library(tidyverse)

library(readxl)

library(openxlsx)

data2021 <- read\_xlsx("./data/2021.xlsx", 1,

 col\_types = c("numeric", "numeric", "text",

 "text", "text", "text", "text", "text",

 "text", "text", "text", "numeric",

 "numeric", "numeric", "numeric",

 "numeric", "text", "text", "text",

 "text", "numeric", "numeric", "numeric",

 "numeric", "text", "numeric", "numeric",

 "numeric", "numeric", "numeric",

 "text"))

data2021 <- left\_join(data2021, dataID,

 by = "Meno a priezvisko") %>%

 select(-Meno, -Priezvisko) %>%

 mutate(ID.x = ID.y) %>%

 select(-ID.y) %>%

 rename(ID = ID.x)

vahy <- read\_xlsx("./data/vahyMUPS.xlsx", 1,

 col\_types = c("text", "text", "text", "numeric")) %>%

 select(Register, vahaMUPS)

# Dopocitanie medianu a MAD

produktivita2021 <- left\_join(data2021, vahy,

 by = c("Agenda" = "Register"))

produktivita2021 <- produktivita2021 %>%

 filter(grepl("^Okresný", `Názov súdu`),

 grepl("^T$|^Tk$|^Tv$|^Nt$|^Pp$|^Tp$|^Ntt$|^C$|^Cpr$|^Ca$|^Csp$|^Cr$|^Ccud$|^Cb$|^CbPv$|^CbR$|^CbBu$|^CbHs$|^Cbi$|^CbVO$|^S$|^P$|^PPOm$|^Pc$|^Ps$|^Po$|^Pu$|^D$|^K$|^R$|^NcKR$|^Nsre$|^Zm$|^CbZm$",

 Agenda))

sum(is.na(produktivita2021$vahaMUPS))

produktivita2021 <- produktivita2021 %>%

 mutate(VazenyPocet1 = Počet1\*vahaMUPS)

produktivita2021 <- produktivita2021 %>%

 filter(grepl("^R$", Typ)) %>%

 group\_by(ID, Rok) %>%

 summarize(Rozhodnute = sum(Počet1, na.rm = T),

 VazeneRozhodnute = sum(VazenyPocet1, na.rm = T)) %>%

 ungroup() %>%

 filter(Rozhodnute > 50) %>%

 mutate(Median = median(VazeneRozhodnute, na.rm = T),

 MAD = mad(VazeneRozhodnute, na.rm = T),

 DeviationScore = (VazeneRozhodnute - Median) / MAD)

# Distribucia

ggplot(produktivita2021, aes(sample = VazeneRozhodnute)) +

 geom\_qq() + geom\_qq\_line()

ggplot(produktivita2021, aes(x = VazeneRozhodnute)) +

 geom\_histogram()

ggplot(produktivita2021, aes( y = VazeneRozhodnute)) +

 geom\_boxplot()

# Filter vzdialenych pozorovani + dopocitanie priemeru a SD

produktivita2021nooutliers <- produktivita2021 %>%

 mutate(Outlier = if\_else(DeviationScore <= 2 & DeviationScore >= -2, 0, 1)) %>%

 filter(Outlier == 0) %>%

 mutate(MeanCisty = mean(VazeneRozhodnute),

 SDCisty = sd(VazeneRozhodnute))

# Distribucia bez vzdialenych pozorovani

ggplot(produktivita2021nooutliers, aes(sample = VazeneRozhodnute)) +

 geom\_qq() + geom\_qq\_line()

ggplot(produktivita2021nooutliers , aes(x = VazeneRozhodnute)) +

 geom\_histogram()

ggplot(produktivita2021nooutliers, aes(y = VazeneRozhodnute)) +

 geom\_boxplot()

# Vypocet z score a bodov

produktivita2021body <- produktivita2021 %>%

 mutate(z\_score = (VazeneRozhodnute - first(produktivita2021nooutliers$MeanCisty))/first(produktivita2021nooutliers$SDCisty),

 Body = case\_when(z\_score < -1.96 ~ 0,

 z\_score >= -1.96 & z\_score < -1.96 + ((2\*1.96)/9) ~ 1,

 z\_score >= -1.96 + ((2\*1.96)/9) & z\_score < -1.96 + 2\*((2\*1.96)/9) ~ 2,

 z\_score >= -1.96 + 2\*((2\*1.96)/9) & z\_score < -1.96 + 3\*((2\*1.96)/9) ~ 3,

 z\_score >= -1.96 + 3\*((2\*1.96)/9) & z\_score < -1.96 + 4\*((2\*1.96)/9) ~ 4,

 z\_score >= -1.96 + 4\*((2\*1.96)/9) & z\_score < -1.96 + 5\*((2\*1.96)/9) ~ 5,

 z\_score >= -1.96 + 5\*((2\*1.96)/9) & z\_score < -1.96 + 6\*((2\*1.96)/9) ~ 6,

 z\_score >= -1.96 + 6\*((2\*1.96)/9) & z\_score < -1.96 + 7\*((2\*1.96)/9) ~ 7,

 z\_score >= -1.96 + 7\*((2\*1.96)/9) & z\_score < -1.96 + 8\*((2\*1.96)/9) ~ 8,

 z\_score >= -1.96 + 8\*((2\*1.96)/9) & z\_score < -1.96 + 9\*((2\*1.96)/9) ~ 9,

 z\_score >= 1.96 ~ 10))

produktivita2021body <- left\_join(produktivita2021body, dataID2,

 by = "ID") %>%

 relocate(`Meno a priezvisko`, Meno, Priezvisko, .after = ID)

ggplot(produktivita2021body , aes(x = Body)) +

 geom\_bar()