

```

# R course for beginners
# Week 2
# assignment by Amir Mano, id 205779788

#### create and save data ----

# import packages
library(tidyverse)

# creating variables
N = 100
sub_id <- seq(1,N)
sex <- sample(c('m','f'),N,replace=T, prob = c(0.5, 0.5))
age <- runif(N, 15, 40)
has_depression <- rbinom(N,1,0.083)
df <- data.frame(sub_id, sex, age, has_depression)

# adding IQ to data frame
df$iq <- rnorm(N,100,15)
write.csv(df, './HW_2.csv', row.names = FALSE)

#### calculations and output ----

# extracting stats with base
print(paste('rows:', nrow(df)))
print(paste('columns:', ncol(df)))
print(paste('names:', names(df)))
print(paste('range of ages:', floor(min(df$age)), '-', ceiling (max(df$age))))
print(paste('average IQ:', mean(df$iq)))
print(paste('median precentage of depression:', median(df$has_depression)*100))

# extracting stats with dplyr
df|> mutate(has_depression = has_depression*100)

df|> group_by(sex)|>
  select(age, iq, has_depression) |> filter(age>18)|>
  summarize(sum(has_depression!=0)/n()*100)

df|> group_by(sex)|>
  select(age, iq, has_depression) |> filter(age<18)|>
  summarize(sum(has_depression!=0)/n()*100)

```