**HÜ2**

1. Der Datensatz „swiss“ beschreibt die standardisierte Fruchtbarkeitsmessung und sozioökonomische Indikatoren für jede der 47 französischsprachigen Provinzen der Schweiz um 1888. Aus dem Datensatz sind 47 Beobachtungen in Prozent aus 6 Variablen angegeben. Hier werden folgende 5 Variablen analysiert: Fertility (standardisiertes Fertilitätsmaß), Agriculture (Prozent der Männer, die in der Landwirtschaft als Beruf tätig sind), Eduction (Prozent der Wehrpflichtigen mit dem Grundschulabschluss), Catholic (Prozent der Bevölkerung, die katholisch sind) und Infant.Mortality (Prozent der Geburten, die weniger als ein Jahr alt sind).

Fertility Agriculture Education Catholic Infant.Mortality

1 80.2 17.0 12 9.96 22.2

2 83.1 45.1 9 84.84 22.2

3 92.5 39.7 5 93.40 20.2

4 85.8 36.5 7 33.77 20.3

5 76.9 43.5 15 5.16 20.6

6 76.1 35.3 7 90.57 26.6

7 83.8 70.2 7 92.85 23.6

8 92.4 67.8 8 97.16 24.9

9 82.4 53.3 7 97.67 21.0

10 82.9 45.2 13 91.38 24.4

11 87.1 64.5 6 98.61 24.5

12 64.1 62.0 12 8.52 16.5

13 66.9 67.5 7 2.27 19.1

14 68.9 60.7 12 4.43 22.7

15 61.7 69.3 5 2.82 18.7

16 68.3 72.6 2 24.20 21.2

17 71.7 34.0 8 3.30 20.0

18 55.7 19.4 28 12.11 20.2

19 54.3 15.2 20 2.15 10.8

20 65.1 73.0 9 2.84 20.0

21 65.5 59.8 10 5.23 18.0

22 65.0 55.1 3 4.52 22.4

23 56.6 50.9 12 15.14 16.7

24 57.4 54.1 6 4.20 15.3

25 72.5 71.2 1 2.40 21.0

26 74.2 58.1 8 5.23 23.8

27 72.0 63.5 3 2.56 18.0

28 60.5 60.8 10 7.72 16.3

29 58.3 26.8 19 18.46 20.9

30 65.4 49.5 8 6.10 22.5

31 75.5 85.9 2 99.71 15.1

32 69.3 84.9 6 99.68 19.8

33 77.3 89.7 2 100.00 18.3

34 70.5 78.2 6 98.96 19.4

35 79.4 64.9 3 98.22 20.2

36 65.0 75.9 9 99.06 17.8

37 92.2 84.6 3 99.46 16.3

38 79.3 63.1 13 96.83 18.1

39 70.4 38.4 12 5.62 20.3

40 65.7 7.7 11 13.79 20.5

41 72.7 16.7 13 11.22 18.9

42 64.4 17.6 32 16.92 23.0

43 77.6 37.6 7 4.97 20.0

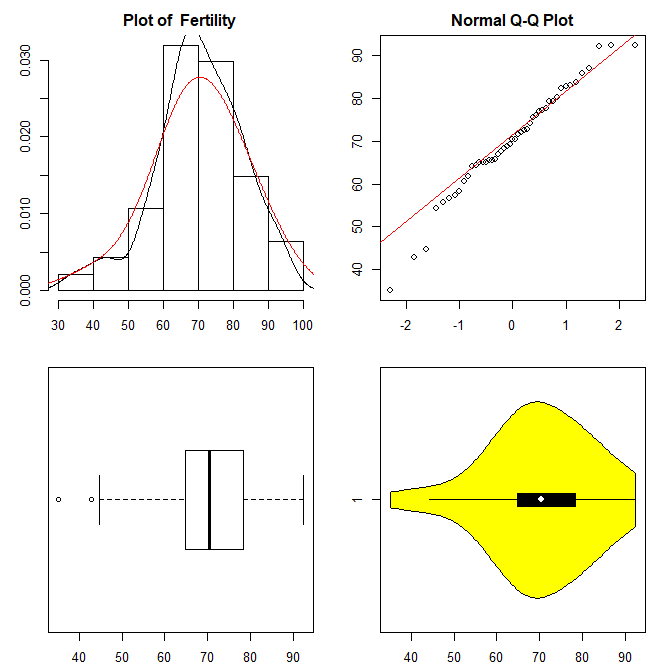
44 67.6 18.7 7 8.65 19.5

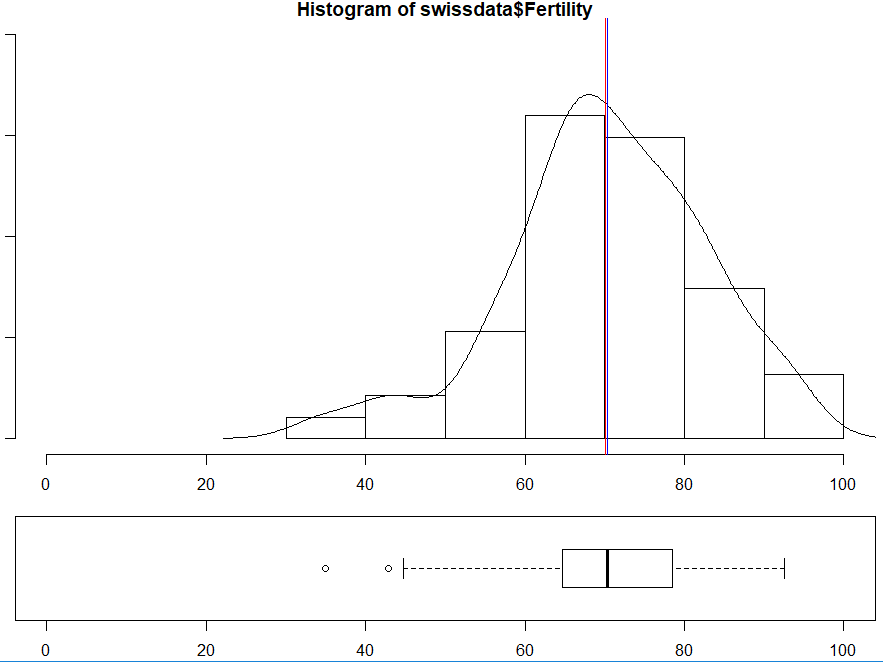
45 35.0 1.2 53 42.34 18.0

46 44.7 46.6 29 50.43 18.2

47 42.8 27.7 29 58.33 19.3

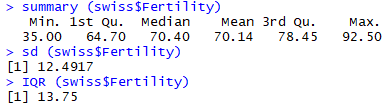
#Stadt statt Zahl?





Überschrift, x & y-Achse beschriften?

Die Kombination aus dem Histogramm und Dichtefunktion (schwarze Dichtelinie) veranschaulicht, dass die Daten unimodal sind. Wenn die Daten unimodal sind, dann darf nur ein Gipfel/eine Spitze zusehen sein (was hier der Fall auch ist). Um genauer zu sein, es handelt sich hier um eine symmetrische sprich annähernde normale Verteilung. Dies bestätigt die Tatsache, dass aus dem Daten der Mittelwert (roter Strich) und Median (blauer Strich) kaum voneinander abweichen.

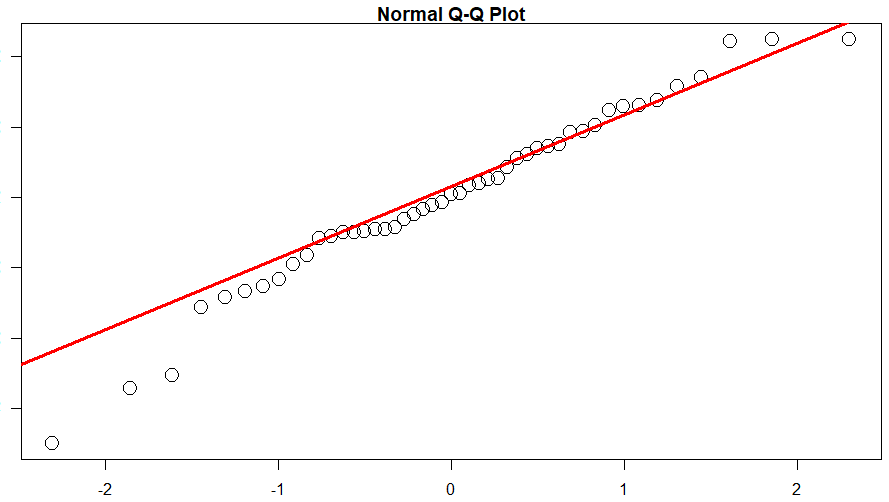


All diese Hinweise machen den Mittelwert und die Standardabweichung zu robusten Lageschätzern.

Der Boxplot bietet eine andere optische Darstellung der Daten.

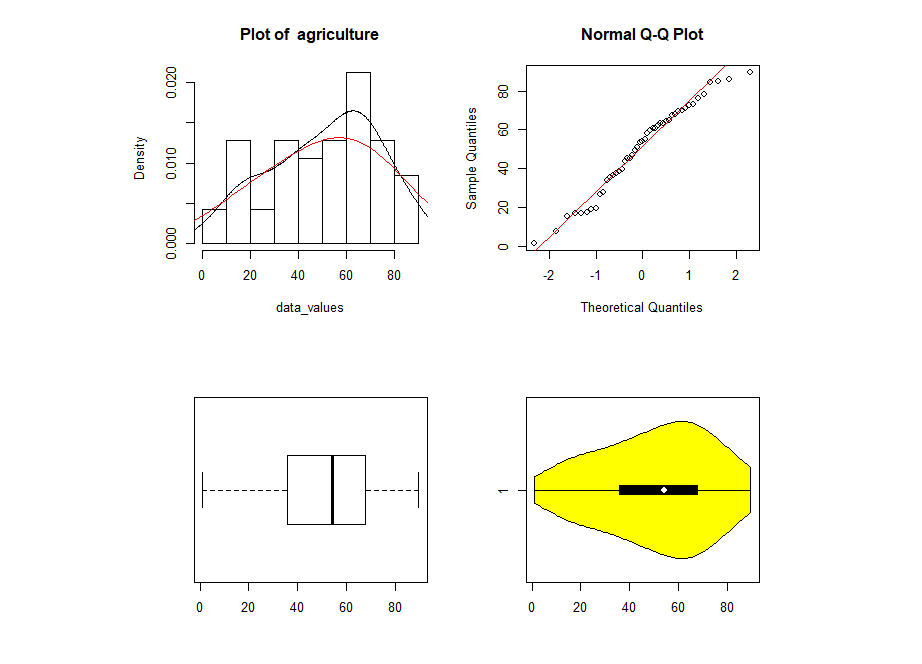
(Aus diesem Boxplot sieht man zwei red card outliers. Da es nur zwei von 47 Werten, somit weniger als <10%, in extreme Zone aufzufinden sind, hat das keine Auswirkung für die Normalverteilung. )

Darüberhinaus, lässt sich auch dem Boxplot eine Zusammenfassung der Daten in Quantilen (Minimum, 25%-Quantil, Median, 75%-Quantil und Maximum) erkennen.

