

# Proportional gestapelte Balkendiagramme

*Peter Baumgartner*

*2017-04-22*

## Contents

<b>Schulform und Repräsentativität</b>	<b>2</b>
<b>Explorative Grafiken</b>	<b>3</b>
Ausstattung . . . . .	3
IKT-Kompetenz . . . . .	4
Erfahrungen mit audiemus . . . . .	4
Förderliche Faktoren (Wünsche differenziert) . . . . .	5
Wünsche differenziert und sortiert . . . . .	7
Interesse an Fortbildung . . . . .	9
Warum wird nicht Audiumus benutzt? . . . . .	9
<b>Förderliche Faktoren neu codiert</b>	<b>10</b>
Wünsche ja/nein mit zentrierte Linie . . . . .	11
Balken-Diagramm: ja/nein Wünsche . . . . .	13
<b>Verschiedene Signifikanz-Tests</b>	<b>14</b>
Chi-Quadrat Homogenitätstest . . . . .	14
Exakter Test nach Fisher . . . . .	14
Odds-Ratio-Test (Chancenverhältnis) . . . . .	16
<b>Literatur</b>	<b>17</b>

---

```
## Loading required package: tidyverse

## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages -----

## filter(): dplyr, stats
## lag():    dplyr, stats

## Loading required package: reshape2

##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths

## Loading required package: readxl

## Loading required package: pander
```

```
## Loading required package: sjmisc
##
## Attaching package: 'sjmisc'
## The following object is masked from 'package:purrr':
##
##   is_empty
## The following object is masked from 'package:tidyr':
##
##   replace_na
## Loading required package: sjPlot
## Visit http://strengjacke.de/sjPlot for package-vignettes.
## Loading required package: forcats
## Loading required package: REdaS
## Loading required package: grid
```

## Schulform und Repräsentativität

```
> schulform <- matrix(c(17, 100, 32, 4, 34, 19, 41, 253, 40, 5, 18, 13),
+                      ncol = 3, byrow = TRUE,
+                      dimnames = list(c("Schulen", "weitergeleitet", "Lehrer/innen", "Rücklauf"),
+                      c("AHS", "NMS", "PTS")))
> schulform <- as.data.frame(schulform)
> schulform.t <- t(schulform)
> schulform.t <- melt(schulform.t)
> schulart <- schulform[c(1,4) , ]
> fisher.test(schulart)
```

Fisher's Exact Test for Count Data

```
data: schulart
p-value = 0.1278
alternative hypothesis: two.sided
> chisq.test(schulart, simulate.p.value = TRUE)
```

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

```
data: schulart
X-squared = 4.0342, df = NA, p-value = 0.1344
```

Die Stichprobe ist daher bezogen auf die Schulform bei beiden Tests nicht repräsentativ (Pearson Chi-Quadrat Test:  $p > .005$ ).

