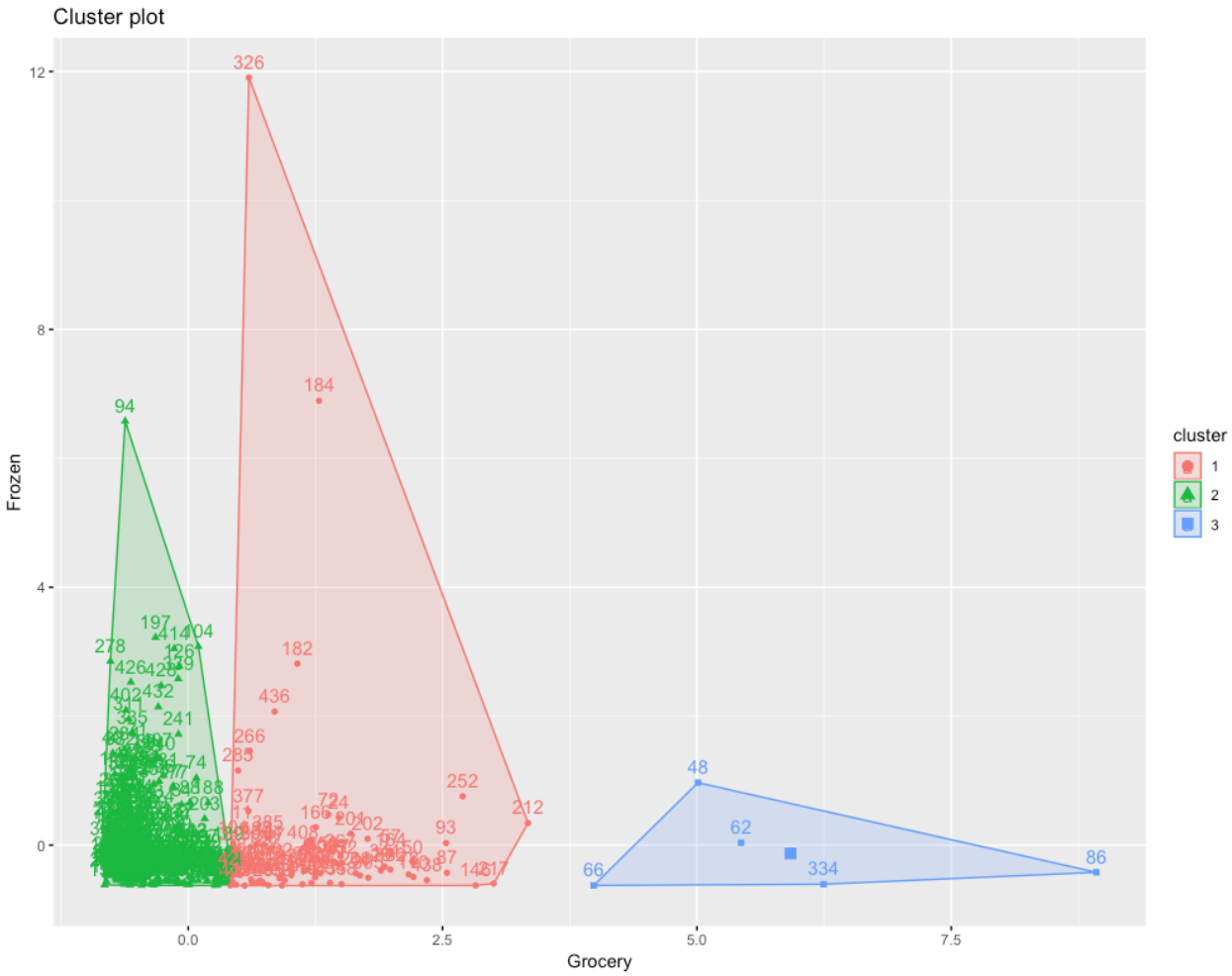


Figure 1.39: Chart for three clusters

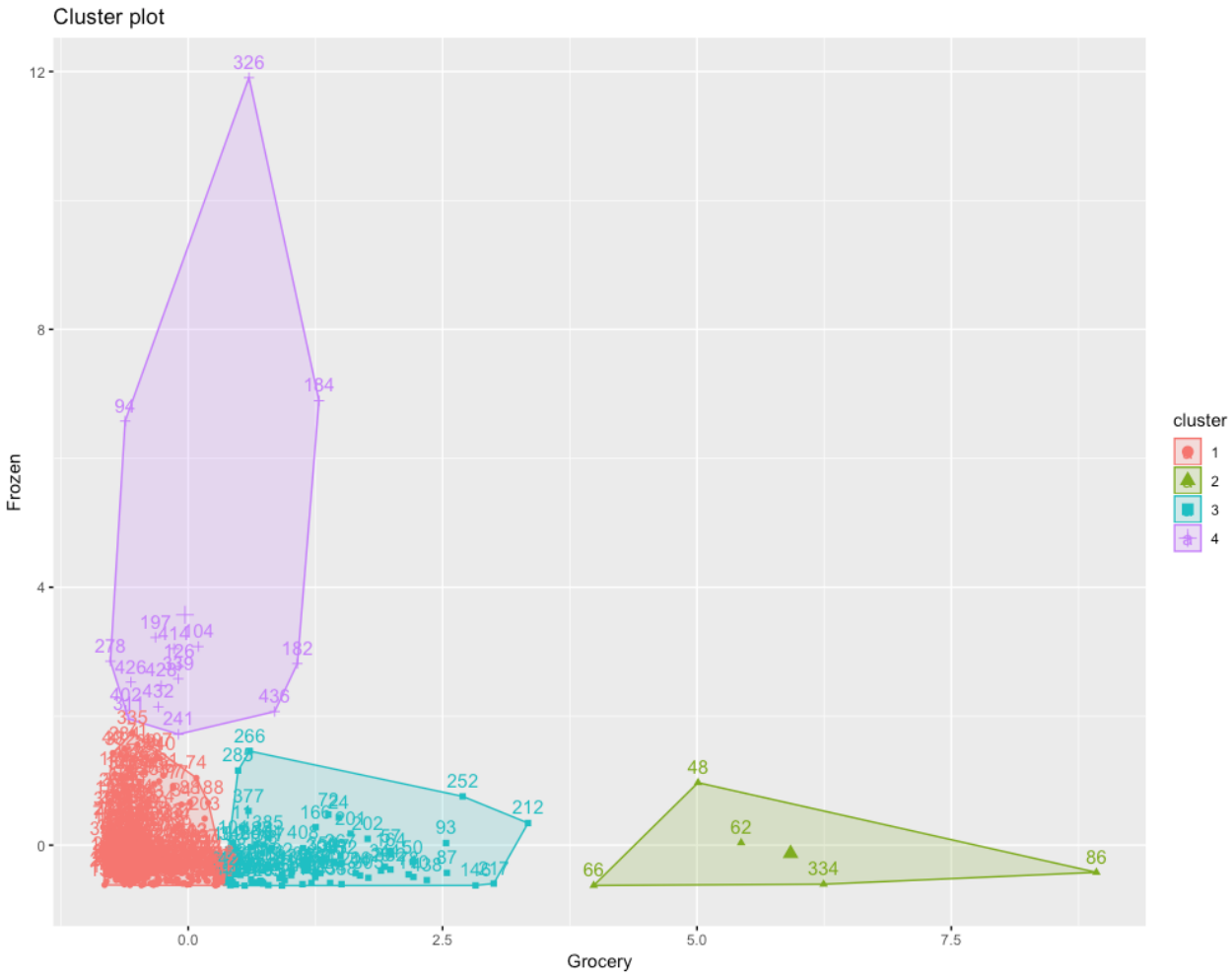


Figure 1.40: Chart for four clusters

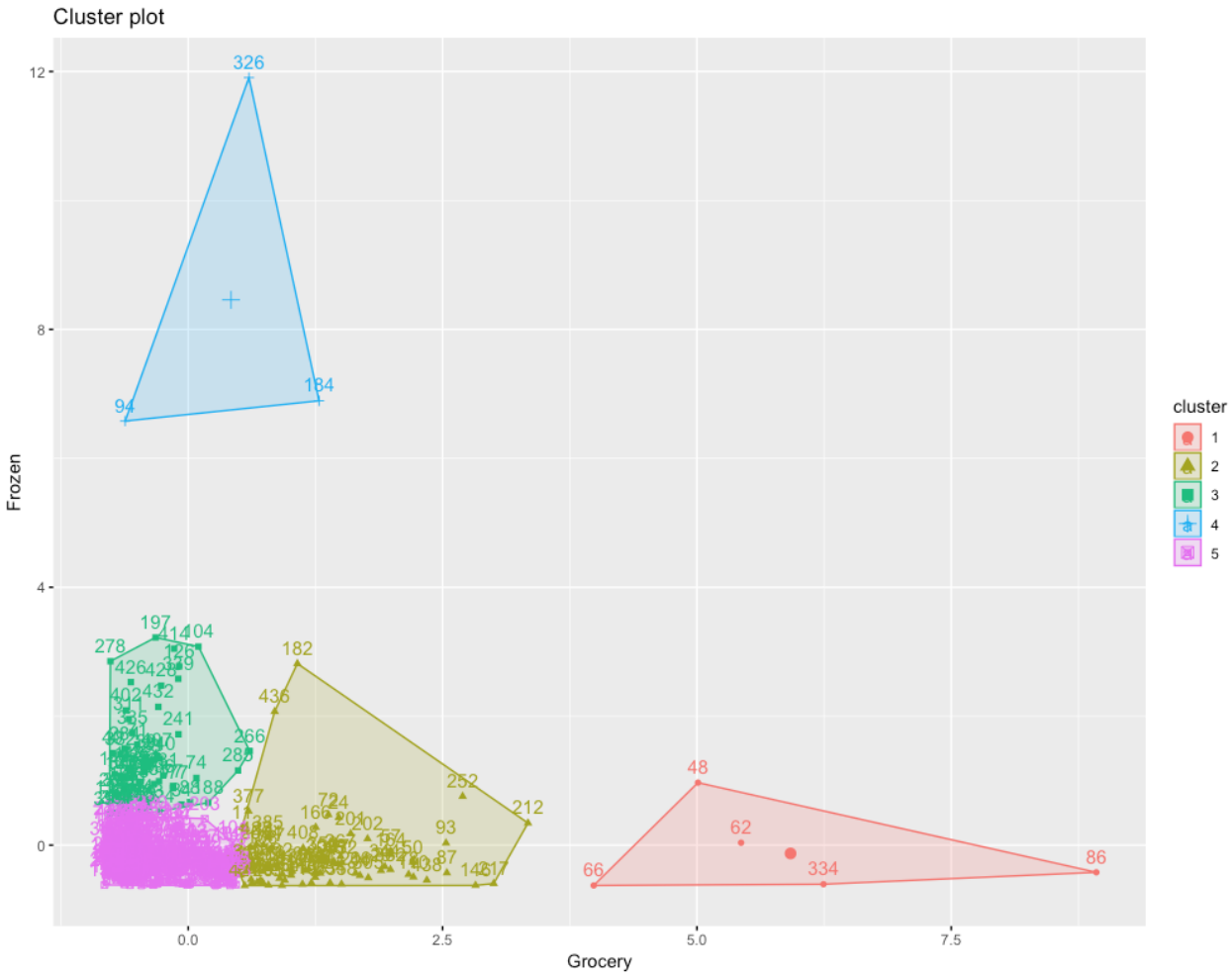


Figure 1.41: Chart for five clusters

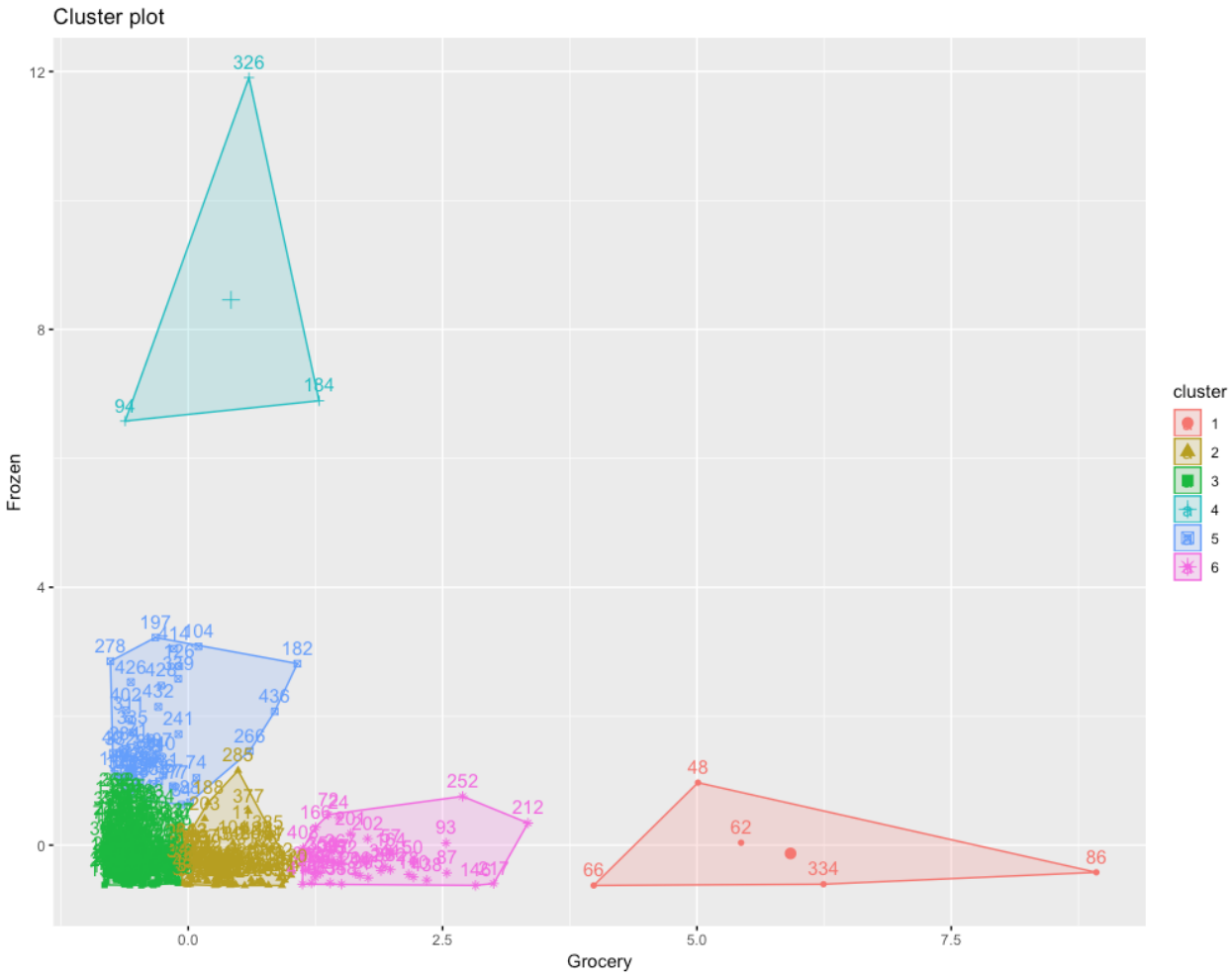


Figure 1.42: Chart for six clusters

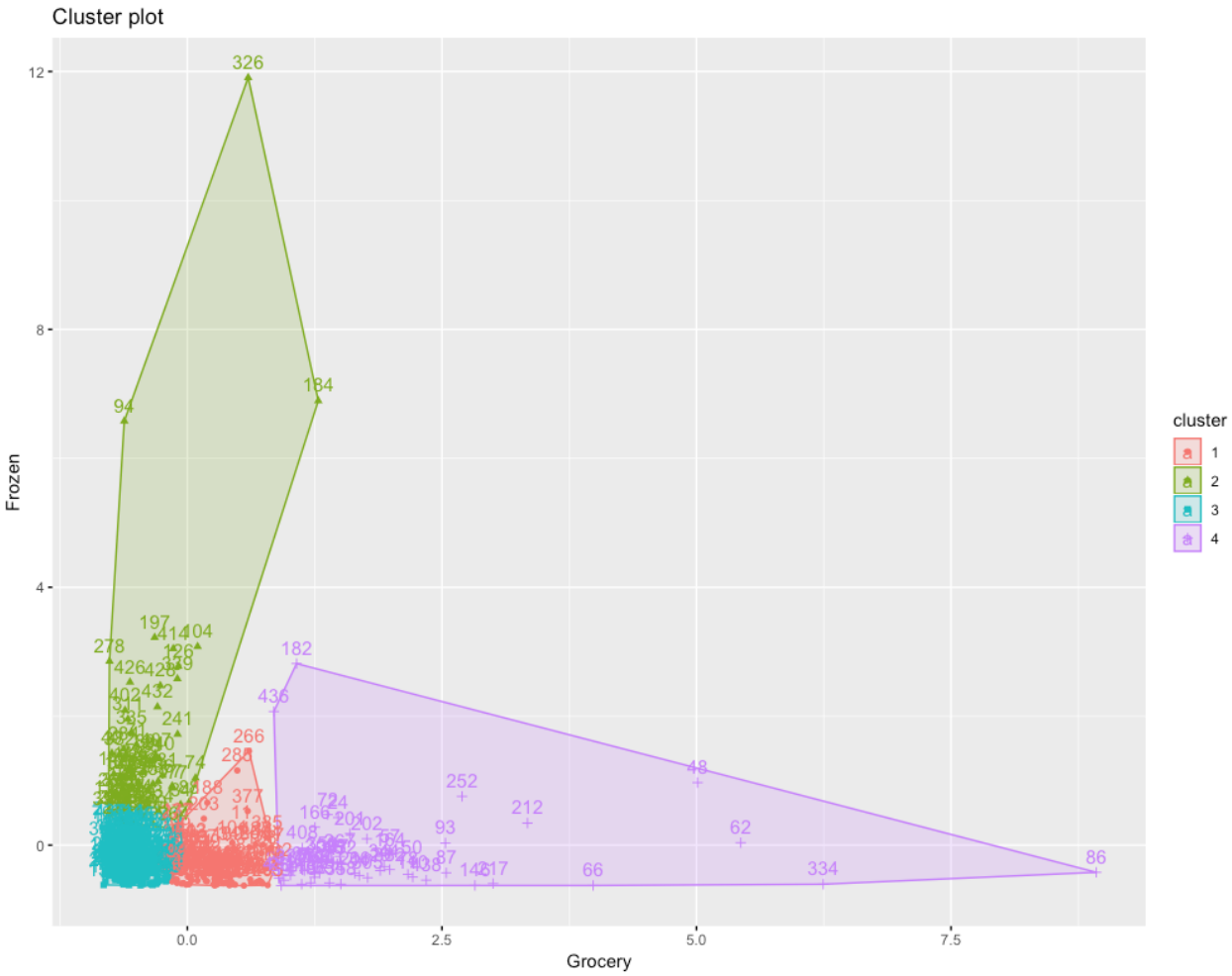


Figure 1.43: K-medoid plot of the clusters

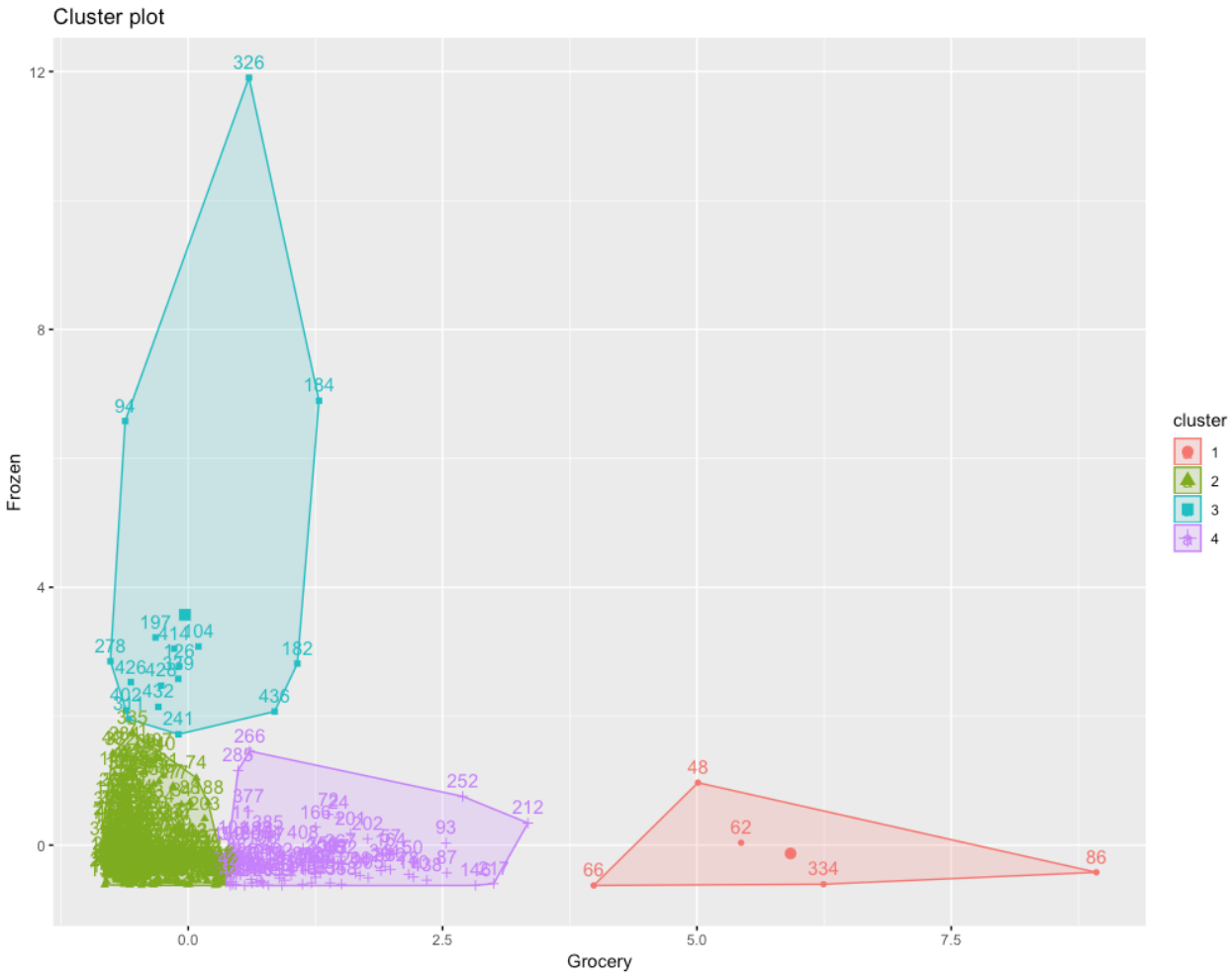


Figure 1.44: K-means plot of the clusters

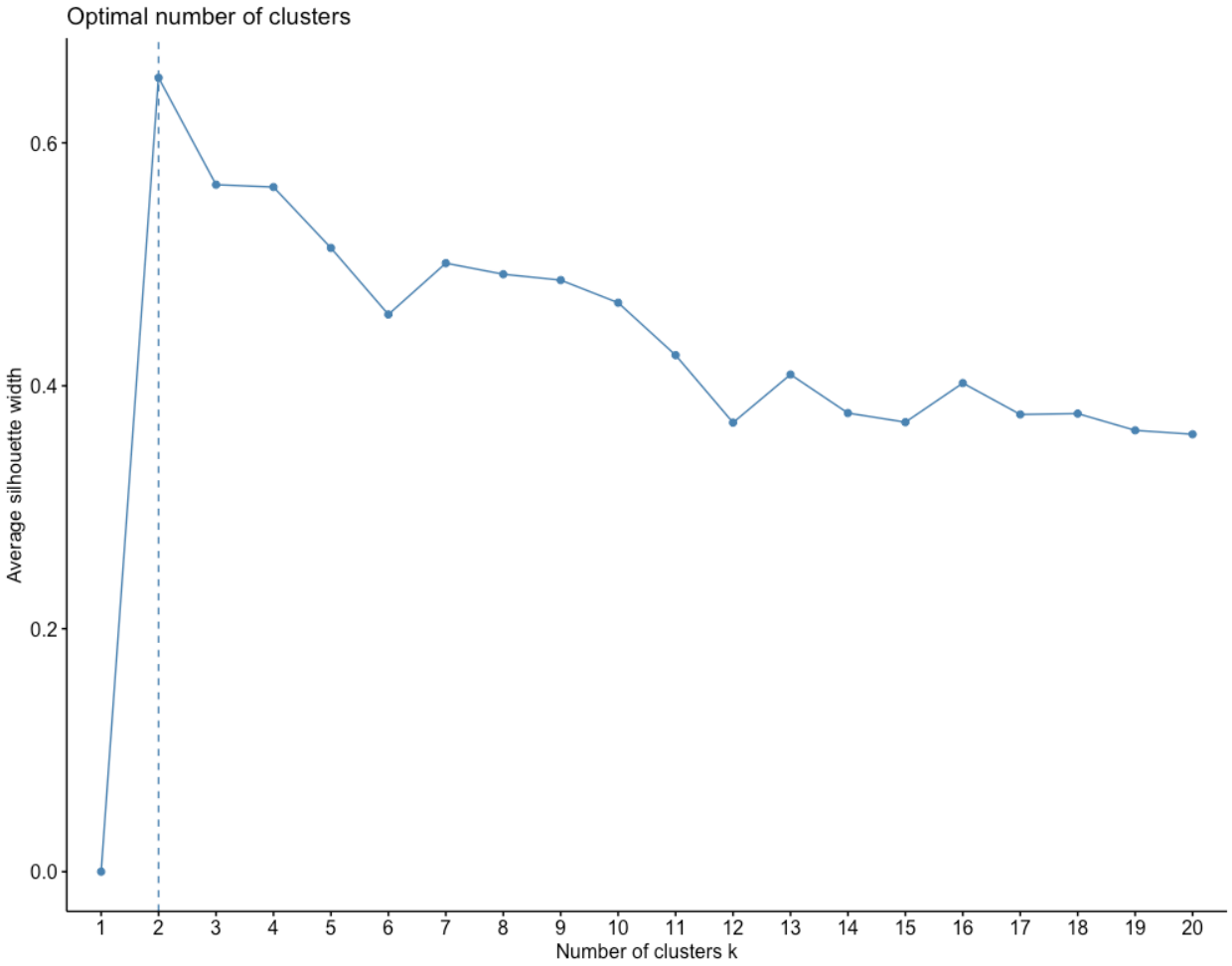


Figure 1.45: Graph representing optimal number of clusters with the silhouette score

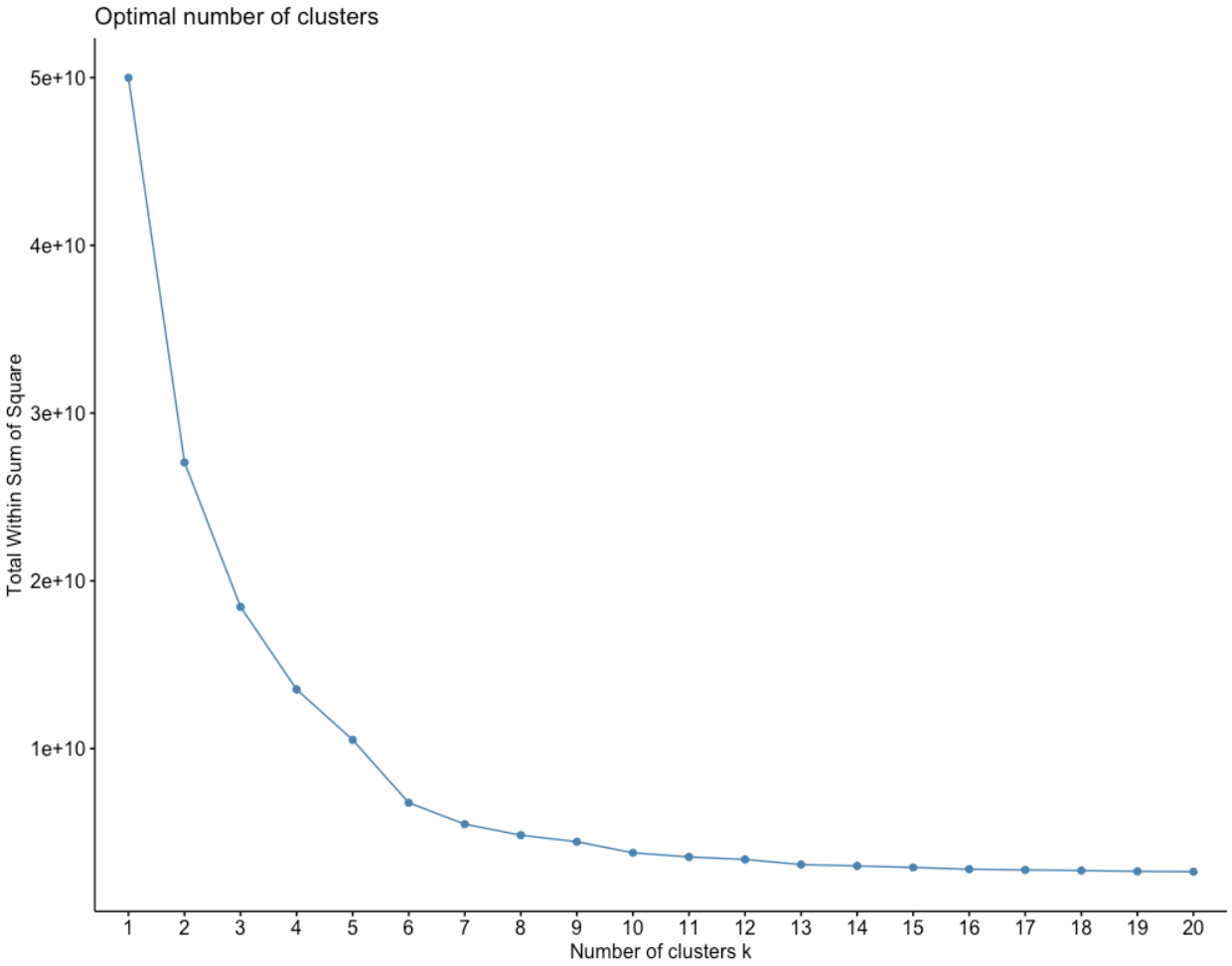


Figure 1.46: Optimal number of clusters with the WSS score

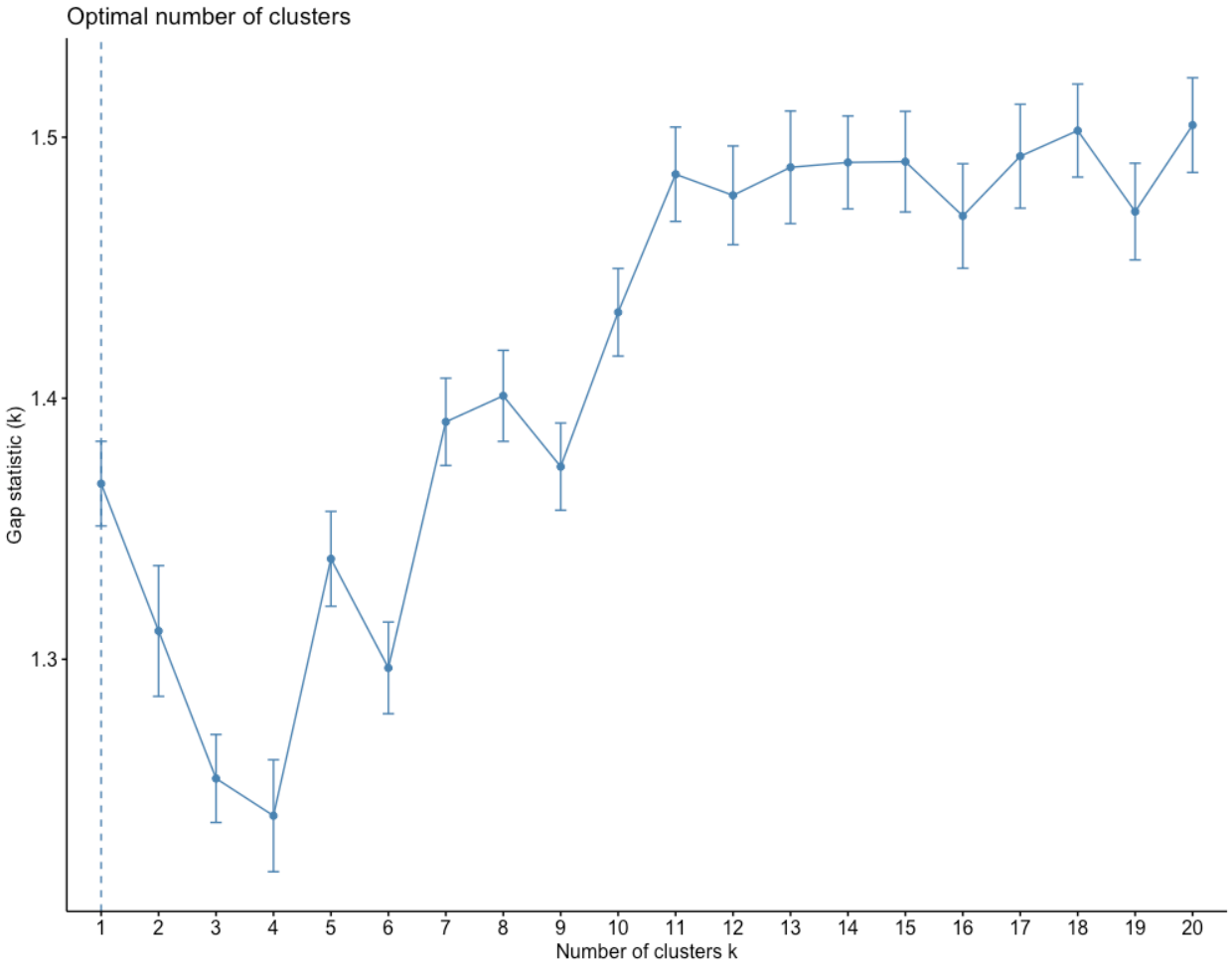


Figure 1.47: Optimal number of clusters with the Gap statistic

```
class    cap.shape cap.surface   cap.color    bruises       odor      gill.attachment
e:4208   b: 452    f:2320      n      :2284   f:4748   n      :3528   a: 210
