Übung 6

Lilli Schuckert

19 5 2021

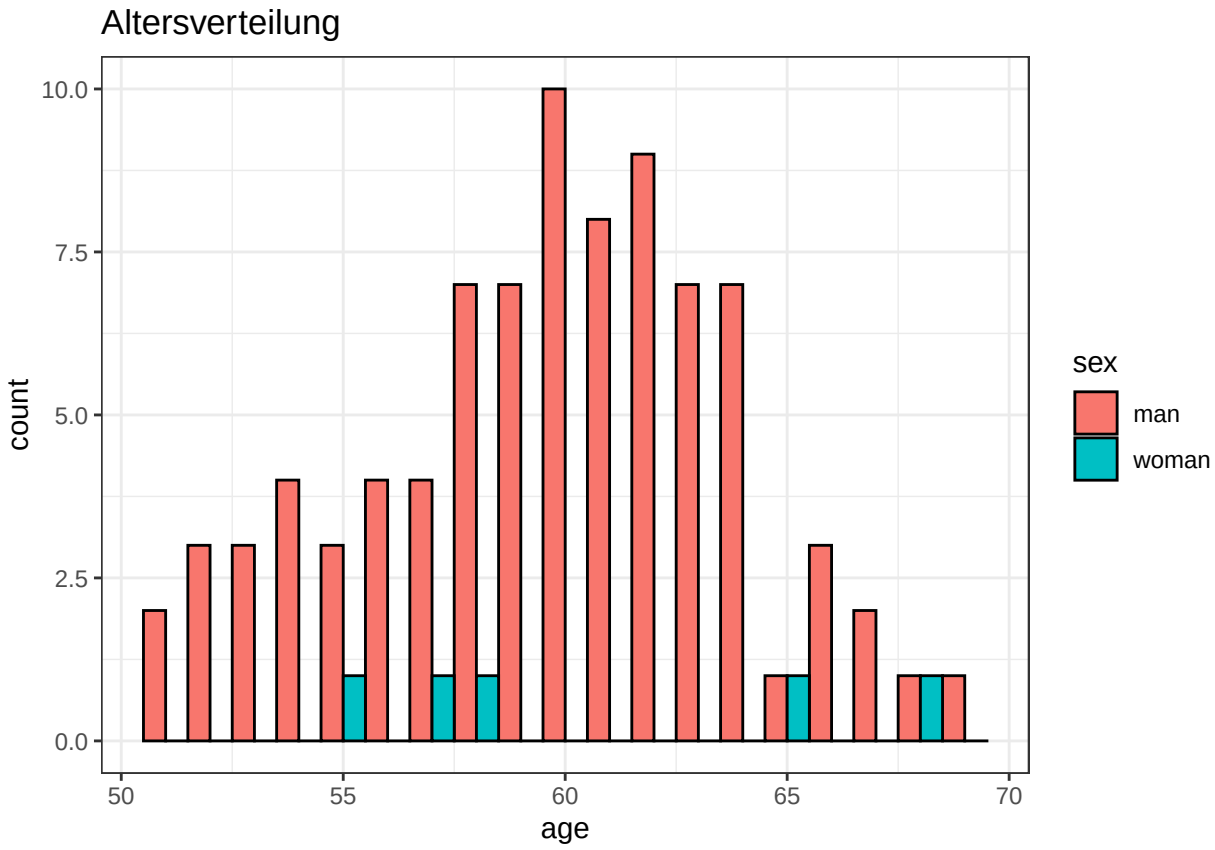
Forschungsfrage:

Gibt es einen Zusammenhang zwischen einer Akupunkturbehandlung und Migräneschmerzen?

```
## ----- Age -----
summary(data$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   51.00   57.00   60.00   59.79   63.00   69.00
```