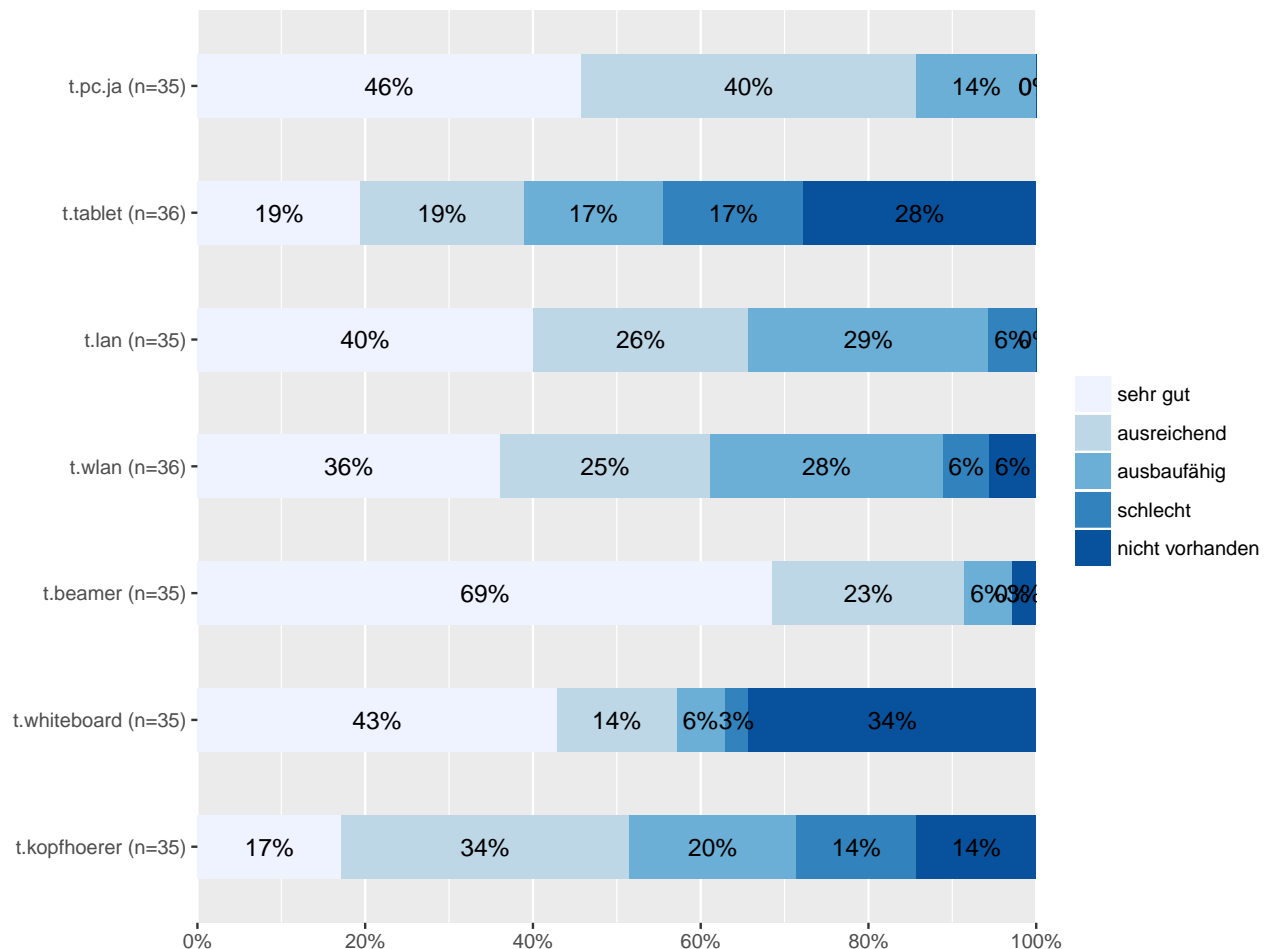
## Explorative Grafiken

Diese Datei stellt die verschiedenen Fragenbatterien als proportional gestapelte Balkendiagramme dar. Das sind Balken, die die prozentuelle Verteilung der Items einer Frage in ihrer Größe anzeigen. Es werden dabei alle Fragen einer Batterie untereinander zum besseren Vergleich dargestellt.

Ich habe die Grafiken derzeit noch nicht hübsche gemacht, d.h. keine detaillierte Beschreibung, bei den Farben, Schriftgrößen gespielt. Das ist vorerst nicht notwendig. Wahrscheinlich werde ich die meisten Grafiken für den Artikel sowieso nicht brauchen. Für ansprechende Folien für den Auftritt (das Referat, die Präsentation) ist ja noch Zeit...

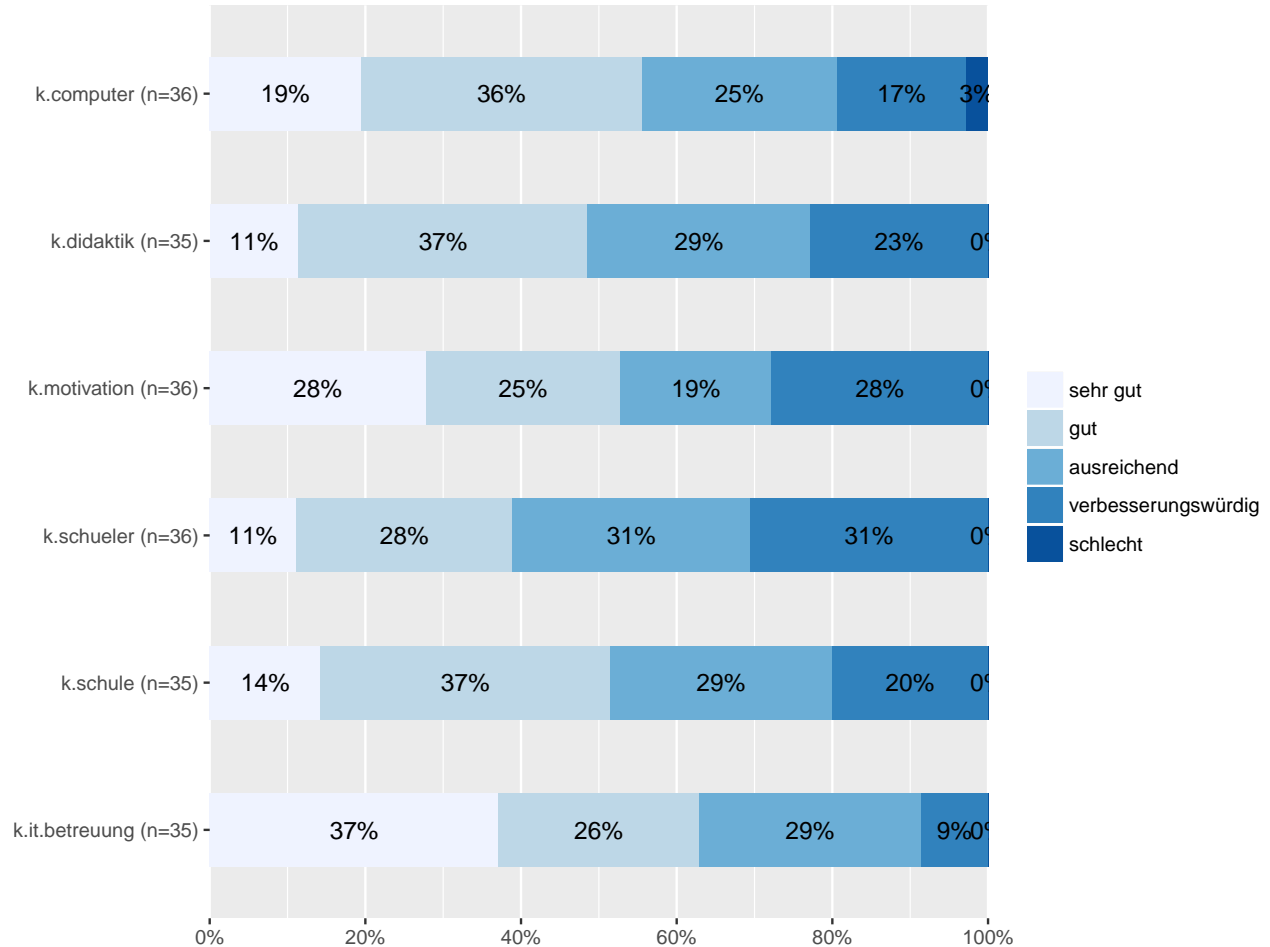
## Ausstattung

```
> library(sjPlot)
> ausstattung <- select(umfrage, c(9:4, 2))
> #names(ausstattung) <- c("PCs", "Tablets", "WLAN")
> sjp.stackfrq(ausstattung, digits = 0)
```



