

Umfrage Windkraft

Sanja Bozic, Friedrich Leisch

Die Umfrage zum Thema *Akzeptanz und Meinung der Österreichischen Bevölkerung bezüglich der Energiegewinnung aus Windkraft* wurde, basierend auf einem Vorschlag von Christian Eberhard und Christoph Stelzer, im Zuge der Lehrveranstaltung “851.307 Vertiefung in statistische Methoden” an der Universität für Bodenkultur Wien im Wintersemester 2015/2016 von Studierenden durchgeführt.

Die Zielgruppe dieser Meinungsumfrage war die Österreichische Bevölkerung ab einem Alter von 16 Jahren (Aktives Wahlrecht).

Im Anhang befinden sich vier Dateien:

- “Fragenkatalog-Windkraft.pdf” enthält den Fragenkatalog zur Akzeptanz von Windkraftanlagen in Österreich.
- “umfrage-windkraft.csv” enthält die empirischen Rohdaten (Primärdaten) der Umfrage.
- “windkraft.RData” enthält die bereinigten Ergebnisse der Umfrage.
- “Windkraft-Abgabe-Vorlage.xlsx” enthält die Namen aller Variablen sowie deren Kodierung.

Zunächst lesen wir die RData-Datei, welche die bereinigten Daten enthält, in R ein. Mit `load()` funktioniert das nur, wenn sich die Datei im aktuellen Arbeitsverzeichnis (“Working Directory”) befindet. Falls das nicht der Fall ist, kann man zum Festlegen des Arbeitsverzeichnisses in R den Befehl `setwd()` benutzen.

```
R> load("windkraft.RData")
```

Demografie der Stichprobe

Im ersten Schritt analysieren wir die Demografie unserer Stichprobe indem wir die Verteilung der erhobenen Daten mit der Verteilung der Bevölkerung von Österreich nach Bundesländern vergleichen. Die Demographischen Daten Österreichs sind auf der Website der Statistik Austria verfügbar: www.statistik.at.

In der Variable `bev13` wird die gesamte Bevölkerungszahl in 100 000 eingegeben. Wir sehen dass 2013 Österreich 8,472 Millionen Einwohner hatte.

```
R> bev13 <- c(B=287, K=555, NO=1621, OO=1421, S=532, ST=1212, T=718, V=373, W=1753)
R> sum(bev13)
```

```
[1] 8472
```

Durch Division der einzelnen Einträge des Vektors mit der Summe der Gesamtbevölkerung, erhalten wir die prozentuellen Werte für jedes Bundesland.

```
R> bev13p <- bev13/sum(bev13)
R> bev13p
```

```

      B      K      NO      OO      S      ST      T
0.03387630 0.06550992 0.19133617 0.16772899 0.06279509 0.14305949 0.08474976
      V      W
0.04402738 0.20691690

```

Im nächsten Schritt erstellen wir eine Häufigkeitstabelle für unsere Stichprobe. Anschließend erstellen wir eine Matrix, indem wir die beiden Vektoren mithilfe der Funktion `cbind()` Spaltenweise verbinden. Die erstellte Matrix `TAB` enthält die prozentuellen Werte der gesamten Bevölkerung und der Stichprobe und wird nun mithilfe der Funktion `barplot()` als Balkendiagramm graphisch dargestellt.

```
R> table(wk$F13.bundesland)
```

```

  B  K  NO  OO  S  ST  T  V  W
37 15 112 42 17 24 15  8 260

```

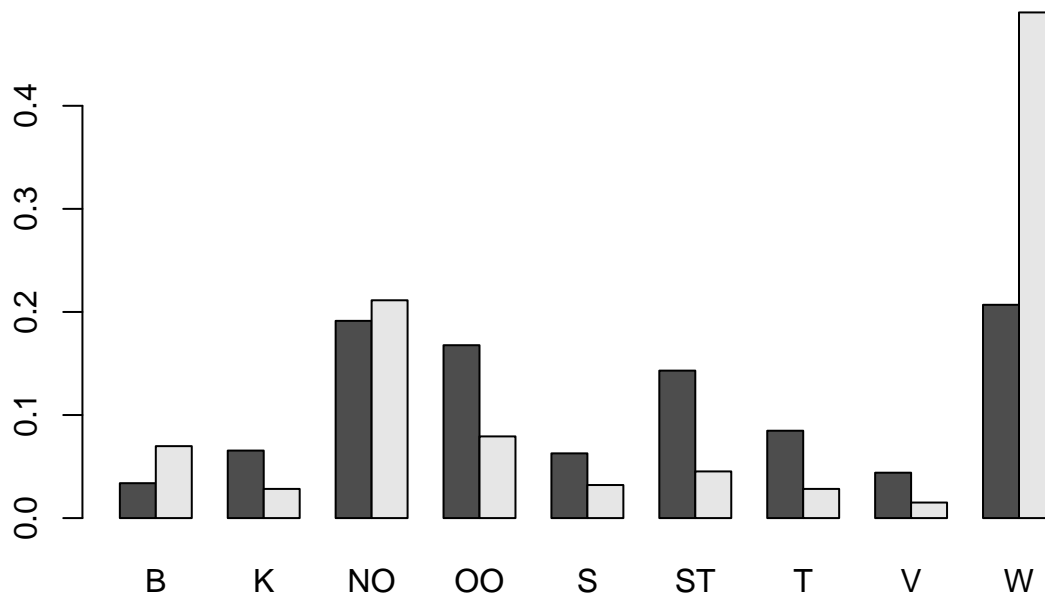
```
R> TAB <- cbind(GesBevProz=bev13p, StichProz=table(wk$F13.bundesland)/nrow(wk))
R> TAB
```

```

      GesBevProz  StichProz
B 0.03387630 0.06981132
K 0.06550992 0.02830189
NO 0.19133617 0.21132075
OO 0.16772899 0.07924528
S 0.06279509 0.03207547
ST 0.14305949 0.04528302
T 0.08474976 0.02830189
V 0.04402738 0.01509434
W 0.20691690 0.49056604

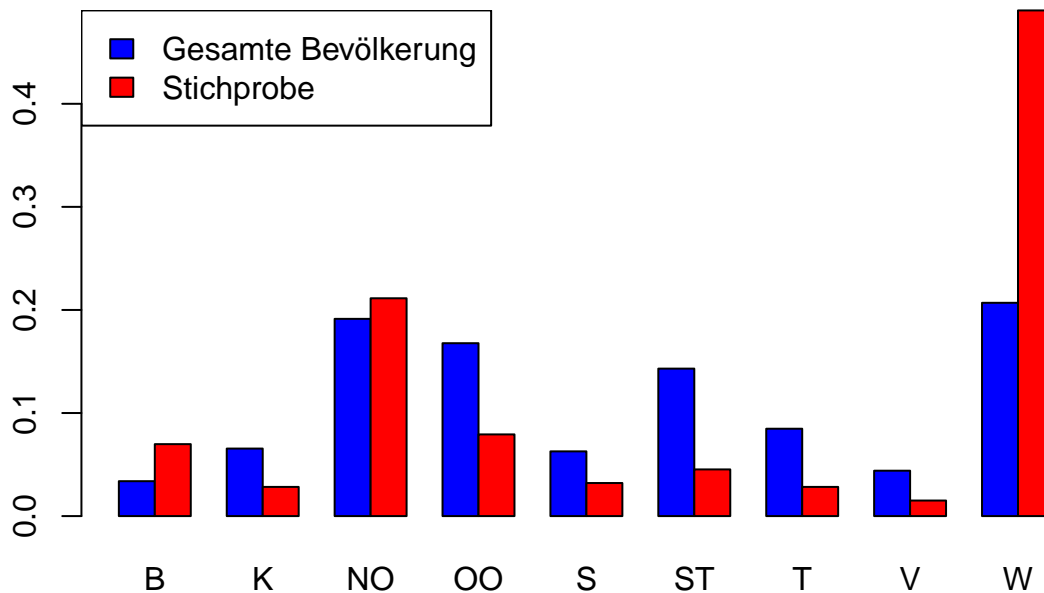
```

```
R> barplot(t(TAB), beside=TRUE)
```



