

Homework 4

For this assignment please turn in to Canvas a plain text file with your GitHub user name. You will then be assigned to peer grade another student's repository.

1. Create a GitHub repository for your class project and push all files needed to execute your data analysis project and produce a report to GitHub. Include a `README.md` file that contains detailed instructions on how to execute your analysis. This README should explicitly spell out what software/packages need to be installed prior to running the analysis and contain bash code that is required to execute the analysis. See the [example repository](#) for an example.

Your repository will be peer-graded for its clarity. You will be asked to download one of your classmates' repository and try to run their code.

Some hints: * Avoid absolute paths. Note that if you run an R session from the command line, then the working directory for the session is whatever folder R was run from. * When peer grading, don't chase down package install errors. If they happen, note them for your classmate, but no need to try harder. * If the Rmd file errors, note the error in your grading.

For this to work, you may need to make your data anonymous. An easy way to do this in R is follows:

```
# say you have a data.frame called dat where rows are individuals
# and columns are features

# set a seed that only you know -- don't save it anywhere
set.seed(124125)
n <- nrow(dat)
new_dat <- dat
for(j in 1:ncol(dat)){
  shuffle_idx <- sample(1:n)
  new_dat[,j] <- dat[shuffle_idx, j]
}

# new dat now contains data formatted like your original data, but
# where each column has been randomly shuffled. use this version in
# your repository
```