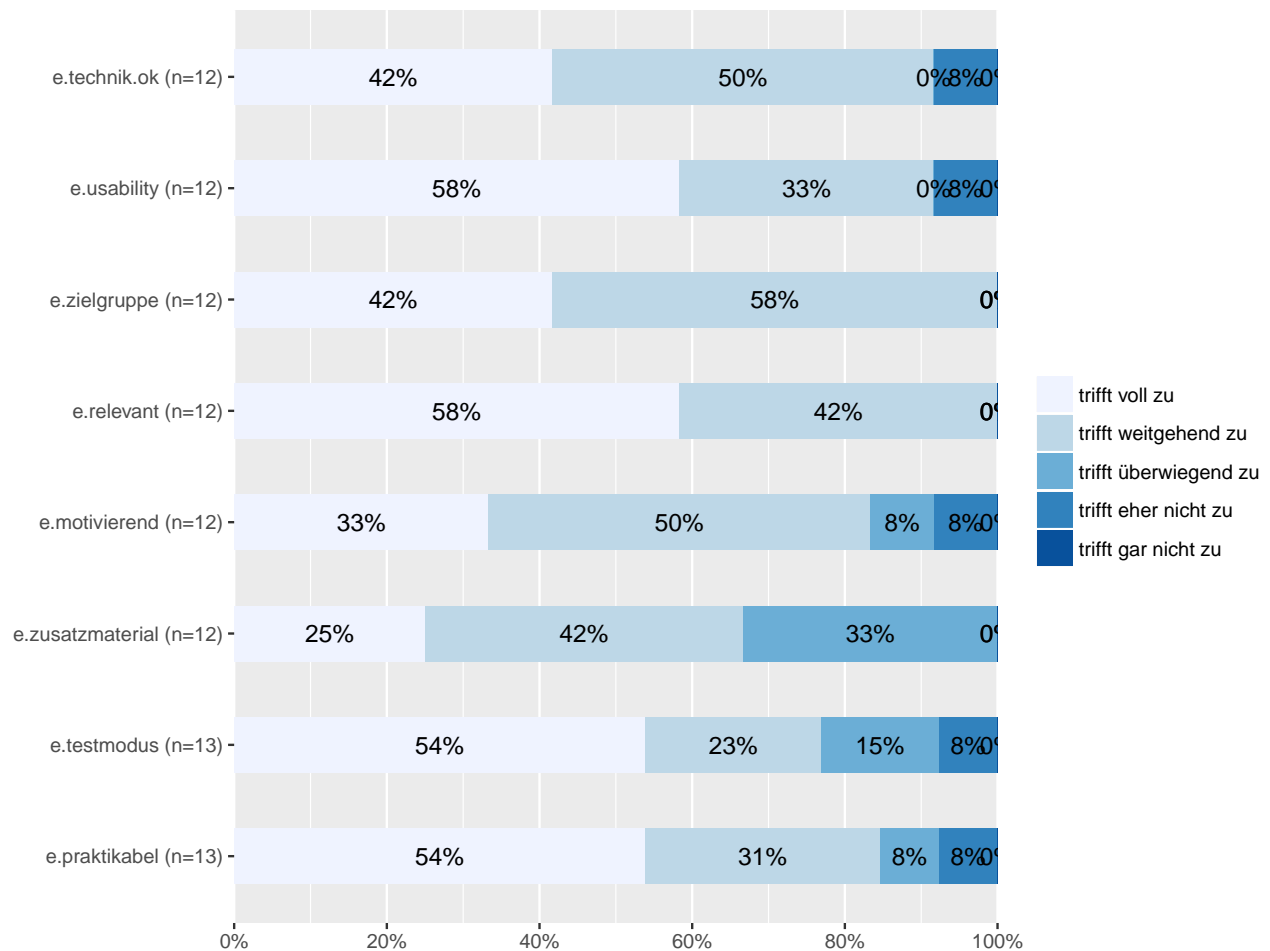
## IKT-Kompetenz

```
> library(sjPlot)
> ikt.kompetenz <- select(umfrage, c(15:10))
> sjp.stackfrq(ikt.kompetenz, digits = 0)
```



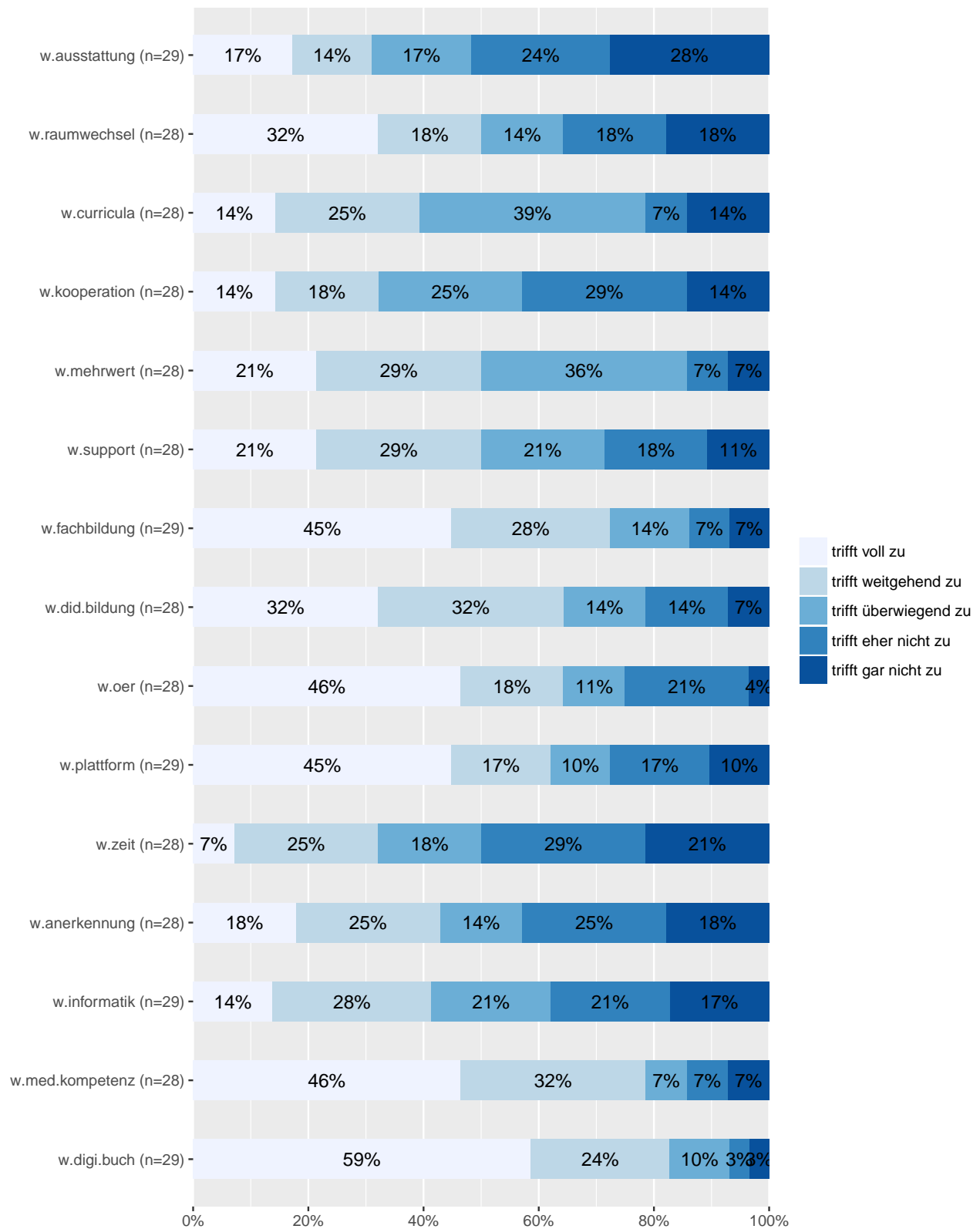
## Erfahrungen mit audiemus

```
> library(sjPlot)
> audiemus.erfahrung <- select(umfrage, c(52:45))
> sjp.stackfrq(audiemus.erfahrung, digits = 0)
```



