mosaic_codealong.R

joycerobbins

2021-08-10

```
# Mosaic Codealong, August 10, 2021


# Get and clean data
library(magrittr)

# CREATE A FREQ COLUMN!

df <- readr::read_tsv("https://raw.githubusercontent.com/jtr13/graphcat21/main/data/age_preweight_gain.
  dplyr::filter(is.na(Notes)) %>%
  dplyr::select(-Notes) %>%
  dplyr::filter(!is.na(Births)) %>%
  dplyr::mutate(Freq = Births) # need a "Freq" column for vcd::mosaic
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Notes = col_logical(),
##   `Age of Mother 9` = col_character(),
##   `Age of Mother 9 Code` = col_character(),
##   `Mother's Pre-pregnancy Weight` = col_character(),
##   `Mother's Pre-pregnancy Weight Code` = col_double(),
##   `Mother's Weight Gain` = col_character(),
##   `Mother's Weight Gain Code` = col_double(),
##   Births = col_double()
## )
```

```
## Warning: 48 parsing failures.
##  row   col           expected                                  actual
## 4224 Notes 1/0/T/F/TRUE/FALSE ---                                      'https://raw.githubusercontent.co
## 4224 NA    8 columns          1 columns                                'https://raw.githubusercontent.co
## 4225 Notes 1/0/T/F/TRUE/FALSE Dataset: Natality, 2016-2019 expanded 'https://raw.githubusercontent.co
## 4225 NA    8 columns          1 columns                                'https://raw.githubusercontent.co
## 4226 Notes 1/0/T/F/TRUE/FALSE Query Parameters:                       'https://raw.githubusercontent.co
## .... ..... .................. ..................................... ..........................
## See problems(...) for more details.
```

```
# NO SPACES IN VARIABLE NAMES! Create short variable names without spaces as vcd::mosaic can't handle s

# Remove "Unknown or Not Stated" and clean up Weight, combine high levels for both variables (be carefu
```

```
df <- df %>%
  dplyr::mutate(Weight = factor(`Mother's Pre-pregnancy Weight`),
         Age = factor(`Age of Mother 9 Code`)) %>%
  dplyr::filter(Weight != "Unknown or Not Stated") %>%
  dplyr::mutate(Weight = factor(stringr::str_remove_all(Weight, "lbs| "))) %>%
  dplyr::mutate(Weight = forcats::fct_lump_n(Weight, w = Freq, n = 3, other_level = "225+")) %>%
  dplyr::mutate(Weight = forcats::fct_relevel(Weight, "75-124")) %>%
  dplyr::mutate(Age = dplyr::recode(Age, "40-44" = "40+", "45-49" = "40+", "50+" = "40+")) %>%
  dplyr::mutate(Gain = `Mother's Weight Gain Code`) %>%
  dplyr::mutate(Gain = floor(Gain/10)*10) %>%
  dplyr::mutate(Gain = ifelse(Gain %in% seq(0, 50, 10), Gain, "60+")) %>%
  dplyr::mutate(Gain = forcats::fct_recode(Gain, "0-10" = "0",
                                           "10-20" = "10",
                                           "20-30" = "20",
                                           "30-40" = "30",
                                           "40-50" = "40",
                                           "50-60" = "50")) %>%
  dplyr::select(Age, Weight, Gain, Freq) %>%
  dplyr::group_by(Age, Weight, Gain) %>%
  dplyr::summarize(Freq = sum(Freq)) %>%
  dplyr::ungroup()
```

