

Analyses master thesis: David Crameri

07/2022

sample description (n = 622)

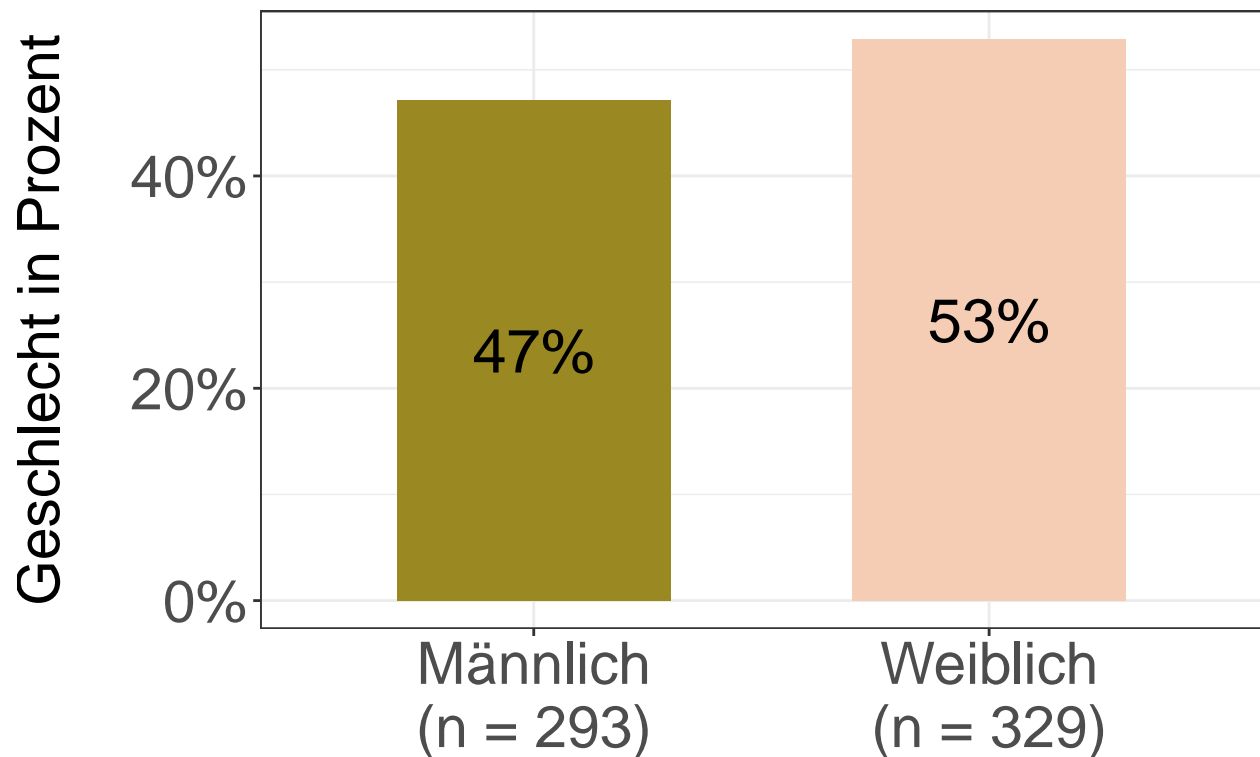
```
# plot gender
data %>%
  filter(!duplicated(ID)) %>%
  group_by(gschlecht_m_w) %>%
  summarise(tot = n()) %>%
  mutate(pct = tot / sum(tot)) %>%
  ungroup() %>%
  mutate(gender = if_else(gschlecht_m_w == "m", "Männlich", "Weiblich")) %>%
  mutate(xlab_1 = paste0("(n = ", tot, ")"),
         xlab = paste(gender, xlab_1, sep = "\n")) -> dp
```

```
#print df
print(dp)
```

```
## # A tibble: 2 x 6
##   geschlecht_m_w   tot   pct gender   xlab_1      xlab
##   <chr>         <int> <dbl> <chr>   <chr>      <chr>
## 1 m             293 0.471 Männlich (n = 293) "Männlich\n(n = 293)"
## 2 w             329 0.529 Weiblich (n = 329) "Weiblich\n(n = 329)"
```

```
#plot
p <- ggplot(data = dp,
            aes(x = xlab, y = pct, fill = gender)) +
  geom_bar(stat = "identity", width = .6) +
  labs(x = "", y = "Geschlecht in Prozent") + # does dot sound right?
  scale_y_continuous(labels = scales::percent) +
  scale_fill_manual(values = wes_palette("Royal2")) +
  guides(fill = "none") + # check for further info: https://ggplot2.tidyverse.org/reference/guide_legend.html
  geom_text(label = scales::percent(dp$pct, accuracy = 1), size = 8, position = position_stack(vjust = 1)) +
  mytheme

print(p)
```



```
#save plot
ggsave(filename = here::here("output/plots/202207_overall_sample_gender.pdf"),
  plot = p,
  width = 6,
  height = 6,
  dpi = 300,
  device = cairo_pdf
)
```

```
#age at operation
data %>%
  filter(!duplicated(ID)) -> dp
psych::describe(dp$alter_beim_opt)
```

```
##      vars   n mean   sd median trimmed  mad min  max range skew kurtosis   se
## X1      1 622 11.48 2.11   11.1   11.32 2.22 7.9 17.2   9.3 0.59   -0.52 0.08
```

```
#unique persons in sample
sum(!duplicated(data$ID))
```

```
## [1] 622
```