p:3916   c:   4    g:   4      g      :1840   t:3376   f      :2160   f:7914
         f:3152    s:2556      e      :1500            s      : 576
         k: 828    y:3244      y      :1072            y      : 576
         s:  32                w      :1040            a      : 400
         x:3656                b      : 168            l      : 400
                               (Other): 220            (Other): 484
gill.spacing gill.size   gill.color   stalk.shape stalk.root stalk.surface.above.ring
c:6812       b:5612    b      :1728   e:3516      ?:2480     f: 552
w:1312       n:2512    p      :1492   t:4608      b:3776     k:2372
                       w      :1202               c: 556     s:5176
                       n      :1048               e:1120     y:  24
                       g      : 752               r: 192
                       h      : 732
                       (Other):1170
stalk.surface.below.ring stalk.color.above.ring stalk.color.below.ring veil.type veil.color
f: 600                     w      :4464           w      :4384           p:8124    n:  96
k:2304                     p      :1872           p      :1872                     o:  96
s:4936                     g      : 576           g      : 576                     w:7924
y: 284                     n      : 448           n      : 512                     y:   8
                           b      : 432           b      : 432
                           o      : 192           o      : 192
                           (Other): 140           (Other): 156
ring.number ring.type spore.print.color population habitat
n:  36      e:2776    w      :2388      a: 384     d:3148
o:7488      f:  48    n      :1968      c: 340     g:2148
t: 600      l:1296    k      :1872      n: 400     l: 832
            n:  36    h      :1632      s:1248     m: 292
            p:3968    r      :  72      v:4040     p:1144
                      b      :  48      y:1712     u: 368
                      (Other): 144                 w: 192
```