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

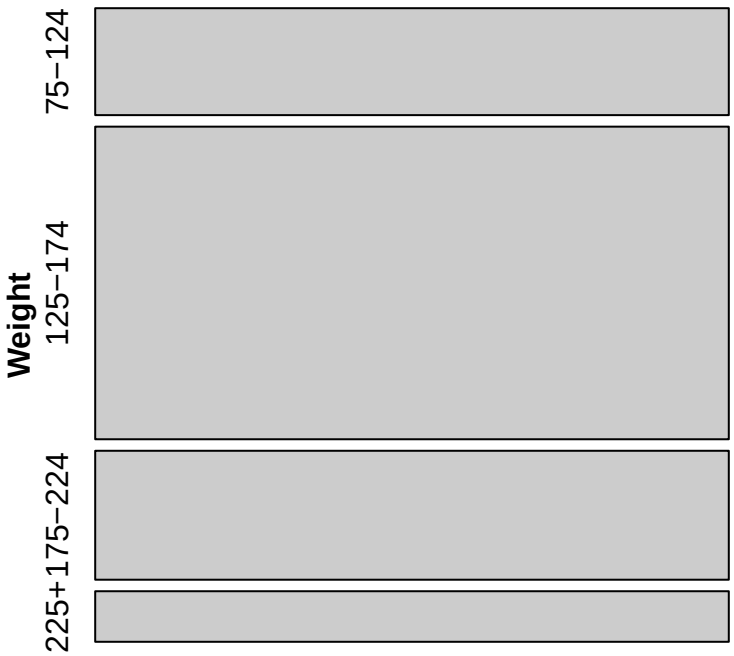
```
# Ready to graph

dplyr::glimpse(df)
```

```
## Rows: 195
## Columns: 4
## $ Age    <fct> 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15,~
## $ Weight <fct> 75-124, 75-124, 75-124, 75-124, 75-124, 75-124, 75-124, 125-174~
## $ Gain   <fct> 0-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60+, 0-10, 10-20, 20-3~
## $ Freq   <dbl> 41, 126, 203, 182, 96, 50, 48, 87, 161, 200, 165, 114, 45, 45, ~
```

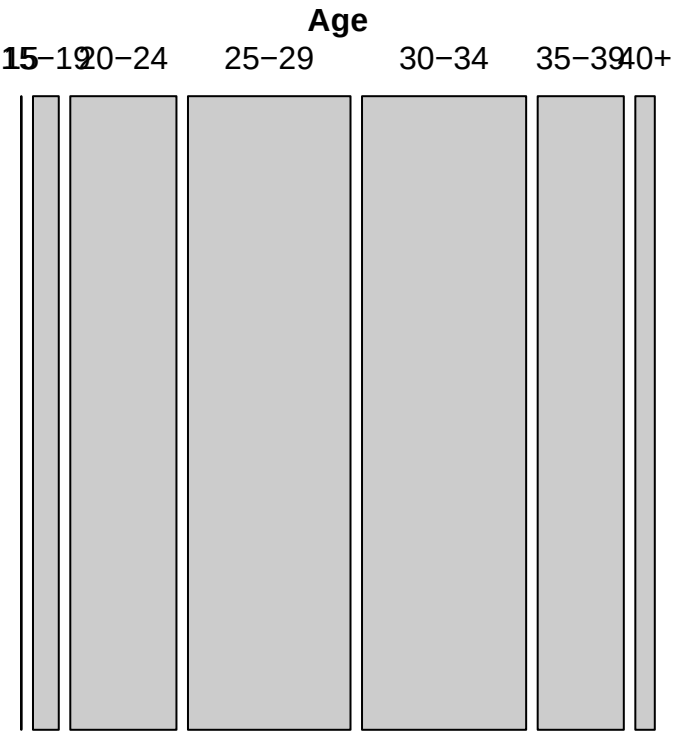
```
head(df)
```

```
## # A tibble: 6 x 4
##   Age   Weight Gain   Freq
##   <fct> <fct>  <fct> <dbl>
## 1 15    75-124 0-10     41
## 2 15    75-124 10-20   126
## 3 15    75-124 20-30   203
## 4 15    75-124 30-40   182
## 5 15    75-124 40-50    96
## 6 15    75-124 50-60    50
```

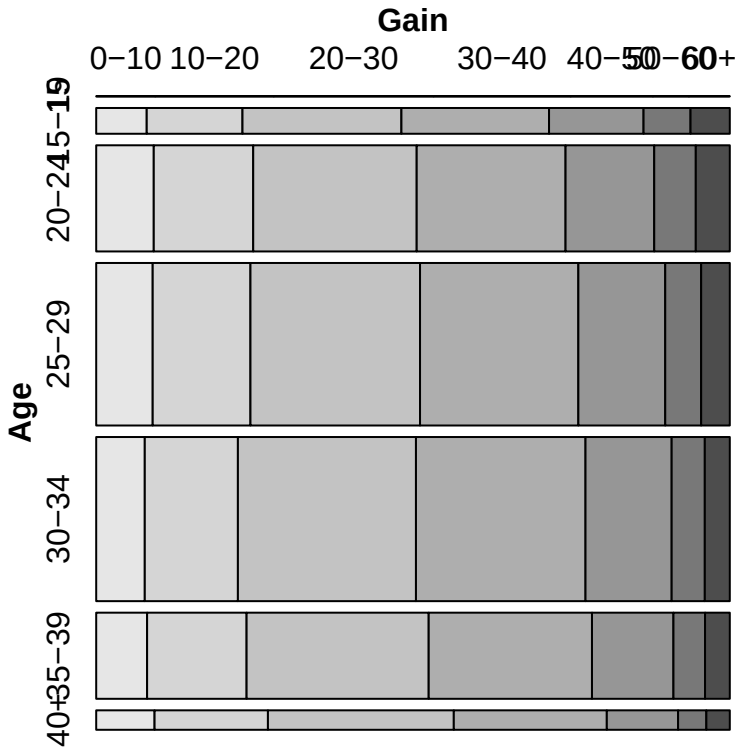
```
# START SMALL

vcd::mosaic(~Weight, data = df)
```