FF1: final sample description (n = 563)

```
#1. gender
# extractet cases were eliminated (n = 9): #is that correct @david?
data %>%
  #no duplicates
  filter(!duplicated(ID)) %>%
  #age older than 9
```

```

filter(alter_beim_opt >= 9) %>%
# extrahierte zähne = nein
filter(is.na(extrahierte_zahne)) -> dp

#filter gender overall
dp %>%
  group_by(gschlecht_m_w) %>%
  summarise(tot = n()) %>%
  mutate(pct = tot / sum(tot)) %>%
  ungroup() %>%
  mutate(gender = if_else(gschlecht_m_w == "m", "Männlich", "Weiblich")) %>%
  mutate(xlab_1 = paste0("(n = ", tot, ")"),
         xlab = paste(gender, xlab_1, sep = "\n")) -> dp_s

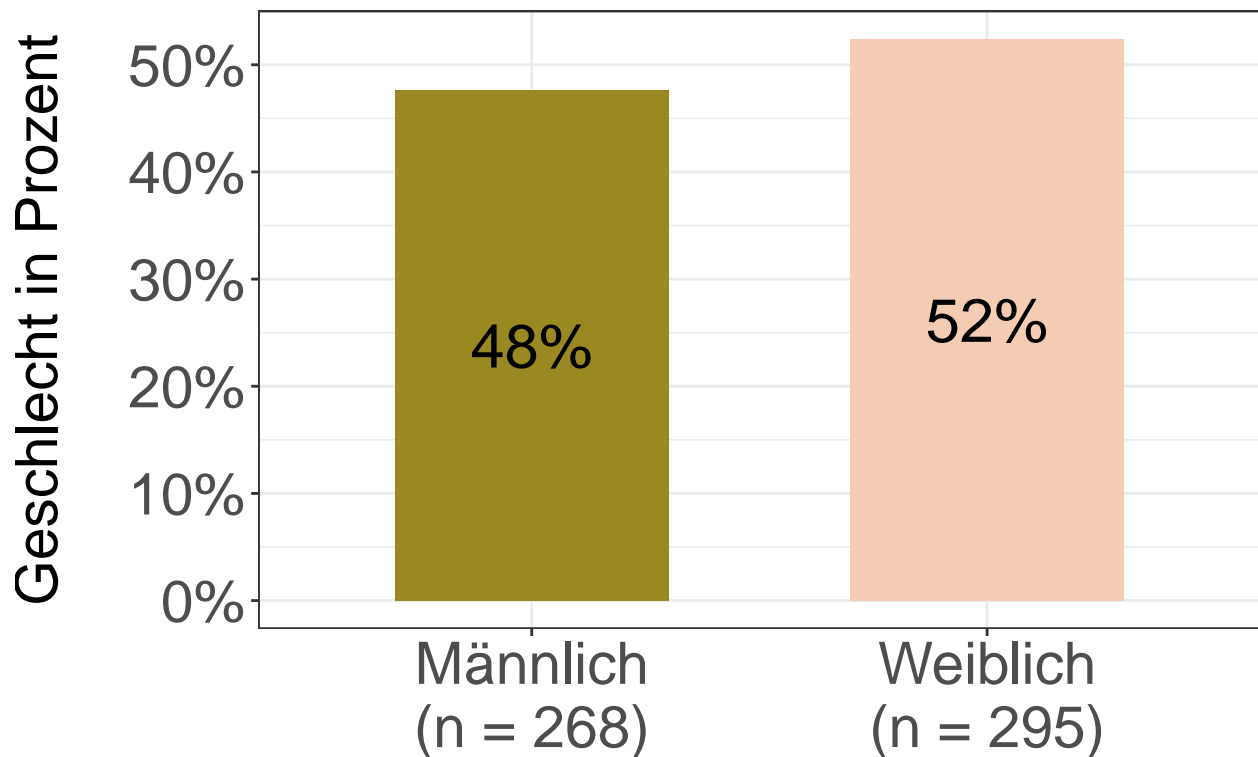
print(dp_s)

## # A tibble: 2 x 6
##   geschlecht_m_w tot   pct gender xlab_1 xlab
##   <chr>          <int> <dbl> <chr>   <chr>   <chr>
## 1 m              268 0.476 Männlich (n = 268) "Männlich\n(n = 268)"
## 2 w              295 0.524 Weiblich (n = 295) "Weiblich\n(n = 295)"

#plot
pq <- ggplot(data = dp_s,
             aes(x = xlab, y = pct, fill = geschlecht_m_w)) +
  geom_bar(stat = "identity", width = .6) +
  labs(x = "", y = "Geschlecht in Prozent") +
  scale_y_continuous(labels = scales::percent) +
  geom_text(label = scales::percent(dp_s$pct, accuracy = 1), size = 8, position = position_stack(vjust = 1)) +
  guides(fill = "none") +
  scale_fill_manual(values = wes_palette("Royal2")) +
  mytheme

print(pq)

