hw01: Welcome to the Jungle

STAT 385, Spring 2018

Due: Friday, February 9th, 2018 at 11:59 PM

Overview

Please see the homework policy for detailed instructions and some grading notes. Failure to follow instructions will result in point reductions. In particular, make sure to commit each exercise as you complete them.

> Hofstadter's Law: "It always takes longer than you expect, even when you take into account Hofstadter's Law."
>
> — **Douglas Hofstadter**, Gödel, Escher, Bach: An Eternal Golden Braid

Objectives

The objectives behind this homework assignment are as follows:

- Create an *RMarkdown* document and write using *Markdown* syntax;
- Apply the principles of literate programming by embedding code and describing outcomes;
- Read in data;
- Provide data overviews;
- Clone a `git` Repository;
- Commit and Push changes to a `git` repository;
- Become familiar with the homework procedures of the course.

Grading

The rubric CAs will use to grade this assignment is:

Task	pts
Link to GitHub Repository	2
Agree to the homework policy	2
Verifying computing environment is setup	2
Listing where to get help in person	2
Writing in *RMarkdown*	10
Working with Baby Name Data	10
Exploring the Excellent Teachers list at UIUC	12
At least one commit per exercise (more is better!)	5
Commit messages that describe what changed	5
Total	50

Note on Markdown

If you need help with markdown syntax, please go to the "Help" menu and select the *Markdown Quick Reference* guide. This will open in the **Help** tab on the *lower-right* corner of *RStudio*. For more examples, please see the literate programming slides and the in class examples of writing in *RMarkdown*.

Assignment

(2 points) Exercise 0: Get aboard the GitHub Bus!

Place a link to your `hw01` GitHub repository here.

(2 points) Exercise 1: Homework Policy

Please uncomment the following statement when you have read and agreed to the homework policy. To *uncomment* a statement in *RMarkdown* delete the `<!-- -->` surrounding the content.

(2 points) Exercise 2: Help! I need somebody

Please answer:

1. Who is part of the STAT 385 instructional staff?
2. Where are **all** STAT 385's Office Hours?
3. When do the office hours take place during the week?

Answers to these questions can be found on the main page of the course website:

http://stat385.thecoatlessprofessor.com/

(2 Points) Exercise 3: Know Thine Environment

Please take screenshots of the following and include them in your *RMarkdown* document:

1. the RStudio Cloud STAT 385 Workspace.
2. the STAT 385 Discussion Forum
    - To get access to this forum, we *must* know your GitHub Username.
    - Please fill out this survey: https://goo.gl/forms/n0U7Q87walHvNAhi2
    - Having issues logging in? Please see the walkthrough guide or the FAQ entry

To take a screenshot press:

- Windows: `Windows Key` + `PrtScn` or use the Snipping tool
- macOS: `Command` + `Shift` + `3` or use `Command` + `Shift` + `4` for part of your screen.

(10 Points) Exercise 4: Who I Am

If you need help with markdown syntax, please go to the "Help" menu and select the *Markdown Quick Reference* guide. This will open in the **Help** tab on the *lower-right* corner of *RStudio*. For more examples, please see the literate programming slides and the in class examples of writing in *RMarkdown*.

- Create a self-portrait of yourself by either taking a picture or sketching it. Include this self-portrait within the *RMarkdown* document.
    - Make sure to upload and commit your photo!
- Make a 7 by 2 table in markdown that has a header row containing "Overview" and "Who I Am".
    - Under the "Overview" column, please write entries using bold text for: Full Name, NetID, Birthday, Year in School, Major, and Expected Graduation Date
    - Under the "Who I Am" column, please fill in your personal information.
- Compile **ordered** lists for each of your favorite:
    - foods;

$$f\left(k \mid \lambda\right) = \frac{\lambda^k \exp\left(-\lambda\right)}{k!}$$

Figure 1: LaTeX Formual

  - TV shows;
  - movies;
  - music (add links to music videos on either YouTube or Vimeo).
- Devise *two* **unordered** lists that contain your most recent memorable events and where you typically spend your free time.
- Write the following formula as an *inline* equation.
  - For help writing in LaTeX, see the following symbol's guide: https://artofproblemsolving.com/wiki/index.php/LaTeX:Symbols and see `lab00` solutions.
- What is the name of your favorite mathematical formula? Include the formula itself using *display mode* and a link to its wikipedia entry.
  - For inspiration, check out Wikipedia's Mathematical Formula list!
  - **Note:** You *cannot* select the pythagorean theorem, golden ratio, or quadratic formula as those were given as examples in `lab00`.

Commit and push your work onto GitHub.

(10 Points) Exercise 5: Got baby?

[**2 Points**] **(a)** Install and load the `babynames` package. **Comment** out the installation command in your `.Rmd` file. (If you do not comment installation commands out, then they will be run every time you knit your `.Rmd` file.)

[**2 Points**] **(b)** Open up the help documentation for `babynames` (the data set), find where the variables for the data set are listed and write in your *RMarkdown* document *how* the `prop` variable was created.

[**6 Points**] **(c)** Provide summary information on the `babynames` by:

- writing a sentence that *dynamically* describes the dimensions of the data;
- showing the last six observations in the data set; and,
- providing a summary overview of the data to understand its contents.

(12 Points) Exercise 6: Excellency at UIUC

Under this exercise, we will explore the "Teachers Ranked As Excellent" data at UIUC from Fall 1993 to Present as compiled by Wade Fagen-Ulmschneider. Please download the data from:

https://raw.githubusercontent.com/wadefagen/Teachers-Ranked-As-Excellent-UIUC/master/TRE-UIUC-AllYears.csv

This data has a file extension of **CSV** form. Contained in the data are the following variables:

- `term`: Two letter semester code (`sp`, `su`, `fa`, or `wi`) followed by a four digit year.
  - Examples: `sp2017`, `fa2013`, `su2012`.

- `unit`: The CITL-supplied headers for the unit teaching the course.
  - Examples: `ACCOUNTANCY`, `SOCIAL WORK`, `LINGUISTICS`, `NUCLEAR, PLASMA & RAD. ENGR.`
- `lname`: The last name of the teacher.
  - Examples: `FAGEN-ULMSCHNEIDER`, `FLANAGAN`, `FLECK`
- `fname`: The first letter of the first name of the teacher.
  - Examples: `W`, `K`, `M`
- `role`: `Instructor` or `TA`
- `ranking`: `Excellent` or `Outstanding`
- `course`: The course the teacher was ranked as excellent. If no course is given, the `course` field is set to `?` (this includes cases when the raw data lists the course as `0`, `000`, or `999`).
  - Examples: `199`, `225`, `560`, `?`

[2 points] (a) Import into *R* the data in `TRE-UIUC-AllYears.csv` with the variable name `tre_data`. As `course` contains a value that is *not* `NA`, which is how *R* represents missing values, you must use the parameter `na.strings = c("?", "NA")` in the appropriate `read.*()` function. If you need help, please see the appropriate help documentation via `?`.

NB The `*` in the `read.*()` statement is a placeholder for the type of file you want to read in.

[2 points] (b) Retrieve the dimensional information for this data using only one function.

[4 points] (c) Perform a summary of the `tre_data`. Are there any summaries that are surprising? What might have caused this?

Hint Consider looking at the structure of `tre_data`.

[4 points] (d) Fix the following code so that it:

1. doesn't error;
2. produces a graph; and
3. hides the code.

Hint for the last exercise look at different code chunk options.

```
ggplot(tre_data, aes(ranking)) +
  geom_bar() +
  facet_wrap( ~role) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(y = "Frequency",
       x = "Excellency Rating")
```