

# CBB752 (Spr. '17) - Biomedical Data Science: Mining and Modeling

 Search this site ▾

[Home](#) [Announcements](#) [Syllabus](#) [Instructors](#) [Grading](#) [Section Readings](#) **[Assignments](#)** [Quiz Archive](#) [Final Project](#)

## Assignments

### [Homework 0](#)

posted Dec 3, 2015, 10:17 AM by Donghoon Lee [ updated Jan 12, 2017, 11:51 AM by Paul Muir ]

Welcome to the 20th iteration of "Biomedical Data Science: Mining and Modeling" course. Homework 0 is only meant for you to actively check-in to the class and get involved in a crowd-sourced activity.

HW 0 will be due on 1/23/2017. Please complete this assignment at your earliest convenience.

HW 0 consists of TWO parts, completing a poll to check you into the class and crowd-sourced comments and edits for "List of Study Topics in Bioinformatics" for both fun and educational purposes.

1. Please complete the following poll:

#### [Assignment 0 poll](#)

2. First, please make sure you are logged onto Google with Yale email address. This will ensure us to track who made comments and edits. Once logged in, please make "comments and edits" for the following Google document:

#### [Assignment 0 google doc](#)

---

### [Homework 1](#)

posted Dec 2, 2015, 2:00 PM by Xiu Huang [ updated Dec 21, 2016, 1:31 PM by Mengting Gu ]

## **DUE: Monday February 20th 11:59PM**

### **Non-Programming Module**

**You can get the non-programming part from either of the following three ways:**

1. Download the attached HW1.pdf file containing all the information about Homework 1. You only need to complete the non-programming modules. Submit the completed file through Canvas.
2. If you want to learn about RMarkdown, you can also download the attached zip file and complete the homework using the .Rmd file. You can find more about RMarkdown file below. Submit the completed .Rmd file through Canvas.
3. If you want to learn about RMarkdown and also turn in your homework using Github, you are more than welcome to get access to the assignment and set up your own homework repository using Github. To do so, just follow the instruction in the following Programming module. You are only required to complete the non-programming module. Remember to commit and sync your changes to the Github server.

### **Programming Module**

#### **Start up with Homework 1 & Submission Instruction**

##### **1. Accept the assignment**

- Click on the [Invitation Link](#) [NEED TO UPDATE LINK] to accept the homework repository at GitHub (If you have already done so before 2/2/2017, **please re-accept the invitation link** since the startup directory has been changed, and please make sure you are working on from the latest homework directory). Also, please note that the programming module part 2 has been modified. Update your code accordingly to the specification described.

## 2. Do you have Github account already?

- If YES, sign in.
- If NO, register an [educational GitHub account](#); it has the added perk of giving you some free private repositories for a couple of years. You probably want a *student, individual* account. Remember to use your Yale mailbox to sign up to get the educational discount.

## 3. Authorize and accept the assignment

- You will now have the access to a private repository for your homework assignment under the [CBB752Spring2016 Github classroom](#) [NEED TO UPDATE LINK].

## General Policy

### Language and Program

- You can either use [R](#) or [Python](#) to complete the assignment.
- For [R](#), we recommend using version 3.2.2 and up, and [Rstudio](#) to work on the homework.
- For [Python](#), we recommend using version 2.7 and up.
- If you want to use other languages, please consult with the TAs and request for instructor's permission.

### Using R

- **R Markdown**
  - **Start working on HW1 by editing the HW1.Rmd file** following the instructions. Your homework is written in the format of [R Markdown](#). Don't worry, it is just a normal txt file with the file extension `.Rmd`.
  - HW1.Rmd is the start-up sample skeleton of the final submitted report. You can download it to your local repository and edit it using Rstudio. Or you can edit it directly using the web UI by clicking on the "edit this file" icon at top right.
- **HTML**
  - Compile your homework to Markdown (file extension should be `.md`) and then to HTML (file extension should be `.html`).
  - RStudio's "Knit HTML" button will do this
  - Notice that the intermediate Markdown files (cache or figures) are required to present your full report.

### Using Python

- **Start working on HW1 by following instructions inside the HW1.Rmd file.**
- Please commit both python code file(s) \*.py and a README to GitHub.

### What to put (or not put) into your Git(Hub) repository

*This is rather specific to CBB 752 and may not necessarily reflect your workflow in the future and in other contexts.*

- Do not commit the input data to your repository.
- Locally, you are of course encouraged to keep the file in some logical place within the homework assignment's directory. But list the names of such data files in your top-level [.gitignore](#) file, so that Git ignores it. We do this so that TAs don't end up with 50 copies of the input data when they mark your work.
- Commit the intermediate Markdown (`.md`) file and the figures generated.
- (For R users) Commit the end product HTML (`.html`) file.
- You may not want to commit the Markdown and HTML until the work is fairly advanced, maybe even until submission. Once these enter the repo, you really should recompile them each time you commit changes to the R Markdown source, so that the Git history reflects the way these files should evolve as an ensemble.
- (For R users) **Never ever** edit the Markdown or HTML "by hand". Only edit the R Markdown source and then regenerate the downstream products from that.

### How to "turn in" your homework

- (For R users) Commit the main R markdown (`.Rmd`) file that constitutes your solution. Commit early, commit often!
- (For Python users) Commit the markdown (`.Rmd`) file that constitutes your solution (if applicable). Please also include a brief README file describing how to run your code. Commit early, commit often!
- Commit all the other relevant files (\*.r or \*.py).
- If you work on the copy locally:

- Make sure you have committed all the files associated with your solution in your local Git repository.
- Make sure you have pushed the current state of your local repo to GitHub (Sync).

## Additional tips for R users

---

### Make it easy for others to run your code

- In exactly one, very early R chunk, load any necessary packages, so your dependencies are obvious.
- In exactly one, very early R chunk, import anything coming from an external file. This will make it easy for someone to see which data files are required, edit to reflect their locals paths if necessary, etc. There are situations where you might not keep data in the repo itself.
- Pretend you are someone else. Clone a fresh copy of your own repo from GitHub, fire up a new RStudio session and try to knit your R markdown file. Does it "just work"? It should!

### Make pretty tables

There are a few occasions where, instead of just printing an object with R, you could format the info in an attractive table. Some leads:

- Consider the `kable()` function from `knitr`. via *Rod Docking* This is fairly primitive, one step up from just printing the object.
- Consider use the `xtable` package for making pretty HTML tables.

### Some more resources

- Ten minutes "Hello World" guide to Github: <https://guides.github.com/activities/hello-world/>
- R Markdown Reference Guide: <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf?version=0.99.764&mode=desktop>

## Additional tips for Python users

---

- For reference, I would recommend [Python Cookbook, 3rd Edition](#). Recipes for Mastering Python 3. By David Beazley, Brian K. Jones. Publisher: O'Reilly Media. You can view [Online copy at Yale Library](#) and [Github source code from the book](#).

*Some of the materials were borrowed from [stat540\\_2014/homework/hw\\_submission-instructions.md](#)*

Attachments:  [hw1.pdf](#)

1-2 of 2

