

How to merge multiple data files in R

Cong Wang, Ph.D.

Loading the pre and post semester data

```
setwd("/Users/congwang/Desktop/Real-world Data/sample data")
library(readxl)
prelc <- read_excel("PreLearningClimate.xlsx")
postlc <- read_excel("PostLearningClimate.xlsx")
dim(prelc)
dim(postlc)
```

Renaming the variables to make sure that the names of variables are unique across different files except for the key variable.

```
library(dplyr)
prelc.renamed <- rename(prelc,
  "preQ1"="Q1",
  "preQ2"="Q2",
  "preQ3"="Q3",
  "preQ4"="Q4",
  "preQ5"="Q5",
  "preQ6"="Q6")

postlc.renamed <- rename(postlc,
  "postQ1"="Q1",
  "postQ2"="Q2",
  "postQ3"="Q3",
  "postQ4"="Q4",
  "postQ5"="Q5",
  "postQ6"="Q6")

names(prelc.renamed)
names(postlc.renamed)
```

Merging the pre and post data by the key variable (i.e., ID).

```
prepostdata <- merge(prelc.renamed, postlc.renamed, by="ID", all=FALSE) # Keeping the cases having both pre
and post data
allpredata <- merge(x = prelc.renamed, y = postlc.renamed, by="ID", all.x=TRUE) # Keeping all cases from pr
e data and adding the variables from post data
allpostdata <- merge(x = prelc.renamed, y = postlc.renamed, by="ID", all.y=TRUE) # keeping all cases from p
ost data and adding the variables from pre data
alldata <- merge(prelc.renamed, postlc.renamed, by="ID", all=TRUE) # Keeping all cases
dim(prepostdata)
dim(allpredata)
dim(allpostdata)
dim(alldata)
```

Saving the merged data as a new excel file

```
install.packages("writexl")
library("writexl")
write_xlsx(alldata, "alldata.xlsx")
```