```r
# R course for beginners
# Week 6
# assignment by Lia Weissbach, ID 323118877



# Part A - Loading and merging files

#1

file_names <- dir("C:/Users/Lia/Documents/Economics and Data
Science/6_sem/R_course/week_6/stroop_data/stroop_data")
df <- data.frame()
for (f in file_names) {
  temp <- read.csv(paste0("C:/Users/Lia/Documents/Economics and Data
Science/6_sem/R_course/week_6/stroop_data/stroop_data/", f))
  df <- rbind(df, temp)
}

#2

num_sub = print(length(unique(df$subject)))
print(length(unique(df$condition)))

#3

missing_count <- sum(is.na(df$rt))
missing_percent <- (missing_count/nrow(df))*100

#4

library(ggplot2)

ggplot(df, aes(x = rt)) +
  geom_histogram(bins = 30, fill = "skyblue", color="black") +
  facet_wrap(~condition) +
  labs(title = "Reaction Times by Condition",x = "Reaction Time (ms)",y = "Count") +
  theme_minimal()



# Part B - Data Cleaning and Finding Outliers

#1

find_Outliers <- function(x) {
  av = mean(x)
  sd = sd(x)
```

```
  logic <- x > av + 2 * sd
  return(logic)
}


#2

all_out <- c()

for (i in num_sub)
  {rts <- df$rt[df$subject == i]
  Out_i = find_Outliers(rts)
  all_out <- c(all_out, Out_i)
  }

df$outlier <- all_out




# Part C - Data Presentation

#1

library(dplyr)

df_summary <- df |>
  group_by(subject) |>
  summarise(num_outliers = sum(outlier))

print(df_summary)


#2

df_clean <- df |>
  filter(outlier == FALSE)

ggplot(df_clean, aes(x = rt)) +
  geom_histogram(bins = 30, fill = "skyblue", color = "black") +
  facet_wrap(~condition) +
  labs(title = "Reaction Times by Condition (Outliers Removed)",
       x = "Reaction Time (ms)",
       y = "Count") +
  theme_minimal()
```