David_Ray_Review

Amy Pollpeter

10/21/2019

Review of David Ray's R Project and R-Markdown by Amy Pollpeter.

Running the R-Markdown file

Ran the entire R-Markdown file. ### First error encountered: Error: '/Users/Harrisvaccines/Documents/R_Assignment/R_Assignment_Data_Sets/fang_et_al_genotypes.txt' does not exist.

This error is occuring because the absolute path used to read in the original data file does not exist on my computer. I would suggest using a web address, either to the class git repository, or to your own git repository to refer to the file. For example:

fang_data <- read_tsv("https://github.com/EEOB-BioData/BCB546X-Fall2019/blob/master/assignments/UNIX_Assignment/fang_et_al_genotypes.txt (https://github.com/EEOB-BioData/BCB546X-Fall2019/blob/master/assignments/UNIX_Assignment/fang_et_al_genotypes.txt)")

I tried to use the web address to call the file, but it didn't read in as a data frame. Instead I just used the absolute path to the file on my computer to read it in correctly.

Second Error encountered

After making the corrections to get the original files into the projects, the R-Markdown ran to line 115 and encountered another error - also due to the fact that the absolute path did not exist on my computer. This same error was encountered at line 115 and 127 (for maize and teosinte data respectively). "cannot open file '/Users/Harrisvaccines/Documents/R_Assignment/R_Assignment_Output/maize_in1.csv': No such file or directoryError in file(file, ifelse(append,"a","w")) : cannot open the connection"

Third Error encountered:

Again at line 301 the program encountered an error due to the absolute path. I changed the path to a path on my computer and continued running the program.

Overall assessment of R-Markdown:

The only errors I encountered were due to absolute path issues. I tried to use a relative path and it often still did not work, the code would run only when I replaced the absolute path with an absolute path on my own computer. Other than getting around this issue, the code ran completely.

Review of Graphs Portion of the Assignment

On the scatter plots of # of SNPs per Chromosome, it would be helpful to add a legend so it is clear as to which graph is maize and which is teosinte.

The code for the "Distribution of SNPs on Chromosome (Lines 209-228) did not produce an output in the R-Markdown file itself.
However, because these were saved as PDF output files, I was still able to look at the graphs (great idea!). The one suggestion I have on the graphs is to add a title to them to show which is maize data, and which is teosinte data.

The "Amount of Heterozygosity" code also did not produce a graph in the r-markdown when I ran it on my computer, however, I was able to look at the PDFs of these graphs.

Again, I would add a title to each graph showing which is maize and which is teosinte. Because the amount of homozygosity is so small relative to heterozygosity, it is somewhat difficult to see much detail on the graphs other than that the heterozygosity is much longer than the homozygosity. There might be a more descriptive way to show this data using a different type of graph.

The graph of "Your own visulization" was interesting. Again, a descriptive title would be very helpful. It is a little difficult to tell what exactly this plot is showing that is different from the scatter plots of SNP distribution above, but it is an interesting graph.

Overall Suggestion:

My main suggestion would be to create a new project, and then pull your Markdown file into that project and run it using the "Run -> all all" function. This would be a quick way to see if any absolute paths are throwing errors.

Great Job!