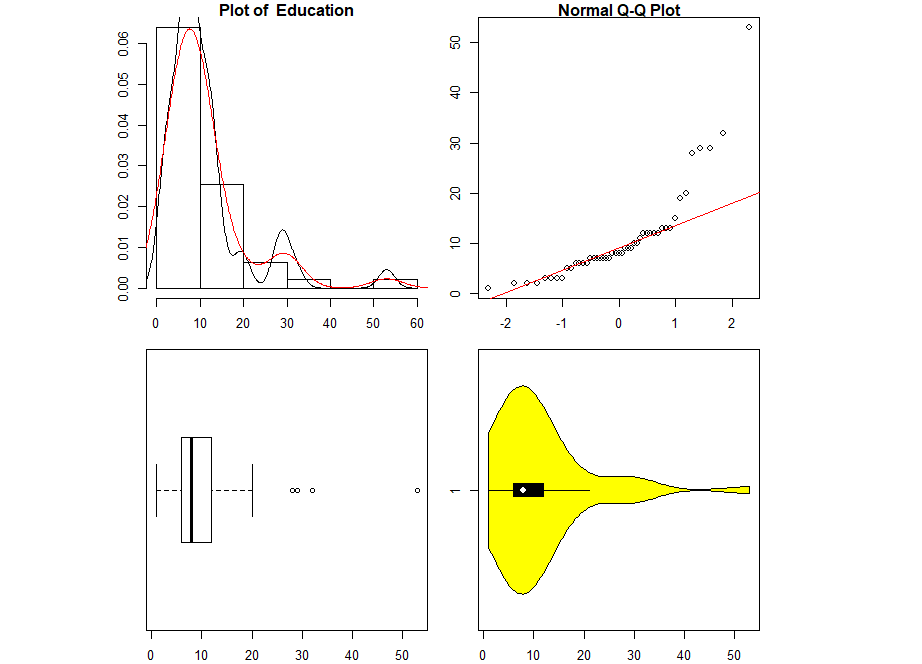


Überschrift, x & y-Achse beschriften?

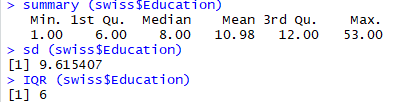
Der Normal Q-Q Plot lässt erkennen, dass sowohl unten als auch oben schwere Ränder gibt. Nichtsdestotrotz, die andere Werte zeigen keine Tendenz auf (weder leichte als schwere) Ränder auf.



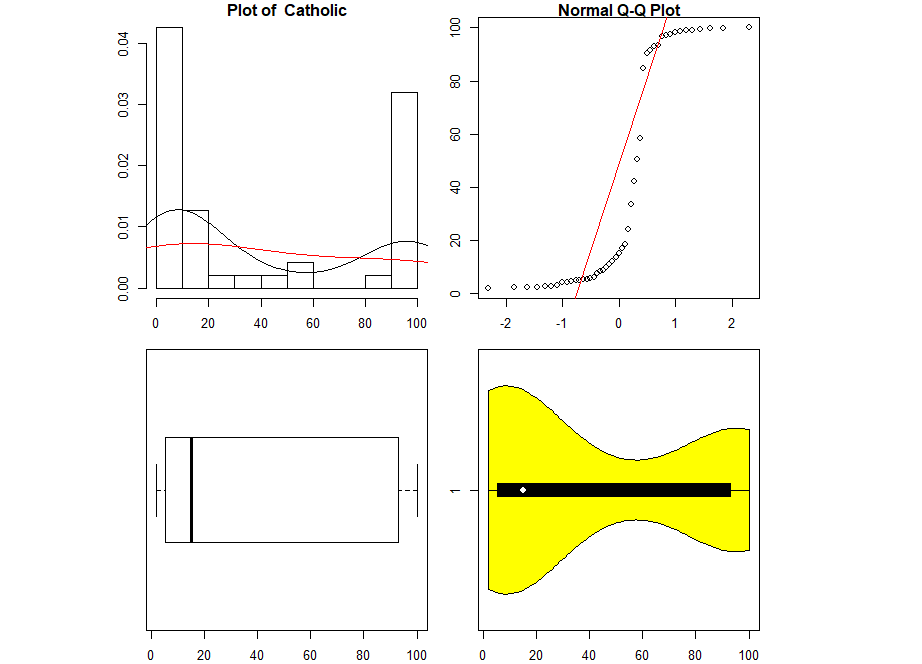
Die Kombination aus dem Histogramm und Dichtefunktion (schwarze Dichtelinie) veranschaulicht, dass die Daten bimodal sind. Wenn die Daten unimodal sind, dann darf nur ein Gipfel/eine Spitze zusehen sein. Hier ist keine eindeutige Spitze zu sehen und somit sind die Daten deutlich nicht normalverteilt. Diese Annahme wird auch durch den normalen QQ Plot veranschaulicht und bestätigt. Daraus lässt sich auch erkennen, dass die Daten anhand dem leicht „S-förmigen“ Verlauf bimodal sind. Aus diesem Grund machen die Lageschätzer (Mittelwert, Median, Boxplot) als Darstellung in diesem Fall keinen Sinn.



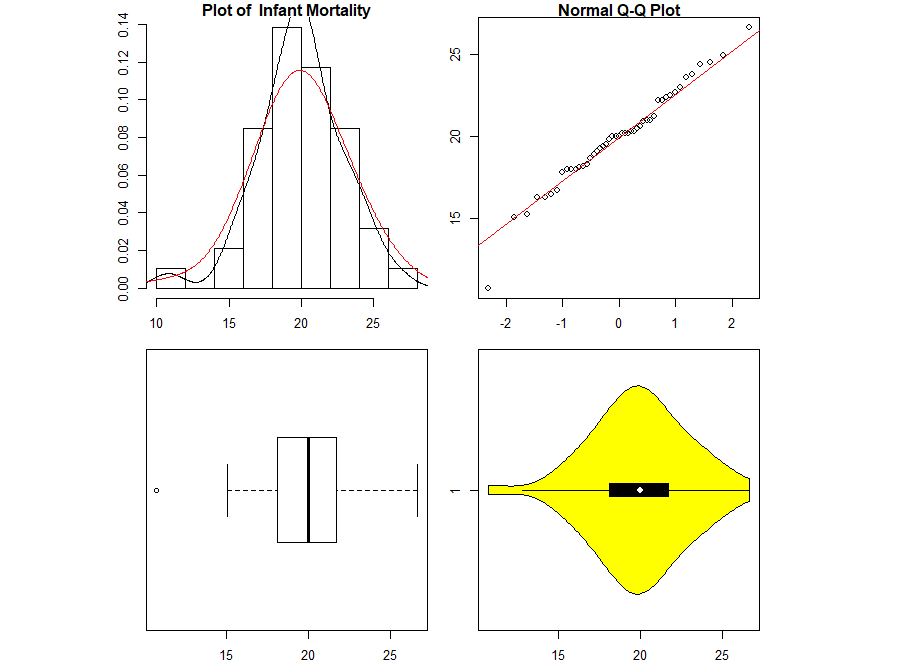
Die Kombination aus dem Histogramm und Dichtefunktion (schwarze Dichtelinie) veranschaulicht, dass die Daten unimodal sind. Wenn die Daten unimodal sind, dann darf nur ein Gipfel/eine Spitze zu sehen sein. In diesem Fall handelt sich um eine rechts schiefe (auf dem Graph markieren) Verteilung, was man auch eindeutig aus dem normalen QQ-Plot erkennt, und somit keine Normalverteilung.



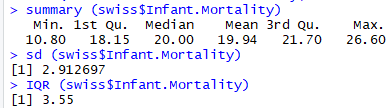
Darüberhinaus, zeigt der Boxplot drei yellow-card outlier und zwei red card outlier an. Die Klassifizierung der Outlier erfolgte über folgende Berechnung (nach oben): Yellow card outlier: 3rd Quantile +1,5 \*IQR= 12+1,5\*6=21. Das heißt alle Werte die ≥ 21 sind gehören zu yellow card outlier-Gruppe. Red card outlier: 3rd Quantile +3 \*IQR= 12+3\*6=30. Das bedeutet alle Werte ≥ 30 sind, sind red card outlier.



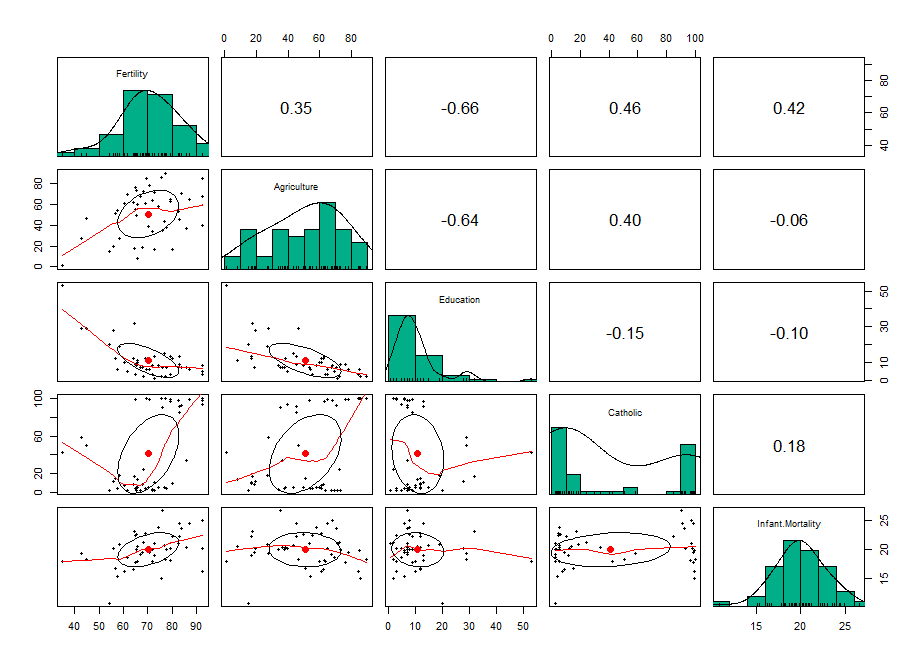
Aus allen drei Darstellungen (Histo, Boxplot, normal QQ-Plot) ist es eindeutig ersichtlich, dass es hier um keine Normalverteilung, sondern um die Bimodalität handelt. Darüberhinaus, lässt sich aus dem normalen QQ-Plot erkennen, dass die Daten sowohl links als auch rechts leichte Ränder haben. Lageschätzer (Mittelwert, Median, Boxplot) als Darstellung in diesem Fall keinen Sinn.