## Förderliche Faktoren (Wünsche differenziert)

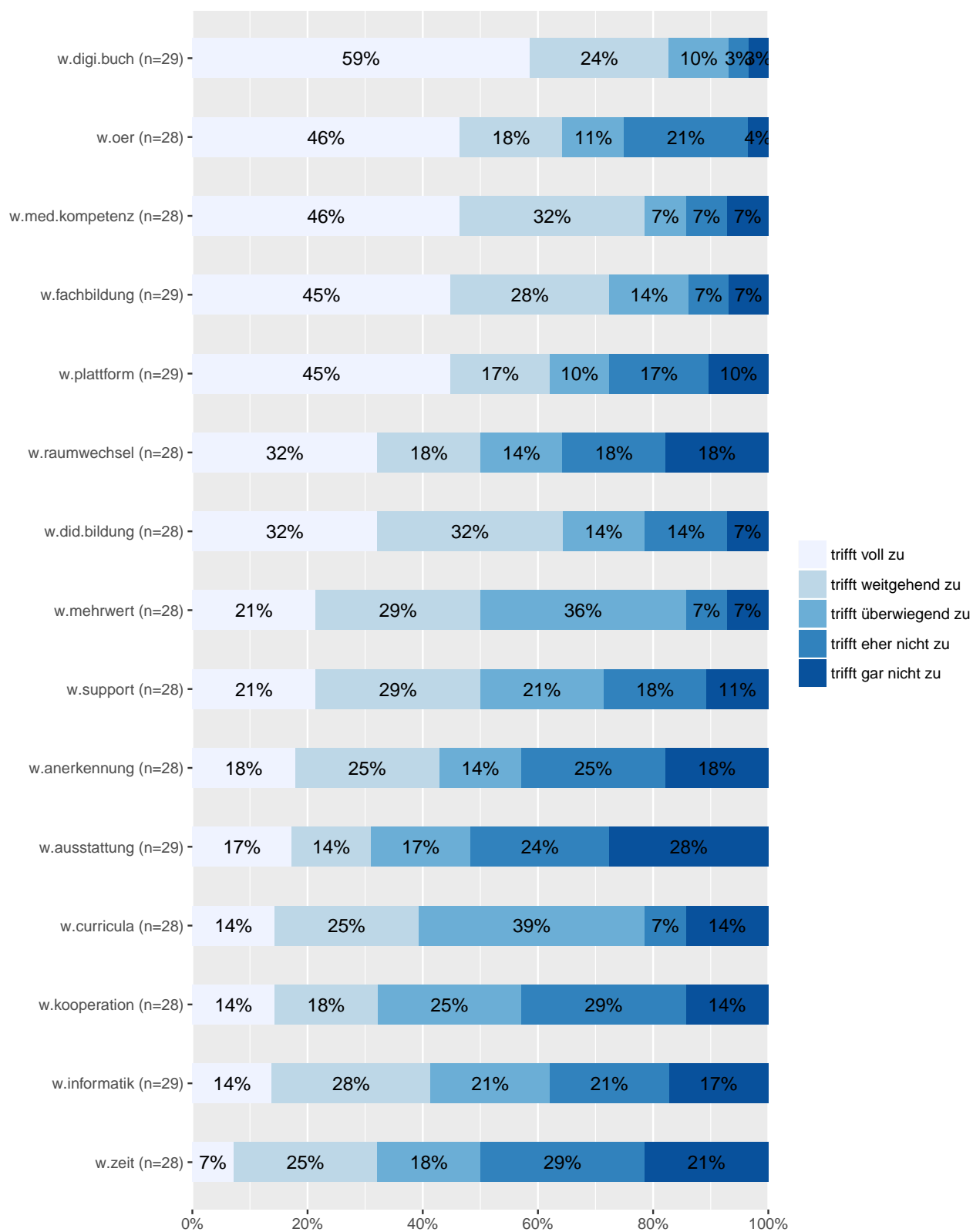
```
> library(sjPlot)
> wunsch <- select(umfrage, c(30:16))
> sjp.stackfrq(wunsch, digits = 0, geom.colors = "Blues")
```



## Wünsche differenziert und sortiert

Zum Vergleich nochmals die Wünsche differenziert, diesmal jedoch sortiert.