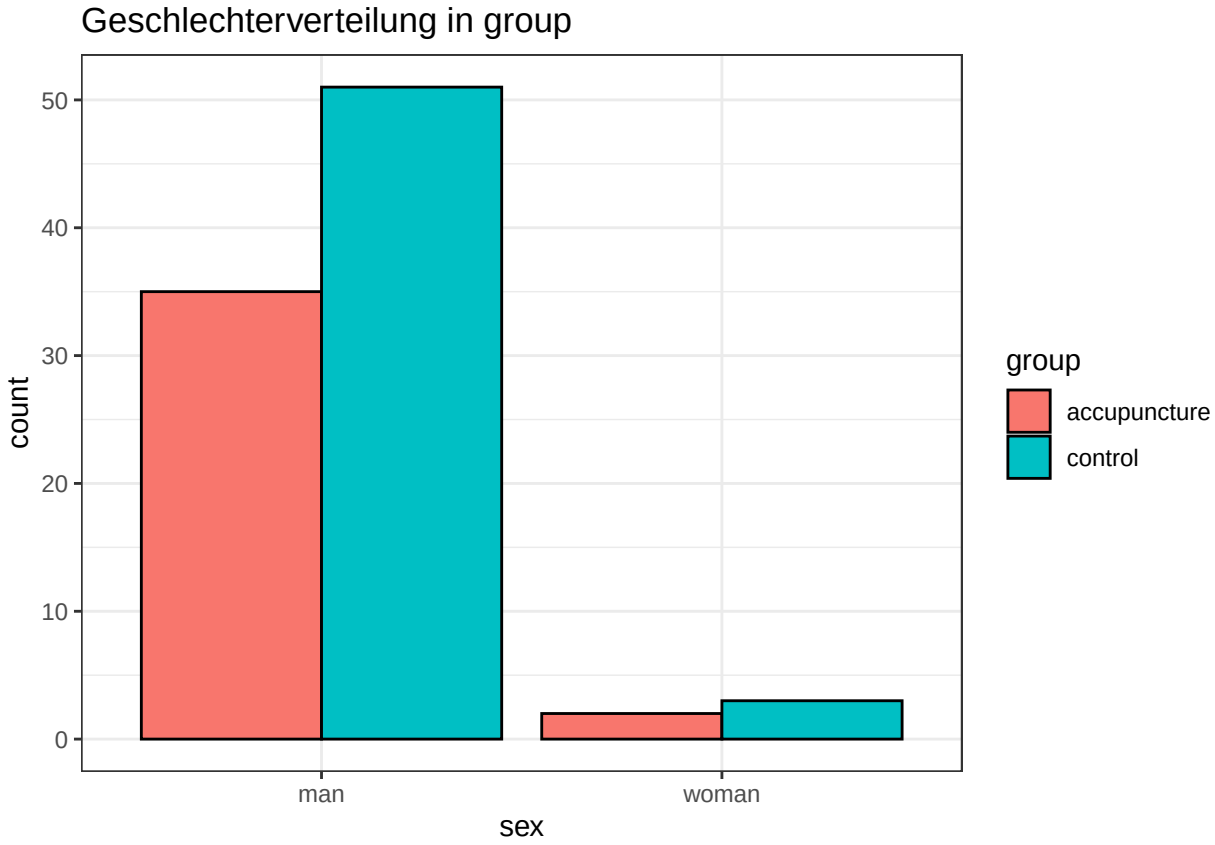
```
ggplot(data, aes(age)) +
  geom_histogram(aes(fill = sex), binwidth = 1, color = "black", position = "dodge")+
  labs(title = "Altersverteilung")+
  theme_bw()
```




```
## ----- Sex -----
table(data$sex)
```

```
##
##   man woman
##    86     5
```

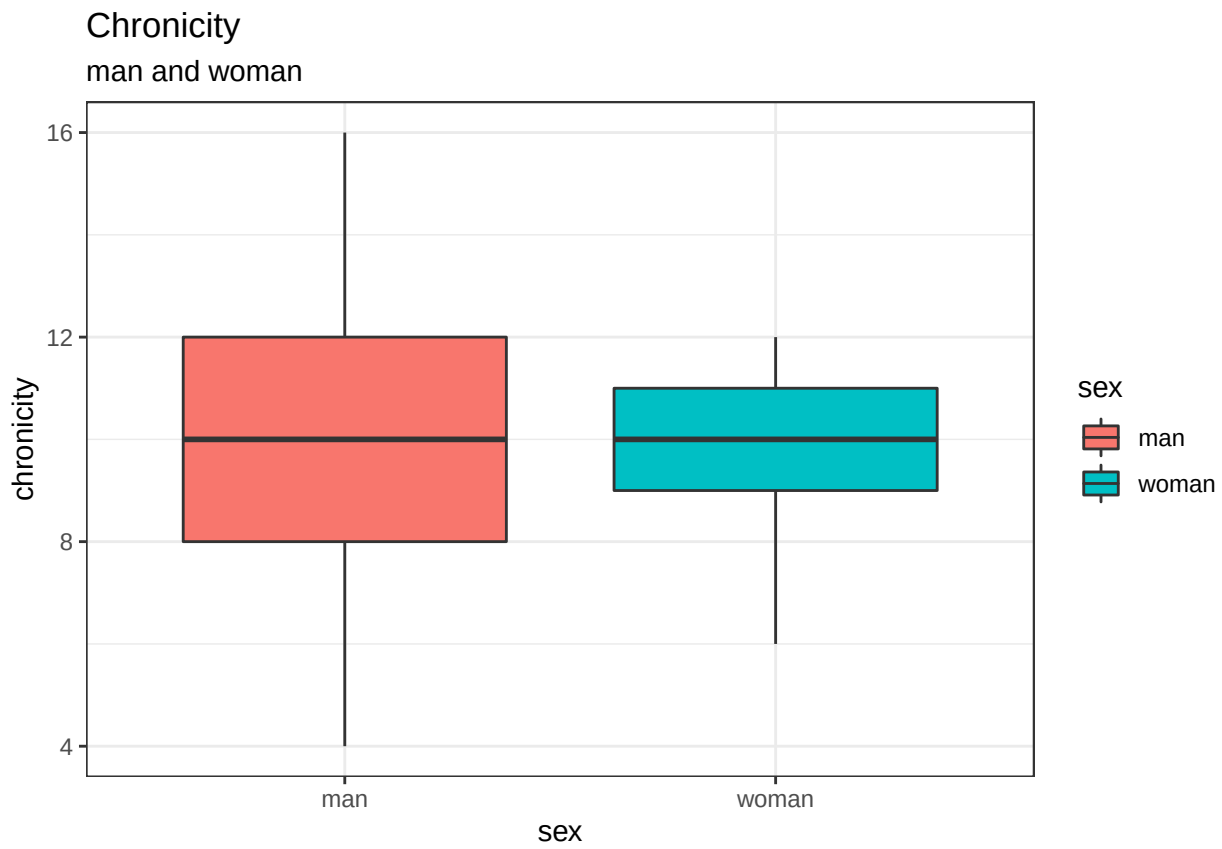
```
ggplot(data, aes(x=sex, y = stat(count), group_by(group), fill = group))+
  geom_bar(position = "dodge", color = "black")+
  theme_bw()+
  labs(title = "Geschlechterverteilung in group")
```



