

Introduction Problem Set

insert your name here

insert date here

Welcome to Data Management and Manipulation Using R!

In this problem set, you will practice “knitting,” loading data from a url, and loading data from absolute/relative filepaths.

- Note: Change author to your name & date to current date above in the YAML header (we’ll learn more about headers later on. . .)
- If you’re experiencing errors while working on this assignment (or any assignment), start fresh by closing out R (when prompted to “save your workspace” hit “don’t save) and re-opening the assignment.
- You may also be experiencing errors for this particular assignment because of running R chunks for questions below out of order (run Question 1, then Question 2, then Question 3)

Step 1: Practice “knitting” the problem set Rmd file!

- 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_ps.Rmd` file (hint it should be located in: `problemsets/introduction` folder)
- Open the `introduction_ps.Rmd` file in RStudio
- At the top of the `introduction_ps.Rmd` file, insert your name and the date in the first few lines of this .Rmd file where indicated for you
- Now select the “Knit” tab (icon with blue yarn ball) or use the drop down menu next to the yard ball and select “Knit to PDF”
- The goal of this step is to make sure you are able to knit 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

- In the R Chunk Below...
- View current/working directory first using the `getwd()` function
 - The current working directory when running from an R Chunk will revert to the location of this introduction_ps.Rmd file; which should be in the problemsets folder within the introduction sub-folder
 - If ever you don't know what a "function" does; you can ask R directly using the syntax `? before the function.`
- 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?raw=true
- When the data is stored online/in a repository, we can load R datasets (their file extensions are .Rdata) directly into R using the `load()` + `url()` functions, or `load(url())` when combined, by using the link!
- So the "generic" syntax will look like this:
 - `load(url("INSERT THE LINK HERE"))`
- Whereas the filled in syntax for this part of the question will look like this:
 - `load(url("https://github.com/ksalazar3/HED696C_RClass/blob/master/data/recruiting/recruit_ps1_small.Rdata?raw=true"))`

Solutions provided; see how I followed the instructions to get the syntax answers in the R Chuck below

```
getwd()
```

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

```
?getwd()
```

```
?load
```

```
load(url("https://github.com/ksalazar3/HED696C_RClass/blob/master/data/recruiting/recruit_ps1_small.Rdata?raw=true"))
```

Question 2: Set your working directory using an absolute filepath; then load the data directly from your computer

- Download .Rdata using link from Question 1 above
- Move the `recruit_ps1_small.Rdata` file (should be in your downloads folder) in the recruiting data folder (HED696C_RClass/data/recruiting)
- Then, in the R Chunk Below...
- View your current/working directory `getwd()` first
 - This will revert to the location of the introduction_ps.Rmd file; which should be in the problemsets folder within the introduction sub-folder
- Then use the `setwd()` function to set your working directory to where the .Rdata file is saved using the **absolute filepath**
 - Hint: This should be pointing to the recruiting folder where the .Rdata file is located, it should not include the name of the Rdata file.
 - For example, my absolute file path is `"/Users/karinasalazar/Dropbox/HED696C_RClass/data/recruiting"`
 - So MY solution for this part of the question would be `setwd("/Users/karinasalazar/Dropbox/HED696C_RClass/data/recruiting")`

- After setting your working directory you can use `getwd()` again to see how they changed!
- After setting the your working directory using the absolute filepath, load the .Rdata file by using the following generic syntax `load(INSERT NAME OF DATAFILE HERE.Rdata)`
- Using the `names(data frame name)` function, show the column names in this dataset (Hint: data frame name for the `recruit_ps1_small.Rdata` is called `df_recruiting`)

Write your solutions in the R Chunk below

```
getwd()

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

setwd("/Users/karinasalazar/Dropbox/HED696C_RClass/data/recruiting")
list.files()

## [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"

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

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

Question 3: Set your working directory using a relative filepath; then load the data directly from your computer**

- In the R Chunk below...
- View your current/working directory `getwd()`
 - This will revert to the location of the `introduction_ps.Rmd` file; which should be in the `problemsets` folder within the `introduction` sub-folder
- Then use the `setwd()` function to set your working directory to where the .Rdata file is saved using the **relative filepath**
 - Hint: you are currently in the `problemsets/introduction` folder; so you will need to move “up” or “back” before you can get to the `data/recruiting` folder
 - For example, if my current working directory is: `“/Users/karinasalazar/Dropbox/HED696C_RClass/problemsets/introduction”`
 - So my solution for this part of the question would be `setwd("../..data/recruiting")`
- After setting the your working directory using the absolute filepath, load the .Rdata file by using the following generic syntax `load(INSERT NAME OF DATAFILE HERE.Rdata)`

- Show the column names in this dataset (Hint: data frame name is df_recruiting)

Write your solutions in the R Chunk below

```
getwd()

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

setwd("../..data/recruiting")
list.files()

## [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"

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

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

Question 4: Without first changing the working directory, load the dataset using relative filepath

- View your current/working directory `getwd()`
- Load data without setting directory `load("relativefilepath/recruit_ps1_small.Rdata")`
- Hint: this will look like a combination of the two steps in Question 3

Write your solutions in the R Chunk below

```
getwd()

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

load("../..data/recruiting/recruit_ps1_small.Rdata")
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_introduction_ps.Rmd”