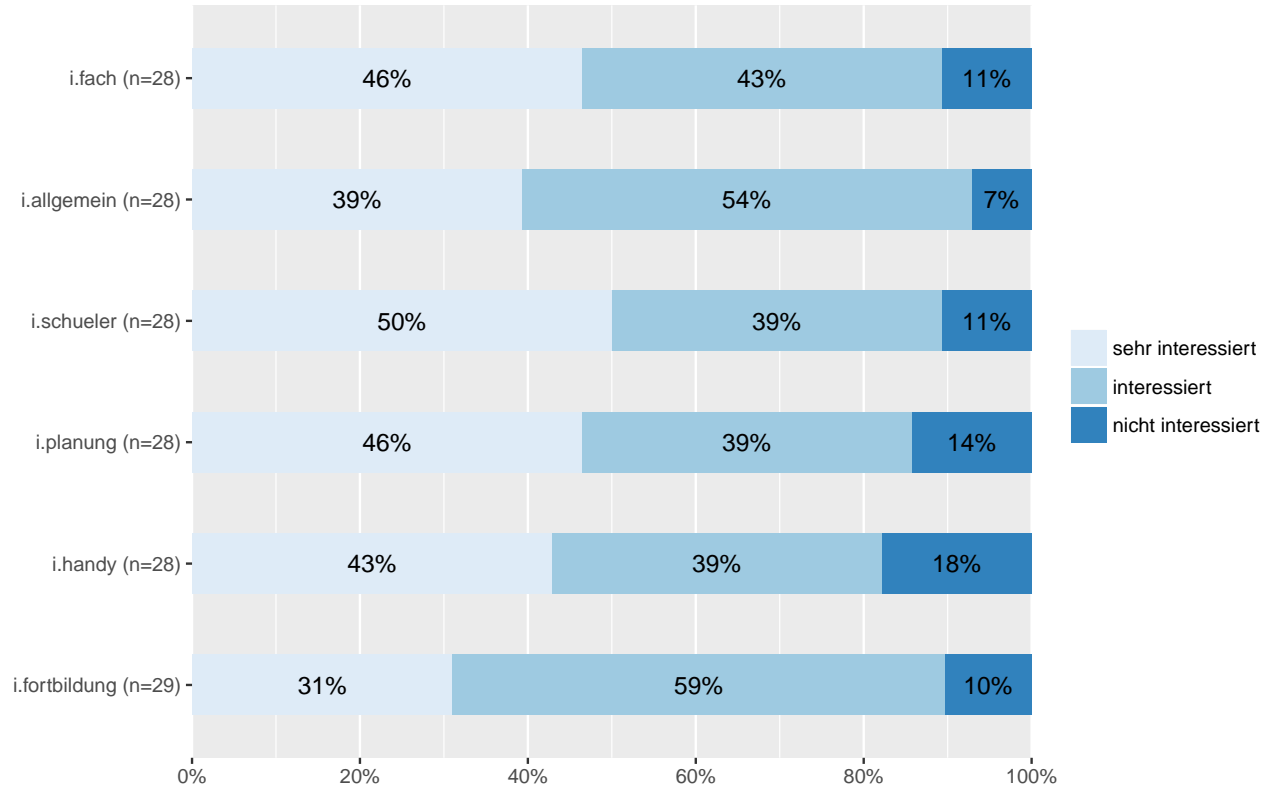
```
> library(sjPlot)
> wunsch <- select(umfrage, c(30:16))
> sjp.stackfrq(wunsch, digits = 0, geom.colors = "Blues", sort.frq = "first.asc")
```





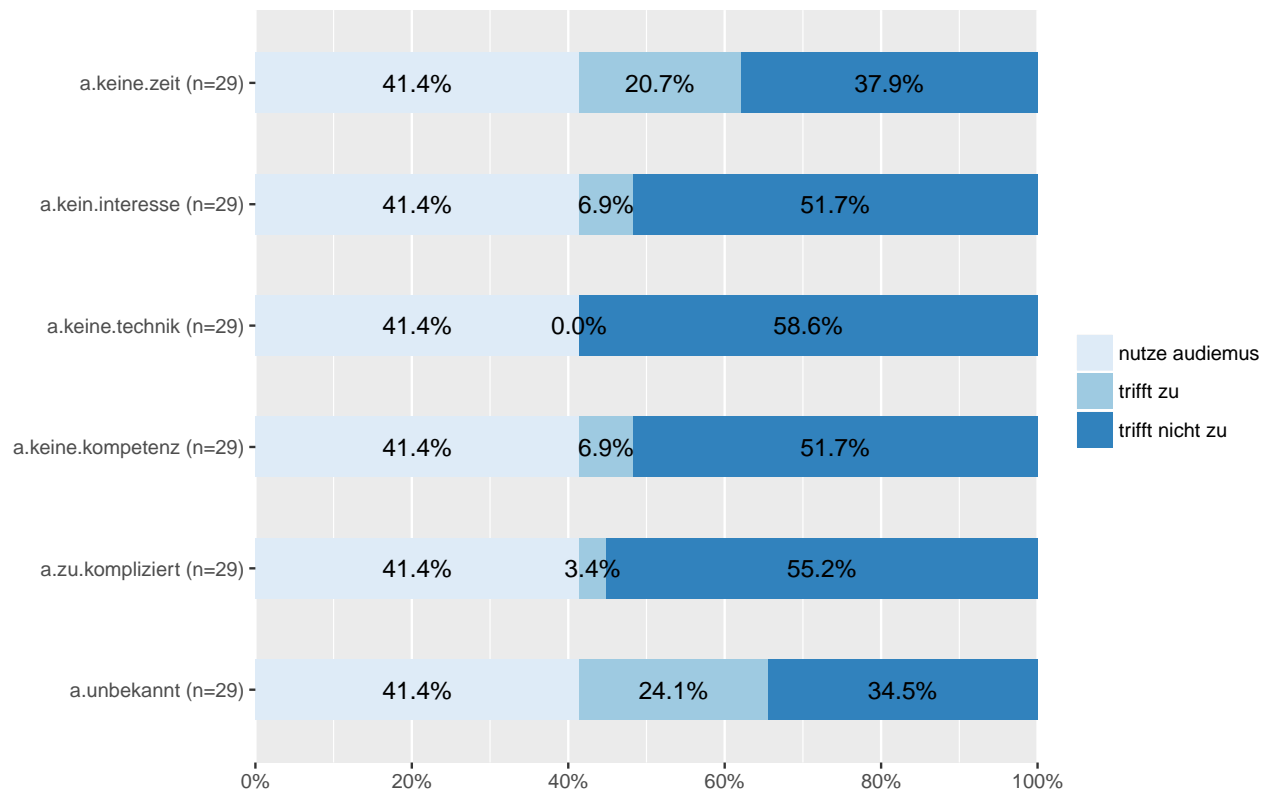
## Interesse an Fortbildung

```
> library(sjPlot)
> interesse <- select(umfrage, c(36:31))
> sjp.stackfrq(interesse, digits = 0)
```