Mit ein bisschen mehr Code kann man ein Balkendiagramm in Farbe und mit einer Legende erstellen:

```
R> barplot(t(TAB), beside=TRUE, col=c("blue","red"))  
R> legend("topleft", legend=c("Gesamte Bevölkerung", "Stichprobe"), fill=c("blue","red"))
```



Es ist leicht zu erkennen, dass sich die beiden Verteilungen der Einwohner nach Bundesland sehr stark unterscheiden. In unserer Stichprobe wohnen sehr viele Befragte in Wien, Burgenland und Niederösterreich, während die Bewohner der anderen Bundesländer stark unterrepräsentiert sind.

Um den Unterschied auch statistisch zu überprüfen, wurde ein χ^2 -Goodness-of-Fit-Test gemacht, welcher in diesem Fall überprüft ob unsere Stichprobe der Verteilung der Gesamtbevölkerung folgt.

Unsere Hypothesen lauten:

Nullhypothese: Die Verteilung der Einwohner nach Bundesland unserer Stichprobe ist der Verteilung der Gesamtbevölkerung gleich. **Alternativhypothese:** Die Verteilung der Einwohner nach Bundesland unserer Stichprobe ist der Verteilung der Gesamtbevölkerung nicht gleich.

```
R> chisq.test(table(wk$F13.bundesland), p=bev13p)
```

```
Chi-squared test for given probabilities
```

```
data: table(wk$F13.bundesland)
X-squared = 336.72, df = 8, p-value < 2.2e-16
```

Da der p-Wert viel kleiner als ein Signifikanzniveau von 0.05 ist, wird die Nullhypothese verworfen. D.h. die Häufigkeitsverteilung unserer Stichprobe unterscheidet sich signifikant von der Häufigkeitsverteilung der Gesamtbevölkerung, wie wir es auch erwartet haben.

Im nächsten Schritt erstellen wir eine Häufigkeitstabelle für die Altersklassen. Wir sehen, dass sehr viele junge Leute an der Umfrage teilgenommen haben, obwohl in Österreich 46-55 das häufigste Alter ist (Babyboom der 1960er Jahre).

```
R> table(factor(wk$F10.alter, labels=c("16-25", "26-35", "36-45", "46-55", "56-65", ">65")))
```

```
16-25 26-35 36-45 46-55 56-65 >65
 164   176    54    77    39    20
```

Hinsichtlich der großen Unterschiede in den Verteilungen der Befragten nach Bundesland und Altersklasse ist unsere Stichprobe **nicht repräsentativ**, jedoch für die Übung der Analyse eines Datensatzes im Rahmen dieser Lehrveranstaltung ausreichend.

Meinungen zu Windkraftanlagen

Nun möchten wir uns die Antworten auf ausgewählte Fragen anschauen.

Frage 1: Erneuerbare Energien sollen in Zukunft in Österreich verstärkt ausgebaut werden.

```
R> table(factor(wk$F1.ausbauerneuerb,
+             labels=c("stimme zu", "stimme eher zu", "stimme eher nicht zu",
+                   "stimme nicht zu", "unbekannt")))

```

```

           stimme zu      stimme eher zu stimme eher nicht zu
           396           108           18
stimme nicht zu      unbekannt
           6             2

```

Frage 3: Windkraft stellt in Österreich eine wichtige erneuerbare Energie dar.

```
R> table(factor(wk$F3.wichtig, labels=c("unbekannt", "stimme zu", "stimme eher zu",
+                                     "stimme eher nicht zu", "stimme nicht zu")))

```

```

           unbekannt      stimme zu      stimme eher zu
           1           212           219
stimme eher nicht zu      stimme nicht zu
           85           12

```

Frage 4: Wohnen Sie in der Nähe einer Windkraftanlage? (5-10km)

```
R> table(factor(wk$F4.naehe, labels=c("ja", "nein")))

```

```

ja nein
115 415

```

Frage 5: Kennen Sie das Verfahren der Bürgerbeteiligung im Zuge der Errichtung von Windkraftanlagen?

```
R> table(factor(wk$F5.kenneverfahren, labels=c("ja", "nein")))

```

```
ja nein
135 395
```

Um die Antworten auf die Fragen 8a - 8f analysieren, erstellen wir eine neue Matrix m.

```
R> m <- as.matrix(wk[, 8:23])
R> summary(m)
```

```
F8a.informiert F8b.ausbauwind F8c.landschaft F8d.beitrag
Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
1st Qu.:2.000 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000
Median :2.000 Median :2.000 Median :2.000 Median :2.000
Mean :2.204 Mean :1.866 Mean :2.356 Mean :1.808
3rd Qu.:3.000 3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:2.000
Max. :4.000 Max. :5.000 Max. :5.000 Max. :5.000
NA's :11 NA's :16 NA's :7 NA's :25

F8e.vielwind F8f.beimir F8g.tiergefahr F8h.foerderung F8i.allggefahr
Min. :1.000 Min. :1.0 Min. :1.000 Min. :1.000 Min. :1.000
1st Qu.:1.000 1st Qu.:1.0 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:2.000
Median :1.000 Median :2.0 Median :2.000 Median :2.000 Median :3.000
Mean :1.498 Mean :2.1 Mean :2.509 Mean :1.656 Mean :3.036
3rd Qu.:2.000 3rd Qu.:3.0 3rd Qu.:3.000 3rd Qu.:2.000 3rd Qu.:4.000
Max. :5.000 Max. :5.0 Max. :5.000 Max. :5.000 Max. :5.000
NA's :32 NA's :30 NA's :90 NA's :18 NA's :115

F8j.aspekte F8k.schatten F8l.wende F8m.wohnhaus
Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
1st Qu.:1.000 1st Qu.:3.000 1st Qu.:1.000 1st Qu.:1.000
Median :2.000 Median :3.000 Median :1.000 Median :2.000
Mean :1.879 Mean :3.133 Mean :1.416 Mean :1.837
3rd Qu.:2.000 3rd Qu.:4.000 3rd Qu.:2.000 3rd Qu.:2.000
Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
NA's :52 NA's :95 NA's :8 NA's :26

F8n.laerm F8o.stoertmich F8p.praegtregion
Min. :1.000 Min. :1.000 Min. :1.000
1st Qu.:2.000 1st Qu.:1.000 1st Qu.:2.000
Median :3.000 Median :2.000 Median :3.000
Mean :2.641 Mean :1.953 Mean :3.114
3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.:4.000
Max. :5.000 Max. :5.000 Max. :5.000
NA's :121 NA's :22 NA's :23
```

