

Removing unneeded columns (variables) from your data file

Cong Wang, Ph.D.

Loading the data

```
setwd("/Users/congwang/Desktop/Real-world Data/sample data")  
mydata.renamed <- readRDS("mydata.renamed") # loading the data  
names(mydata.renamed) #checking the names of the columns in the data
```

Selecting the variables needed for the analysis

```
library(dplyr)  
mydata.needed <- mydata.renamed %>% select(ID,familiarity,contains(c("FL","RGO","PL")),interesting,identification1:external3,Gender,Age,Ethnicity) # selecting the columns needed  
names(mydata.needed)
```

Or you can remove the variables (i.e., columns) that are not needed for the analysis

```
mydata.needed1 <- mydata.renamed %>%select(-c(StartDate:UserLanguage,contains(c("Click","Page")),Q9,Q50_16_T  
EXT:Q54_6))  
names(mydata.needed1) # removing the columns unneeded
```

Saving the subset of the data as a new file named mydata.needed

```
saveRDS(mydata.needed, file="mydata.needed")
```