## Warum wird nicht Audiumus benutzt?

```
> library(sjPlot)
> audiumus.nutzung <- select(umfrage, c(44:39))
> sjp.stackfrq(audiumus.nutzung, digits = 1)
```



## Förderliche Faktoren neu codiert

Für mich sind derzeit die Daten zu den Wünschen die interessanteste Abbildung. Einerseits, weil sie recht umfangreiche mögliche förderliche Faktoren abfragt, andererseits aber auch, weil die Ergebnisse zum Teil überraschend sind. So kommt z.B. recht deutlich heraus, dass der Wunsch nach einer besseren technischen Ausstattung mäßig ist und vor allem Zusatzmaterialien (w.oer und w.plattform) und interaktive digitale Schulbücher gewünscht werden.

In dieser Fragebatterie steckt sehr viel interessantes Material, das mit einer anderen Codierung vielleicht noch deutlicher visualisiert werden kann. Ich dichotomisiere die Skala, indem ich die Antwortvorgaben 1 - 3 ("trifft voll zu", "trifft weitgehend zu", "trifft überwiegend zu") auf "trifft zu" reduziere und die Antwortvorgaben 4 und 5 ("trifft eher nicht zu", "trifft gar nicht zu") in "trifft nicht zu" überführe.

```
> library(sjPlot)
> library(sjmisc)
> library(forcats)
>
> # diese funktion könnte generalisiert werden
> # indem noch weitere Parameter hinzugefügt werden,
> # die aus einer liste von zwei character vektoren bestehen
> # der erste vektor ist der neue (zusammengefasste) Level
> # der andere Vektor beinhaltet die Levels, die zusammengefasst werden
> my_fct_collapse <- function(df) {
+   j = ncol(df)
+   for (i in 1:j) {
+     df[[i]] <- fct_collapse(df[[i]],
+                             "trifft zu" = c("trifft voll zu",
+                                              "trifft weitgehend zu",
```

```

+           "trifft überwiegend zu"),
+           "trifft nicht zu" = c("trifft eher nicht zu",
+                                "trifft gar nicht zu"))
+       }
+       df
+ }
>
>
> wunsch.rec <- my_fct_collapse(wunsch)

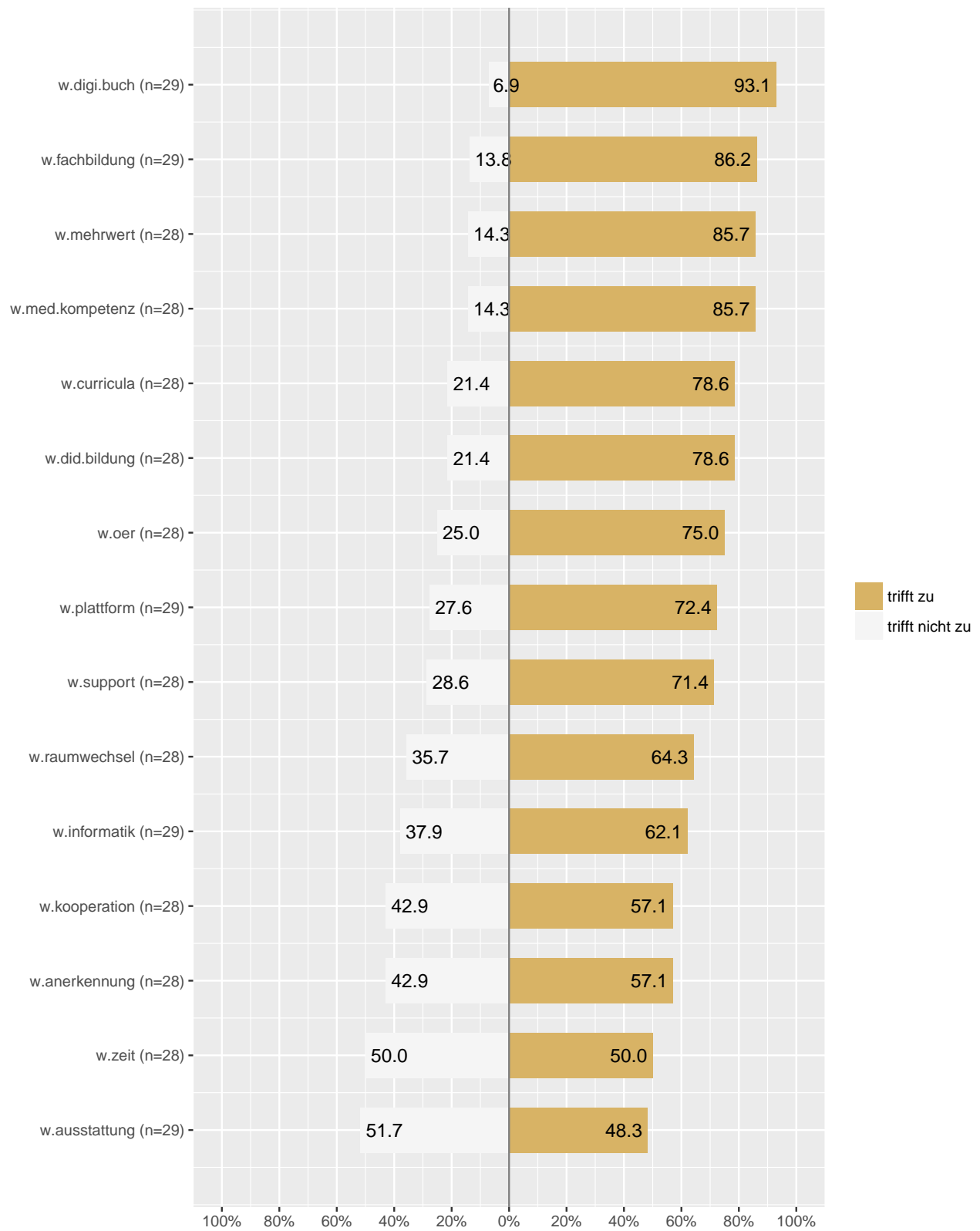
```

