

Problem Set 1

insert your name here

insert date here

Welcome to Data Management and Manipulation Using R!

In this problem set, you will load data from a url, or absolute/relative filepath.

- Note: Change author to your name & date to current date above in the YAML header (we'll learn more about headers later on...)

Step 1: Practice “knitting” the lecture files!

- Open your RStudio via the Rproject (HED96C_RClass.Rproj) we created in class
- Once in RStudio, in the top menu bar, click on File > Open File...
- Navigate to your HED696C_RClass folder and open your saved introduction.Rmd file (hint it should be located in: modules/introduction folder)
- Open the introduction.Rmd file in RStudio
- At the top of the introduction.Rmd file, change my name under “author” to your first and last name
- Now select the “Knit” tab (icon with blue yarn ball) or use the drop down menu next to the yarn ball and select “Knit to PDF”
- The goal of this step is to make sure you are able to knit lecture slides to a PDF without running into errors. But it is OK if you do! We'll figure them out together!

Step 2: Practice with Directories!

Question 1: Load .Rdata directly with url

- View current/working directory first using the `getwd()` function
 - If ever you don't know what a “function” does; you can ask R directly using the syntax `?function` before the function. See my example below in the R chunk
- We're going to open a dataset on my “recruiting” project which is stored in an online repository. Here is the link to that dataset: https://github.com/ksalazar3/HED696C_RClass/blob/master/data/recruiting/recruit_ps1_small.Rdata
- When the data is stored online/in a repository, we can load R datasets (they file extensions are .Rdata) directly into R using the `load()` and `url()` functions by using the link!
 - There's no need to download the file!
- So the syntax will look like this: `load(url("link"))`

```
?getwd()
getwd()
```

```
## [1] "/Users/karinasalazar/Dropbox/HED696C_RClass/problemsets/introduction"
```

```
?load
```

```
load(url("https://github.com/ozanj/rclass/raw/master/data/recruiting/recruit_ps1_small.Rdata"))
```

Question 2: Set your working directory and use absolute filepath to load data

- Download .Rdata using link from above
- Save .Rdata file in the recruiting folder (rclass/data/recruiting)
- View your current/working directory `getwd()`
- Use `setwd()` function to set your working directory to where .Rdata file is saved (rclass/data/recruiting)
- When setting working directory, use the absolute filepath
- Load the .Rdata file `load(name.Rdata)`
- Using the `names(data frame name)` function, report the names in this dataset (Hint: data frame name is `df_recruiting`)

```
getwd()

## [1] "/Users/karinasalazar/Dropbox/HED696C_RClass/problemsets/introduction"

setwd("/Users/karinasalazar/Desktop/rclass/data/recruiting") #absolute file path to data
getwd() #if you run lines entire chunk so you can see how the working directory changes

## [1] "/Users/karinasalazar/Desktop/rclass/data/recruiting"

list.files() #list.files displays the files in the folder

## [1] "recruit_ps1_small.Rdata"

load("recruit_ps1_small.Rdata")
names(df_recruiting)

## [1] "state_code" "school_type" "ncessch" "name" "address"
## [6] "city"
```

Question 3: Set your working directory and use relative filepath to load data

- View your current/working directory `getwd()`
- Use `setwd()` function to set your working directory to where .Rdata file is saved (rclass/data/recruiting)
- When setting working directory, use the relative filepath
- Load the .Rdata file `load()`
- Report the names in this dataset (Hint: data frame name is `df_recruiting`)

```
getwd()

## [1] "/Users/karinasalazar/Dropbox/HED696C_RClass/problemsets/introduction"
# Working directory is in the lecture1 folder, need to get to recruiting folder using relative file path
setwd("../..../data/recruiting")
getwd() #View current directory after setting the directory

## [1] "/Users/karinasalazar/Dropbox/HED696C_RClass/data/recruiting"

list.files() #listing files in the recruiting folder

## [1] "create_recruit_events.R" "create_recruit_school.R"
## [3] "data_hs.csv" "events_data.csv"
## [5] "events_data_dp.csv" "recruit_event_allvars.RData"
## [7] "recruit_event_somevars.RData" "recruit_ps1_small.Rdata"
## [9] "recruit_school_allvars.RData" "recruit_school_somevars.RData"
## [11] "text"
```

```
# Lets load recruit_ps1_small.Rdata
load("recruit_ps1_small.Rdata")
# Use names function to view the names in the data set (data frame)
names(df_recruiting)

## [1] "state_code" "school_type" "ncessch" "name" "address"
## [6] "city"

# Run the entire code chunk
```

Question 4: Load dataset using relative filepath without first changing working directory

- View your current/working directory `getwd()`
- Load data without setting directory `load(filepath/data.Rdata)`

```
getwd()

## [1] "/Users/karinasalazar/Dropbox/HED696C_RClass/problemsets/introduction"
load("../data/recruiting/recruit_ps1_small.Rdata") #Set file path and load data at once
names(df_recruiting)

## [1] "state_code" "school_type" "ncessch" "name" "address"
## [6] "city"
```

Knit to pdf select the “Knit” tab (icon with blue yarn ball) or scroll down and “Knit to PDF”

- Submit both .Rmd and pdf
- Use this naming convention “lastname_firstname_ps#”