Die Kombination aus dem Histogramm und Dichtefunktion (schwarze Dichtelinie) veranschaulicht, dass die Daten unimodal sind. Wenn die Daten unimodal sind, dann darf nur ein Gipfel/eine Spitze zusehen sein (was hier der Fall auch ist). Um genauer zu sein, es handelt sich hier um eine symmetrische sprich schöne normale Verteilung. Aus dem QQ-Plot sind keine leichten oder schweren Rändern erkennbar. Dies bestätigt die Tatsache, dass aus dem Daten der Mittelwert (roter Strich) und Median (blauer Strich) kaum voneinander abweichen. All diese Hinweise machen den Mittelwert und die Standardabweichung zu robusten Lageschätzern.



Darüberhinaus, zeigt der Boxplot einen red card outlier an. Die Klassifizierung der Outlier erfolgte über folgende Berechnung (nach unten): Yellow card outlier: 1st Quantile -1,5 \*IQR= 18,15-1,5\*3,55=12,825. Das heißt alle Werte die ≤ 12,825 sind gehören zu yellow card outlier-Gruppe. Red card outlier: 1st Quantile -1,5 \*IQR= 18,15-3\*3,55=7,5. Das bedeutet alle Werte ≤ 7,5 sind, sind red card outlier.



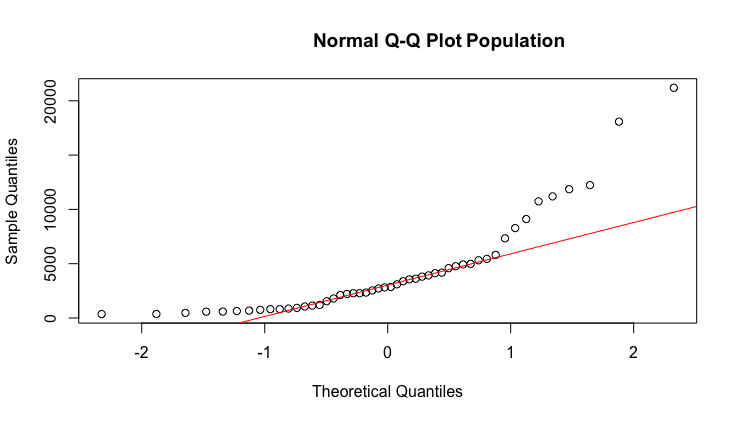
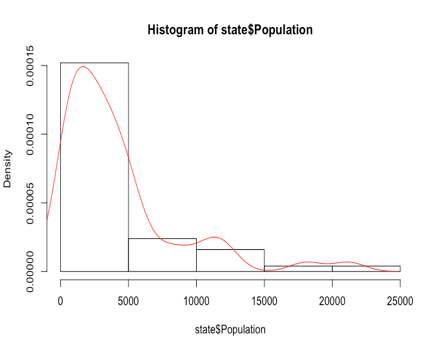
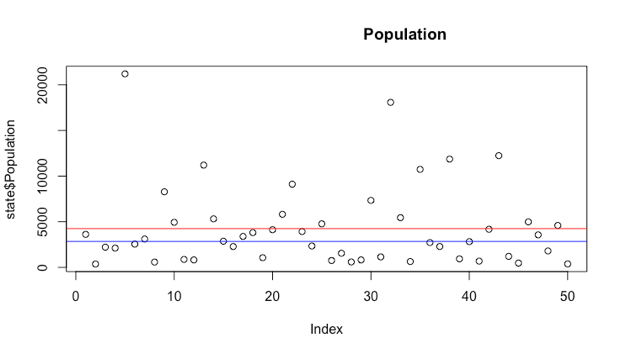
Die Zusammenhänge der einzelnen Datensätzen (Variablen) über einen Scatterplot-matrix bzw. Korrelation feststellen. Je höher die Korrelation umso besser korrelieren die Daten miteinander.

In diesem Fall ist die Korrelation zwischen Fertility & Education; Agriculture & Education am niedrigsten. Das heißt, je höher die Ausbildung der Menschen ist, umso niedriger ist die Fertilitätsrate und umso weniger sind sie Landwirtschaft tätig. Weiters ist es auch ersichtlich, dass es einen Zusammenhang zwischen Fertility-Catholic; Agriculture-Catholic und Fertility-Agriculture gibt. Daraus lässt sich schließen, dass es mehr Kinder gibt in den katholischen Familien, die in der Landwirtschaft tätig sind. Korrelationsfaktor zwischen Fertility-Infanten.Mortality zeigt aber auch dass es im 1887 sehr hohe Sterblichkeitsrate von Neugeborenen (≤ 1 Jahr) gab.

Bsp 2:

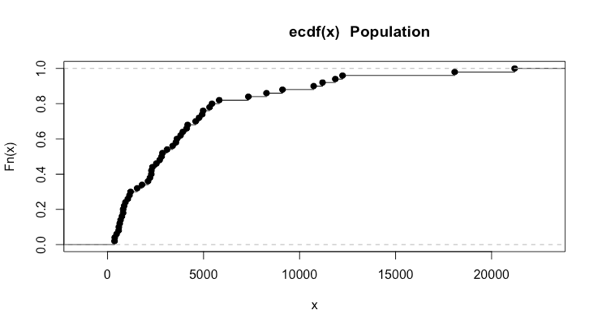
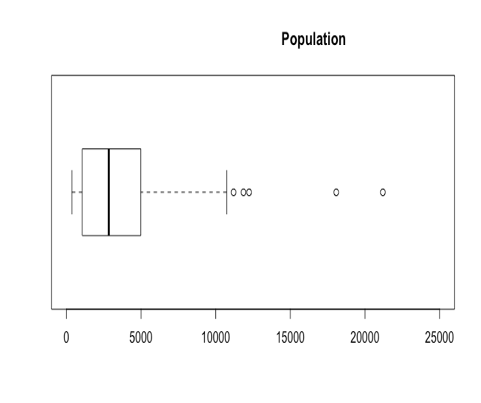
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Population** | **Income** | **Illiteracy** | **Life Exp** | **Murder** | **HS Grad** | **Frost** | **Area** |
| **Alabama** | 3615 | 3624 | 43467 | 69.05 | 43480 | 41.3 | 20 | 50708 |
| **Alaska** | 365 | 6315 | 43586 | 69.31 | 43535 | 66.7 | 152 | 566432 |
| **Arizona** | 2212 | 4530 | 43678 | 70.55 | 43684 | 58.1 | 15 | 113417 |
| **Arkansas** | 2110 | 3378 | 43709 | 70.66 | 43475 | 39.9 | 65 | 51945 |
| **California** | 21198 | 5114 | 43466 | 71.71 | 43534 | 62.6 | 20 | 156361 |
| **Colorado** | 2541 | 4884 | 0.7 | 72.06 | 43683 | 63.9 | 166 | 103766 |
| **Connecticut** | 3100 | 5348 | 43466 | 72.48 | 43468 | 56 | 139 | 4862 |
| **Delaware** | 579 | 4809 | 0.9 | 70.06 | 43502 | 54.6 | 103 | 1982 |
| **Florida** | 8277 | 4815 | 43525 | 70.66 | 43656 | 52.6 | 11 | 54090 |
| **Georgia** | 4931 | 4091 | 2 | 68.54 | 43721 | 40.6 | 60 | 58073 |
| **Hawaii** | 868 | 4963 | 43709 | 73.6 | 43502 | 61.9 | 0 | 6425 |
| **Idaho** | 813 | 4119 | 0.6 | 71.87 | 43529 | 59.5 | 126 | 82677 |
| **Illinois** | 11197 | 5107 | 0.9 | 70.14 | 43534 | 52.6 | 127 | 55748 |
| **Indiana** | 5313 | 4458 | 0.7 | 70.88 | 43472 | 52.9 | 122 | 36097 |
| **Iowa** | 2861 | 4628 | 0.5 | 72.56 | 43526 | 59 | 140 | 55941 |
| **Kansas** | 2280 | 4669 | 0.6 | 72.58 | 43589 | 59.9 | 114 | 81787 |
| **Kentucky** | 3387 | 3712 | 43617 | 70.1 | 43626 | 38.5 | 95 | 39650 |
| **Louisiana** | 3806 | 3545 | 43679 | 68.76 | 43509 | 42.2 | 12 | 44930 |
| **Maine** | 1058 | 3694 | 0.7 | 70.39 | 43648 | 54.7 | 161 | 30920 |
| **Maryland** | 4122 | 5299 | 0.9 | 70.22 | 43593 | 52.3 | 101 | 9891 |
| **Massachusetts** | 5814 | 4755 | 43466 | 71.83 | 43527 | 58.5 | 103 | 7826 |
| **Michigan** | 9111 | 4751 | 0.9 | 70.63 | 43476 | 52.8 | 125 | 56817 |
| **Minnesota** | 3921 | 4675 | 0.6 | 72.96 | 43526 | 57.6 | 160 | 79289 |
| **Mississippi** | 2341 | 3098 | 43557 | 68.09 | 43597 | 41 | 50 | 47296 |
| **Missouri** | 4767 | 4254 | 0.8 | 70.69 | 43533 | 48.8 | 108 | 68995 |
| **Montana** | 746 | 4347 | 0.6 | 70.56 | 5 | 59.2 | 155 | 145587 |
| **Nebraska** | 1544 | 4508 | 0.6 | 72.6 | 43710 | 59.3 | 139 | 76483 |
| **Nevada** | 590 | 5149 | 0.5 | 69.03 | 43596 | 65.2 | 188 | 109889 |
| **New Hampshire** | 812 | 4281 | 0.7 | 71.23 | 43527 | 57.6 | 174 | 9027 |
| **New Jersey** | 7333 | 5237 | 43466 | 70.93 | 43501 | 52.5 | 115 | 7521 |
| **New Mexico** | 1144 | 3601 | 43498 | 70.32 | 43655 | 55.2 | 120 | 121412 |
| **New York** | 18076 | 4903 | 43556 | 70.55 | 43718 | 52.7 | 82 | 47831 |
| **North Carolina** | 5441 | 3875 | 43678 | 69.21 | 43476 | 38.5 | 80 | 48798 |
| **North Dakota** | 637 | 5087 | 0.8 | 72.78 | 43556 | 50.3 | 186 | 69273 |
| **Ohio** | 10735 | 4561 | 0.8 | 70.82 | 43562 | 53.2 | 124 | 40975 |
| **Oklahoma** | 2715 | 3983 | 43466 | 71.42 | 43561 | 51.6 | 82 | 68782 |
| **Oregon** | 2284 | 4660 | 0.6 | 72.13 | 43500 | 60 | 44 | 96184 |
| **Pennsylvania** | 11860 | 4449 | 1 | 70.43 | 43471 | 50.2 | 126 | 44966 |
| **Rhode Island** | 931 | 4558 | 43525 | 71.9 | 43557 | 46.4 | 127 | 1049 |
| **South Carolina** | 2816 | 3635 | 43526 | 67.96 | 43627 | 37.8 | 65 | 30225 |
| **South Dakota** | 681 | 4167 | 0.5 | 72.08 | 43647 | 53.3 | 172 | 75955 |
| **Tennessee** | 4173 | 3821 | 43647 | 70.11 | 11 | 41.8 | 70 | 41328 |
| **Texas** | 12237 | 4188 | 43498 | 70.9 | 43508 | 47.4 | 35 | 262134 |
| **Utah** | 1203 | 4022 | 0.6 | 72.9 | 43589 | 67.3 | 137 | 82096 |
| **Vermont** | 472 | 3907 | 0.6 | 71.64 | 43590 | 57.1 | 168 | 9267 |
| **Virginia** | 4981 | 4701 | 43556 | 70.08 | 43594 | 47.8 | 85 | 39780 |
| **Washington** | 3559 | 4864 | 0.6 | 71.72 | 43528 | 63.5 | 32 | 66570 |
| **West Virginia** | 1799 | 3617 | 43556 | 69.48 | 43652 | 41.6 | 100 | 24070 |
| **Wisconsin** | 4589 | 4468 | 0.7 | 72.48 | 3 | 54.5 | 149 | 54464 |
| **Wyoming** | 376 | 4566 | 0.6 | 70.29 | 43714 | 62.9 | 173 | 97203 |