## Wünsche ja/nein mit zentrierte Linie

```

> sjp.likert(wunsch.rec, digits = 1, value = "sum.inside", geom.colors = "BrBG", sort.frq = "pos.desc")

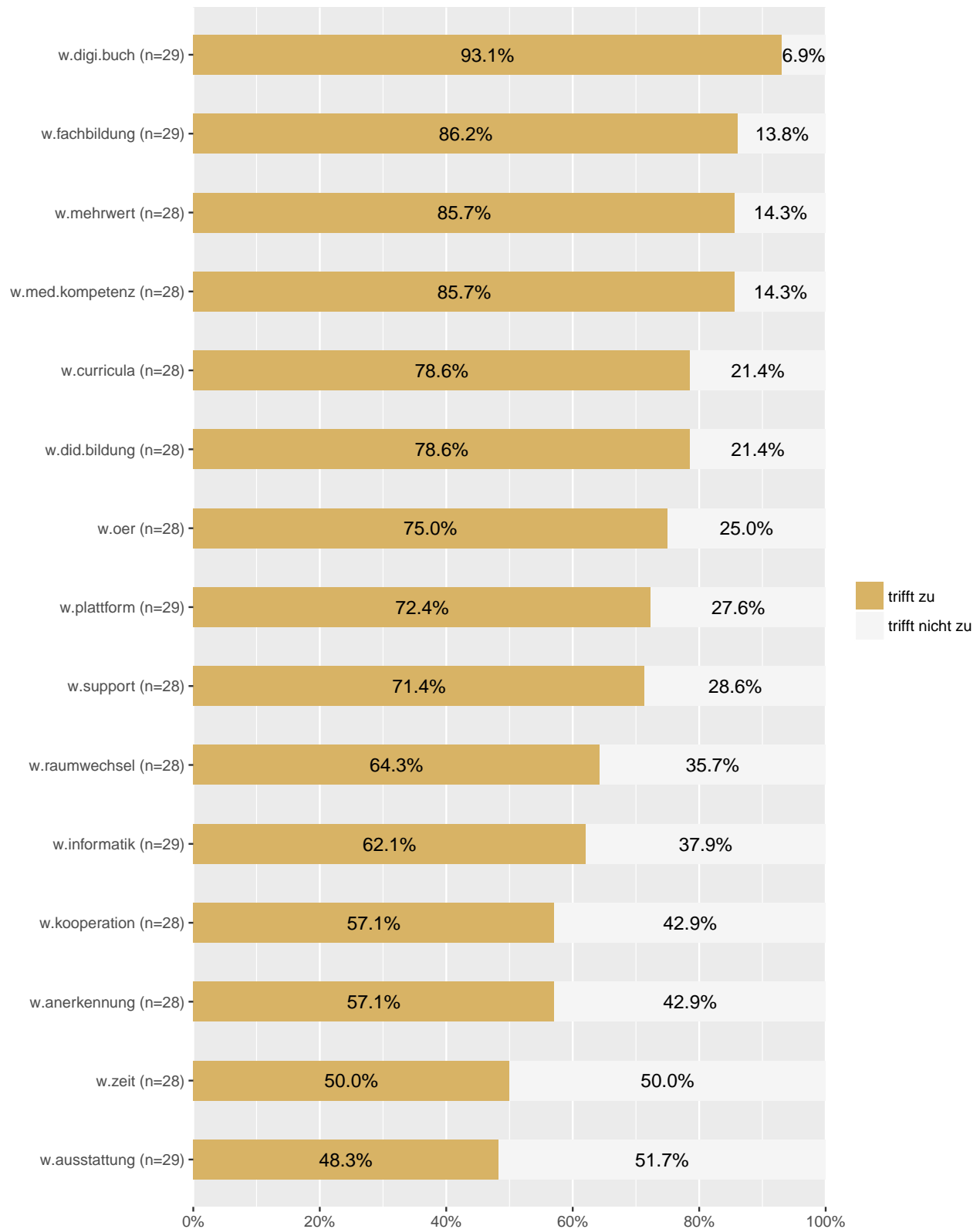
```



Die Ja/nein-Zentrierung à la Likter-Skala ist nicht so erhellend, wir das normale gestapelte Balkendiagramm.

## Balken-Diagramm: ja/nein Wünsche

```
> sjp.stackfrq(wunsch.rec, digits = 1, geom.colors = "BrBG", sort.frq = "first.asc")
```



# Verschiedene Signifikanz-Tests

## Chi-Quadrat Homogenitätstest

Ist die Verteilung von Häufigkeiten in verschiedenen Gruppen gleich, bzw. sind die Unterschiede nur durch Zufälle bestimmt?

Wegen der geringen Zahlen muss ich hier die Monte Carlo Simulation verwenden.

```
> table(wunsch.rec$w.ausstattung, wunsch.rec$w.fachbildung)
```

	trifft zu	trifft nicht zu
trifft zu	13	1
trifft nicht zu	12	3

```
> chisq.test(table(wunsch.rec$w.ausstattung, wunsch.rec$w.fachbildung), simulate.p.value = TRUE)
```

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

```
data: table(wunsch.rec$w.ausstattung, wunsch.rec$w.fachbildung)
X-squared = 1.0067, df = NA, p-value = 0.6172
```

Es ergibt sich hier keine Signifikanz, was mich wundert. Das Problem liegt darin, dass die Fallzahlen in der zweiten Spalte zu klein ist. Als Faustregel gilt, dass die zu erwarteten Häufigkeiten über 5 liegen sollten (Hatzinger, Hornik, Nagel, & Maier, 2014, S. 285).