Manche Fragen wurden mit einer 5 beantwortet, da aber unsere Likert-Skala nur Antworten zwischen 1 und 4 zulässt, werden diese zu NA (not available) umbenannt.

```
R> m[m==5] <- NA
R> summary(m)
```

```
F8a.informiert F8b.ausbauwind F8c.landschaft F8d.beitrag
Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
1st Qu.:2.000 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000
Median :2.000 Median :2.000 Median :2.000 Median :2.000
```

Mean :2.204	Mean :1.847	Mean :2.345	Mean :1.763
3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:3.000	3rd Qu.:2.000
Max. :4.000	Max. :4.000	Max. :4.000	Max. :4.000
NA's :11	NA's :19	NA's :9	NA's :32
F8e.vielwind	F8f.beimir	F8g.tiergefahr	F8h.foerderung
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1st Qu.:1.000	1st Qu.:1.000	1st Qu.:2.000	1st Qu.:1.000
Median :1.000	Median :2.000	Median :2.000	Median :2.000
Mean :1.484	Mean :2.077	Mean :2.378	Mean :1.637
3rd Qu.:2.000	3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:2.000
Max. :4.000	Max. :4.000	Max. :4.000	Max. :4.000
NA's :34	NA's :34	NA's :112	NA's :21
F8i.allggefahr	F8j.aspekte	F8k.schatten	F8l.wende
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1st Qu.:2.000	1st Qu.:1.000	1st Qu.:3.000	1st Qu.:1.000
Median :3.000	Median :2.000	Median :3.000	Median :1.000
Mean :2.921	Mean :1.798	Mean :3.076	Mean :1.409
3rd Qu.:4.000	3rd Qu.:2.000	3rd Qu.:4.000	3rd Qu.:2.000
Max. :4.000	Max. :4.000	Max. :4.000	Max. :4.000
NA's :138	NA's :64	NA's :108	NA's :9
F8m.wohnhaus	F8n.laerm	F8o.stoertmich	F8p.praegtregion
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1st Qu.:1.000	1st Qu.:2.000	1st Qu.:1.000	1st Qu.:2.000
Median :2.000	Median :3.000	Median :2.000	Median :3.000
Mean :1.825	Mean :2.494	Mean :1.935	Mean :3.096
3rd Qu.:2.000	3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:4.000
Max. :4.000	Max. :4.000	Max. :4.000	Max. :4.000
NA's :28	NA's :145	NA's :25	NA's :28

Im nächsten Schritt wollen wir uns anschauen welche Frage wie viele Personen nicht beantwortet haben. Mithilfe der Funktion `apply` wenden wir Spaltenweise (in der Funktion `apply()` steht eine 1 für Zeilenweise oder eine 2 für Spaltenweise Anwendung der angegebenen Funktion) die Funktion `is.na` auf alle Antwortet an und summieren diese. Somit erhalten wir einen Vektor mit der Anzahl der fehlenden Antworten pro Frage.

```
R> # Spalten mit weniger als 50 missing
R> na <- apply(m, 2, function(x) sum(is.na(x)))
R> na
```

F8a.informiert	F8b.ausbauwind	F8c.landschaft	F8d.beitrag
11	19	9	32
F8e.vielwind	F8f.beimir	F8g.tiergefahr	F8h.foerderung
34	34	112	21
F8i.allggefahr	F8j.aspekte	F8k.schatten	F8l.wende
138	64	108	9
F8m.wohnhaus	F8n.laerm	F8o.stoertmich	F8p.praegtregion
28	145	25	28

In der Matrix `m2` werden alle Einträge gespeichert, welche weniger als 50 missings pro Frage (Spalte) haben. Im Vektor `ok` werden boolesche Werte (TRUE oder FALSE) über die Vollständigkeit der Zeilen gespeichert (TRUE wenn keine fehlenden Werte in der Zeile vorkommen). Dieses Eliminationsverfahren zählt zu den gängigen Missing-Data-Techniken. Anschließend werden in der Matrix `m2` nurmehr vollständige Datensätze gespeichert. Die Dimension der Matrix `m2` hat sich von 530 x 11 auf 379 x 11 reduziert. In der summary sehen wir, dass es keine weiteren Na's mehr gibt und dass alle Werte zwischen 1 und 4 liegen.

```
R> m2 <- m[, na<50]
R> ok <- complete.cases(m2)
R> m2 <- m2[ok,]
R> dim(m2)
```

```
[1] 379 11
```

```
R> summary(m2)
```

```
F8a.informiert F8b.ausbauwind F8c.landschaft F8d.beitrag
Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
1st Qu.:1.000 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000
Median :2.000 Median :2.000 Median :2.000 Median :2.000
Mean :2.087 Mean :1.834 Mean :2.319 Mean :1.794
3rd Qu.:3.000 3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:2.000
Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000
 F8e.vielwind F8f.beimir F8h.foerderung F8l.wende
Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000
Median :1.000 Median :2.000 Median :2.000 Median :1.000
Mean :1.493 Mean :2.098 Mean :1.652 Mean :1.406
3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:2.000 3rd Qu.:2.000
Max. :4.000 Max. :4.000 Max. :4.000 Max. :4.000
 F8m.wohnhaus F8o.stoertmich F8p.praegregion
Min. :1.000 Min. :1.000 Min. :1.000
1st Qu.:1.000 1st Qu.:1.000 1st Qu.:2.000
Median :2.000 Median :2.000 Median :3.000
Mean :1.784 Mean :1.921 Mean :3.079
3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:4.000
Max. :4.000 Max. :4.000 Max. :4.000
```

Wir erstellen ein neues dataframe `wk2`, welches alle Daten aus `wk` kopiert bei denen es keine fehlenden Werte in den Fragen 8a - 8f gab.

```
R> wk2 <- wk[ok,]
```

Als nächstes führen wir eine Hauptkomponentenanalyse (Principal Component Analysis - PCA) der 11 Variablen durch. Wir sehen, dass 44,58% der Varianz durch die ersten beiden Komponenten erklärt werden kann, weitere 11,75% durch die dritte. Die erste Komponente wird am stärksten von der Frage `8f.beimir` erklärt.

```
R> # PCA
R> p1 <- prcomp(m2)
R> p1
```

Standard deviations (1, ..., p=11):

```
[1] 1.6071991 1.0595498 0.9883232 0.8677154 0.8477970 0.6594897 0.6520491
[8] 0.6166448 0.5898924 0.5412575 0.5261402
```