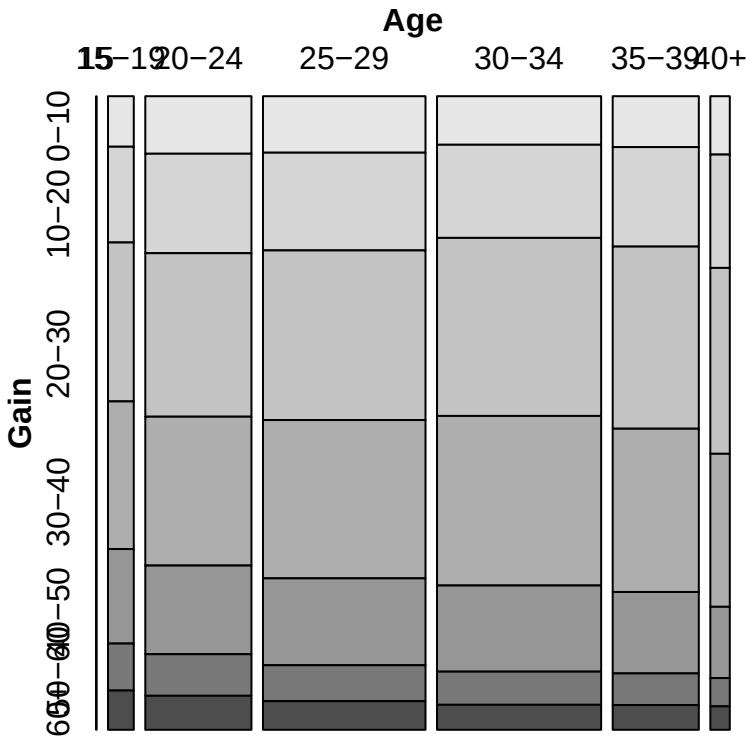

```
vcd::mosaic(~Age, direction = "v", data = df)
```



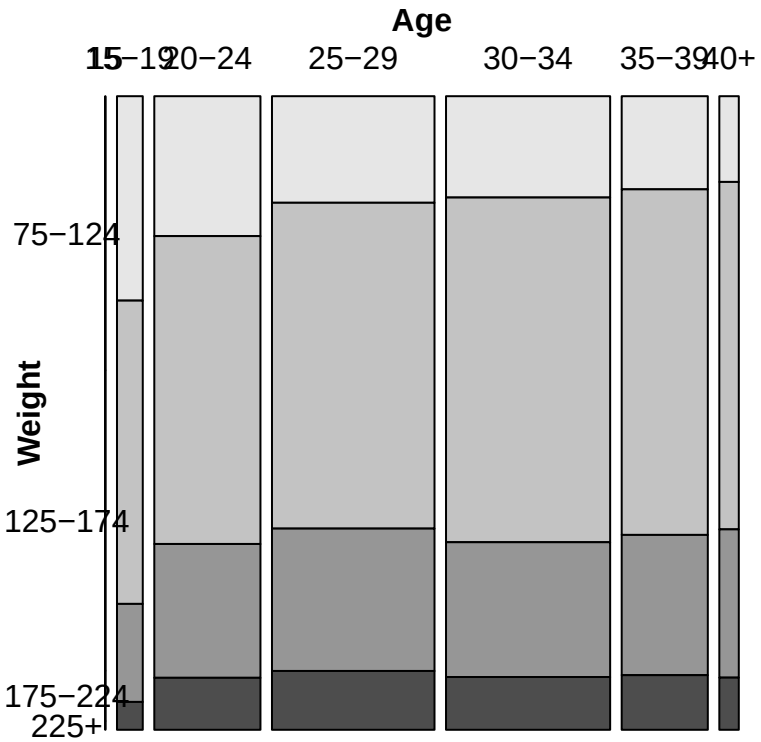

```
# Two variables

vcd::mosaic(Gain~Age, data = df)
```