Datensatz state.x77

Das Histogramm zeigt eine unimodale Verteilung der Bevölkerung, wenngleich auch ein schwerer Rand zu erkennen ist. Dieser wird zusätzlich durch QQPlot das veranschaulicht.

QQPlot: rechts schwerer Rand, links leichter Rand. Hierdurch ist keine Normalverteilung der Daten gegeben. Die Verteilung der Daten wird durch das unten gezeigt Boxplot veranschaulicht.



Boxplot:  
yellow card outlier: 3rd Quartile + 1,5\*IQR: 4968+ 1,5\*3889= 10801,5

Alle Werte ≥10801,5 sind yellow card Ausreißer

Red card outlier: 3rd Quartile + 3\*IQR= 4968+ 3\*3889= 10801,5=16635

Alle Werte ≥16635 sind yellow card Ausreißer

**ECDF: Veranschaulicht die Bevölkerungsverteilung. Die meisten**

Lageparameter:

> summary(state$Population)

Min. 1st Qu. Median Mean 3rd Qu. Max.

365 1080 2838 4246 4968 2119

Streuungsparameter:

Standardabweichung

> sd(state$Population)

[1] 4464.491

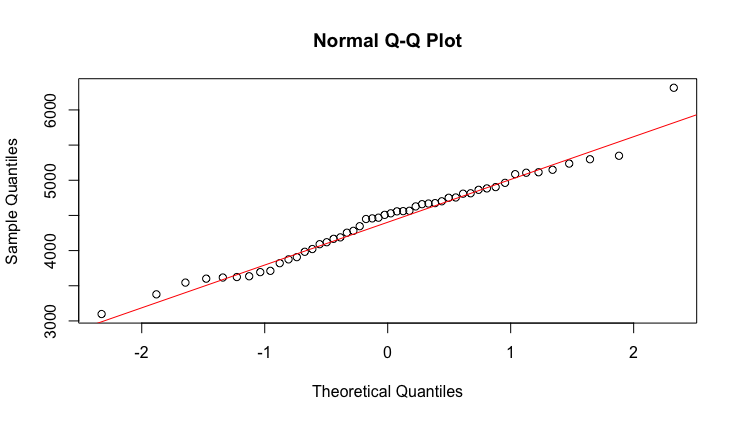
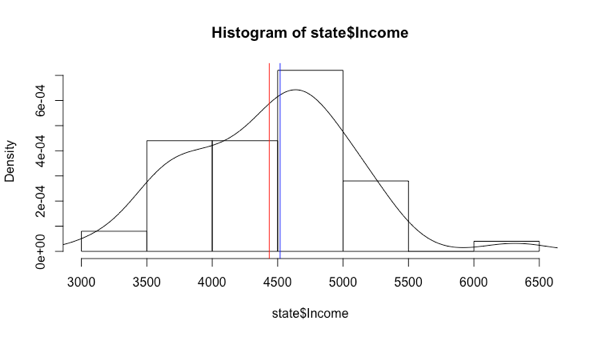
Interquartilsrange

> IQR(state$Population)

[1] 3889

Die Zusammenfassung bestätigt eine Rechtsverschiebung der Daten. Die Besiedelung der Staaten fällt zumeist dünn aus (siehe Scatterplot).

Datensatz Income



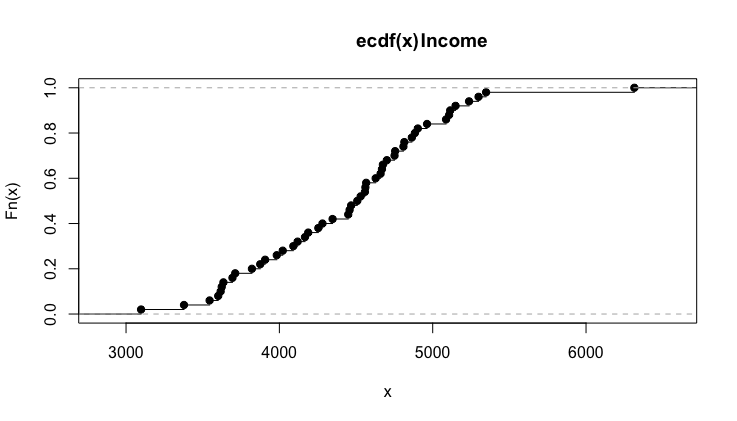
Das Histogramm für „Income“ zeigt eine klare unimodale Verteilung. Mean und Median liegen eng beieinander. Die Daten wirken annähernd normalverteilt, es liegen nur leichte Ränder vor (siehe QQPlot).

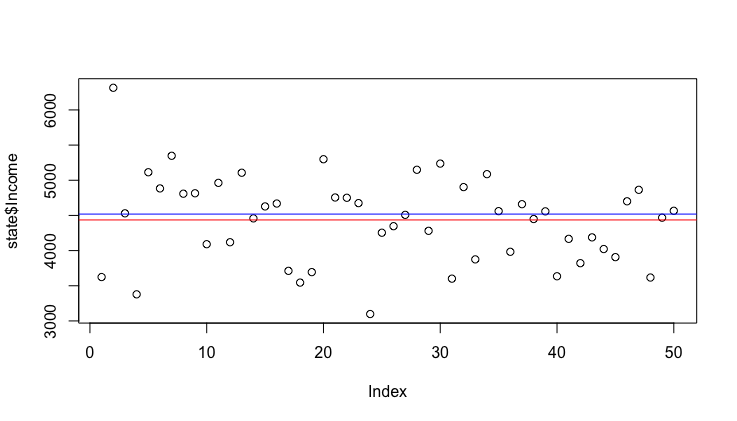
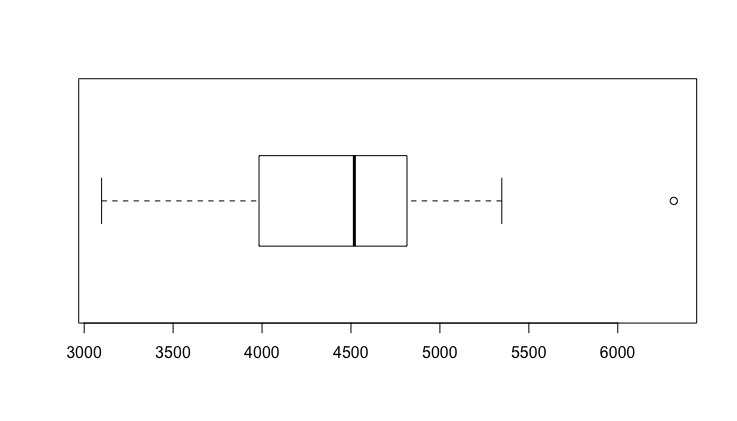
Zusammenfassung der Daten:

> summary(state$Income)

Min. 1st Qu. Median Mean 3rd Qu. Max.

3098 3993 4519 4436 4814 6315





Am Boxplot ist ersichtlich, dass es nur einen Datenpunkt gibt, der sich außerhalb der Whisker befindet. Da es sicher hierbei nur um einen von 50 Punkten handelt, hat dieser keinerlei Auswirkungen auf die Normalverteilung.