Rotation (n x k) = (11 x 11):

```
PC1 PC2 PC3 PC4 PC5
```


F8a.informiert	0.06148647	-0.24303562	-0.35996686	0.06870125	-0.86085531
F8b.ausbauwind	0.35897035	0.07658524	-0.29283820	-0.08533034	0.05354132
F8c.landschaft	-0.43577989	0.06812232	-0.05704809	-0.70317736	-0.20750291
F8d.beitrag	0.22338144	0.06256440	-0.30867325	-0.39788057	0.15212696
F8e.vielwind	-0.09665851	0.16456400	-0.12073099	-0.15423141	0.12402299
F8f.beimir	0.51015619	0.01552946	-0.24922435	0.16642343	0.01611908
F8h.foerderung	0.22733053	0.12013262	-0.25333719	-0.30561833	0.15704861
F8l.wende	0.16566667	0.09306981	-0.16829306	-0.14947499	0.05336597
F8m.wohnhaus	-0.30168563	0.26875870	-0.47714836	0.20060108	0.03235283
F8o.stoertmich	-0.40754840	0.19500321	-0.45876406	0.35344609	0.17623300
F8p.praegtreion	-0.14737563	-0.87512264	-0.28073107	-0.05326514	0.33853050
	PC6	PC7	PC8	PC9	PC10
F8a.informiert	0.06295344	-0.12706485	0.15465486	-0.10078340	0.06480678
F8b.ausbauwind	-0.10524312	0.27115491	0.11681296	0.43151605	-0.37824201
F8c.landschaft	-0.30880004	0.26000676	-0.22210573	0.01069389	-0.19499462
F8d.beitrag	0.36020559	0.32945392	0.08007468	-0.12265640	0.63946032
F8e.vielwind	-0.55883933	-0.51270515	0.39941674	0.06678004	0.36727609
F8f.beimir	-0.56833548	0.19734554	-0.25779003	-0.09055501	0.05335084
F8h.foerderung	0.25613343	-0.31656254	0.38603449	-0.19116796	-0.48941709
F8l.wende	0.09369357	-0.48216744	-0.68266282	-0.29288520	-0.04401099
F8m.wohnhaus	0.20118111	-0.15838681	-0.22784193	0.59814711	0.09029134
F8o.stoertmich	-0.08366476	0.26262564	0.11065685	-0.53858804	-0.13358914
F8p.praegtreion	-0.05188599	-0.07495606	-0.03868118	0.07505939	-0.02445370
	PC11				
F8a.informiert	0.05832055				
F8b.ausbauwind	0.58468227				
F8c.landschaft	-0.11939725				
F8d.beitrag	0.02831419				
F8e.vielwind	0.18915027				
F8f.beimir	-0.45823295				
F8h.foerderung	-0.40091064				
F8l.wende	0.33965888				
F8m.wohnhaus	-0.29050272				
F8o.stoertmich	0.18245900				
F8p.praegtreion	-0.01233242				

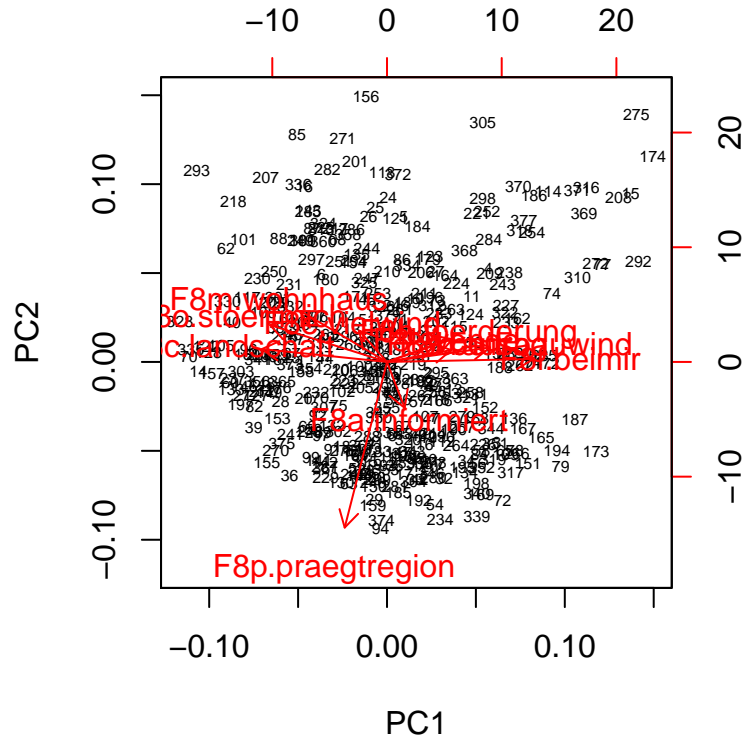
```
R> summary(p1)
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	1.6072	1.0595	0.9883	0.86772	0.84780	0.65949	0.65205
Proportion of Variance	0.3108	0.1351	0.1175	0.09058	0.08647	0.05232	0.05115
Cumulative Proportion	0.3108	0.4458	0.5633	0.65390	0.74037	0.79270	0.84385
	PC8	PC9	PC10	PC11			
Standard deviation	0.61664	0.58989	0.54126	0.5261			
Proportion of Variance	0.04575	0.04186	0.03524	0.0333			
Cumulative Proportion	0.88959	0.93145	0.96670	1.0000			

Wenn man den Biplot für die ersten beiden Komponenten betrachtet, sieht man dass 8b.ausbauwind und 8d.beitrag sowie 8h.foerderung und 8l.wende zu einer Komponente zusammengefasst werden könnten.

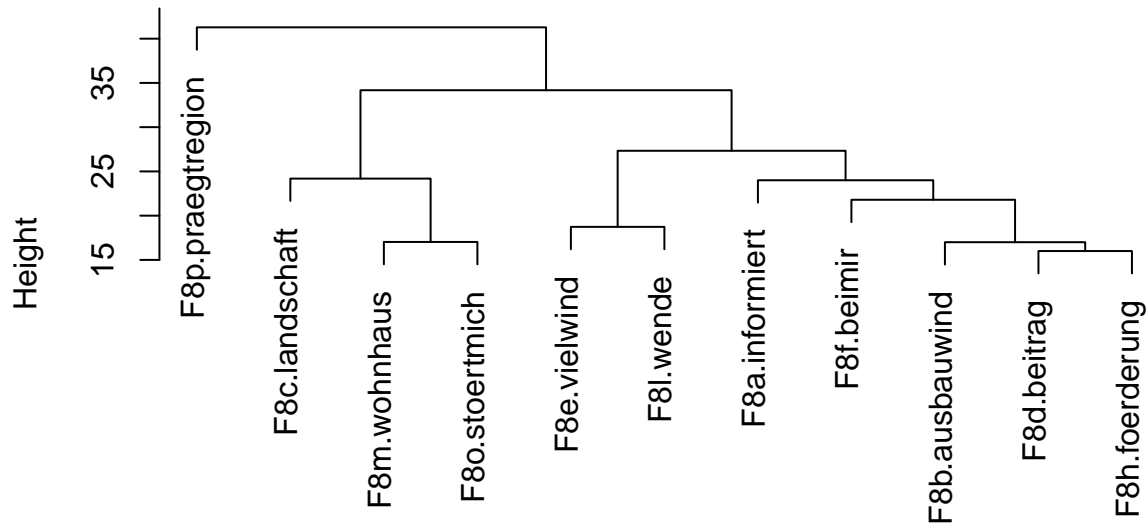
```
R> biplot(p1, cex=c(.5,1))
```



Um eine hierarchische Clusteranalyse durchzuführen, muss unsere Matrix `m2` transponiert werden. Danach berechnen wir die Distanzmatrix und führen anschließend die Hierarchische Clusteranalyse durch. Mit der Funktion `plot()` stellen wir die Ergebnisse graphisch mittels eines Dendrogramms dar. Je niedriger die Verzweigung, desto ähnlicher sind die zwei Cluster. Die Antworten auf die Fragen `F8d.beitrag` und `F8h.foerderung` sind am ähnlichsten.