```



```
#save:
ggplot2::ggsave(filename = "output/plots/202207_gender_final_subset.pdf",
  plot = pq,
  width = 6,
  height = 6,
  dpi = 300,
  device = cairo_pdf
)

#2. gender and age: MEAN
dp %>%
  group_by(gschlecht_m_w) %>%
  summarise(mean = mean(alter_beim_opt, na.rm = TRUE),
    sd = sd(alter_beim_opt), #standard deviation
    se = sd(alter_beim_opt)/sqrt(length(alter_beim_opt))) #standard error of the mean

## # A tibble: 2 x 4
##   geschlecht_m_w mean    sd    se
##   <chr>          <dbl> <dbl> <dbl>
## 1 m             11.6  1.97 0.121
## 2 w             11.8  2.00 0.117

#3. missing teeth overall (n = 39)
dp %>%
  filter(fehlender_zahn_fdi_quadrant_zahnzahl != 0) -> dp_m

head(dp_m,2)

## # A tibble: 2 x 12
##       ID datum_des_opt      alter_beim_opt geschlecht_m_w nichtanlage_ja_nein
```

```
## <dbl> <dtm> <dbl> <chr> <chr>
## 1 38 2021-09-24 00:00:00 15.3 m ja
## 2 67 2006-03-09 00:00:00 14.5 m ja
## # ... with 7 more variables: fehlender_zahn_fdi_quadrant_zahnzahl <dbl>,
## # x1_quadrant_zahnnr <dbl>, x2_quadrant_zahnnr <dbl>,
## # x3_quadrant_zahnnr <dbl>, x4_quadrant_zahnnr <dbl>,
## # extrahierte_zahne <chr>, zusatzzahn <chr>

#export data
write.csv(dp_m, here::here("output/datasets//202207_missing_teeth_overall.csv"))
```