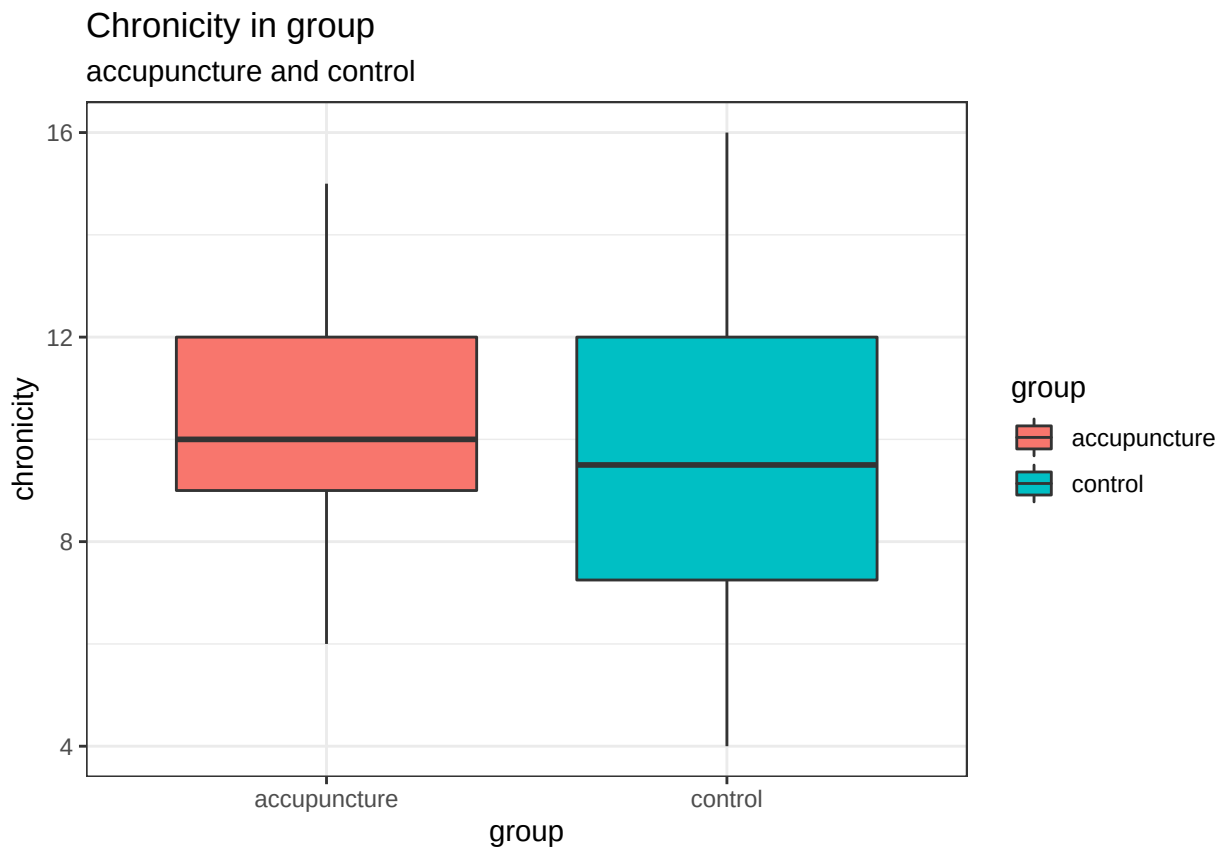

```
## ----- Chronicity -----
summary(data$chronicity)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.00    8.00   10.00   10.02   12.00   16.00
```

```
ggplot(data, aes(x=sex, y =chronicity))+
  geom_boxplot(aes(fill = sex)) +
  labs(title = "Chronicity", subtitle = "man and woman")+
  theme_bw()
```




```
ggplot(data, aes(x=group, y =chronicity))+
  geom_boxplot(aes(fill = group)) +
  labs(title = "Chronicity in group", subtitle = "accupuncture and control")+
  theme_bw()
```




```
## ----- Migraine Score in Accupuncture and Control Group  -----
summary(data$migraine_score)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.00   18.00   20.00   20.67   23.00   29.00
```

```
## Accupuncture Group:
data_migraine_score_accu <- filter(data, group == "accupuncture")
data_accupuncture <- data_migraine_score_accu %>%
  select(migraine_score, group)
summary(data_accupuncture)
```

```
##  migraine_score     group
##  Min.   :21.00   Length:37
##  1st Qu.:23.00   Class :character
##  Median :24.00   Mode  :character
##  Mean   :24.62
##  3rd Qu.:27.00
##  Max.   :29.00
```

```
cat("Mean age in Accupuncture group:", mean(data_migraine_score_accu$age), "\n")
```

```
## Mean age in Accupuncture group: 60.05405
```

```
cat("Mean chronicity in Accupuncture group:", mean(data_migraine_score_accu$chronicity), "\n")
```

```
## Mean chronicity in Accupuncture group: 10.59459
```

```r
table(data_migraine_score_accu$sex)
```

```
##
##   man woman
##    35     2
```

```r
cat("Gender distribution in %: ", "\n", "male:", round(100/37*35, digits=2), "%",
    "\n", "female:", round(100/37*2, digits = 2), "%")
```

```
## Gender distribution in %:
##  male: 94.59 %
##  female: 5.41 %
```

```r
## Control Group:
data_migraine_score_cont <- filter(data, group == "control")
data_control <- data_migraine_score_cont %>%
  select(migraine_score, group)
summary(data_control)
```

```
##  migraine_score     group
##  Min.   :12.00   Length:54
##  1st Qu.:16.00   Class :character
##  Median :18.00   Mode  :character
##  Mean   :17.96
##  3rd Qu.:20.00
##  Max.   :23.00
```

```r
cat("Mean age in Control group:", mean(data_migraine_score_cont$age), "\n")
```

```
## Mean age in Control group: 59.61111
```

```r
cat("Mean chronicity in Control group:", mean(data_migraine_score_cont$chronicity), "\n")
```

```
## Mean chronicity in Control group: 9.62963
```