```
R> h1 <- hclust(dist(t(m2)))
R> plot(h1)
```

Cluster Dendrogram

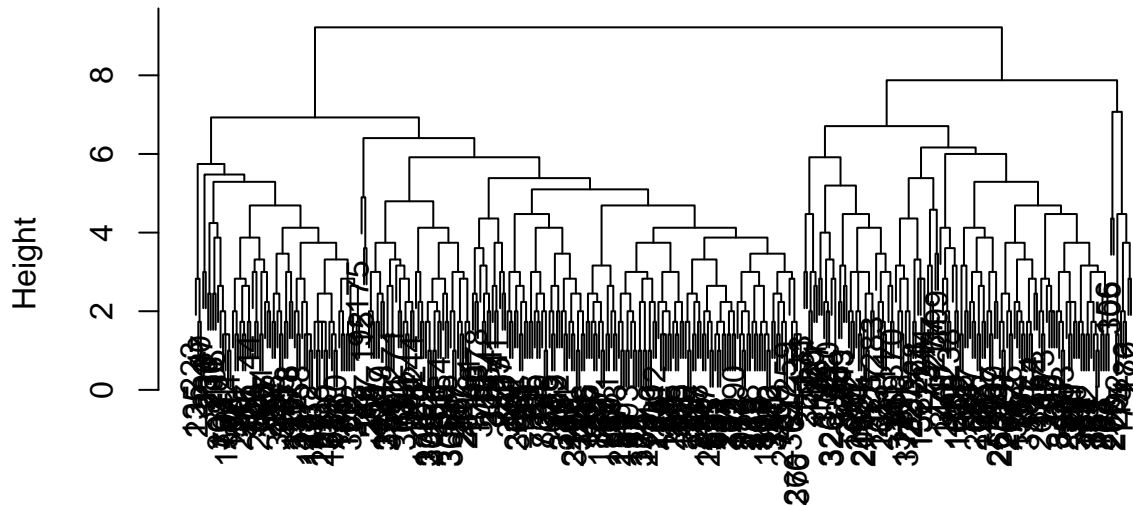


```
dist(t(m2))  
hclust (*, "complete")
```

Wenn man versucht eine Hierarchische Clusteranalyse durchzuführen ohne davor die Matrix zu transponieren, erhält man ein Dendrogramm das schwer lesbar ist und daher wenig Sinn ergibt.

```
R> h2 <- hclust(dist(m2))  
R> plot(h2)
```

Cluster Dendrogram



```
dist(m2)
hclust (*, "complete")
```

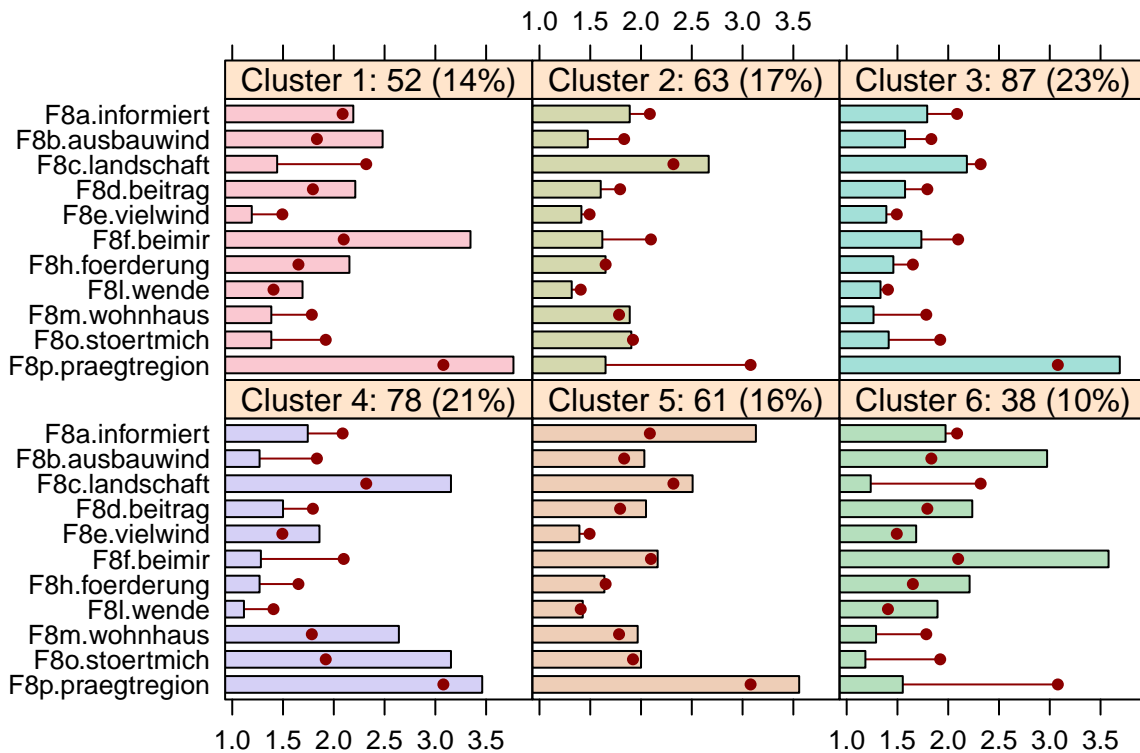
Um die Daten in Cluster einzuteilen, kann z.B. das Paket `flexclust` verwendet werden. Die Funktion `set.seed()` gibt uns die Möglichkeit simulierte Ergebnisse später rekonstruieren zu können. Mit der Funktion `stepcclust` teilen wir unsere Stichprobe in sechs Gruppen und stellen das Ergebnis graphisch dar.

```
R> library(flexclust)
R> set.seed(1801)
R> k1 <- stepcclust(m2, k=6, nrep=10)
```

```
6 : * * * * * * * * * *
```

Mithilfe der Balkendiagramme können wir die Zustimmung zu einzelnen Fragen betrachten. Die roten Punkte sind in allen Gruppen gleich und stellen die Mittelwerte über alle Daten dar. Zur Erinnerung, 1=stimme zu, 2=stimme eher zu, 3=stimme eher nicht zu, 4=stimme nicht zu.