Figure 2.29: Screenshot of the summary of distribution of all columns

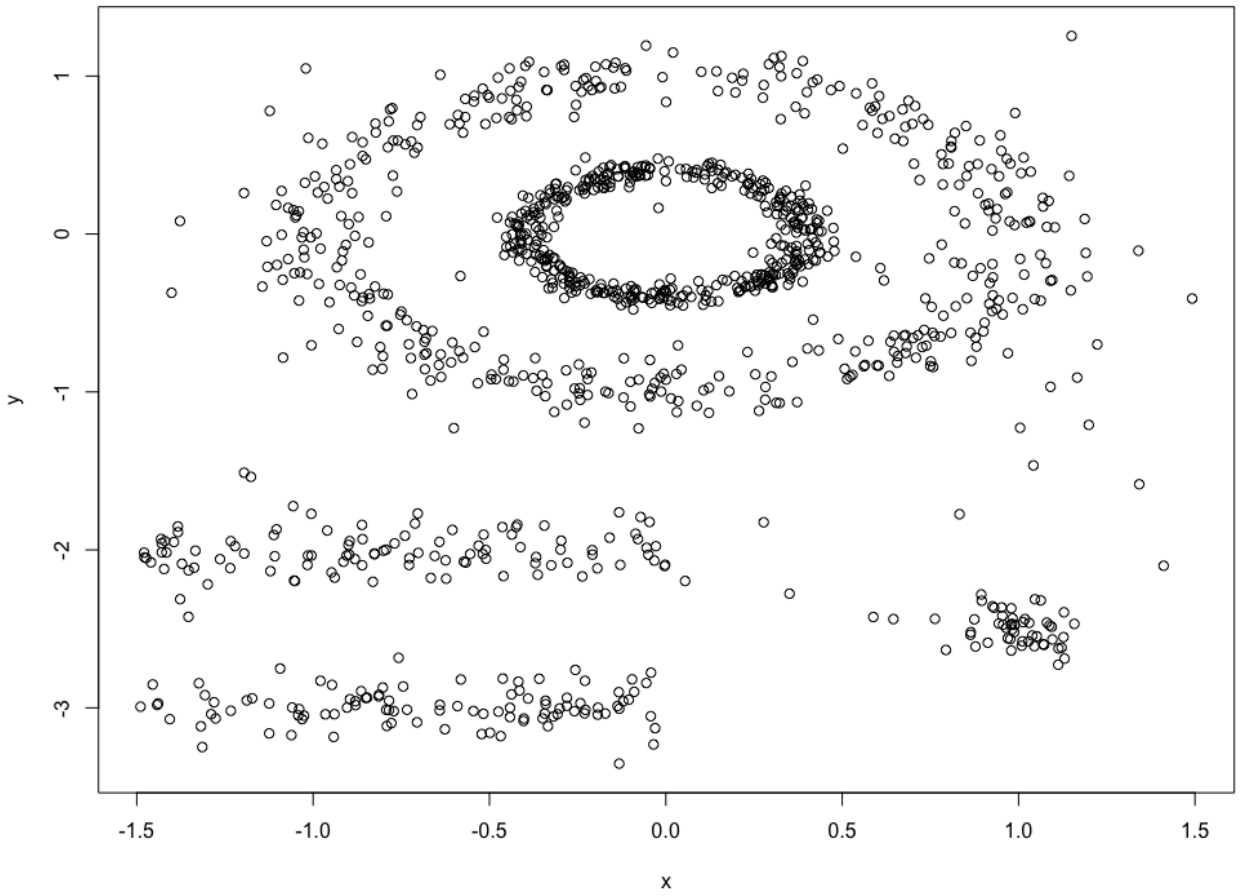


Figure 2.30: Plot of the multishapes dataset

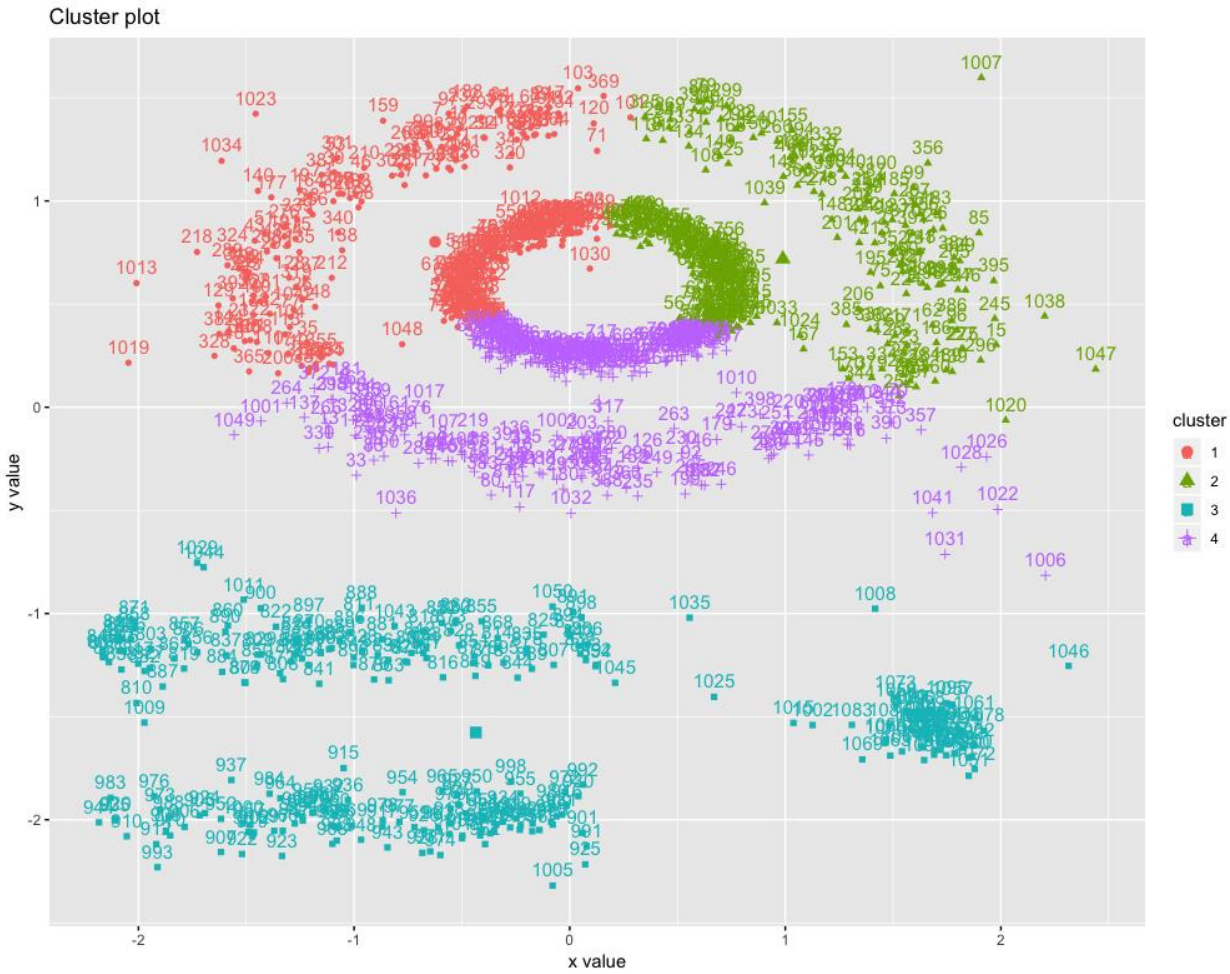


Figure 2.31: Plot of k-means on the multishapes dataset

Figure 2.32: Plot of DBCAN on the multishapes dataset

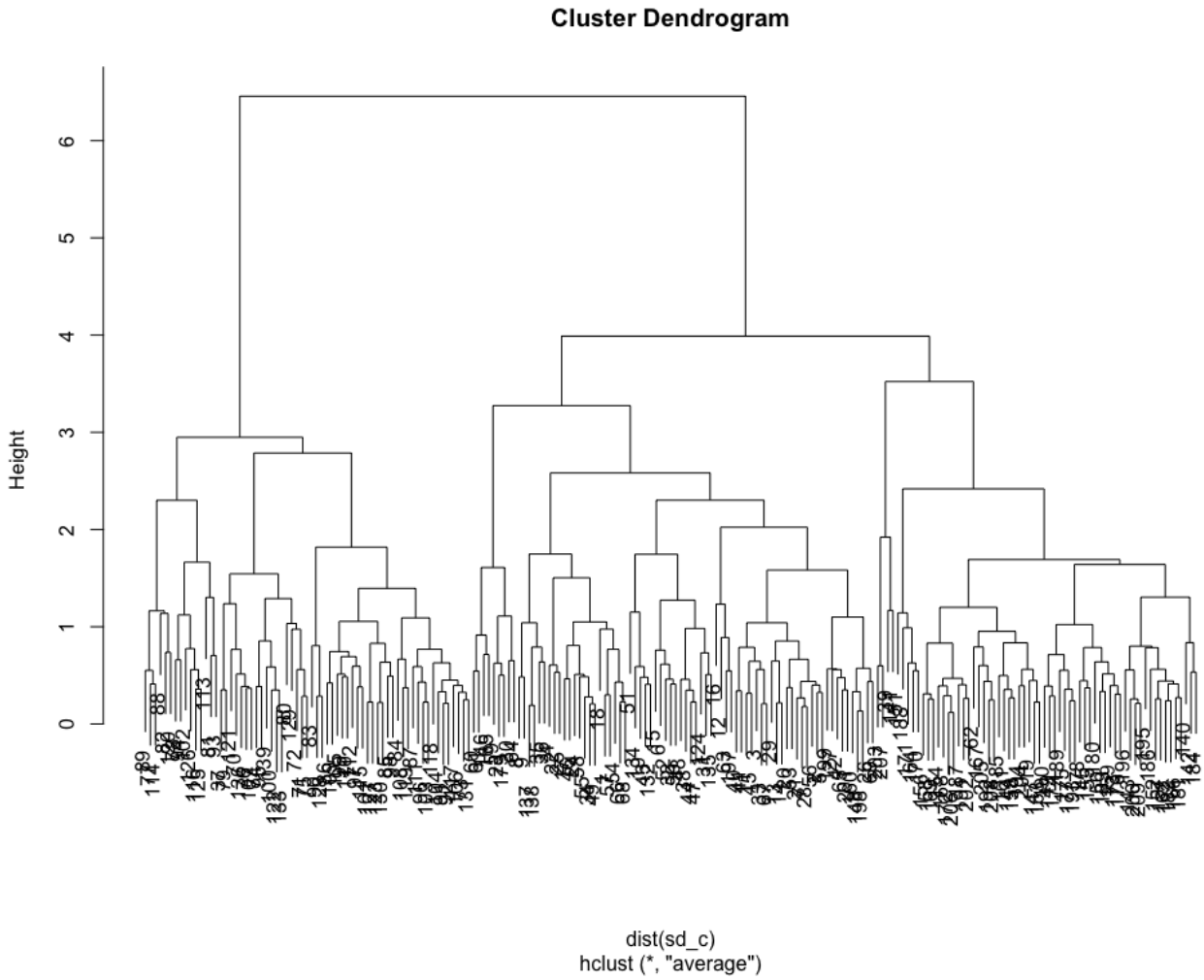


Figure 2.33: Cluster dendrogram

memb			
	1	2	3
1	65	3	1
2	6	0	64
3	9	61	0

Figure 2.34: Table classifying the three types of seeds

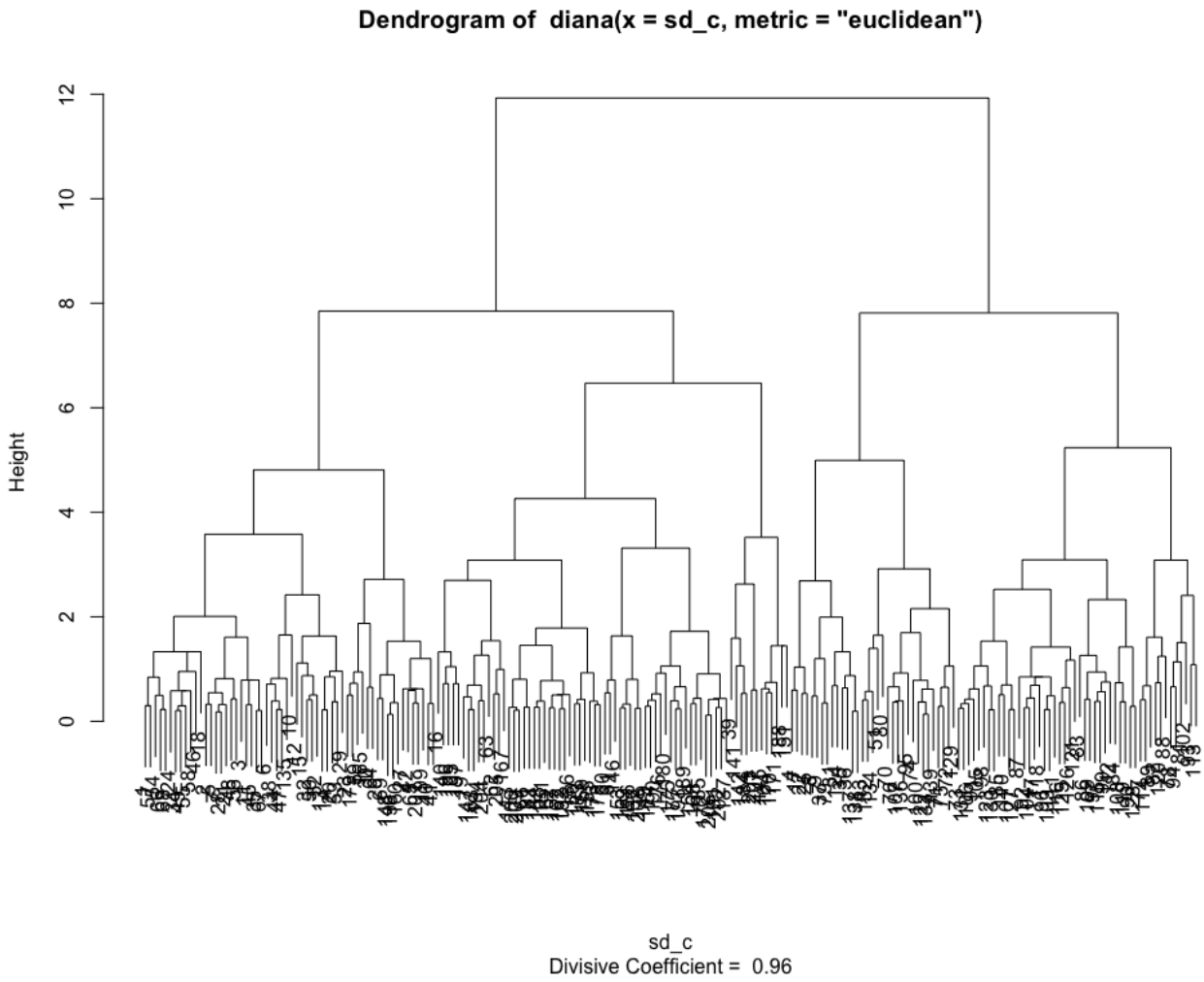


Figure 2.35: Dendrogram of divisive clustering

memb	1	2	3
1	65	3	1
2	6	0	64
3	9	61	0

Figure 2.36: Table classifying the three types of seeds

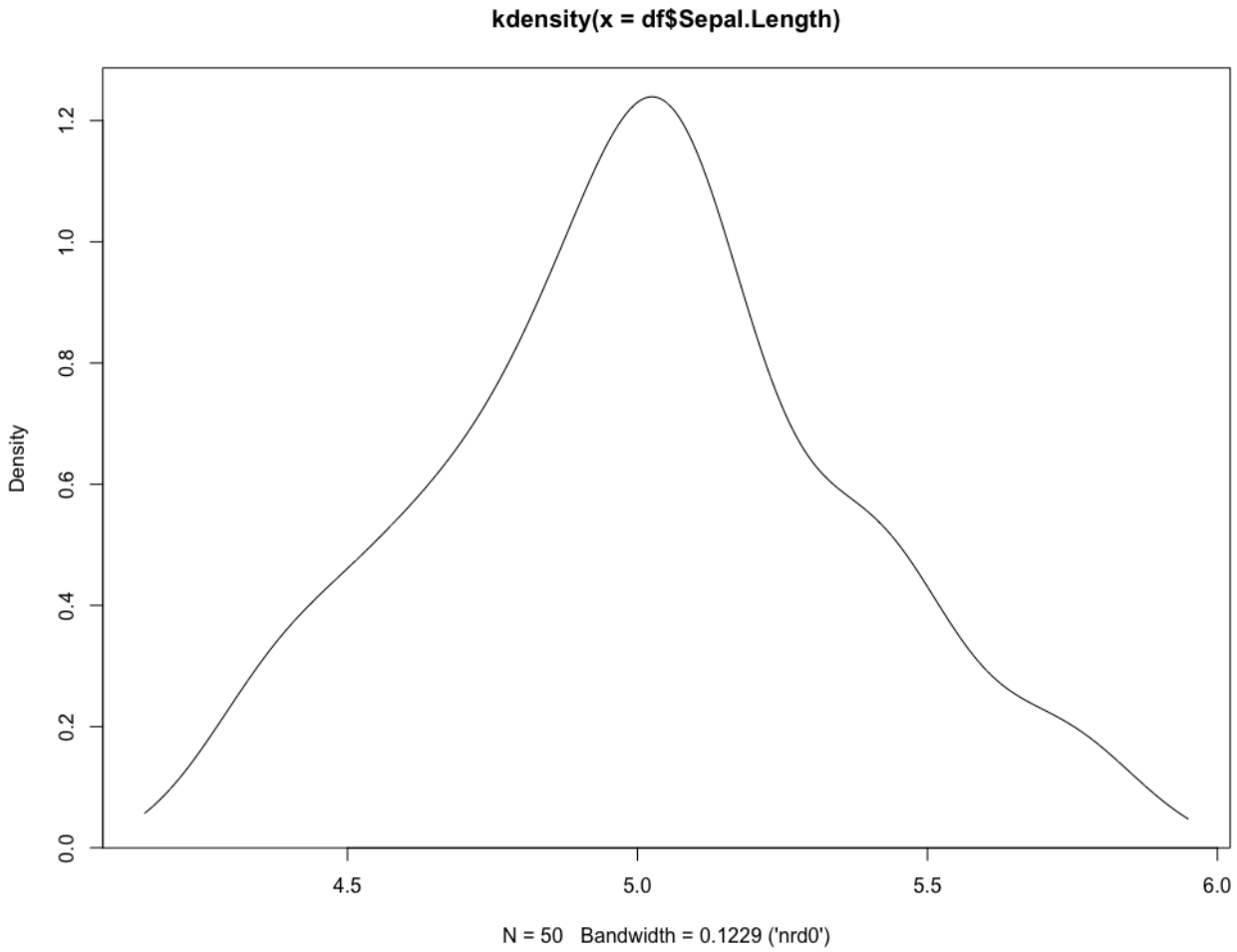


Figure 3.36 Plot of the KDE for sepal length

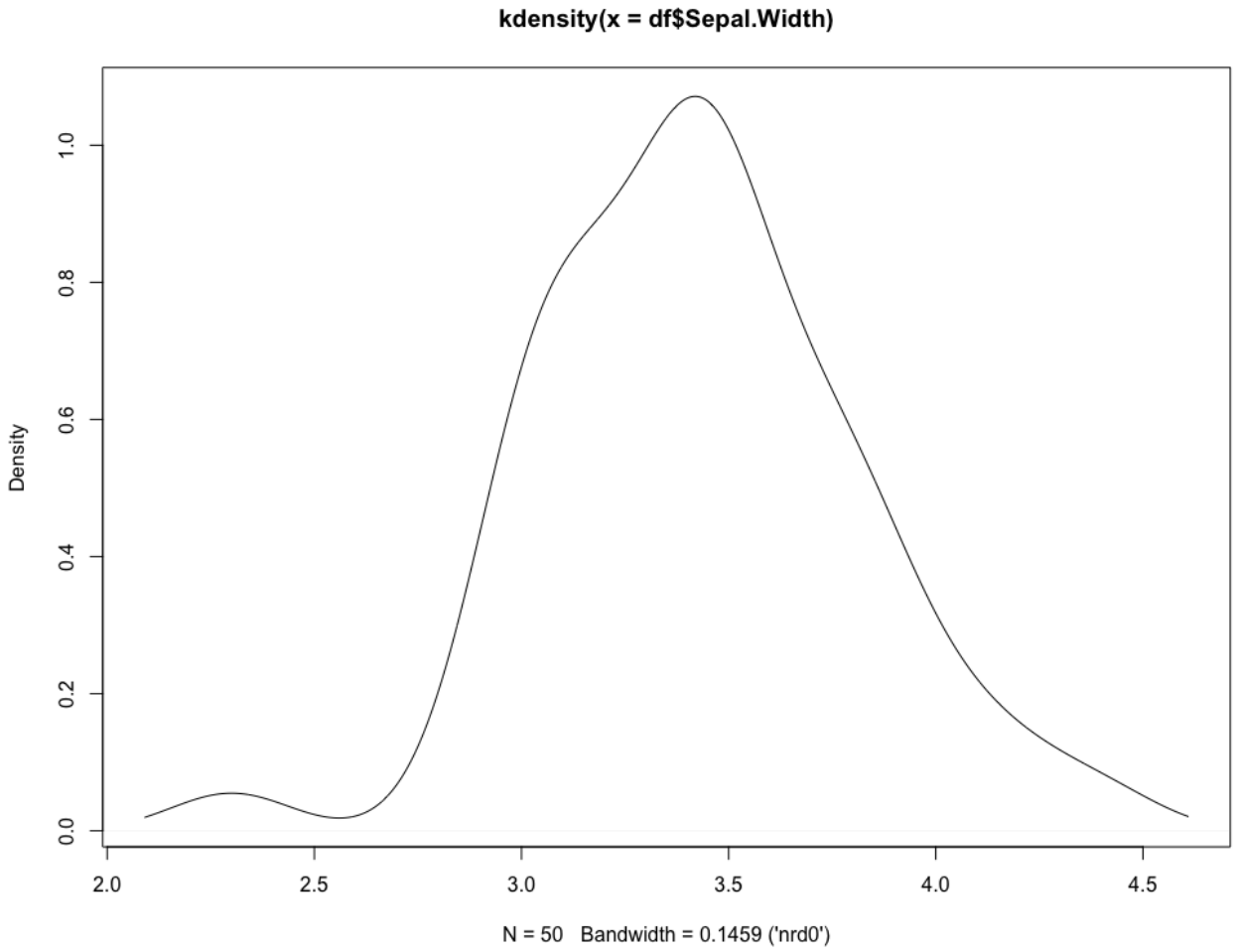


Figure 3.37 Plot of the KDE for sepal width

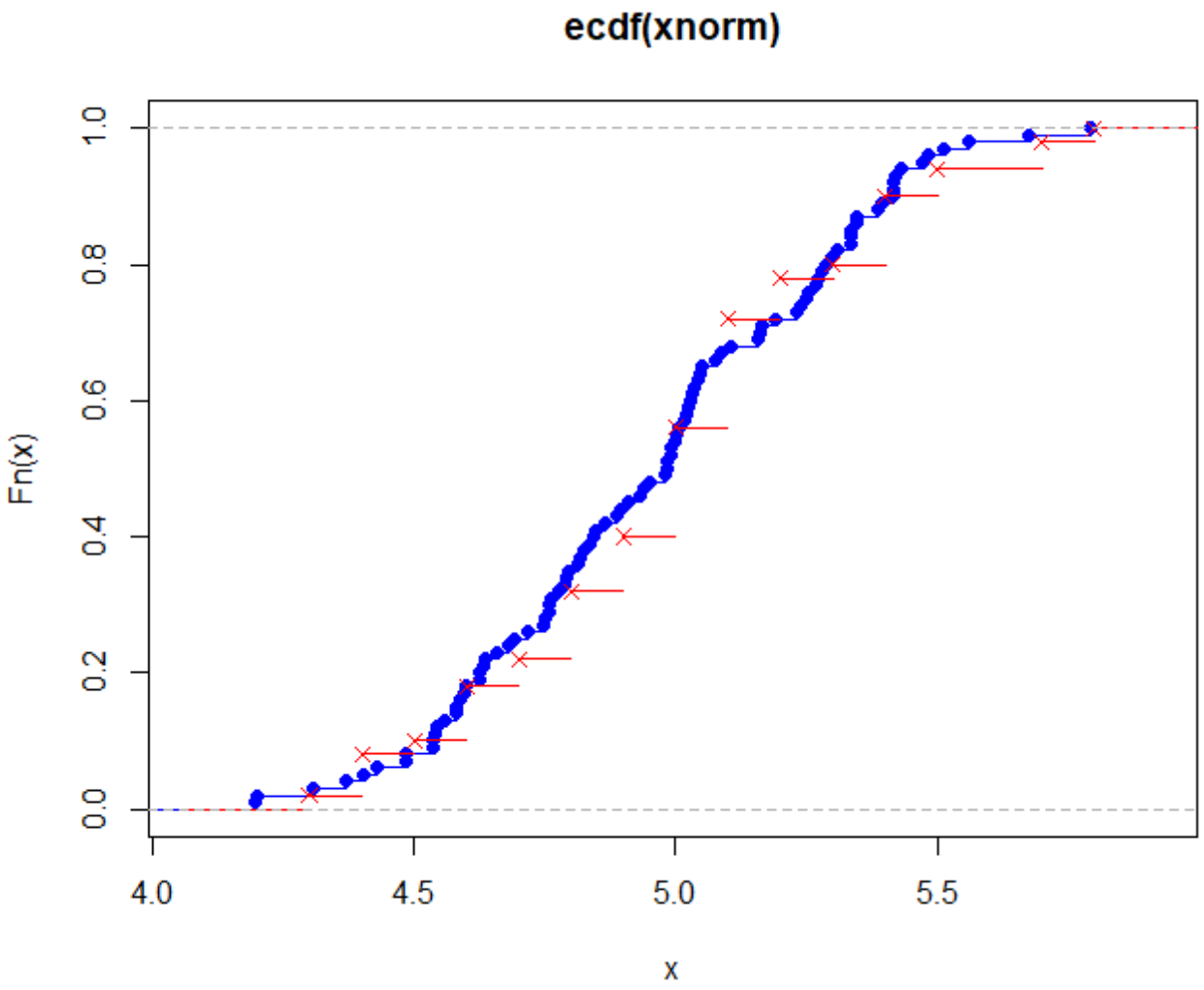


Figure 3.38: The CDF of xnorm and sepal length

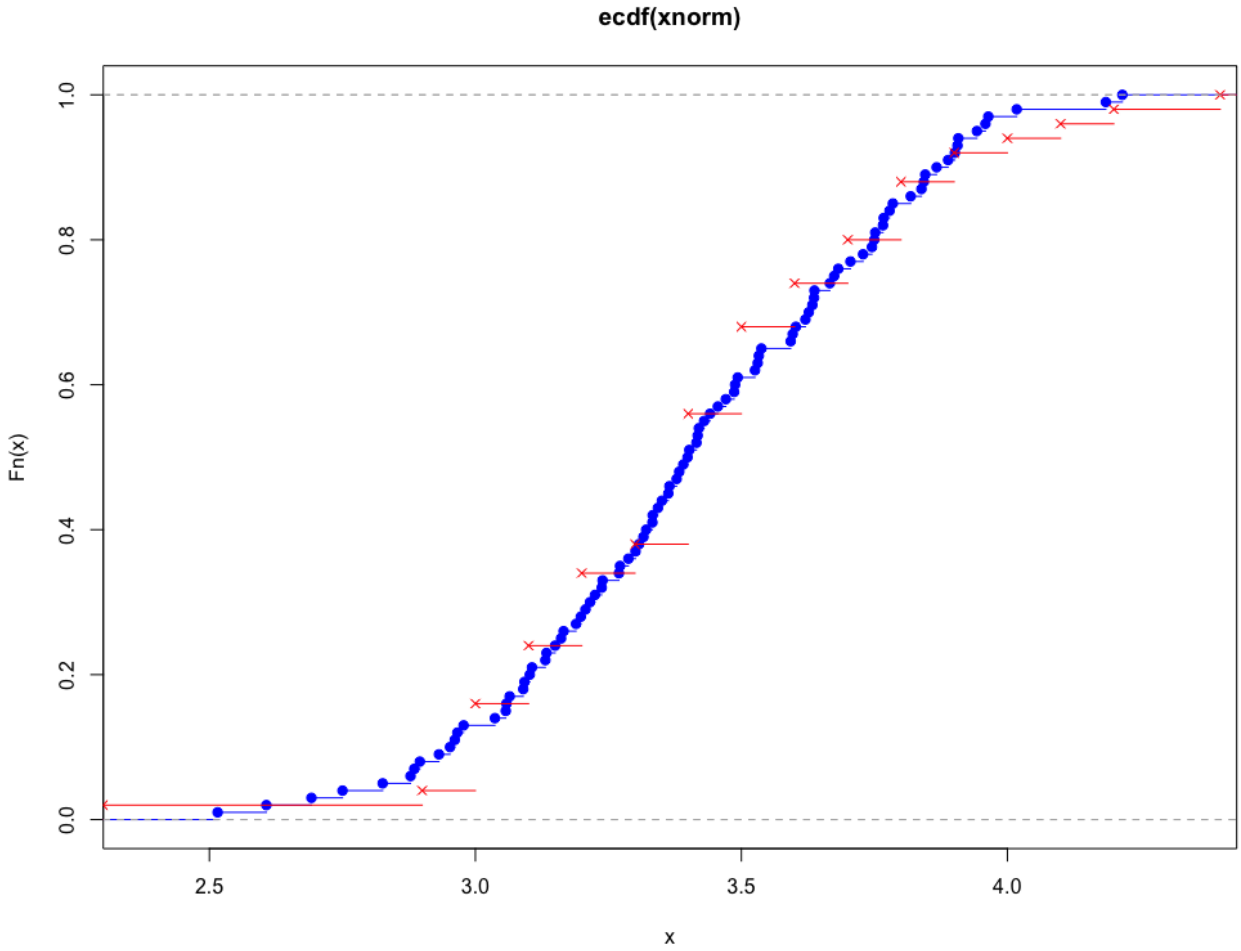


Figure 3.39: CDF of xnorm and sepal width

```
              [,1]         [,2]          [,3]          [,4]          [,5]
 [1,]  2.964674e-02 -0.015003651 -0.0268303390  0.3105083975  0.9159233539
 [2,] -4.492354e-02  0.631037941 -0.7639084529  0.0914984764 -0.0419675589
 [3,]  2.939532e-02 -0.088515154  0.0130376306  0.0541523272 -0.1258473131