> sd(state$Income)

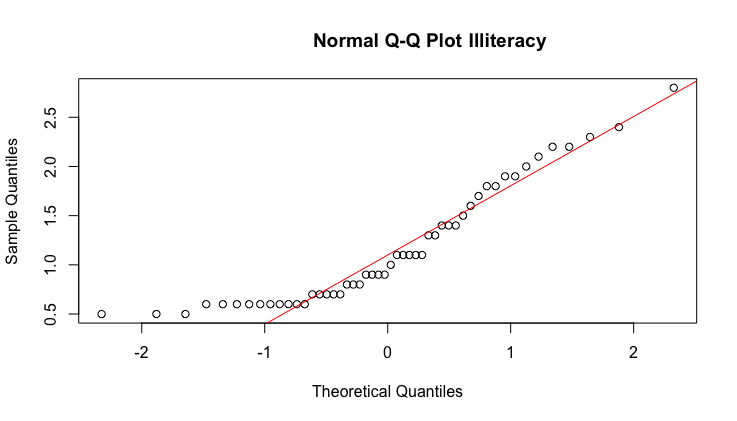
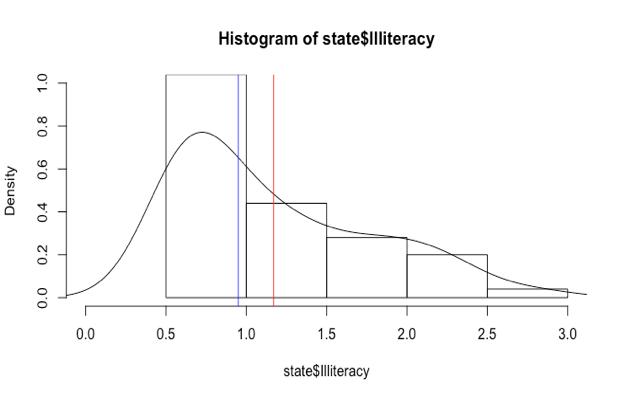
[1] 614.4699

> IQR(state$Income)

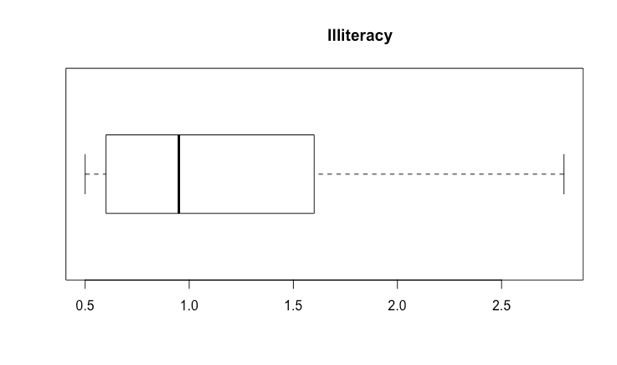
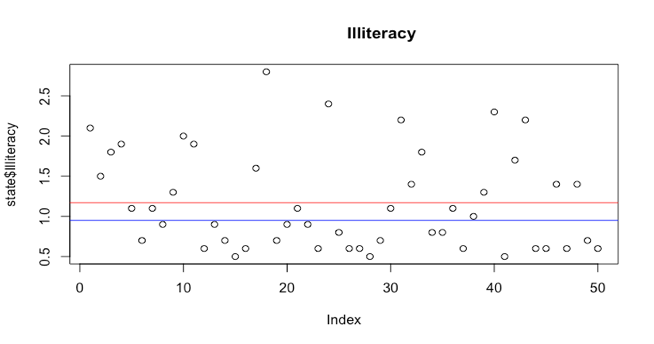
[1] 820.75

Aus dem Scatterplot ist zu erkennen, dass das Einkommen zumeist etwas über dem Mittel liegt. Die Bevölkerungsgruppe kann daher als Mittelstand bezeichnet werden. Das Einkommen soiegelt sich auch im ECDF Plot, bei dem eine starke Gruppierung bis etwas 5000 gegeben ist.

Zusammenfassung der Daten: Illiteracy



Zwar liegt eine Unimodalität vor, jedoch ist auch eine deutliche Linksverschiebung zu erkennen. Es ist somit keine Normalverteilung gegeben.

> summary(state$Illiteracy)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.500 0.625 0.950 1.170 1.575 2.800

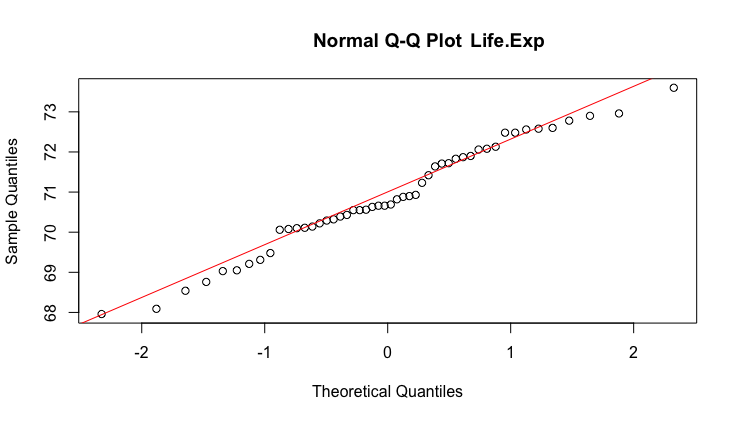
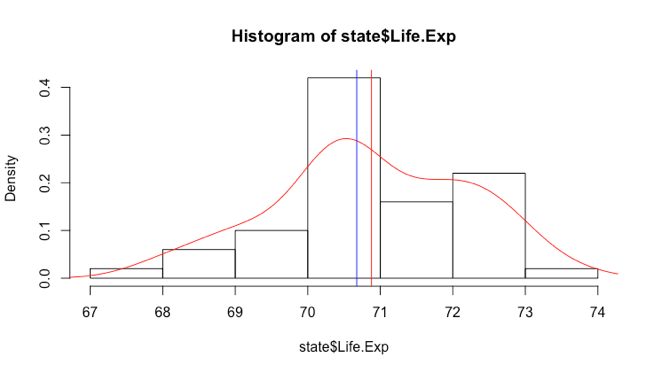
> sd(state$Illiteracy)

[1] 0.6095331

> IQR(state$Illiteracy)

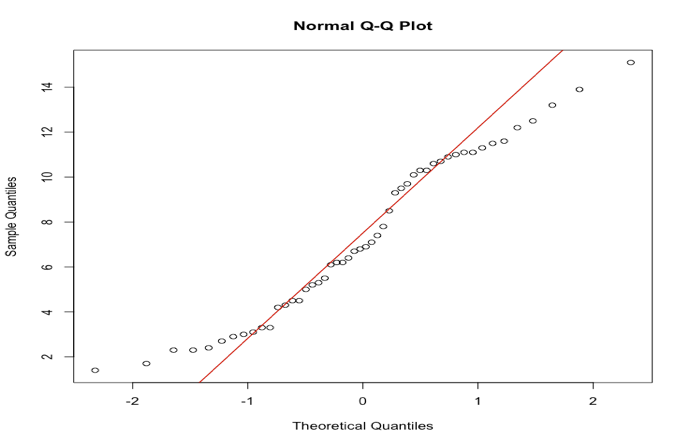
[1] 0.95

**Datensatz Life Expectacy**



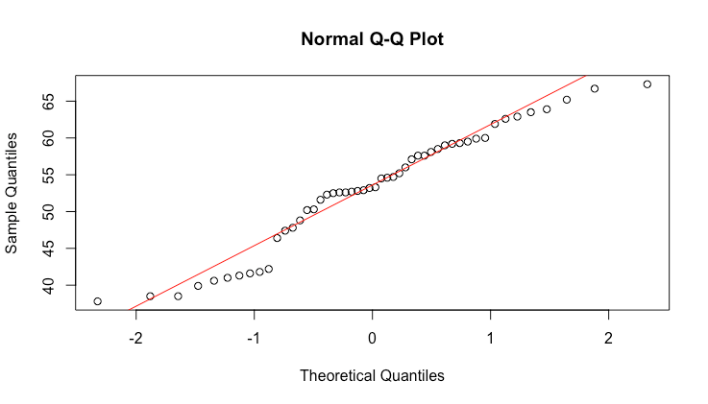
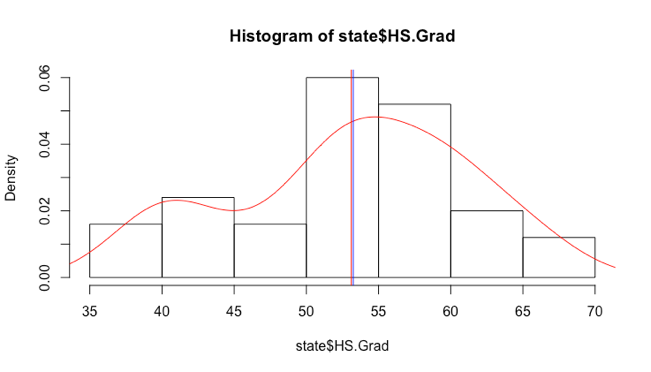
Das Histogramm lässt auf eine Bimodalität schließen. Diese bestätigt sich dann anhand des QQPlots. Die „S“Form ist nur schwach ausgeprägt. Sowohl links wie auch rechts liegen leichte Ränder vor. Es herrascht daher keine Symmetrie vor. Aufgrund der Bimodalität werden keine Lageschätzer (Mittelwert, Median, Boxplot) bestimmt.

Datensatz Murder

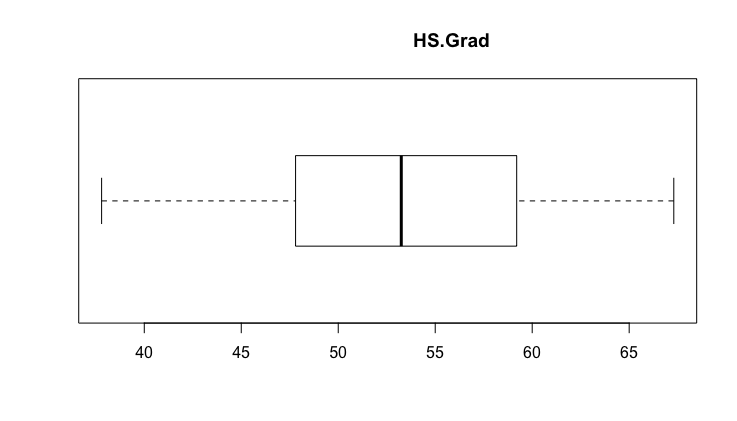


Anhand des Histogramms ist deutlich zu erkennen, dass hierbei eine Bimodalität und keine Normalverteilung und damit auch keine Symmetrie vorliegt. Auch das QQplot zeigt die hierfür charakteristische „S“ Form. Sowohl links wie auch rechts sind leichte Ränder zu erkennen. Da hierbei Bimodalität vorliegt, machen Lageschätzer (Mittelwert, Median oder Boxplot) keinen Sinn.