```
R> barchart(k1)
```



Personen die in die **erste Gruppe** fallen, sind stark gegen den Bau in Ihrer Region, da sie der Meinung sind Windkraftanlagen beeinträchtigen das Landschaftsbild, aber prägen das Landschaftsbild ihrer Region nicht.

Personen aus der **vierten Gruppe** wären mit dem Bau von Windkraftanlagen in ihrer Region einverstanden und sind der Meinung dass Windkraftanlagen in Zukunft in Österreich verstärkt errichtet werden sollen. Außerdem stimmen sie der Aussage, dass Windkraft eine unerschöpfliche Energiequelle ist und daher Förderungswürdig ist zu. Sie meinen Windkraftanlagen beeinträchtigen und prägen das Landschaftsbild ihrer Region nicht und sollten in der Nähe von Wohnhäusern errichtet werden .

Die Personen aus **Gruppe sechs** sind stark gegen den Bau von Windkraftanlagen in ihrer Region und der Meinung Windkraftanlagen sollen in Zukunft in Österreich nicht verstärkt errichtet werden, Windkraftanlagen sollten nicht in der Nähe von Wohnhäusern errichtet werden. Eine Windkraftanlage in unmittelbarer Nähe zu ihrem Wohnhaus würde sie stören. Sie glauben Windkraftanlagen prägen und beeinträchtigen das Landschaftsbild ihrer Region.

Wie man sieht, sind die Gruppen unterschiedlich groß.

```
R> c1 <- clusters(k1)
R> table(c1)
```

```
c1
 1  2  3  4  5  6
52 63 87 78 61 38
```

Mithilfe von Mosaikdiagrammen bzw. Mosaik-Plots können wir Datensätzen mit zwei oder mehreren qualitativen Variablen graphisch darstellen um Zusammenhänge zwischen den Merkmalen zu erkennen.

Achtung Es sind sowohl die Höhe, als auch die Breite der einzelnen Felder zu betrachten, da diese proportional zur Anzahl der Beobachtungen sind.

Auf den folgenden Mosaikdiagrammen wurden jeweils ein Merkmal und die sechs Gruppen betrachtet. Deshalb wird die Breite der Felder in allen Plots gleich sein. Gruppe 3 wird immer die breitesten Felder haben, da sich mehr Personen in dieser Gruppe als in den anderen befinden.

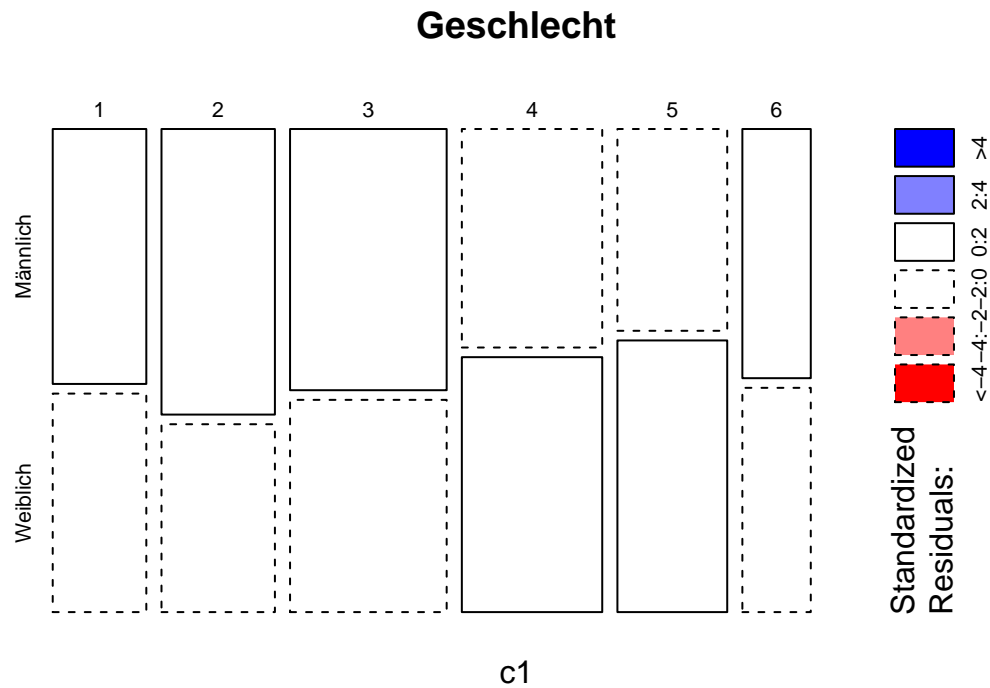
Die einzelnen Felder werden rot, weiß oder blau gefärbt, in Abhängigkeit davon ob die standardisierten Residuen positiv oder negativ sind. Standardisierte Residuen sind berechnete Größen und messen den standardisierter Abstand zwischen Beobachtung und unter Null erwartetem Wert. Falls die beobachteten Häufigkeiten größer sind als die erwarteten, werden die Felder blau oder weiß gefärbt und haben eine durchgezogene Linie als Rand. Falls die beobachteten Häufigkeiten kleiner sind als die erwarteten, werden die Felder rot oder weiß gefärbt und der Rand der Felder ist eine gestrichelte Linie.

Um leichter lesbare Mosaik-Plots zu erstellen, müssen wir die Namen der möglichen Ausprägungen für die kategorischen Daten die uns interessieren umbenennen.