 [4,] -5.070965e-05 -0.000906056 -0.0009292491 -0.0055661084  0.0009662470
 [5,]  4.615941e-04 -0.001817055 -0.0006921813  0.0002758662 -0.0003315697
 [6,] -1.225450e-03  0.005008013 -0.0063668357 -0.0463339423  0.0208901648
 [7,]  8.630861e-02 -0.752355714 -0.6397040816 -0.0812146848 -0.0035253723
 [8,] -6.760251e-03  0.044759063 -0.0017451705  0.0325596672 -0.0278668129
 [9,]  4.670538e-02  0.002571526  0.0181608586 -0.0234532588  0.2379011114
[10,]  9.926372e-01  0.101541990  0.0199846897 -0.0305981550 -0.0272793198
[11,]  5.910888e-03 -0.011370960  0.0329866611  0.0589679625 -0.0184675768
[12,]  2.321636e-02 -0.096883535 -0.0409221274  0.4586710733 -0.0858251068
[13,] -2.565800e-02  0.076330017 -0.0528757677 -0.8167640810  0.2778687767
              [,6]         [,7]         [,8]         [,9]         [,10]
 [1,] -0.155644295 -0.158326612  0.111224831  0.025486508 -0.0227333012
 [2,] -0.038206577 -0.006260425 -0.057018732  0.017656920  0.0318490448
 [3,] -0.860395716 -0.075885802 -0.465215293  0.009900762 -0.0977504102
 [4,] -0.006737114  0.005445449 -0.006327226 -0.009680569 -0.0052830397
 [5,] -0.004868223  0.001270763 -0.003446759 -0.015765167  0.0154002725
 [6,]  0.011854493 -0.009388874 -0.006015218 -0.009814883 -0.0007255794
 [7,]  0.078477626 -0.061457820 -0.007848599  0.009174481 -0.0238738299