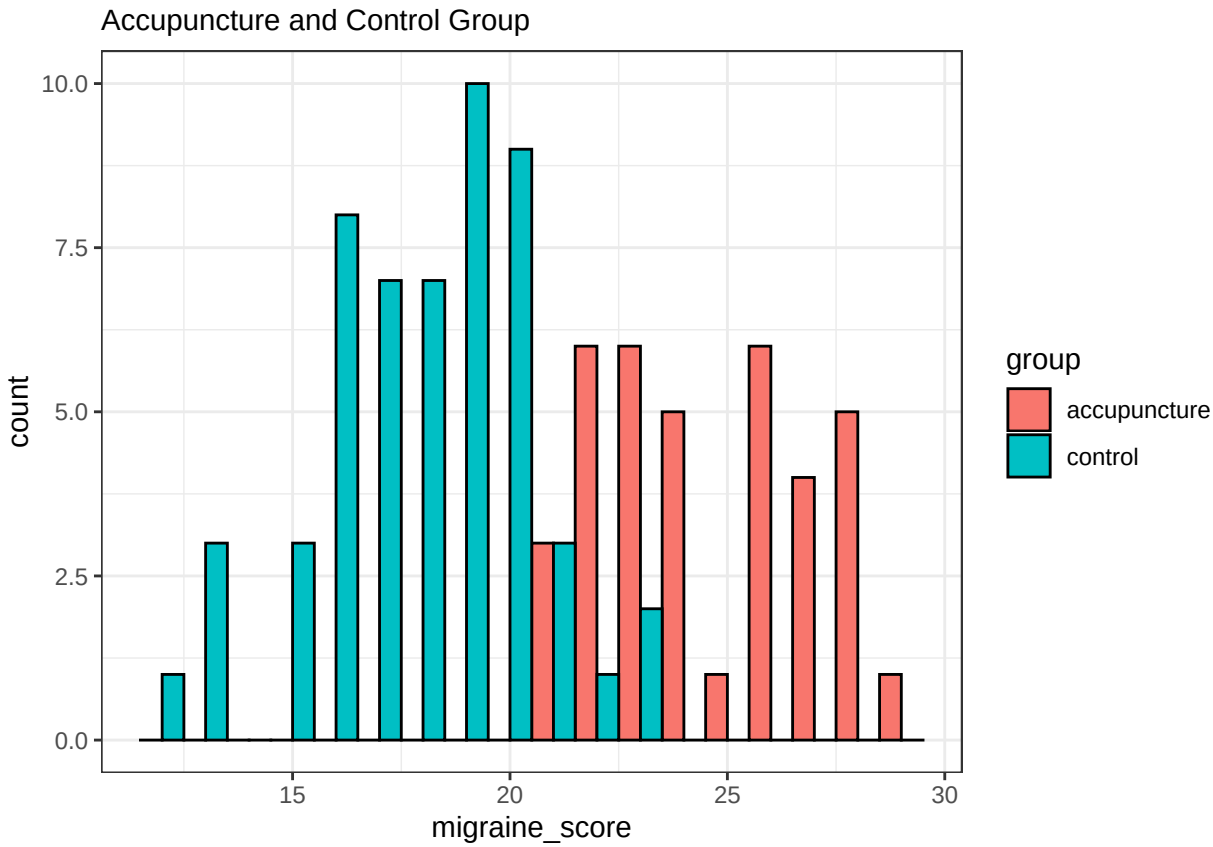
```r
table(data_migraine_score_cont$sex)
```

```
##
##   man woman
##    51     3
```

