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

#### PRE-PROCCESSING COLLECTED DATA ----

#clear environment and set up data
rm(list = ls())
unzip("stroop_data.zip", exdir = "stroop_data" )
setwd("stroop_data")
library(dplyr)
library(tools)

#binding data to collected_data
collected_data <- data.frame()
for (subjects_data in dir()) {
  if(file_ext(subjects_data) == "csv") {
    collected_data <- rbind(collected_data, read.csv(subjects_data))
  }
  else{
    message(paste("Skipping non-CSV file:", subjects_data))
  }
}

#converting to raw data
raw_data <- collected_data |>
  mutate(
    congruency = factor(ifelse(grepl("_cong", condition), "congruent", "incongruent")),
    task = factor(ifelse(grepl("reading", condition), "word_reading", "ink_labeling")),
    accuracy = if_else(correct_response == participant_response, 1, 0),

    trial = as.numeric(trial),
    block = as.numeric(block),
    rt = as.numeric(rt)
  ) |>
  select(subject, block, trial, congruency, task, accuracy, rt)

#checking for conversion problems on independent variables
ifelse(sum(is.na(raw_data$congruency)) > 0, "congruency variable generation failed",
"congruency variable generation worked")
ifelse(sum(is.na(raw_data$task) > 0), "task variable generation failed", "task variable
generation worked")

#assigning contrasts
contrasts(raw_data$congruency) <- contr.treatment(levels(raw_data$congruency), base =
which(levels(raw_data$congruency) == "congruent"))
contrasts(raw_data$task) <- contr.treatment(levels(raw_data$task), base =
which(levels(raw_data$task) == "word_reading"))

#save raw_data
save(raw_data, file = "raw_data.RData")
```