 [8,]  0.110677220 -0.041812471  0.013583151  0.116127372 -0.9839320643
 [9,]  0.358117027  0.407606104 -0.793550182 -0.126111107 -0.0106763230
[10,]  0.003158506 -0.016277550  0.044752857  0.001579001 -0.0006268944
[11,]  0.088474600 -0.046285998 -0.146719744  0.973113016  0.1268199497
[12,] -0.170432847  0.807481699  0.287200268  0.071157157 -0.0279765721
[13,] -0.224240115  0.378038741  0.178558244  0.130435940 -0.0545477337
              [,11]         [,12]         [,13]
 [1,] -3.579451e-03 -1.386274e-03  7.168469e-04
 [2,] -2.722109e-03 -1.079790e-04  1.021309e-05
 [3,]  1.260538e-02  8.190011e-03 -4.138848e-03
 [4,] -2.085826e-02 -9.995483e-01 -1.406496e-02
 [5,] -3.828993e-04 -1.392499e-02  9.996394e-01
 [6,]  9.982814e-01 -2.054663e-02  2.645536e-05
 [7,] -5.357718e-03  1.068461e-03 -8.705608e-04
 [8,]  7.023865e-04  2.505075e-03  1.772891e-02
 [9,] -1.221601e-02  6.746324e-03 -3.164339e-03
[10,]  7.268209e-05 -4.231324e-04 -4.147717e-05
[11,]  1.046531e-02 -1.078571e-02  1.320768e-02
[12,]  3.536485e-02 -6.651794e-05  3.293986e-04
[13,] -3.579491e-02  6.976225e-03  2.455507e-03
```

Figure 4.17: Eigenvectors of the covariance matrix

```
 [1] 2.881882e+04 8.260671e+02 2.673629e+02 7.984006e+01 4.733876e+01
 [6] 1.706442e+01 1.359537e+01 8.961253e+00 2.761729e+00 1.103303e+00
[11] 2.233296e-01 5.914172e-02 2.930149e-03
```