```
R> # In factoren umwandeln
R> wk2Factor <- list()
R> wk2Factor$F9.geschlecht <- factor(wk2$F9.geschlecht, labels = c("Männlich", "Weiblich"))
R> wk2Factor$F10.alter <- ordered(wk2$F10.alter,
+                               labels = c("16-25", "26-35", "36-45", "46-55", "56-65", ">65"))
R> wk2Factor$F11.schule <- factor(wk2$F11.schule,
+                               labels = c("Pflichtschule", "Fachschule", "Matura",
+                               "Akademiker", "sonstiges"))
R> wk2Factor$F12.beruf <- factor(wk2$F12.beruf,
+                               labels = c("Angestellter", "Selbständig", "ohne Beschäftigung",
+                               "Arbeiter", "Schueler", "leitender Ang", "Beamter",
+                               "Haushalt", "Pension", "Landwirt", "Sonstiges"))
R> wk2Factor$F13.bundesland <- factor(wk2$F13.bundesland)
```

Auf dem ersten Mosaikdiagramm sehen wir, dass die Anzahl der Männer in den Gruppen 1, 2, 3 und 6 größer als die Anzahl der Frauen ist.

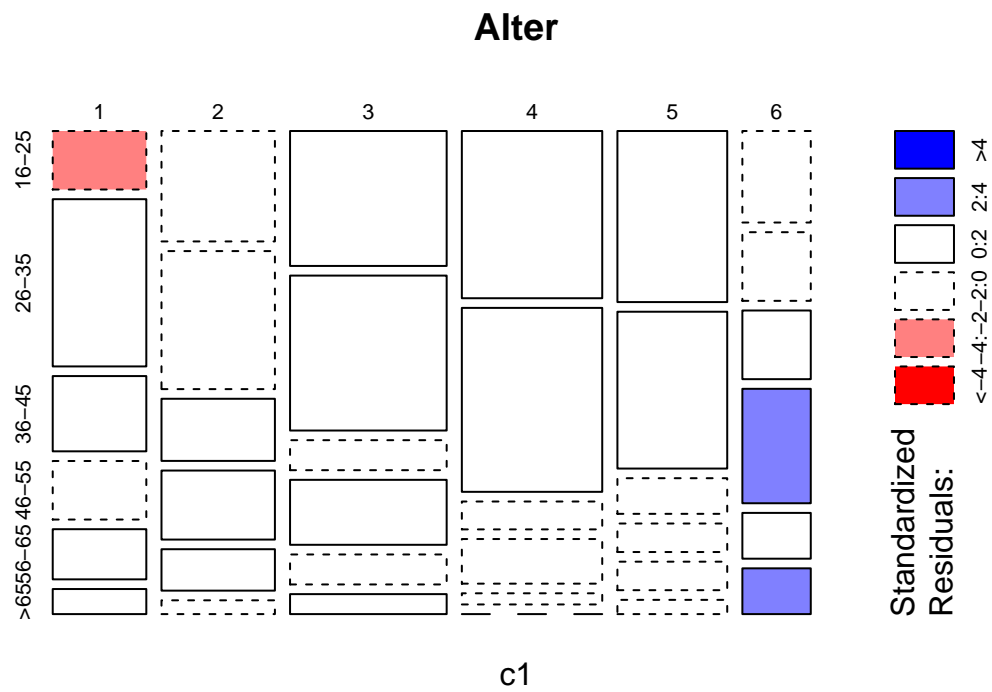
```
R> mosaicplot(table(c1, wk2Factor$F9.geschlecht), shade=TRUE, main="Geschlecht")
```



In diesem Mosaikdiagrammen sehen wir wieder, dass sehr viele junge Personen an der Umfrage teilgenommen haben. Besonders auffällig ist es in den Gruppen 3, 4 und 5. In der Gruppe 6 hingegen sind die Alterskategorien gleichmäßiger. Die Anteile von Personen mit einem Alter zwischen 46 und 55 Jahren sowie Personen die mehr als 65 Jahre alt sind, ist in dieser Gruppe höher als erwartet.

Zur Erinnerung, die Gruppen 1 und 6 waren eher gegen den Bau von Windkraftanlagen in ihrer Region als Gruppe 4, welche mit dem Bau einverstanden wäre.

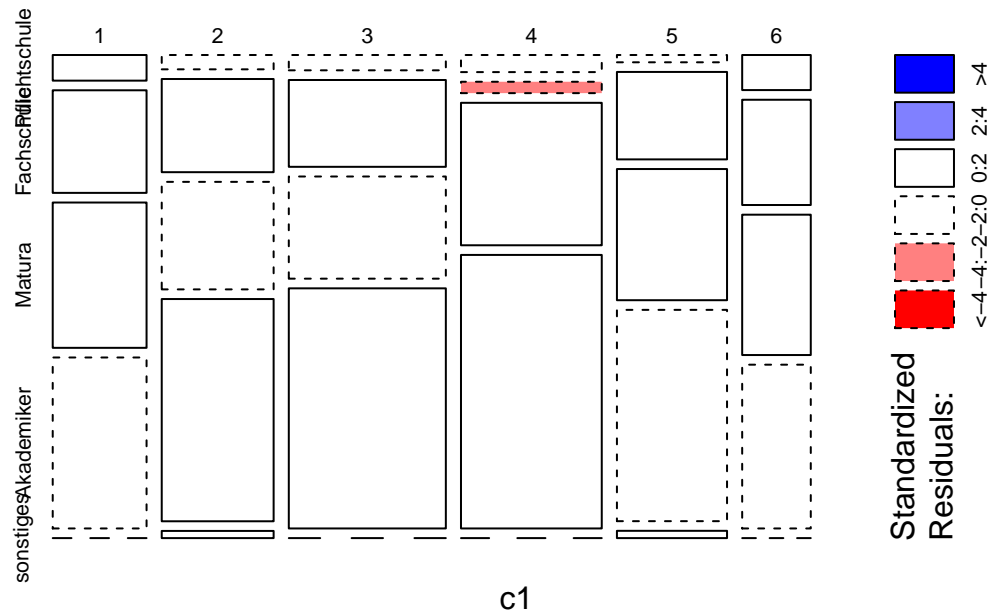
```
R> mosaicplot(table(c1, wk2Factor$F10.alter), shade=TRUE, main="Alter")
```



Bei der Aufteilung nach dem höchsten Schulabschluss sind Akademiker in allen Gruppen am häufigsten vertreten.

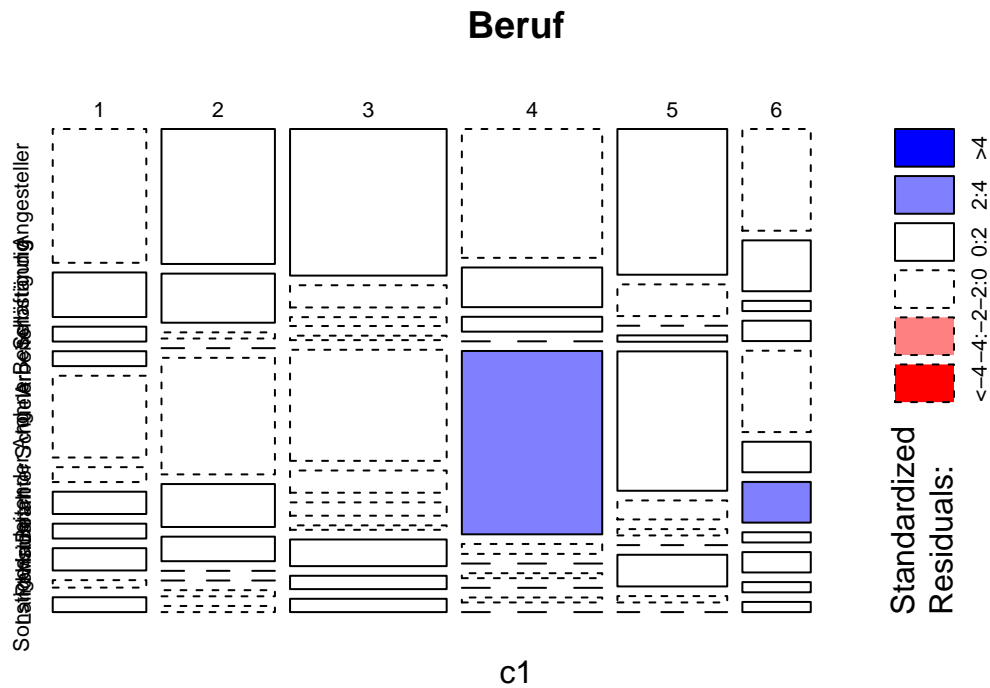
```
R> mosaicplot(table(c1, wk2Factor$F11.schule), shade=TRUE, main="Schulabschluss")