FF2: Nichtanlage vs. Gesund

```
#nach geschlecht
data %>%
  #age older than 9
  filter(alter_beim_opt >= 9) %>%
  #drop all cases with extrahierte zähne
  filter(is.na(extrahierte_zahne)) %>%
  filter(!duplicated(. $ID)) -> dp_d #duplicated cases are excluded

#summaries per gender and nichtanlage
dp_d %>%
  group_by(gschlecht_m_w, nichtanlage_ja_nein) %>%
  summarise(tot = n()) %>%
  mutate(pct = tot / sum(tot)) %>%
  ungroup() -> df

## `summarise()` has grouped output by 'gschlecht_m_w'. You can override using the
## `.groups` argument.

#Add text to xlab
txt <- df %>%
  group_by(gschlecht_m_w) %>%
  summarise(gender_tot = sum(tot)) %>%
  mutate(gender = if_else(gschlecht_m_w == "m", "Männlich", "Weiblich")) %>%
  mutate(xlab_1 = paste0("n = ", gender_tot, " "),
         xlab = paste(gender, xlab_1, sep = "\n"))

#merge back
df_plot <- df %>%
  left_join(., txt, by = "gschlecht_m_w")

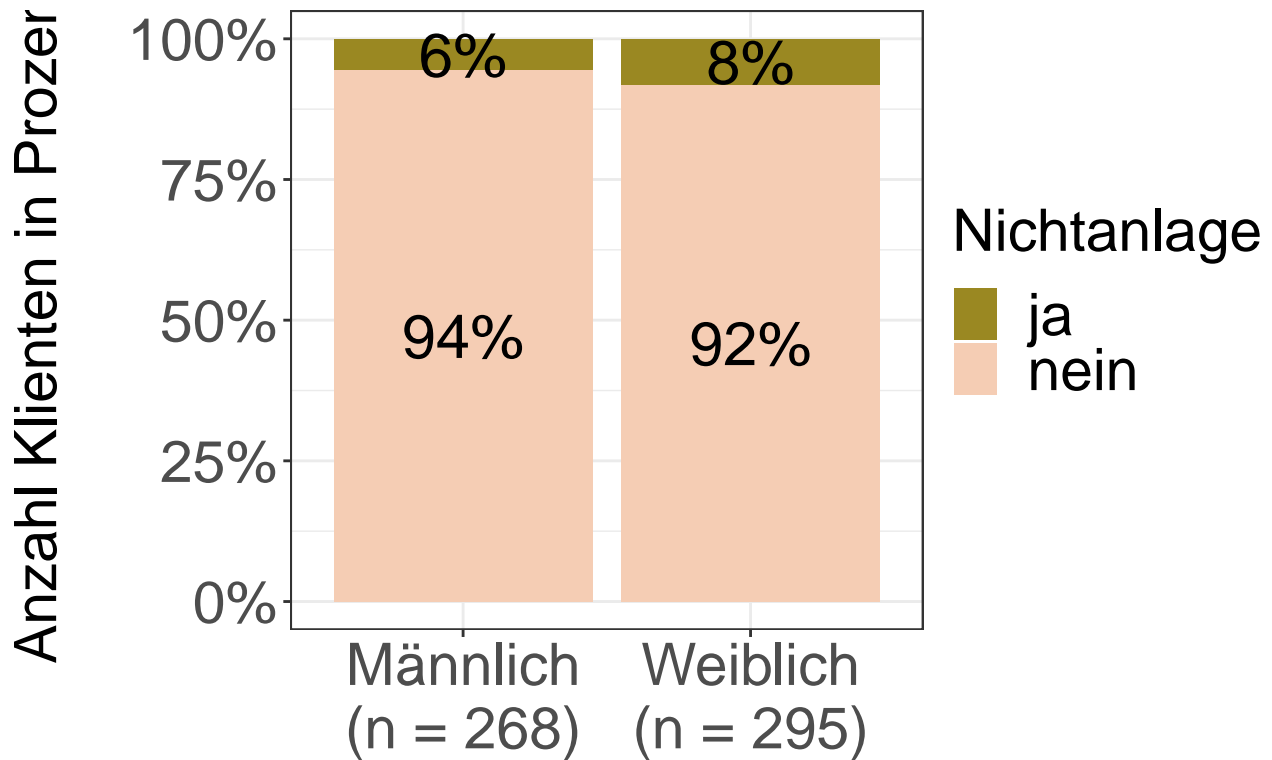
print(df)

## # A tibble: 4 x 4
##   gschlecht_m_w nichtanlage_ja_nein tot pct
##   <chr>         <chr>          <int> <dbl>
## 1 m           ja              15 0.0560
## 2 m           nein             253 0.944
## 3 w           ja              24 0.0814
## 4 w           nein             271 0.919

#plot
p <- ggplot(df_plot, aes(x = xlab, y = pct, fill = nichtanlage_ja_nein)) +
  geom_bar(stat = "identity", position = "stack") +
```

```
labs( x = "", y = "Anzahl Klienten in Prozent") + #title ok like this?
guides(fill = guide_legend(title = "Nichtanlage")) +
scale_y_continuous(labels = scales::percent) +
scale_fill_manual(values = wes_palette("Royal2")) +
geom_text(label = scales::percent(df_plot$pct, accuracy = 1), position = position_stack(vjust = .5),
mytheme
```

```
print(p)
```



```
#save
ggplot2::ggsave(filename = "output/plots/202207_gender_nichtanlage.pdf",
  plot = p,
  width = 7,
  height = 6,
  device = cairo_pdf)

#do that correlation differ from a random result
#from results first into a matrix
xtabs(~ nichtanlage_ja_nein + geschlecht_m_w, data = as.data.frame(dp_d)) -> df_chi #which datas et show

chi_sq <- chisq.test(df_chi) # both data sets does not show any assimilation between gender and nichtanlage
chi_sq
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: df_chi
## X-squared = 1.0375, df = 1, p-value = 0.3084
```

