

```

# R course for beginners
# Week 7
# assignment by Omri Sabato, id 322777988

#### CREATING FILTERED DATA

#clearing environment and loading data
rm(list = ls())
setwd("./stroop_data")
load("raw_data.RData")
library(dplyr)

print(paste("number of subjects is:", length(unique(raw_data$subject))))

#filtering out NA data
filtered_data <- na.omit(raw_data)

#filtering out rt times
filtered_data <- filtered_data |>
  filter(rt > 300, rt < 3000)

#statistics of removed data
raw_data |>
  group_by(subject) |>
  summarise(original_amount = n()) |>
  full_join(
    filtered_data |>
      group_by(subject) |>
      summarise(filtered_amount = n())
  ) |>
  reframe(
    subject,
    trial_after_filter_percentage = filtered_amount/original_amount,
    mean_filtered = mean(filtered_amount/original_amount),
    sd_filtered = sd(filtered_amount/original_amount)
  )

#save filtered data
save(filtered_data, file = "filtered_data.RData")

```