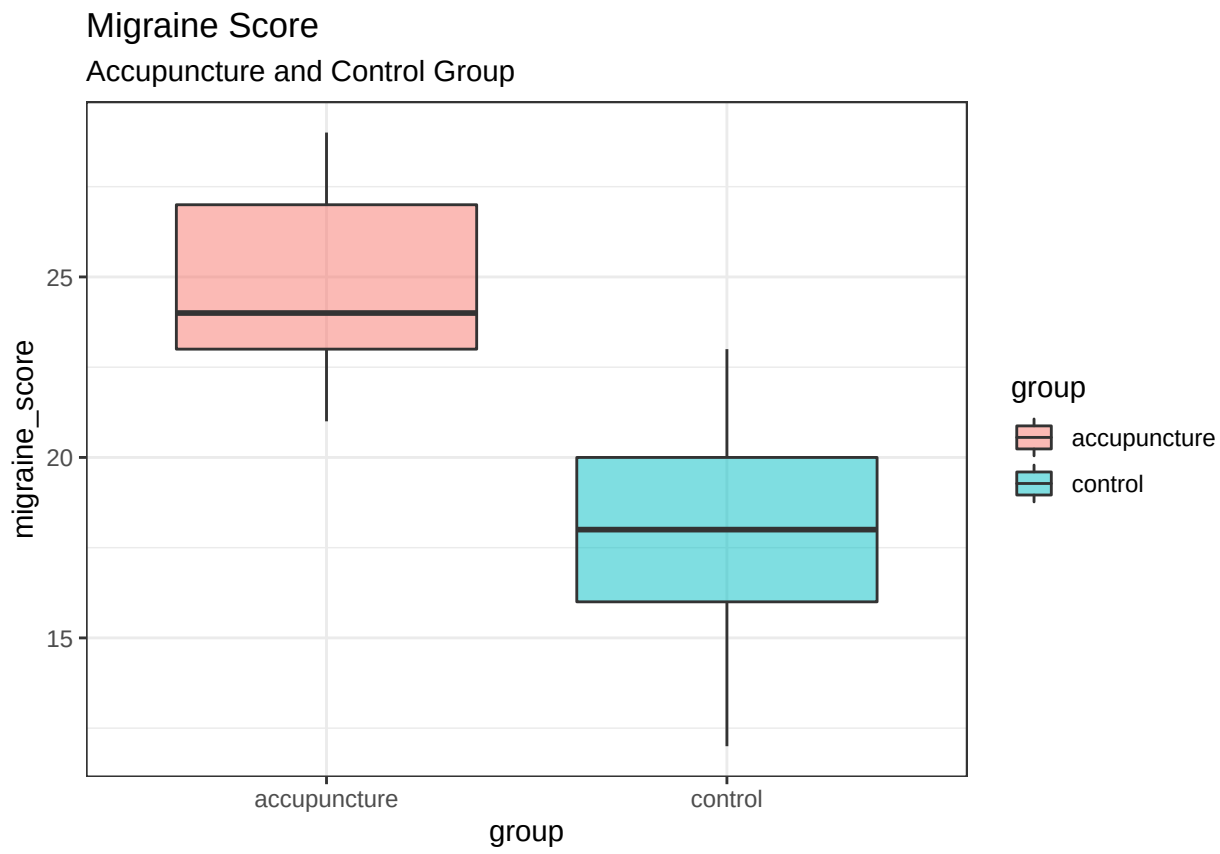
```r
cat("Gender distribution in %: ", "\n", "male:", round(100/54*51, digits=2), "%",
    "\n", "female:", round(100/54*3, digits = 2), "%")
```

```
## Gender distribution in %:
##  male: 94.44 %
##  female: 5.56 %
```

```r
## ----- Migraine_score visualization -----
ggplot(data, aes(migraine_score)) +
  geom_histogram(aes(fill = group), binwidth = 1, color = "black", position = "dodge")+
  labs("Migraine Score", subtitle = "Accupuncture and Control Group")+
  theme_bw()
```