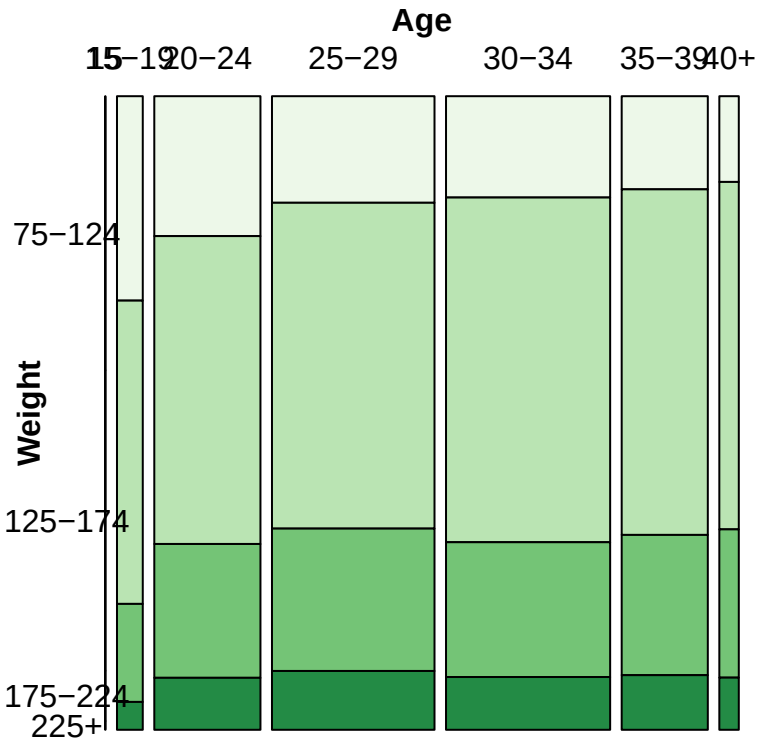

```
# Change direction of cuts (last should be horizontal)

vcd::mosaic(Gain~Age, data = df, direction = c("v", "h"))
```



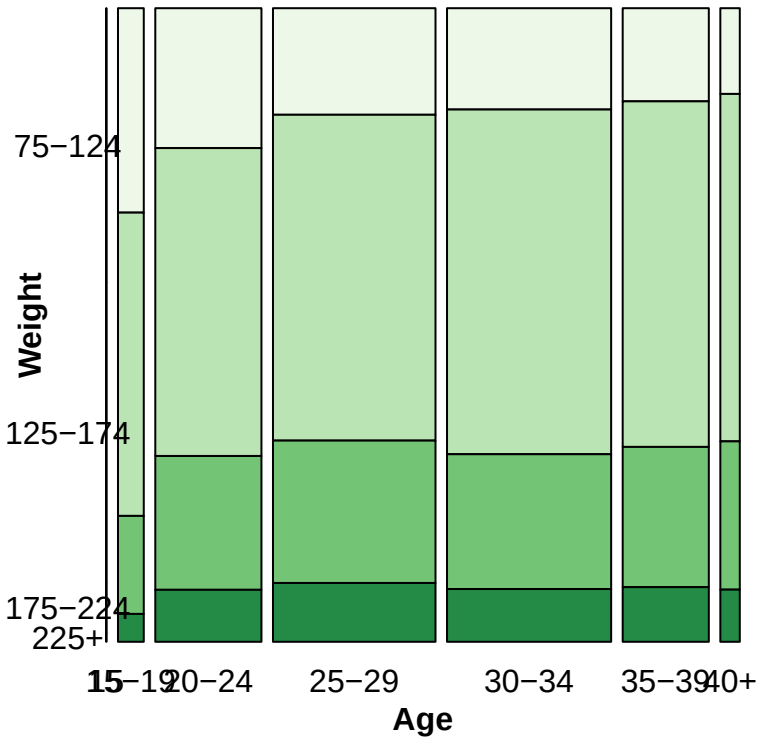

```
# Rotate labels

vcd::mosaic(Weight~Age, data = df, direction = c("v", "h"),
       rot_labels = c(0, 0, 0, 0))  # top, right, bottom, left
```



