

# PH251D Fall 2019 - Project 1

*FirstName MI LastName*

*2019-MM-DD*

Download this Rmarkdown template (PH251D2019\_LastName\_Project1.Rmd) and edit.

Create a project folder called `project1` on your computer. You will put all your Project 1 files in this folder.

Use this template and R Markdown to demonstrate the following skills:

## 1. Using the source function

Create a R script file named `problem1.R` and save it into your `project1` folder. Type `print("Hello World")` and `source` this file.

```
## R code to source file here
```

## 2. Read an ASCII data set

The Evans data set (`evans.txt`) is here: <https://github.com/taragonmd/data>.

Alternatively, here is the raw Evans data set: <https://raw.githubusercontent.com/taragonmd/data/master/evans.txt>.

Demonstrate reading the Evans data file (`evans.txt`) to create a data frame, and use the `str` function to explore the structure of the data set. The data dictionary is in Appendix C of the PHDS book.

Show the R code chunk and results below.

```
## explore structure of Evan County data
```

## 3. Discretizing a continuous variable into a categorical variable

Use the `cut` function to discretize age into the following age categories and make a table of counts and a table of proportions.

- 30-39
- 40-49
- 50-59
- 60-69
- $\$ > \$70$

Be sure to pay attention to age interval transitions.

```
## discretizing age into categories
```

## 4. Working with dates and times

President Donald Trump was elected on "November 8, 2016". Convert this character string into a R date object. Show how to use R to display (a) the Julian date; (b) the day of the week, and (c) the week of the year.

```
## working with dates
```

## 5. Simple two-way analysis

Create a simple 2x2 table of smoking (`smk`) and coronary heart disease (`chd`). Use the `fisher.test` on this 2x2 table and describe your findings.

```
## 2x2 table
```

## 6. Write your own function

Now, write a function to calculate the risk ratio of your 2x2 table above. The exposure is smoking status and the outcome is coronary heart disease.

```
## function to calculate the risk ratio
```

## 7.

Now, use the `xtabs` function to create a 3-D array object of `chd`, `hpt`, and `smk`. Now use the `addmargins` function on this object.

```
## demo of xtabs function below
## demo of addmargins function below
```

## 8. Create a PNG graph and save file

From the Evans data create a histogram of age (`age`). Label with a title and axis labels. Output to a PNG file using the `png` function. Hint is provided.

```
png(file = "myplot.png") # start PNG device
## enter code below to create histogram
dev.off() # close device
```

```
## pdf
## 2
```

## 9. Display PNG file in your Rmarkdown document

Using Rmarkdown syntax, display the PNG file you created above. Hint: use the `include_graphics` function from the `knitr` package.

```
## enter code below
```

## 10. Using regular expressions

Here are the California counties: <https://raw.githubusercontent.com/taragonmd/data/master/calcounty.txt>

Read in data using the `scan` function. Hint provided below.

Remove the “California” entry.

Use regular expressions to identify and display the County names that start with "San " and end with "o".

```
cac <-
scan('https://raw.githubusercontent.com/taragonmd/data/master/calcounty.txt',
     what="")
## enter code below
```