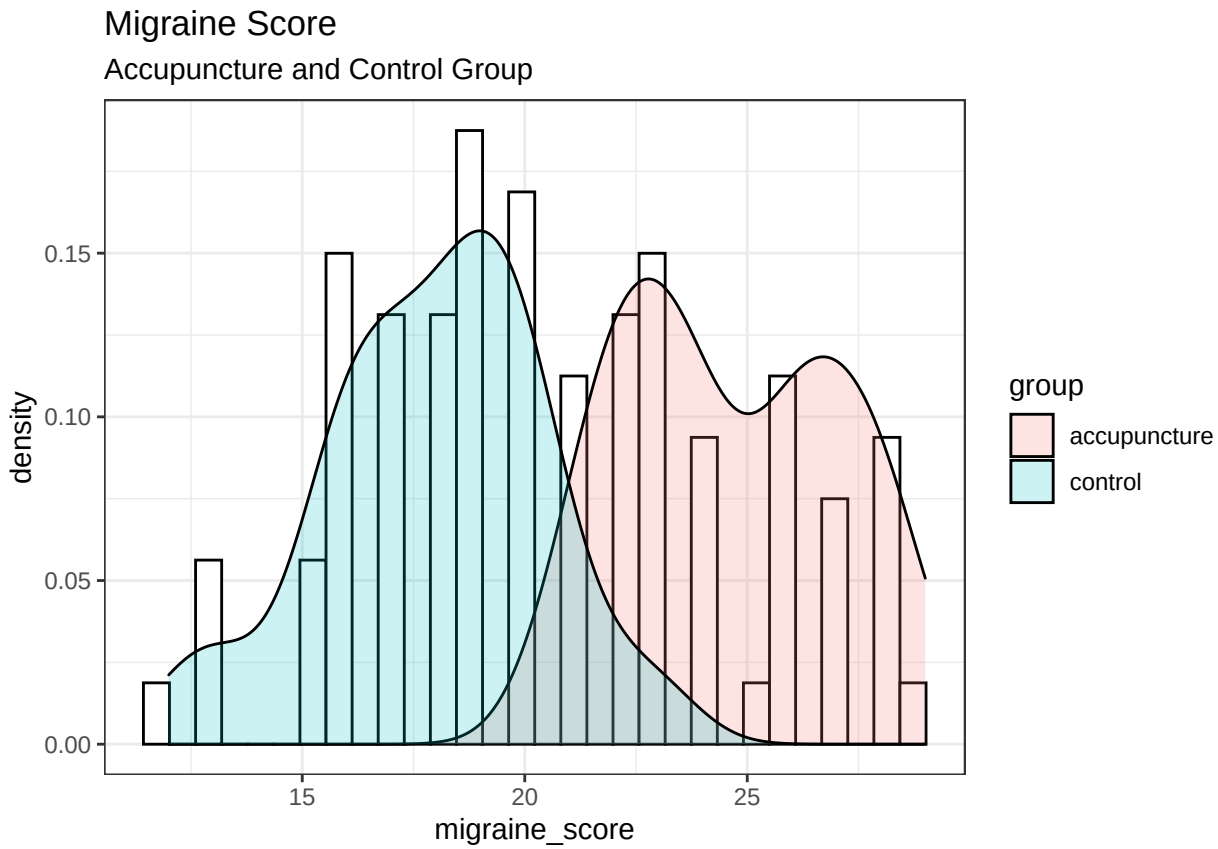



```
ggplot(data = data, aes(x = group, y = migraine_score, fill = group))+
  geom_boxplot(alpha = 0.5) +
  labs(title = "Migraine Score", subtitle = "Accupuncture and Control Group") +
  theme_bw()
```




```
ggplot(data = data, aes(x=migraine_score, fill=group)) +
  geom_histogram(aes(y=..density..), colour="black", fill="white")+
  geom_density(alpha = 0.2)+
  labs(title = "Migraine Score", subtitle = "Accupuncture and Control Group") +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Statistische Hypothesen:

H0: Es gibt keinen Zusammenhang zwischen dem Migränescore und einer Akkupunkturbehandlung H1: Es gibt einen Zusammenhang zwischen dem Migränescore und einer Akkupunkturbehandlung

Aufgabe 1

Rechnen und präsentieren Sie Ihre Prüfung Statistik für Biowissenschaften I und Daten mit der Methode„t-Test“.