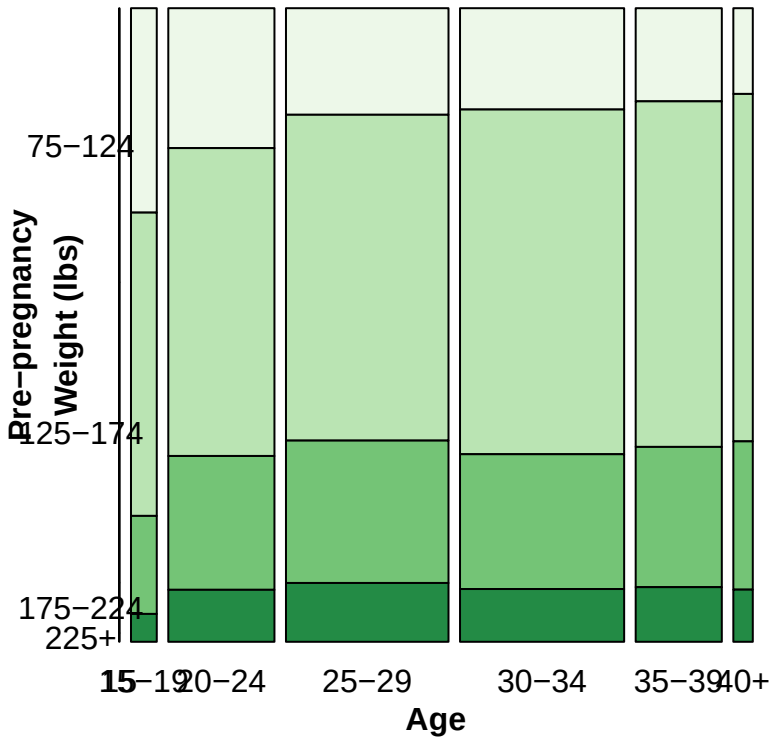

```
# Add fill color

mycolors <- RColorBrewer::brewer.pal(4, "Greens")

vcd::mosaic(Weight~Age , data = df,
            direction = c("v", "h"),
            rot_labels = c(0, 0, 0, 0), # top, right, bottom, left
            highlighting_fill = mycolors)
```



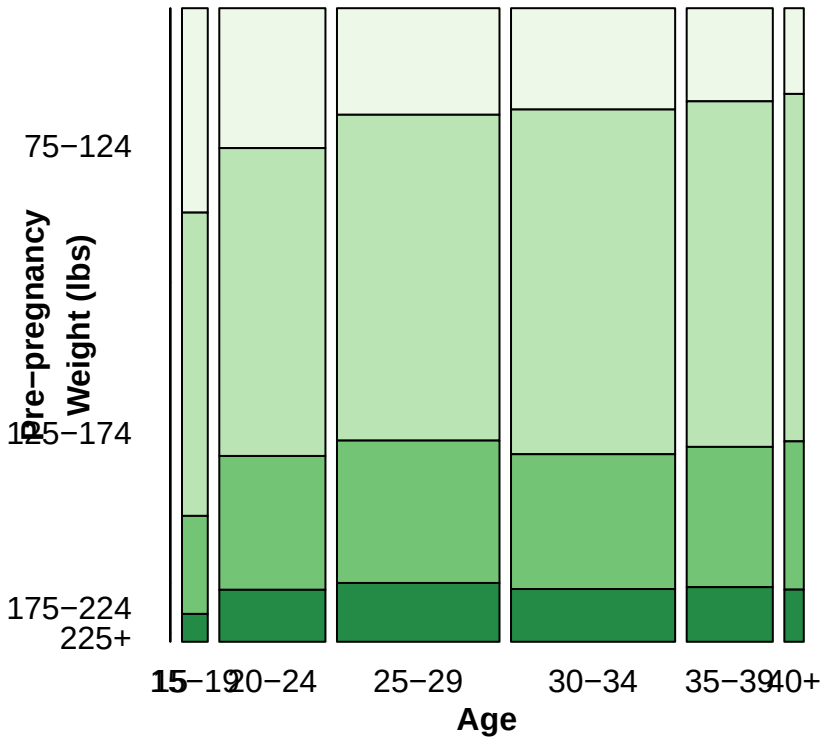

```
# Move labels to bottom

vcd::mosaic(Weight~Age, data = df, direction = c("v", "h"),
       rot_labels = c(0, 0, 0, 0), # top, right, bottom, left
       highlighting_fill = mycolors,
       # top = FALSE, left = TRUE
       labeling_args = list(tl_labels = c(FALSE, TRUE)))
```




```
# Change variable names

vcd::mosaic(Weight~Age, data = df, direction = c("v", "h"),
       rot_labels = c(0, 0, 0, 0), # top, right, bottom, left
       highlighting_fill = mycolors,
       labeling_args = list(tl_labels = c(FALSE, TRUE),
          set_varnames = c(Weight = "Pre-pregnancy\nWeight (lbs)")))
```