FF3 & FF4: missing teeth description

attention: missing teeth do not correspond with the sample size!

```
#missing single tooth overall
data %>%
  #age older than 9
  filter(alter_beim_opt >= 9) %>%
  #drop all cases with extrahierte zähne
  filter(is.na(extrahierte_zahne)) %>%
  filter(fehlender_zahn_fdi_quadrant_zahnzahl > 0) -> dp_d #exclude all persons without missing teeth

#missing teeth per person
dp_d %>%
  group_by(ID) %>%
  summarise(teeth_miss = n()) %>%
  ungroup() -> df

#aggregate data set from above on individual level: overall----
dp_ind <- df %>%
  group_by(teeth_miss) %>%
  summarise(tot_persons = n()) %>%
  ungroup() %>%
  mutate(pct = tot_persons / sum(tot_persons))

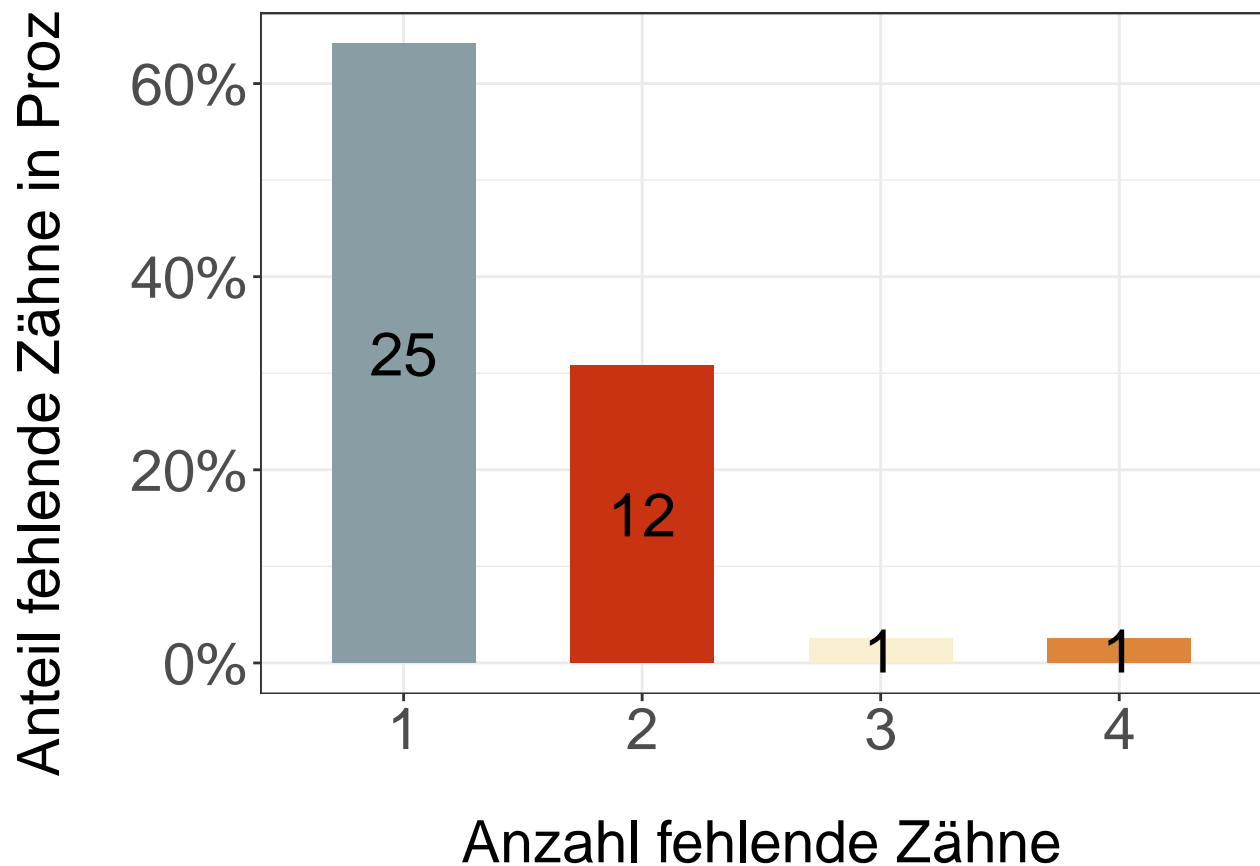
#attention here are all cases again summed up
#in other words 39 person have 56 missing teeth
print(dp_ind)

## # A tibble: 4 x 3
##   teeth_miss tot_persons    pct
##   <int>      <int>  <dbl>
## 1         1         25 0.641
## 2         2         12 0.308
## 3         3          1 0.0256
## 4         4          1 0.0256

write.csv(file = "output/datasets/202207_table_missing_teeth_overall.csv", dp_ind)

#missing teeth per person: meaning most people misses 1 and 2 teeth
p <- ggplot(dp_ind, aes(x = as.factor(teeth_miss), y = pct, fill = as.factor(teeth_miss))) +
  #ev. as stacked bars? however data needs to be re-adjusted
  geom_bar(stat = "identity", position = position_dodge(), width = .6) +
  geom_text(label = dp_ind$tot_persons, position = position_stack(vjust = .5), size = 8) +
  scale_y_continuous(labels = scales::percent) +
  scale_fill_manual(values = wes_palette("Royal1")) +
  labs(y = "Anteil fehlende Zähne in Prozent", x = "Anzahl fehlende Zähne") +
  guides(fill = "none") +
  mytheme

print(p)
```



```
#save plot
ggplot2::ggsave(filename = "output/plots/202207_missing_teeth.pdf",
  device = cairo_pdf,
  width = 6,
  height = 5,
  plot = p)

#another table: number of missing tooth buds
dp_d %>%
  filter(fehler_zahn_fdi_quadrant_zahnzahl > 0) %>%
  group_by(fehler_zahn_fdi_quadrant_zahnzahl) %>%
  summarise(tot = n()) %>%
  #calculate percentage
  mutate(pct_number_missing = tot / sum(tot) * 100) %>%
  ungroup() %>%
  #divide by number of persons in data set: 39
  mutate(pct_ind_hypodontia = tot / 39 * 100) -> df_e

#save
write.csv(file = "output/datasets/202207_table_number_missing_tooth.csv", df_e)
```