**Datensatz HS Grade**



Anhand des Histogramms ist zu erkennen, dass es sich um unimodale Daten handelt. Mean und Median liegen nahe beieinander Aus dem QQ Plot lassen sich keine schweren Ränder erkennen. Alle Punkte sind nahe der Linie.



Aus dem Boxplot lassen sich keine Outlier erkennen.

> summary(state$HS.Grad)

Min. 1st Qu. Median Mean 3rd Qu. Max.

37.80 48.05 53.25 53.11 59.15 67.30

> sd(state$HS.Grad)

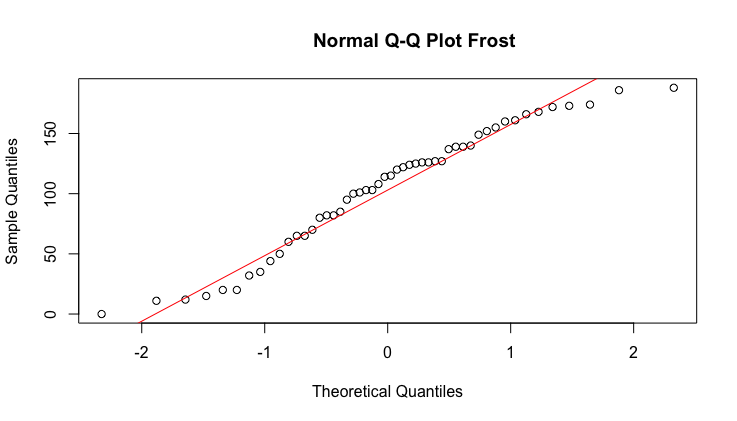
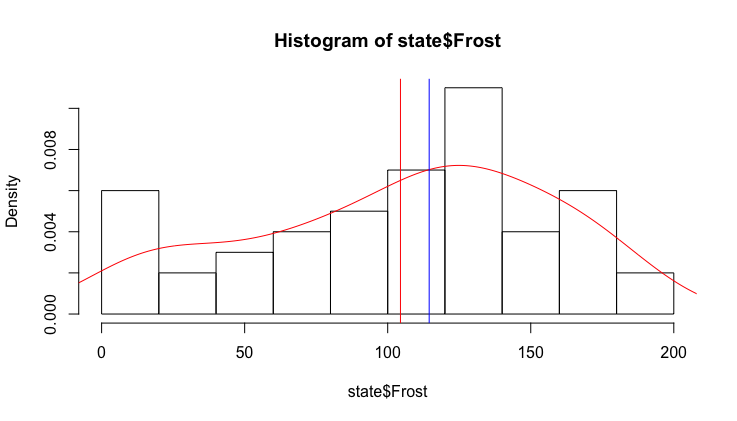
[1] 8.076998

> IQR(state$HS.Grad)

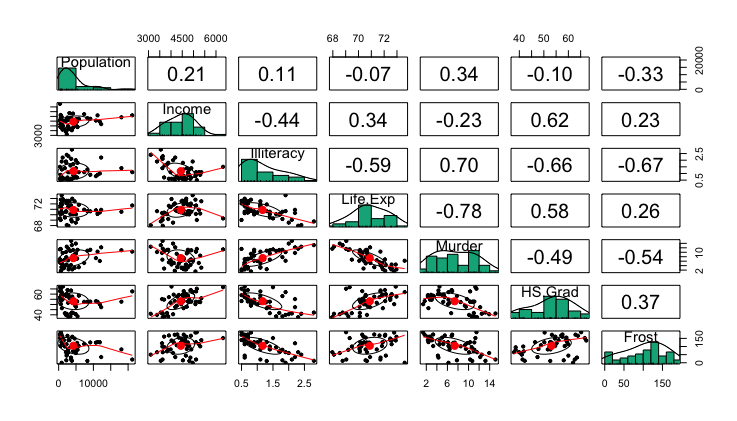
[1] 11.1

Wie auch schon aus dem Histogramm ersichtlich, liegen Median und Mean nahe zusammen. Dies bestätigt die Annahme, dass es sich hierbei um eine symmetrische Normalverteilung handelt.

Datensatz Frost



Sowohl das Histogramm wie auch der QQ Plot verdeutlichen, dass es sich hierbei um keine unimodale Daten handelt. Der leicht „S“ förmige Verlauf im QQ Plot deutet auf Bimodalität hin. Die Erläuterung von Lageschätzer (Mittelwert, Median, Boxplot) sind daher nicht sinnvoll.



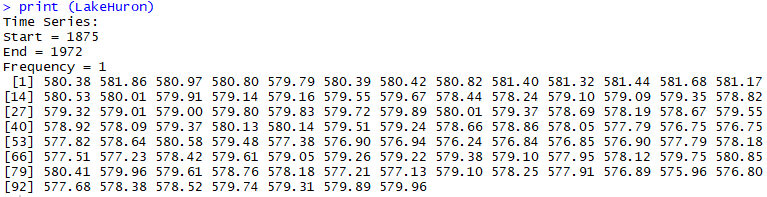
In dieser Grafik sind die Korrelationen der einzelnen Datensätze zueinander dargestellt. Die Wahrscheinlichkeiten auf der rechten Seite steigen mit höherer Korrelation.

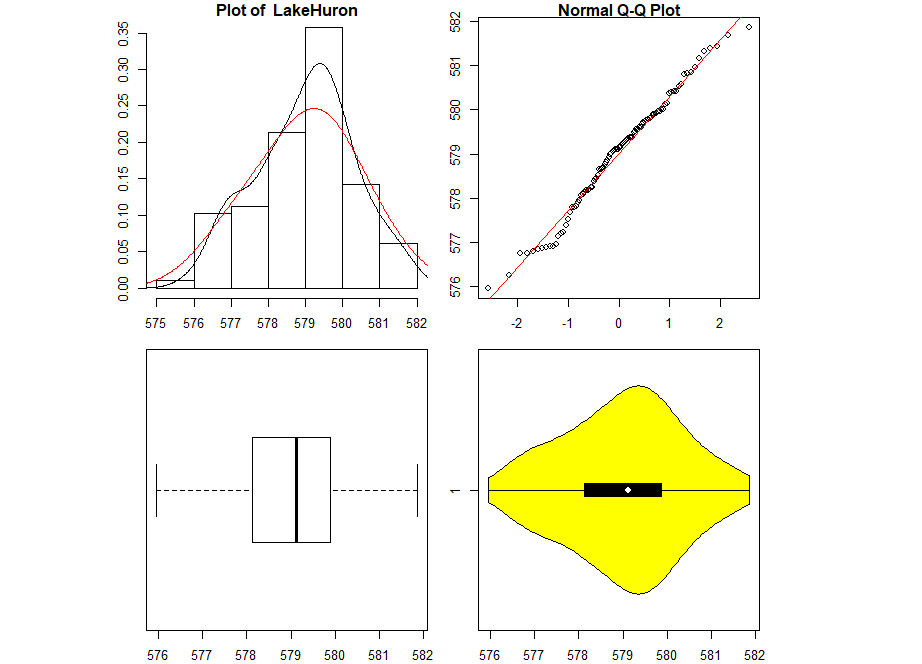
Die Korrelation zwischen Income und Illiteracy, zwischen Illiteracy und Life Exp, Life exp and Murder sowie zwischen Murder und HS Grad sind gering. Eine höhere Korrelation hingegen ist gegeben zwischen Population und Murder, Income und Life Exp, Income und HS Grad (starke Korrelation), Illiteracy und Murder (starke Korrelation) und zwischen Life Exp. und HS Grad.

Daraus lässt sich rückschließen, dass ein höheres Einkommen die durchschnittliche Lebenserwartung steigert, ein höheres Einkommen mit dem Bildungsgrad zusammenhängt und die Mordrate mit Analphabetismus in Verbidnung steht.

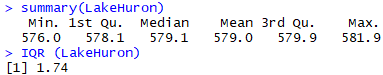
Bsp3:

Der Datensatz „Lake Huron“ enthält die Messungen über die Tiefe von 1875 bis 1972. Ein Messpunkt wurde einmal pro Jahr durchgeführt.





Die Kombination aus dem Histogramm und Dichtefunktion (schwarze Dichtelinie) veranschaulicht, dass die Daten unimodal sind. Wenn die Daten unimodal sind, dann darf nur ein Gipfel/eine Spitze zusehen sein (was hier der Fall auch ist). Um genauer zu sein, es handelt sich hier um eine symmetrische sprich schöne normale Verteilung. Aus dem QQ-Plot sind keine leichten oder schweren Rändern erkennbar. Dies bestätigt die Tatsache, dass aus dem Daten der Mittelwert (roter Strich) und Median (blauer Strich) kaum voneinander abweichen. Im Violin Plot ist es auch eindeutig ersichtlich, dass der Lagemaß sinnvoll ist und das heißt es handelt sich hier um robusten Lageabschätzer.

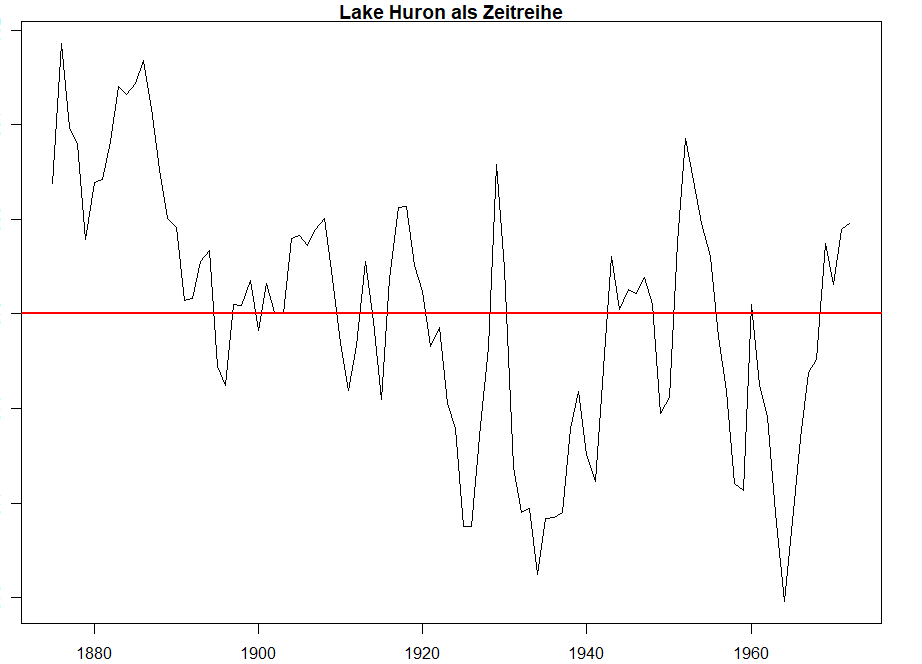




Der Verlauf der Messpunkte über die Zeit wird mithilfe von Plot visualisiert. Auf der x-Achse sind Messpunktjahre und auf der y-Achse sind die Daten der Tiefe des Lake Huron dargestellt.