Figure 4.18: Eigenvalues of the covariance matrix

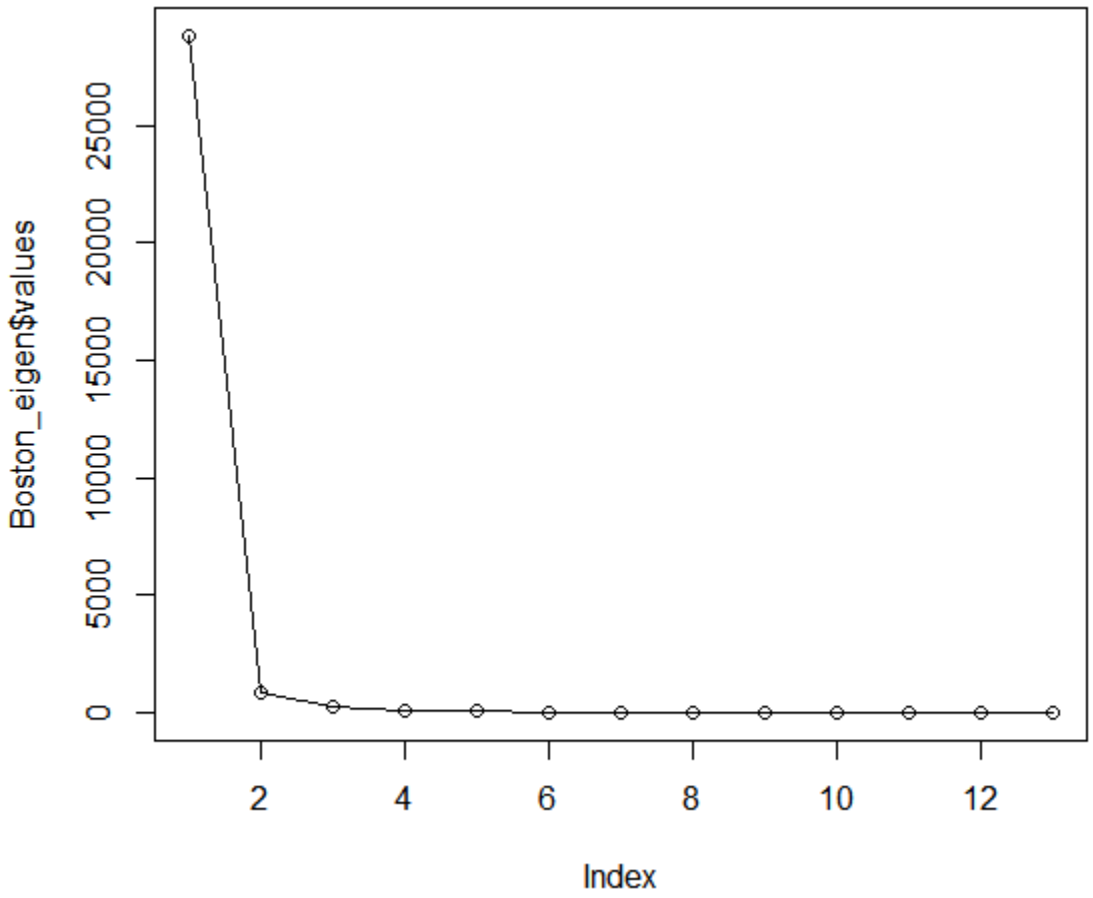


Figure 4.19: Plot of the eigenvalues

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1	3	6	0.2588933	0.9776119	1.917332
[2,]	1	3	8	0.2509881	0.9477612	1.018189
[3,]	1	3	10	0.2411067	0.9104478	1.693701
[4,]	1	3	14	0.2332016	0.8805970	1.761194
[5,]	1	3	15	0.2371542	0.8955224	1.791045
[6,]	1	3	18	0.2015810	0.7611940	1.254606

Figure 4.20: Output of the three-item basket

Figure 5.12: Matrix of borges_signature

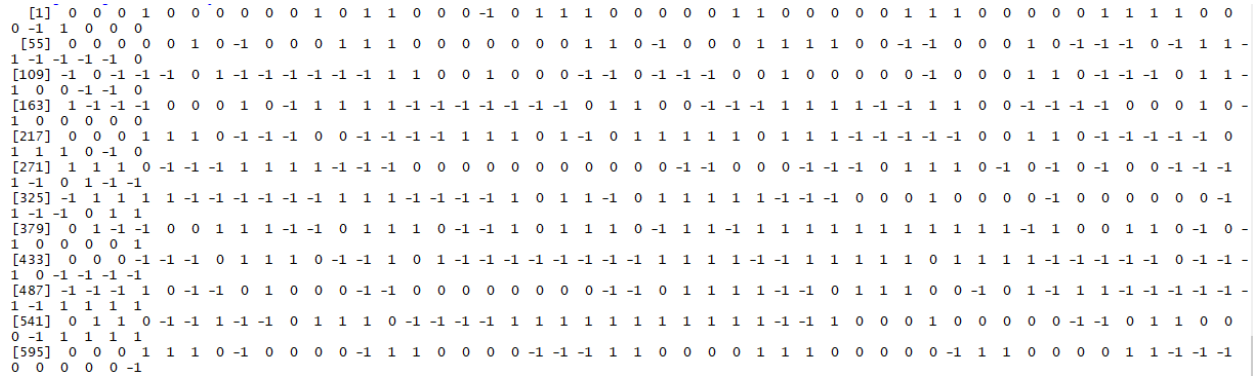


Figure 5.13: Matrix of borges_signature_ninebynine

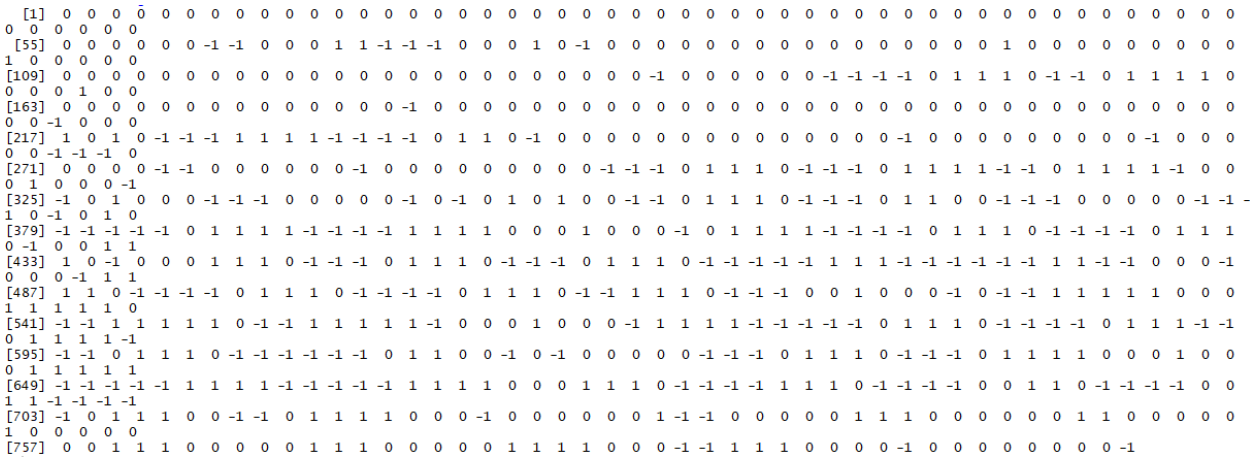


Figure 5.14: Signature of watermarked image

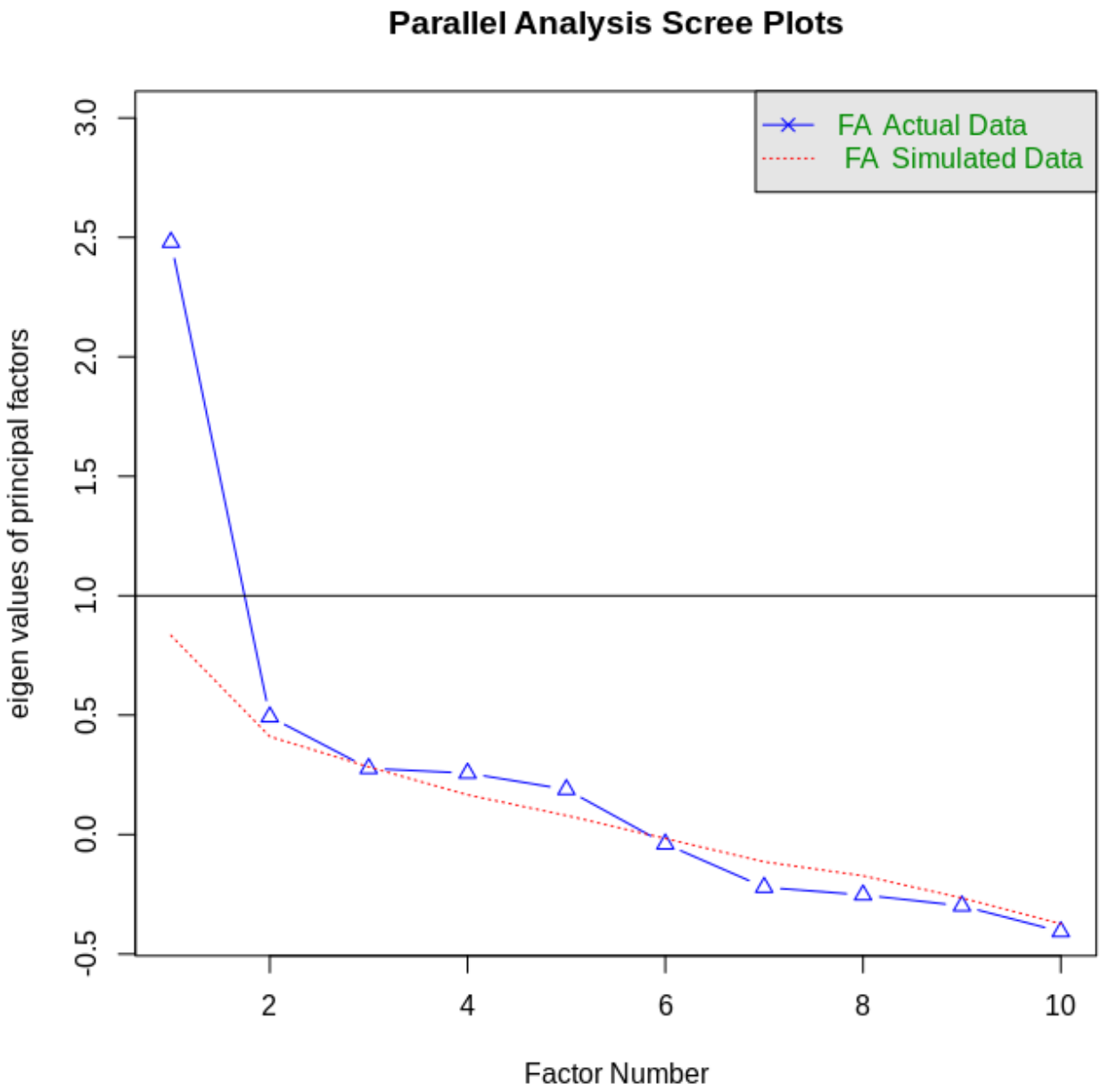


Figure 5.15: Parallel Analysis Scree Plots

```
Factor Analysis using method =  minres
Call: fa(r = ratings_cor, nfactors = 1)
Standardized loadings (pattern matrix) based upon correlation matrix
              MR1      h2   u2 com
Category.1  -0.02 0.00027 1.00   1
Category.2   0.16 0.02454 0.98   1
Category.3   0.68 0.46025 0.54   1
Category.4   0.30 0.08942 0.91   1
Category.5   0.43 0.18654 0.81   1
Category.6   0.61 0.37424 0.63   1
Category.7   0.88 0.77276 0.23   1
Category.8  -0.13 0.01718 0.98   1
Category.9   0.05 0.00257 1.00   1
Category.10 -0.74 0.55225 0.45   1

                MR1
SS loadings    2.48
Proportion Var 0.25
```