```


Schulabschluss



Auf dem nächsten Mosaik-Plot sehen wir, dass in allen sechs Gruppen Angestellte und Schüler bzw. Studenten die Mehrheit bilden. Dies ist nachvollziehbar, da in unserer Stichprobe sehr viele junge Menschen sind. In Gruppe 4 befinden sich mehr Schüler als erwartet.

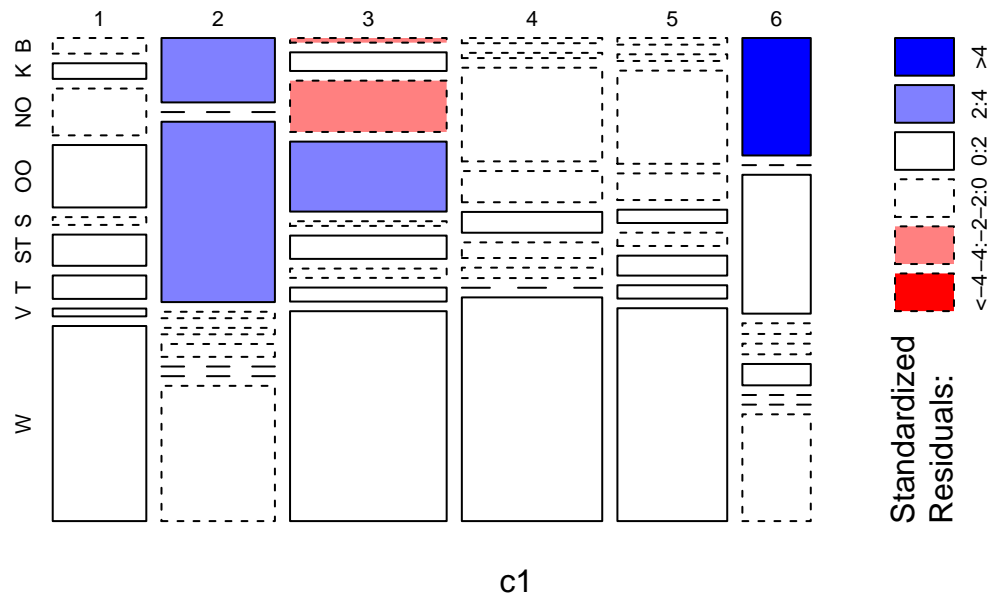
```
R> mosaicplot(table(c1, wk2Factor$F12.beruf), shade=TRUE, main="Beruf")
```



Auf der folgenden Graphik ist zu erkennen, dass in fast allen Gruppen die meisten Befragten in Wien wohnen. In den Gruppen 2 und 6 sind auffällig viele Niederösterreicher und Burgenländer, vor allem in Gruppe 6 haben wir deutlich mehr Burgenländer als erwartet. In Gruppe 3 befinden sich weniger Burgenländer und Niederösterreicher, dafür aber mehr Oberösterreicher als erwartet.

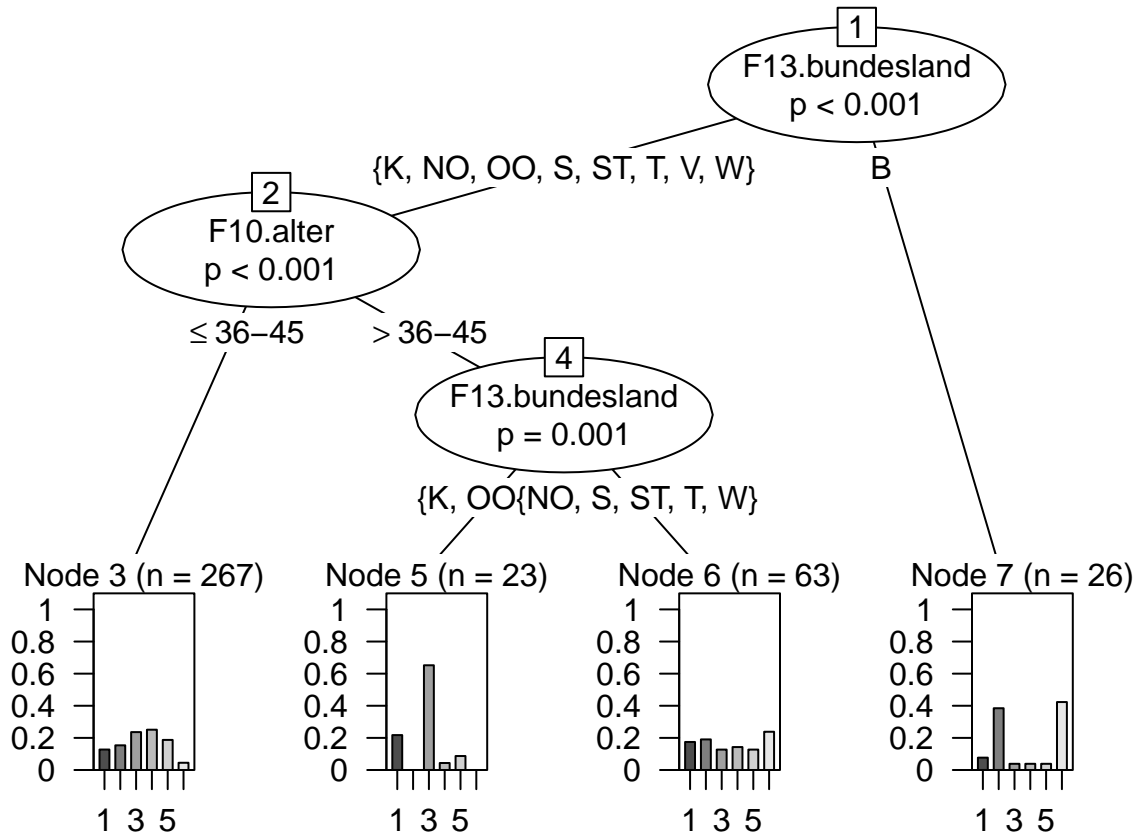
```
R> mosaicplot(table(c1, wk2Factor$F13.bundesland), shade=TRUE, main="Bundesland")
```

Bundesland



Zu guter Letzt wollen wir uns anschauen welche Eigenschaften Personen haben, die in eine der Gruppen fallen. Dazu erstellen wir einen Klassifikationsbaum mit der Funktion `ctree()` aus dem Paket `party`.

```
R> library(party)
R> b1 <- ctree(as.factor(c1) ~ F9.geschlecht + F10.alter + F11.schule + F12.beruf + F13.bundesland,
+             data=wk2Factor)
R> plot(b1)
```



Im ersten Knoten werden die Befragten nach Bundesland aufgeteilt. Wir können sehen dass die meisten der 26 Burgenländer, die an der Umfrage teilgenommen haben, in die Gruppen 2 und 6 fallen. Im zweiten Knoten wird nach Alter unterschieden. Personen die weniger als 45 Jahre alt sind, sind seltener in den Gruppen 1 und 6 vertreten, welche eher gegen den Bau von Windkraftanlagen in ihrer Region waren. Im letzten Knoten werden Personen die über 45 Jahre alt sind erneut nach Bundesland gesplittet. Von den Personen aus Kärnten, Oberösterreich und Vorarlberg fallen sehr viele in Gruppe 3. Ältere Personen aus den anderen Bundesländern sind in alle Gruppe gleich häufig.