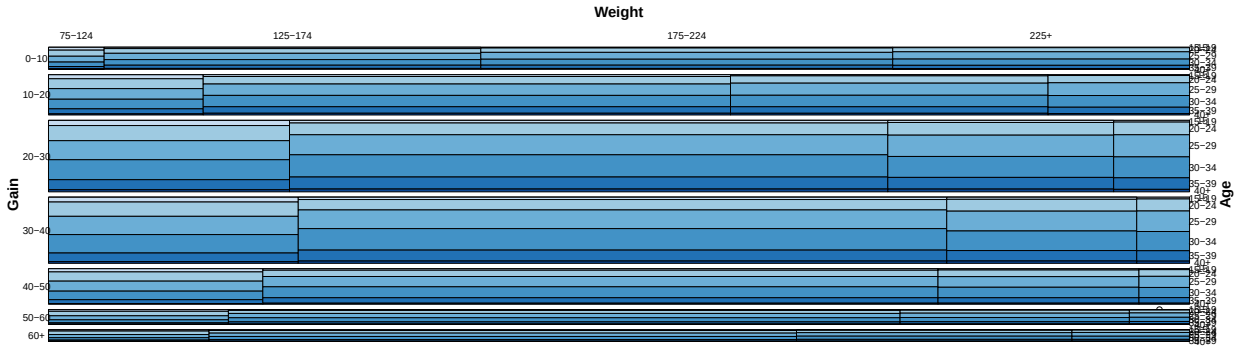



```
# Move variable names (increase margin), justify labels

vcd::mosaic(Weight~Age, data = df, direction = c("v", "h"),
       rot_labels = c(0, 0, 0, 0), # top, right, bottom, left
       highlighting_fill = mycolors,
       labeling_args = list(tl_labels = c(FALSE, TRUE),
          set_varnames = c(Weight = "Pre-pregnancy\nWeight (lbs)"),
          offset_varnames = c(0, 0, 0, 1),
          just_labels = c("center", "center", "center", "right")))
```




```
# THREE VARIABLES
```

```
vcd::mosaic(Age~Gain+Weight, data = df,
          #  direction = c("v", "v", "h"),
            rot_labels = c(0,0,0,0),
            gp_labels = grid::gpar(fontsize = 8),
            spacing = vcd::spacing_dimequal(c(.3, 0, 0)),
            highlighting_fill = RColorBrewer::brewer.pal(7, "Blues"))
```




```
# ADVANCED: Example using labeling_cells, data must be in table form
```

```
df <- df %>% dplyr::mutate(Weight = forcats::fct_rev(Weight))
dftab <- xtabs(Freq~Weight+Age+Gain, data = df)
vcd::mosaic(dftab,
       direction = c("v", "v", "h"),
       rot_labels = c(0,0,0,0),
       spacing = vcd::spacing_dimequal(c(.3, 0, 0)),
       labeling_args = list(abbreviate = c(Age = -1),
                            labeling = vcd::labeling_cells(text = labels)),  gp_labels = grid::gpar(fon
       highlighting = "Gain",
       highlighting_fill = RColorBrewer::brewer.pal(7, "Blues"), pop=FALSE)
labels <- ifelse(dftab > -1, NA, NA)
labels[3,,4] <- paste(levels(factor(df$Age)), "\n years")
vcd::labeling_cells(text = labels, margin = 0,
                    gp_text = grid::gpar(fontsize = 8))(dftab)
```