ADS2 Problem Set 1

MI Stefan

5 September 2019

We expect this problem set to take around an hour to complete. But professors are sometimes wrong![citation missing] If this or future problem sets are too long, please let us know, so we can adjust and plan accordingly.

Git, GitHub refresher

Remember Git and GitHub? As we did for IBI1, we will distribute problem sets and data files for this course via GitHub.

In this course, there is no need to keep a practical portfolio or a portfolio of your responses to the problem sets, but it may be useful to you to use git to keep track of your work.

The course repository can be found here: https://github.com/MelanieIStefan/ADS2_2019-20

Check if you can remember how to do the following:

- Forking the ADS_2019-20 repository in GitHub
- Opening GitKraken
- Cloning the ADS_2019-20 repo in GitKraken
- Creating a file in this repo, adding it to git
- Making changes to the file and committing the changes

There is also in-built git functionality in RStudio. You are not required to use it in this course, but you may find it helpful. This allows you to run git commands from the RStudio interface, so you don't have to switch back and forth between RStudio and GitKraken while you work. While RStudio can do simple things like commit, push, pull, and examine the history of a git project, it cannot do more complex things like branching or merging. (You would still have to use GitKraken for that.) If you would like to try out the Git functionality in RStudio, try the following steps:

- Click on File -> New Project. If you want to clone a new repository from GitHub, select "Version Control" and follow the instructions. If on the other hand you want to work with an already existing git directory on your computer, select "Existing Directory" and navigate to the directory of your choice.
- There should now be a new tab named "Git" in the "Environment, History, ..." pane in RStudio (for many people, this will be on the upper right side, but this may vary). If you click on it, you will see it has a list of files that have recently changed, as well as symbols for common git commands, such as "Commit", "Pull", and "Push". You can also click on the little clock icon to see the git log of the project. Try making changes to a file and committing, and see whether you can see the commit in the log.
- Importantly, GitKraken and RStudio are not mutually exclusive - you can see that commits you make in RStudio show up in GitKraken and commits you make in GitKraken show up in RStudio. They are just two different interfaces to talk to the same underlying version control system. You can choose how you want to use them in your own workflow. For instance, you may want to use RStudio for convenient quick commits while you are working on a project, but then use GitKraken when you want to do more complex things like create new branches.

Looking at student grades

In an introductory biomedical informatics class at the University of Awesome, the mean course mark is 86, with a standard deviation of 5.0

- Use this information to generate a distribution of marks that might be expected for a class with 100 students.
- Plot that distribution as a histogram with informative axis labels and in a pretty colour
- In this distribution, how many students have a mark higher than 91 or lower than 81? How many students have a mark higher than 96 or lower than 76?
- Discuss one potential problem with creating a predicted grade distribution in this way

Getting good grades

Oh no! There is an exam that you have not prepared for! The exam consists of 20 multiple-choice questions. You know none of the answers! Each question has four answer choices, of which exactly one is correct. You need to answer 10 or more questions correctly in order to pass. The examiners have used a random number generator to determine which of the answer choices (A, B, C, or D) should have the correct answer.

- Which of the following strategies gives you a higher chance of passing?
  - Randomly selecting one of the four answers for each question
  - Selecting "A" for every answer
  - They are the same
- Often, exams are set in such a way that the choice of the correct answer (A, B, C, or D) is random, but with the additional constraint that each of the four choices should be used equally often for the correct answer. If the exam is set in such a way, does this change your answer? Why or why not?