## Exakter Test nach Fisher

Abhilfe schafft hier der exakte Test nach Fischer. Dabei “werden alle denkbaren Kreuztabellen mit denselben Randsummen wie in der Stichprobe bestimmt, die mindestens so stark wie die der Stichprobe gegen die Nullhypothese sprechen” (Hatzinger et al., 2014, S. 285).

```
> ff1 <- matrix(c(25, 4, 14, 15), nr = 2, byrow = TRUE)
> rownames(ff1) <- c("Fachdidaktik", "Ausstattung")
> colnames(ff1) <- c("Ja", "Nein")
> fisher.test(ff1)
```

Fisher's Exact Test for Count Data

```
data: ff1
p-value = 0.004459
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.641096 32.106631
sample estimates:
odds ratio
 6.463707
```

Hier gibt es nun ein deutliches signifikantes Ergebnis. Allerdings ist die händische Eingabe mit der Funktion `c` etwas umständlich.

```
> ff2 <- matrix(c(25, 4, 24, 4), nr = 2, byrow = TRUE)
> rownames(ff2) <- c("Fachdidaktik", "Medienkompetenz")
```

```
> colnames(ff2) <- c("Ja", "Nein")
> fisher.test(ff2)
```

Fisher's Exact Test for Count Data

```
data: ff2
p-value = 1
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.1727182 6.2728904
sample estimates:
odds ratio
 1.040936
```

```
> ff3 <- matrix(c(25, 4, 22, 6), nr = 2, byrow = TRUE)
> rownames(ff3) <- c("Fachdidaktik", "Didaktik")
> colnames(ff3) <- c("Ja", "Nein")
> fisher.test(ff3)
```

Fisher's Exact Test for Count Data

```
data: ff3
p-value = 0.5045
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.3475197 9.2523329
sample estimates:
odds ratio
 1.688656
```

```
> ff4 <- matrix(c(22, 6, 14, 15), nr = 2, byrow = TRUE)
> rownames(ff4) <- c("Didaktik", "Ausstattung")
> colnames(ff4) <- c("Ja", "Nein")
> fisher.test(ff4)
```

Fisher's Exact Test for Count Data

```
data: ff4
p-value = 0.02768
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.085319 15.160749
sample estimates:
odds ratio
 3.83034
```

```
> ff5 <- matrix(c(24, 4, 14, 15), nr = 2, byrow = TRUE)
> rownames(ff5) <- c("Medienkompetenz", "Ausstattung")
> colnames(ff5) <- c("Ja", "Nein")
> fisher.test(ff5)
```

Fisher's Exact Test for Count Data

```

data: ff5
p-value = 0.004459
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.570342 30.936647
sample estimates:
odds ratio
 6.2076

```

## Odds-Ratio-Test (Chancenverhältnis)

Hier wird ebenfalls gefragt, ob sich zwei Gruppen in ihren Häufigkeiten unterscheiden. Der Chi-Quadrat und Fisher-Test sagt bloß aus, ob es signifikante Unterschiede gibt, aber nicht wie groß die Unterschiede sind. Das bedeutet, dass bei großen Stichproben selbst kleine Unterschiede (Effektgrößen) signifikant sind.

Mit dem Test zum Odds-Ratio-Test wird nach dem Verhältnis der Anteile gefragt, der unabhängig von der Stichprobe ist.

```

> or.ff1 <- odds_ratios(as.table(ff1))
> summary(or.ff1)

```

Odds Ratios

	Ja	Nein
Fachdidaktik	25	4
Ausstattung	14	15

```

Odds-Ratio = 6.696
Log(Odds-Ratio) = 1.902, Standard Error = 0.654
z-value = 2.906, p-value = .00366

```

```

> or.ff2 <- odds_ratios(as.table(ff2))
> summary(or.ff2)

```

Odds Ratios

	Ja	Nein
Fachdidaktik	25	4
Medienkompetenz	24	4

```

Odds-Ratio = 1.042
Log(Odds-Ratio) = 0.041, Standard Error = 0.763
z-value = 0.054, p-value = .95731

```

```

> or.ff3 <- odds_ratios(as.table(ff3))
> summary(or.ff3)

```

Odds Ratios

	Ja	Nein
Fachdidaktik	25	4
Didaktik	22	6

```

Odds-Ratio = 1.705
Log(Odds-Ratio) = 0.533, Standard Error = 0.709

```



z-value = 0.753, p-value = .45169

```
> or.ff4 <- odds_ratios(as.table(ff4))
> summary(or.ff4)
```

Odds Ratios

	Ja	Nein
Didaktik	22	6
Ausstattung	14	15

Odds-Ratio = 3.929

Log(Odds-Ratio) = 1.368, Standard Error = 0.592

z-value = 2.312, p-value = .02077

```
> or.ff5 <- odds_ratios(as.table(ff5))
> summary(or.ff5)
```

Odds Ratios

	Ja	Nein
Medienkompetenz	24	4
Ausstattung	14	15

Odds-Ratio = 6.429

Log(Odds-Ratio) = 1.861, Standard Error = 0.656

z-value = 2.838, p-value = .00453

Hier zeigt sich nun, dass der Unterschied von didaktischer Ausbildung zur Ausstattung zwar noch auf dem Niveau  $p < 0,05$  signifikant ist, aber diese Effektstärke im Vergleich zum Chancenverhältnis von Fachdidaktik und Ausstattung nur etwa halb so groß ist.

## Literatur

Hatzinger, D. R., Hornik, K., Nagel, D. H., & Maier, M. J. (2014). *R: Einführung durch angewandte Statistik* (2nd ed.). Hallbergmoos: Pearson Studium.