Figure 5.16: Result of factor analysis

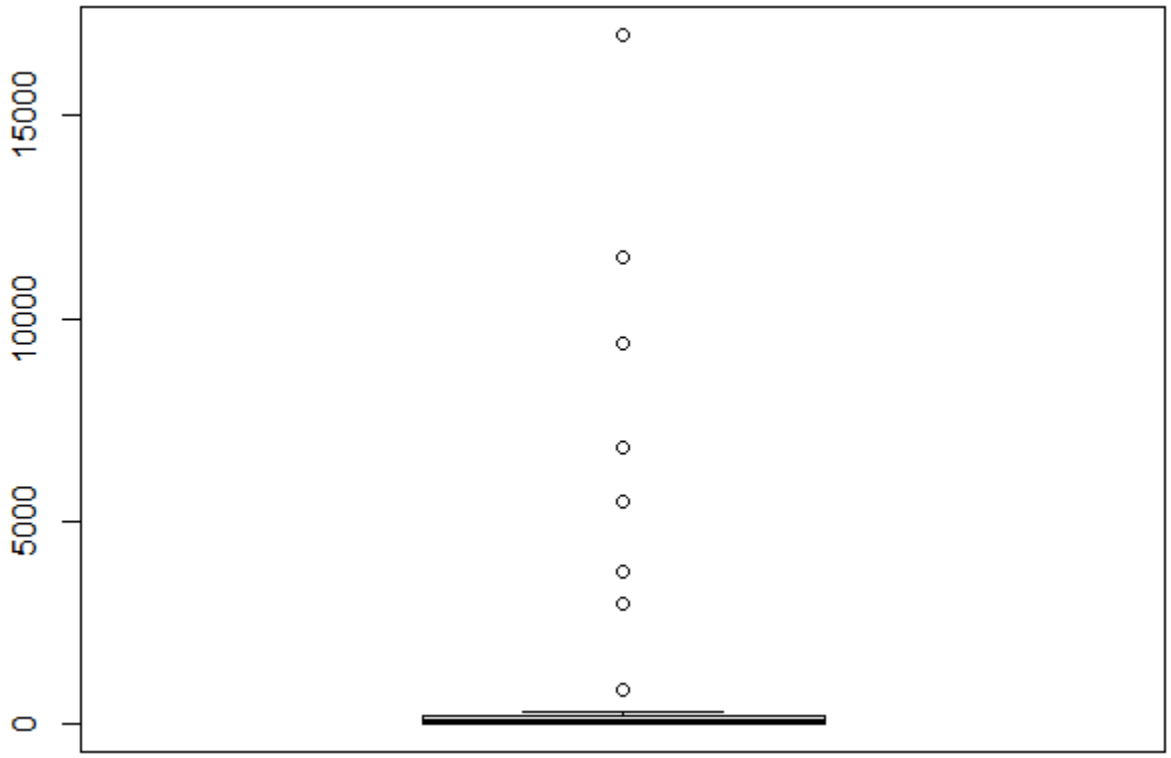


Figure 6.21: Boxplot of the islands dataset

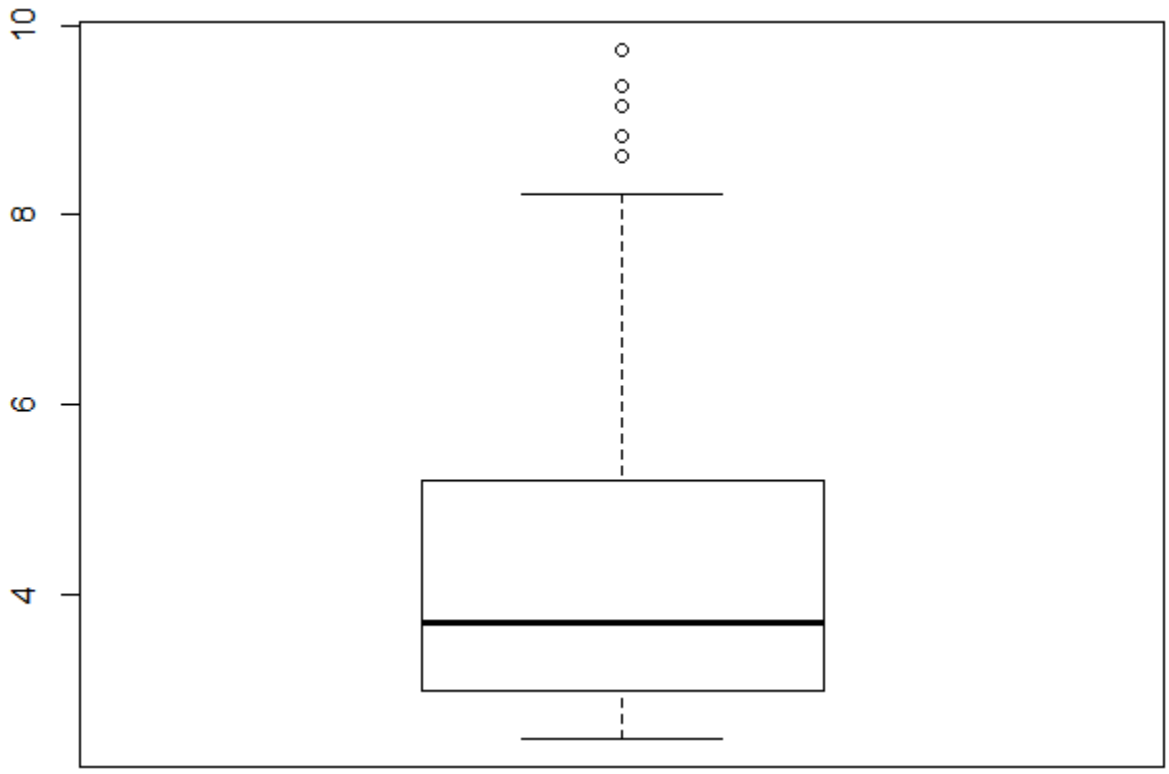


Figure 6.22: Boxplot of log-transformed dataset

Africa	Antarctica	Asia	Australia	Europe	Greenland	North America	South America
11506	5500	16988	2968	3745	840	9390	6795

Figure 6.23: Non-transformed outliers

Africa	Antarctica	Asia	North America	South America
11506	5500	16988	9390	6795

Figure 6.24: Log-transformed outliers

Africa	Asia	North America
11506	16988	9390

Figure 6.25: Screenshot of the outliers

```
  Africa     Antarctica           Asia North America South America
9.350624       8.612503       9.740262       9.147401       8.823942
```

Figure 6.26: Log-transformed outliers

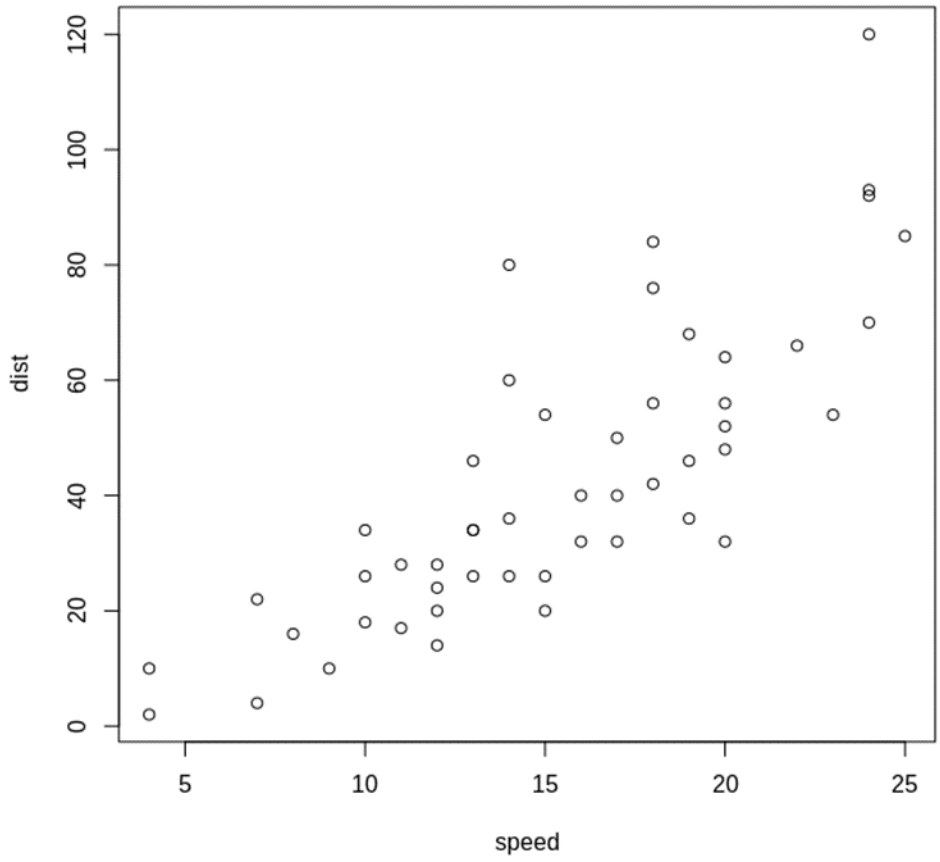


Figure 6.27: Plot of the cars dataset

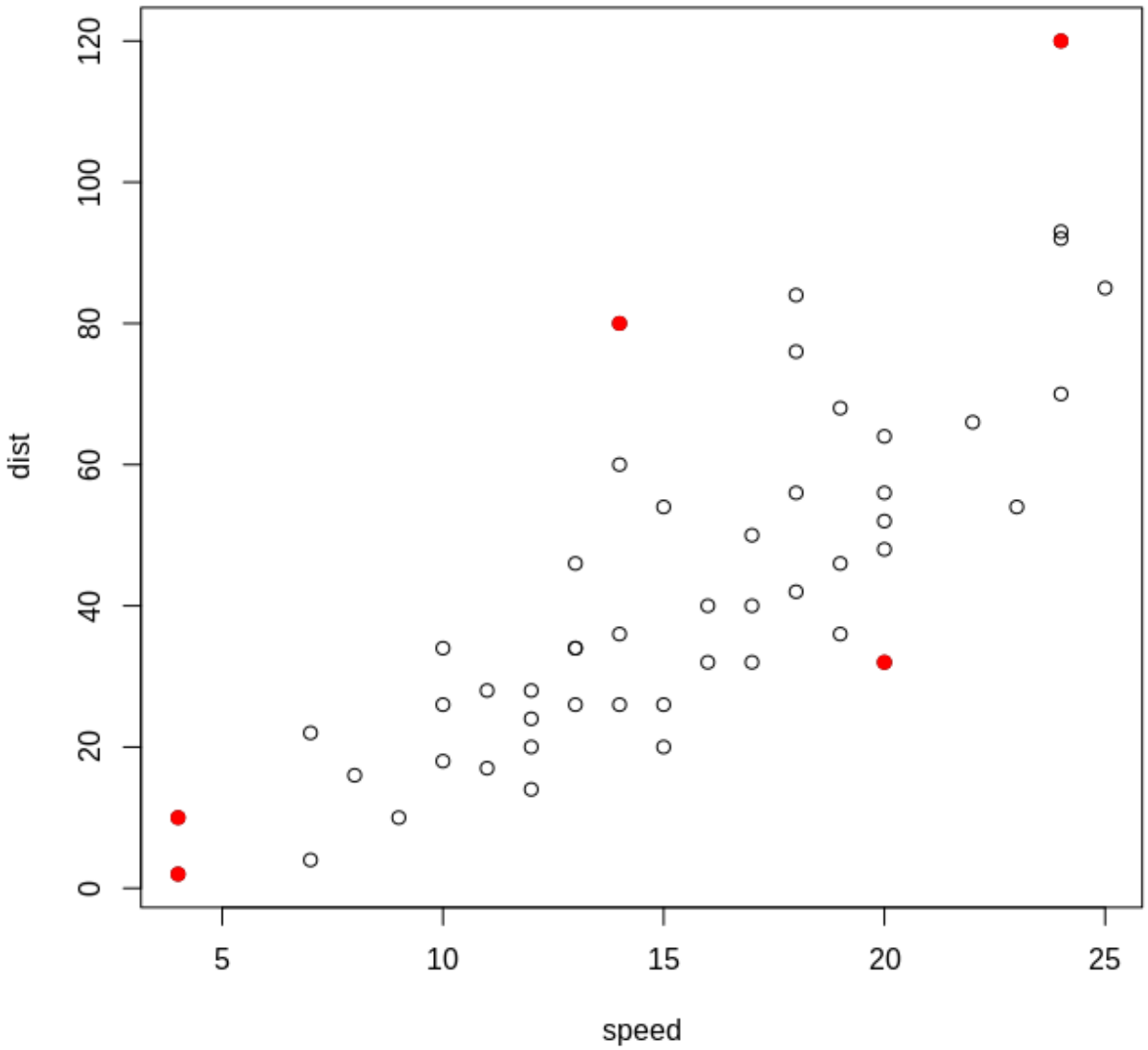


Figure 6.28: Plot with outliers marked