```
summary(data$group)
```

```
##    Length     Class      Mode
##        91 character character
```

```
mosaic::var(migraine_score ~ group, data = data)
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2
```

```
## accupuncture      control
##     5.797297     5.960867
```

```
t.test(migraine_score ~ group, data = data)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  migraine_score by group
## t = 12.885, df = 78.222, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  5.629858 7.687459
## sample estimates:
## mean in group accupuncture      mean in group control
##                   24.62162                   17.96296
```

```
t.test(migraine_score ~ group, data = data, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  migraine_score by group
## t = 12.851, df = 89, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  5.629109 7.688208
## sample estimates:
## mean in group accupuncture      mean in group control
##                   24.62162                   17.96296
```

```
##t-Wert: signifikant
##p-Wert < 0.05 --> wir kÃ¶nnen Nullhypothese ablehnen
##Effektschätzer beim t-test Mittelwertsunterschied

effektschätzer = 24.62 - 17.96
effektschätzer
```

```
## [1] 6.66
```

```
# 0 ist nicht im Intervall --> H0 ablehnen
```

Aufgabe 2

Rechnen und präsentieren Sie Ihre Prüfung Statistik für Biowissenschaften I und Daten mit der Methode„Chi-Quadrat-Test“.

```
##Dichotomisierung von migraine_score in binäre Variablen mit
##  cutpointr(data, migraine_score, group) --> >=21
data1$migraine_score <- ifelse(data1$migraine_score >=21,1,0)
data1$group[data1$group == "accupuncture"]  <- 1
data1$group[data1$group == "control"]  <- 0
view(data1)

## 2x2 Tafel migraine_score ~ group
migraineXgroup <- xtabs(~migraine_score + group, data1)[2:1,2:1]
migraineXgroup
```

```
##                group
## migraine_score  1  0
##              1 37  6
##              0  0 48
```

```
## --> group 1 und migraine_score 0 enthält 0 Probanden
## --> mathematisch genaue Berechnung des Chi-Quadrattests nicht möglich
## --> hier: Fisher-Test
```

```r
## Manipulated 2x2 Tafel migraine_score ~ group
migraineXgroupXmanipulated <- matrix(c(32, 5, 6, 48), nrow = 2, byrow = FALSE)
rownames(migraineXgroupXmanipulated) <- c("1", "0")
colnames(migraineXgroupXmanipulated) <- c("1", "0")
migraineXgroupXmanipulated
```

```
##    1  0
## 1 32  6
## 0  5 48
```

```r
## Chi-Quadrat-Test
chisq.test(migraineXgroupXmanipulated , correct = FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  migraineXgroupXmanipulated
## X-squared = 51.29, df = 1, p-value = 7.966e-13
```

```r
##    --> df = (2-1)*(2-1) = 1
##    --> reject H0: p-value = 1.801*10^-5 < alpha = 0,05

##-------------------------------------------------------------------------------
##Konfidenzintervalle
xlaccu <- mean(data_accupuncture$migraine_score) - 1.96*(sd(data_accupuncture$migraine_score)/sqrt(37))
xraccu <- mean(data_accupuncture$migraine_score) + 1.96*(sd(data_accupuncture$migraine_score)/sqrt(37))

xlcon <- mean(data_control$migraine_score) - 1.96*(sd(data_control$migraine_score)/sqrt(54))
xrcon <- mean(data_control$migraine_score) + 1.96*(sd(data_control$migraine_score)/sqrt(54))

cat("KI Accupuncture:[", xlaccu, " ; ", xraccu, "]", "\n","KI Control:[", xlcon, " ; ", xrcon, "]")
```

```
## KI Accupuncture:[ 23.84579  ;  25.39745 ]
##  KI Control:[ 17.31176  ;  18.61416 ]
```

```r
## --> zu 95%iger Wahrscheinlichkeit liegt der wahre Mittelwert der Grundgesamtheit in dem KI

##-------------------------------------------------------------------------------
##Effektschätzer
OR <- function(a, b, c, d)
{
  return ((a*d)/(c*b))
}
cat("OR: ", OR(32, 6, 5, 48))
```

```
## OR:  51.2
```

```r
## --> OR:  51.2 --> "Risiko" : Akkupunktur erhöht das Risiko für einen höheren Migraine Score um das 5
```

##Aufgabe 3 Rechnen und präsentieren Sie Ihre Prüfung Statistik für Biowissenschaften I und Daten mit der Methode „Gaussian linear regression“

```r
gauss_lin_reg <- lm(migraine_score ~ group, data = data_gaussian, family = gaussian)
```

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'family' will be disregarded
```

```
gauss_lin_reg %>% tidy(conf.int = TRUE)
```

```
## # A tibble: 2 x 7
##   term         estimate std.error statistic  p.value conf.low conf.high
##   <chr>           <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
## 1 (Intercept)     24.6      0.399      61.7 8.04e-75    23.8      25.4
## 2 groupcontrol    -6.66     0.518     -12.9 5.48e-22    -7.69     -5.63
```

```
gauss_lin_reg %>% summary
```