FF5: distribution of missing teeth

```
#overall, attention some persons are double
data %>%
```



```

#age older than 9
filter(alter_beim_opt >= 9) %>%
#drop all cases with extrahierte zähne
filter(is.na(extrahierte_zahne)) %>%
filter(fehler_zahn_fdi_quadrant_zahnzahl > 0) %>%
# create new variables
mutate(quadrant = stringr::str_sub(fehler_zahn_fdi_quadrant_zahnzahl, end = 1),
       position = stringr::str_sub(fehler_zahn_fdi_quadrant_zahnzahl, start = 2)) %>%
group_by(quadrant, position) %>%
summarise(tot = n()) %>%
mutate(pct = tot / sum(tot)) %>%
ungroup() -> dp

```

`summarise()` has grouped output by 'quadrant'. You can override using the
`.groups` argument.

```
print(dp)
```

```

## # A tibble: 8 x 4
##   quadrant position  tot    pct
##   <chr>    <chr>    <int> <dbl>
## 1 1      2          7 0.636
## 2 1      5          4 0.364
## 3 2      2          7 0.583
## 4 2      5          5 0.417
## 5 3      1          3 0.158
## 6 3      5         16 0.842
## 7 4      1          1 0.0714
## 8 4      5         13 0.929

```

```

#xlab
txt <- dp %>%
  group_by(quadrant) %>%
  summarise(tot_q = sum(tot)) %>%
  ungroup() %>%
  mutate(xlab1 = paste0("n = (", tot_q, ")"))