> plot(LakeHuron, main = "Lake Huron als Zeitreihe", ylab = "Lake Level (Feet)", xlab="Year")

> abline(h=mean(LakeHuron),col=2,lwd=2)



Die Tiefe des Lake Huron war vor 1920 meistens über dem Mittelwert (rote Linie).

Bsp. 4:

> ftable(Titanic)

Survived: No Yes

Class Sex Age

1st Male Child 0 5

Adult 118 57

Female Child 0 1

Adult 4 140

2nd Male Child 0 11

Adult 154 14

Female Child 0 13

Adult 13 80

3rd Male Child 35 13

Adult 387 75

Female Child 17 14

Adult 89 76

Crew Male Child 0 0

Adult 670 192

Female Child 0 0

Adult 3 20

> Titanic

, , Age = Child, Survived = No

Sex

Class Male Female

1st 0 0

2nd 0 0

3rd 35 17

Crew 0 0

, , Age = Adult, Survived = No

Sex

Class Male Female

1st 118 4

2nd 154 13

3rd 387 89

Crew 670 3

, , Age = Child, Survived = Yes

Sex

Class Male Female

1st 5 1

2nd 11 13

3rd 13 14

Crew 0 0

, , Age = Adult, Survived = Yes

Sex

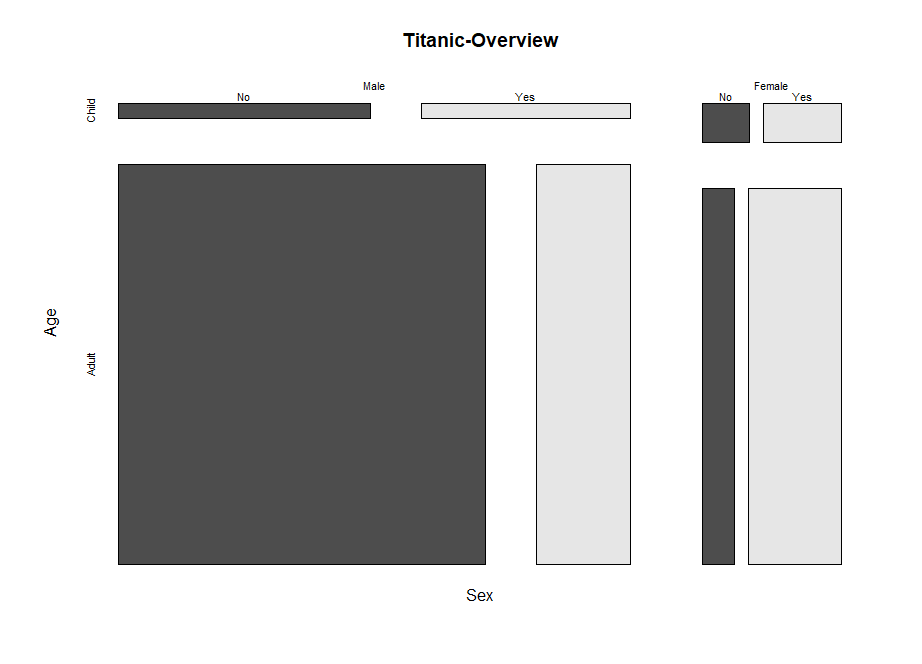
Class Male Female

1st 57 140

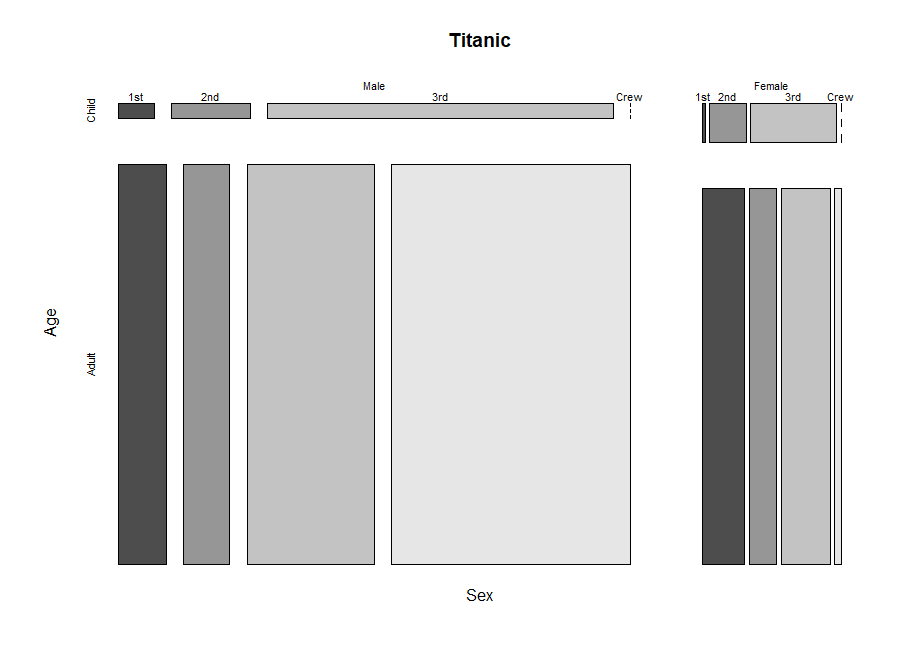
2nd 14 80

3rd 75 76

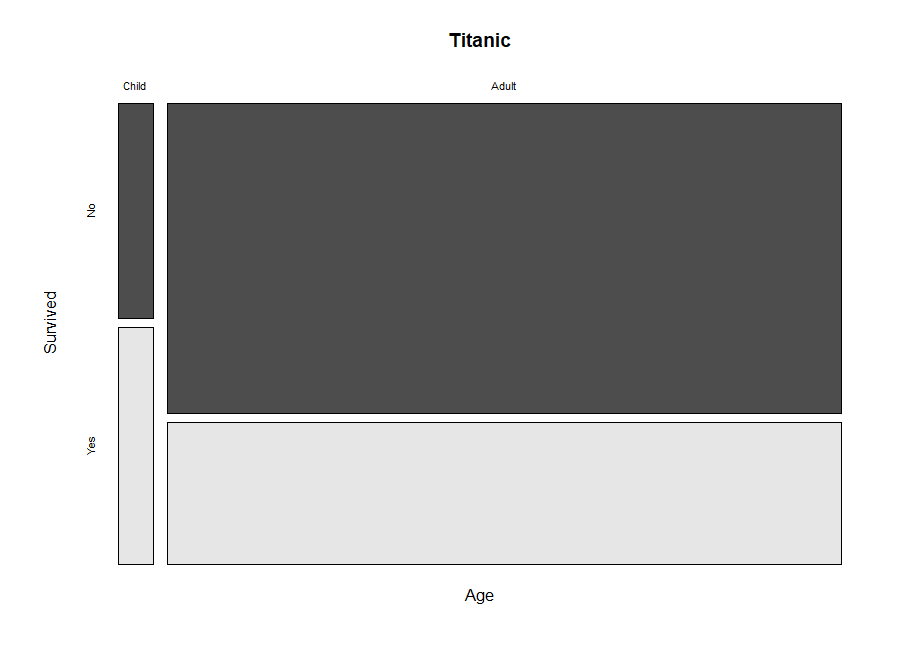
Crew 192 20



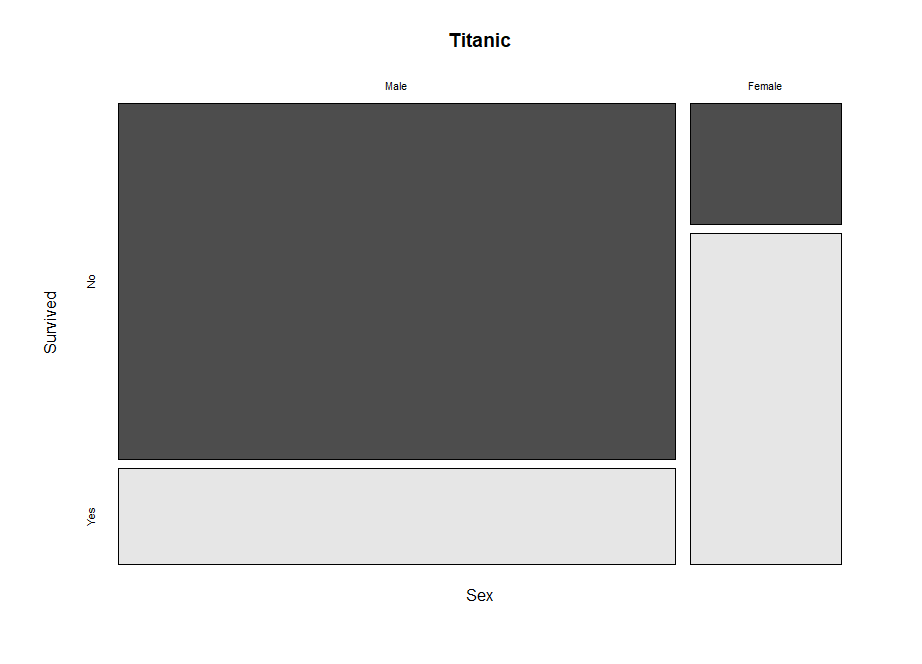
Diese Mosaik-Darstellung zeigt den Überblick von allen drei Datensätzen (Geschlecht, Alter, und Klasse) das Überleben bzw. Nicht-Überleben. Dabei lässt sich erkennen, dass die Männer Anteilsmäßig in etwa 60 % der Gesamtpopulation von Titanic ausgemacht hat. Diese hatten auch die niedrige Überlebensrate.