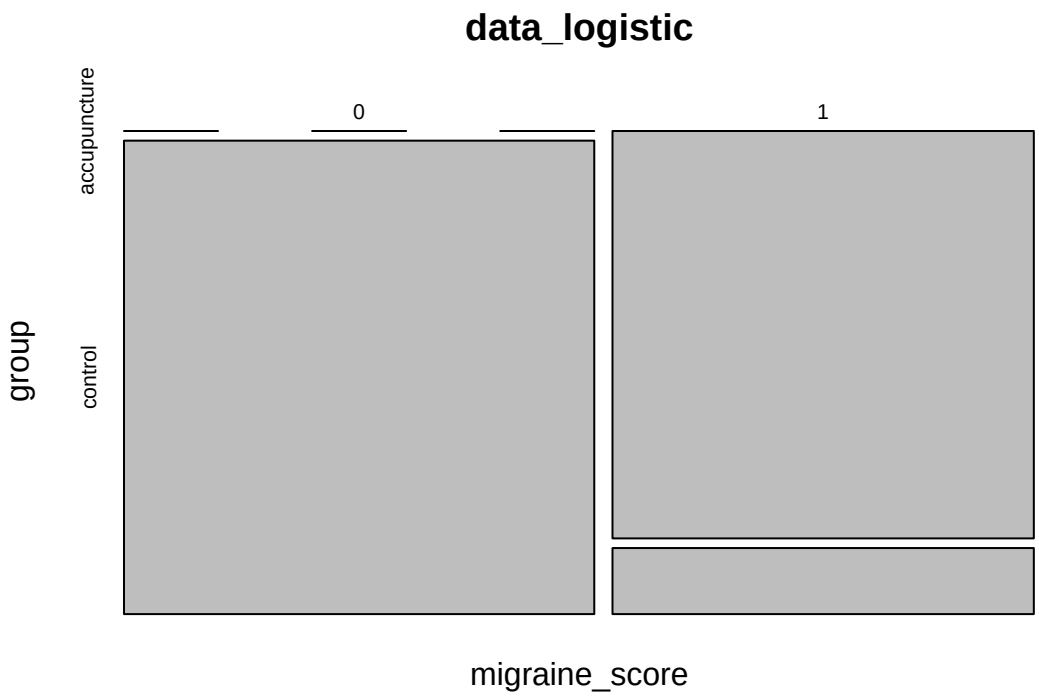
```
##
## Call:
## lm(formula = migraine_score ~ group, data = data_gaussian, family = gaussian)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.963 -1.963  0.037  2.037  5.037
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   24.6216     0.3991   61.69   <2e-16 ***
## groupcontrol  -6.6587     0.5181  -12.85   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.428 on 89 degrees of freedom
## Multiple R-squared:  0.6498, Adjusted R-squared:  0.6459
## F-statistic: 165.1 on 1 and 89 DF,  p-value: < 2.2e-16
```

```
##p-Wert --> signifikant --> Ho ablehnen
## R-squared = 0.51 --> die Varianz von der abhÃ¤ngigen Variable (migraine_score) wird zu 51% von der e

##zu 95% ist der Mittelwert der Grundgesamtheit zwischen 23.2 und 24.84
```

Aufgabe 4

Rechnen und präsentieren Sie Ihre Prüfung Statistik für Biowissenschaften I und Daten mit der Methode „Logistic linear regression“

```
## cutpointr(data_logistic, migraine_score, group) --> >=21
data_logistic$migraine_score <- ifelse(data_logistic$migraine_score >=21,1,0)
data1$group[data_logistic$group == "accupuncture"]  <- 1
data1$group[data_logistic$group == "control"]  <- 0

mosaicplot(migraine_score ~ group, data = data_logistic)
```




```
log_lin_reg <- glm(migraine_score ~ group, data = data_logistic,
                   family = binomial) %>%
  tidy(exponenciate = TRUE)

log_lin_reg %>% summary()
```

```
##      term              estimate          std.error      statistic
##  Length:2           Min.   :-22.646   Min.   :2915   Min.   :-0.0077690
##  Class :character   1st Qu.:-11.843   1st Qu.:2915   1st Qu.:-0.0040628
##  Mode  :character   Median : -1.040   Median :2915   Median :-0.0003567
##                     Mean   : -1.040   Mean   :2915   Mean   :-0.0003567
##                     3rd Qu.:  9.763   3rd Qu.:2915   3rd Qu.: 0.0033495
##                     Max.   : 20.566   Max.   :2915   Max.   : 0.0070556
##     p.value
##  Min.   :0.9938
##  1st Qu.:0.9939
##  Median :0.9941
##  Mean   :0.9941
##  3rd Qu.:0.9942
##  Max.   :0.9944
```

```
log_lin_reg
```

```
## # A tibble: 2 x 5
##   term         estimate std.error statistic p.value
##   <chr>           <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)      20.6     2915.   0.00706   0.994
```

```
## 2 groupcontrol    -22.6     2915.  -0.00777   0.994
```

```
## p-Wert --> signifikant --> Ho ablehnen
## ist 1 enthalten oder nicht
#effektschätzer
##kleiner 1 --> protektiv
#größer 1 --> risiko
```