```

```

#merge back data
df_plot <- dp %>%
  left_join(., txt, by = "quadrant") %>%
  mutate(xlab = paste(quadrant, xlab1, sep = "\n"))

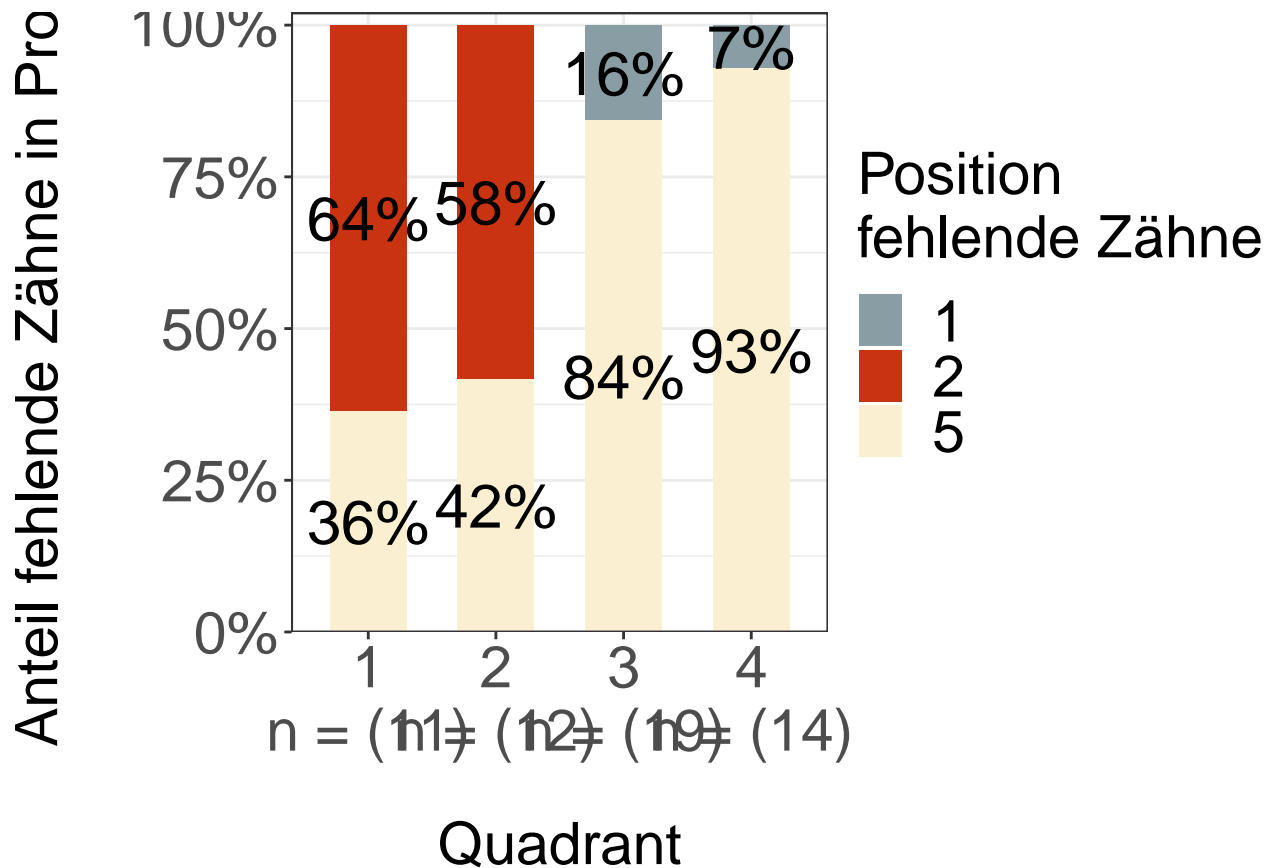
```

```

#plot
p <- ggplot(df_plot, aes(x = xlab, y = pct, fill = position)) +
  geom_bar(stat = "identity", width = .6) +
  scale_y_origin(label = scales::percent) +
  guides(fill = guide_legend(title = "Position \nfehlende Zähne")) +
  geom_text(label = scales::percent(dp$pct, accuracy = 1), position = position_stack(vjust = .5), size = 10) +
  scale_fill_manual(values = wes_palette("Royal1")) +
  labs(x = "Quadrant", y = "Anteil fehlende Zähne in Prozent") +
  mytheme

```

```
print(p)
```



```
#save plot
ggplot2::ggsave(filename = "output/plots/202207_missing_teeth_per_quadrants.pdf",
  device = cairo_pdf,
  width = 9,
  height = 8,
  plot = p)
```

some additional tables

um keine verzerrungen zu bekommen müsste man mit 56 cases und nicht 39 personen arbeiten!
hab nun die 56 als gesamtheit genommen (anstelle die 39 personen)

```
#overall, attention some persons are double!
data %>%
  #age older than 9
  filter(alter_beim_opt >= 9) %>%
  #drop all cases with extrahierte zähne
  filter(is.na(extrahierte_zähne)) %>%
  filter(fehlender_zahn_fdi_quadrant_zahnzahl > 0) -> df_t

#table gender and chi-square
df_t %>%
  group_by(gschlecht_m_w) %>%
  summarise(tot_g = n()) %>%
  mutate(pct = tot_g / sum(tot_g)) -> df
```

```

print(df)

## # A tibble: 2 x 3
##   gschlecht_m_w tot_g   pct
##   <chr>         <int> <dbl>
## 1 m             21 0.375
## 2 w             35 0.625

write.csv(file = "output/datasets/202207_gender_hypodontia.csv", df)

#binomial test!
binom.test(15, 39) #gender and prevalence of hypodontia do not differ statistically significant

##
## Exact binomial test
##
## data: 15 and 39
## number of successes = 15, number of trials = 39, p-value = 0.1996
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2336393 0.5538094
## sample estimates:
## probability of success
##           0.3846154

binom.test(21, 56) #wibliche patienten haben eine höhere prevalence von hypodontie

##
## Exact binomial test
##
## data: 21 and 56
## number of successes = 21, number of trials = 56, p-value = 0.08143
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2491841 0.5145217
## sample estimates:
## probability of success
##           0.375

#further info about binomial test: https://www.statology.org/binomial-test-r/

#table unter/oberkiefer
df_t %>%
  mutate(quadrant = stringr::str_sub(fehlender_zahn_fdi_quadrant_zahnzahl,
                                     end = 1)) %>%
  mutate(quadrant = as.numeric(quadrant)) %>%
  group_by(ID) %>%
  summarise(tot = sum(quadrant)) %>%
  ungroup() %>%
  # mutate(tot = if_else(tot == 3, 2.9, tot)) %>%
  mutate(kiefer = case_when((tot == 3 | tot == 1 | tot == 2) ~ "oben",
                            (tot == 7 | tot == 3 | tot == 4) ~ "unten",
                            TRUE ~ "beide")) -> df_kiefer