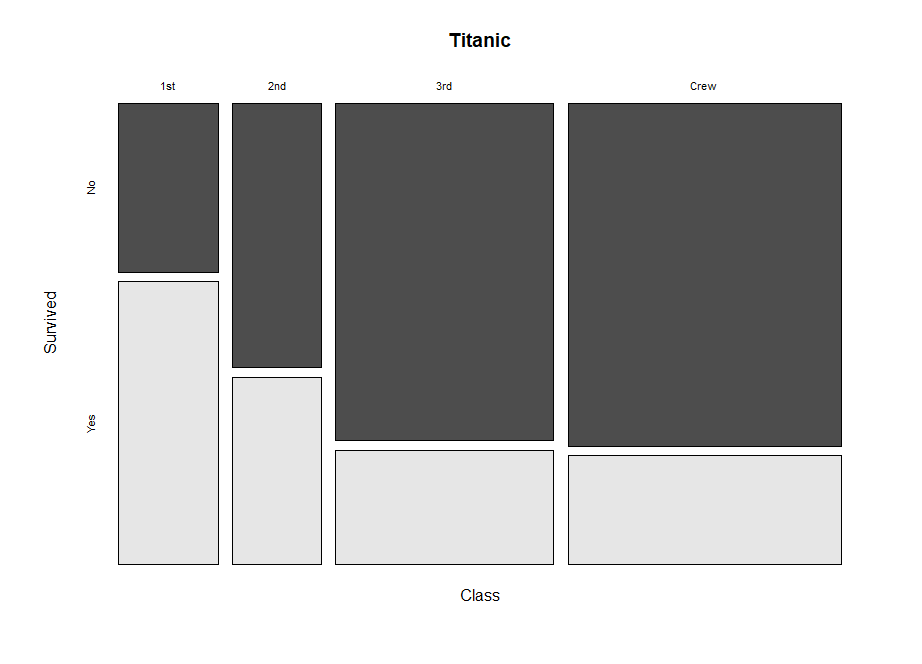
Verteilung der Klassen (1., 2., 3.Klasse sowie die Crew) in Geschlecht ist darstellt. Anteilsmäßig war die Männer-Gruppe aus der Crew-Klasse die größte, gefolgt von den Männern aus der 3.Klasse. Klasse 3 der Frauen war die größte Klassen der Frauen.



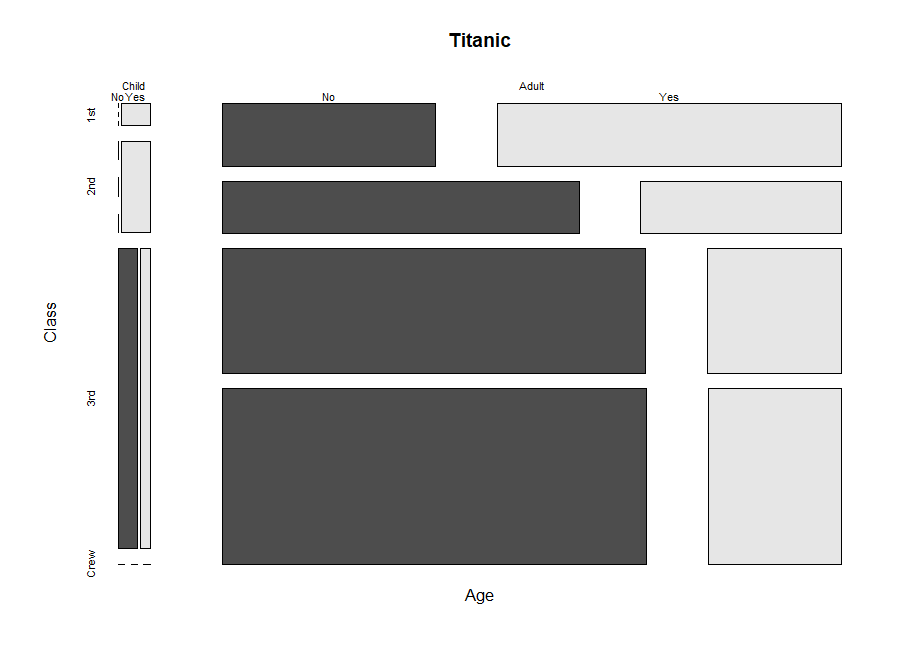
Dargestellt wird der Einfluss des Alters auf die Überlebensrate. Die Kinder hatten höhere Überlebenschance als die Erwachsene.



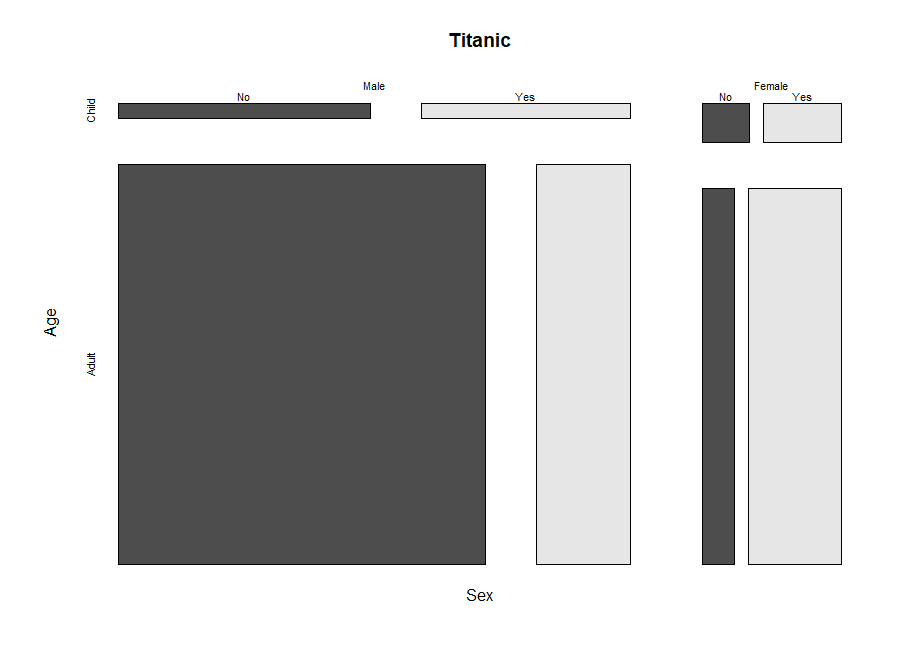
Dargestellt wird der Einfluss des Geschlechts auf die Überlebensrate. Die Frauen hatten bessere und höhere Überlebenschance als die Männer.



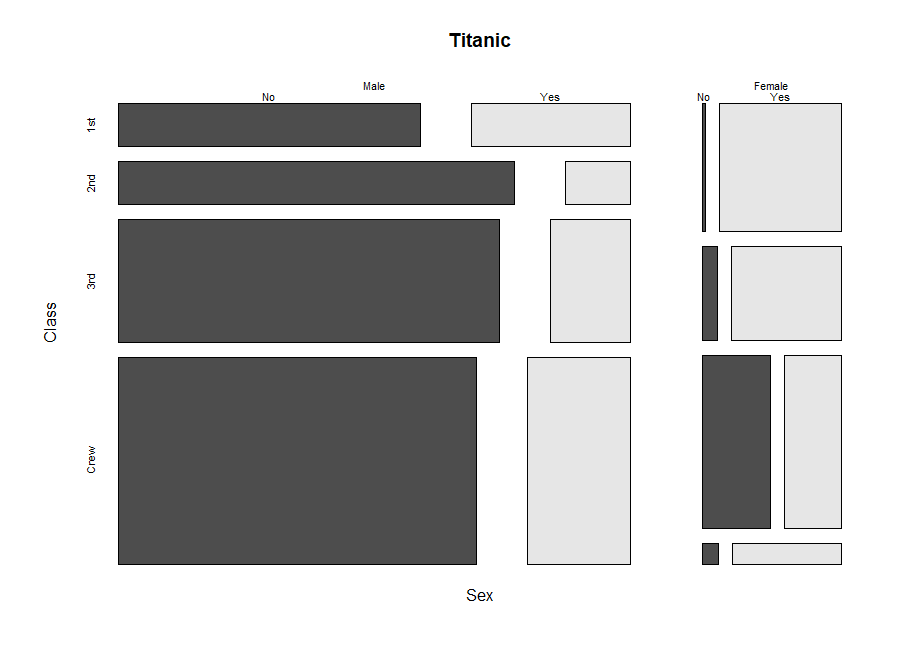
Dargestellt wird der Einfluss der Klasse auf die Überlebensrate. Dabei ist es klar ersichtlich, dass die Menschen mit höheren Klassen bessere Überlebenspotential hatten. Außerdem sieht man, dass die Besatzungsmitglieder die niedrigste Überlebensrate hatten.



Dargestellt wird der Einfluss bzw. der Vergleich zwischen Alter und Klasse auf die Überlebensrate. Dabei ist es klar ersichtlich, dass die Kinder bzw. die Erwachsene mit höheren Klassen bessere Überlebenspotential hatten.



Dargestellt wird der Einfluss bzw. der Vergleich zwischen Alter und Geschlecht auf die Überlebensrate. Dabei ist es ersichtlich, dass die Kinder und Frauen im Vergleich zu Männern höheren Überlebenspotential hatten. Die weiblichen Kinder hatten im Vergleich zu Frauen geringe Überlebenspotential.



Dargestellt wird der Einfluss bzw. der Vergleich zwischen Klasse und Geschlecht auf die Überlebensrate. Interessanterweise zeigt diese Darstellung, dass die Männer der 1.Klasse, 3.Klasse und die männlichen Besatzungsmitglieder höhere Überlebenschancen hatten als die Männer der 2.Klasse. Darüberhinaus, hatten die weiblichen Besatzungsmitglieder höhere Überlebenschancen als die Frauen der 3.Klasse. Bisher wurde angenommen, dass je höher die Klasse umso höher hatte man Überlebenschancen. Wie ist das möglich? Es handelt sich hier um Simpson-Paradoxon. Verteilungsmäßig gesehen, waren die Frauen der Klasse 3 die größte und die der Frauen der Crew die kleinste Klassengruppe der Frauen. Daraus lässt sich aussagen, dass durch den geringen Anteil der Frauen von Besatzungsmitglieder-Klasse die überlebt haben viel einfacher auf einen höheren Prozentsatz der Überlebungsrate kommen als die Frauen der 3.Klasse. Bei Männern kann man dieses Effekt nicht erklären, da Verteilungsmäßig die Besatzungsklasse und dann die 3.Klasse den größten Prozentsatz hatten als die 2.Klasse der Männer. Die Männer der Besatzungsklasse und 3.Klasse haben sich besser zurecht gefunden in der Situation als die Männer der 2.Klasse. In diesem Fall ist das Geschlecht die Variable, die für die Umkehrung und Verwirrung sorgt.