Domača naloga 10

Alen Kahteran

13. 11. 2020

Definicija funkcije

```r
generate_linear_data <- function(n, L){
    ### returns tibble in long format, where data is of size n represented by L groups.
    ### we calculate x and b from U[0, 1] where b is calculated by group.
    ### we also add noise which is calculated from N(0, 1)
    ###
    ### then we calculate y from equation (by b group):
    ###     y = 2 + 5*x + 10*b + noise
    ###
    ### we then calculate mean Y by group
    ### then arrange by b(group) and y
    ### then add index, calculate min, max index
    ### calculate min, max y
    ### so we can create shadows correctly.
    ###
    ### input:
    ###     n - size
    ###     L - number of groups

    # create grp by creating L groups of random sizes that sum up to n size
    data_lin <- tibble(grp=sample(1:L, n, replace = TRUE)) %>%
        # group by grp and the calculate value b which is unique to group
        group_by(grp) %>%
        mutate(b=runif(1)) %>%
        ungroup() %>%
        # create x values, create noise, and then calculate y
        mutate(x=runif(n),
               noise=rnorm(n),
               y=2+5*x+10*b+noise) %>%
        # group by grp to calculate mean value by group
        group_by(grp) %>%
        mutate(mean_y=mean(y)) %>%
        ungroup() %>%
        # arrange by group
        arrange(grp, y) %>%
        # add index after sorting so the drawing process is correct
        mutate(i=1:n) %>%
        # group by grp and then calculate min, max index and min, max y
        group_by(grp) %>%
        mutate(imin=min(i),
               imax=max(i),
```

```
                ymin=min(y),
                ymax=max(y)) %>%
        ungroup()

    # return tibble
    return(data_lin)
}
```

Izris

```
# choice of size and groups
n <- 1000
L <- 10

# setting seed for reproducibility
set.seed(8)

# generating data
data_lin <- generate_linear_data(n, L)

# drawing
ggplot(data_lin, aes(x=i, y=y, col=as.factor(grp))) +
    # add shadows
    geom_rect(aes(xmin=imin,
                  xmax=imax,
                  ymin=ymin,
                  ymax=ymax,
                  group=grp),
              alpha=.2,
              colour=NA,
              fill="lightgrey") +
    # add points
    geom_point() +
    # add horizontal lines
    geom_line(aes(y=mean_y, group=grp),
              linetype="dashed",
              color="black",
              size=0.5) +

    # add labels
    labs(x="i",
         y="y",
         title=paste0("n = ", n,", L = ", L)) +
    # remove legend name and allign title to middle
    theme(legend.title=element_blank(),
          plot.title = element_text(hjust = 0.5))
```

n = 1000, L = 10