```

```

#one case needs to be change by hand (because sum 1, 3 is 4)
df_kiefer[df_kiefer$ID == 472,]$kiefer <- "beide"

#group again
df_kiefer %>%
  group_by(kiefer) %>%
  summarise(tot = n()) %>%
  mutate(pct = tot / sum(tot)) -> df

print(df)

## # A tibble: 3 x 3
##   kiefer   tot   pct
##   <chr> <int> <dbl>
## 1 beide     4 0.103
## 2 oben    23 0.590
## 3 unten    12 0.308

# TODO David: Zahlen überprüfen
write.csv(file = "output/datasets/202207_kiefer_hypodontia.csv", df)

#chisquare test
chisq.test(df$tot) # ist signifikant, aber auch nur deshalb, weil

##
## Chi-squared test for given probabilities
##
## data: df$tot
## X-squared = 14, df = 2, p-value = 0.0009119

# in der condition "beide" nur 4 personen beinhaltet sind
binom.test(12, 39) # lässt man die konditino "beide" weg, zeigt

##
## Exact binomial test
##
## data: 12 and 39
## number of successes = 12, number of trials = 39, p-value = 0.0237
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.1701959 0.4756914
## sample estimates:
## probability of success
##           0.3076923

# der binomial test eine signifikanz auf,
#d.h. es scheint, dass im oberkiefer mehr zähne fehlen als unten
# @David: macht das sinn?

#table links - rechts
df_t %>%
  mutate(quadrant = stringr::str_sub(fehrender_zahn_fdi_quadrant_zahnzahl, end = 1)) %>%
  mutate(quadrant = as.numeric(quadrant)) %>%

```

```

group_by(ID) %>%
  summarise(tot = sum(quadrant)) %>%
  ungroup() %>%
  mutate(tot = if_else(tot == 3, 2.5, tot)) %>% #weil summe beider seiten 5 ergeben
  mutate(seite = case_when((tot == 5 | tot == 1 | tot == 4) ~ "rechts",
                           (tot == 4.5 | tot == 2 | tot == 2.5) ~ "links",
                           TRUE ~ "beide")) -> df_seite

#some cases need to be changes manually
df_seite[df_seite$ID == 67, ]$seite <- "beide" #sum = 3
df_seite[df_seite$ID == 185, ]$seite <- "beide" #sum = 3
df_seite[df_seite$ID == 243, ]$seite <- "beide" #sum = 3
df_seite[df_seite$ID == 419, ]$seite <- "beide" #sum = 3
df_seite[df_seite$ID == 472, ]$seite <- "beide" #sum = 4
df_seite[df_seite$ID == 480, ]$seite <- "beide" #sum = 3
df_seite[df_seite$ID == 508, ]$seite <- "beide" #sum = 3

#group again
df_seite %>%
  group_by(seite) %>%
  summarise(tot = n()) %>%
  mutate(pct = tot / sum(tot)) -> df

print(df)

## # A tibble: 3 x 3
##   seite    tot  pct
##   <chr> <int> <dbl>
## 1 beide     13 0.333
## 2 links     14 0.359
## 3 rechts    12 0.308

# TODO David: Zahlen überprüfen
write.csv(file = "output/datasets/202207_seite_hypodontia.csv", df)

#binomial test!
chisq.test(df$tot) # nicht signifikant

##
## Chi-squared test for given probabilities
##
## data: df$tot
## X-squared = 0.15385, df = 2, p-value = 0.926

```