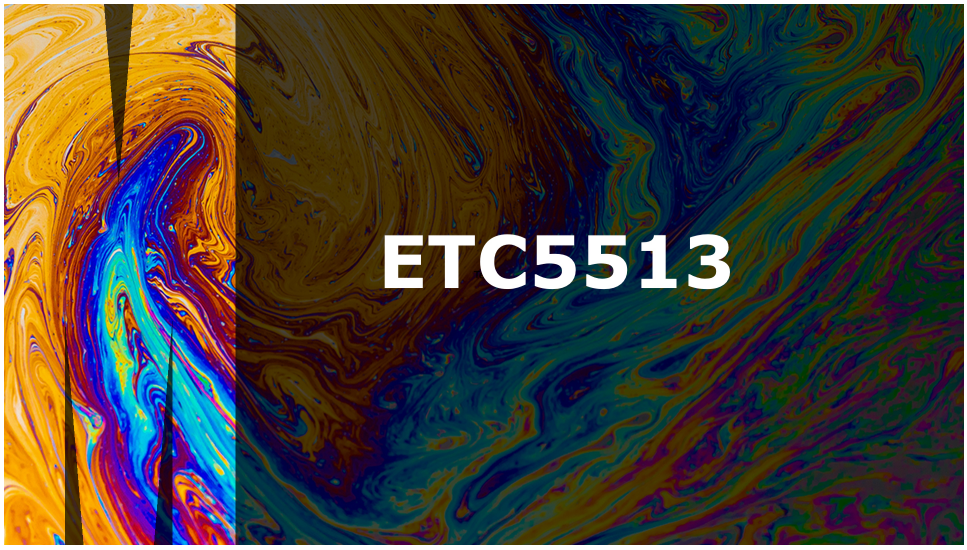


Reproducible and Collaborative Practices

Assignment 1

Contents

Assignment objectives

Demonstrate that you are capable to create a reproducible report using R, Rstudio, Git and GitHub employing all the tools that we have learned in Weeks 1 to 4. In particular:

- Working on a reproducible Rstudio project
- Produce a html report using an Rmd file and explore YAML themes to define your prefer template
- Practice markdown syntax
- Practice R coding
- Explore R chunk options to customize your report template
- Create html tables, add table captions and refer to them in the report text as described in Week 4
- Create simple figures to visualize the data, add figure captions and refer to the figures in the report text as you learned in Week 4
- Demonstrate that you are able to clone a GitHub repository locally and synchronize the changes between your local and remote repositories.
- Show that you can create meaningful Git commits so that the changes and the history of the project can be recorded and tracked.

This is an INDIVIDUAL assignment.

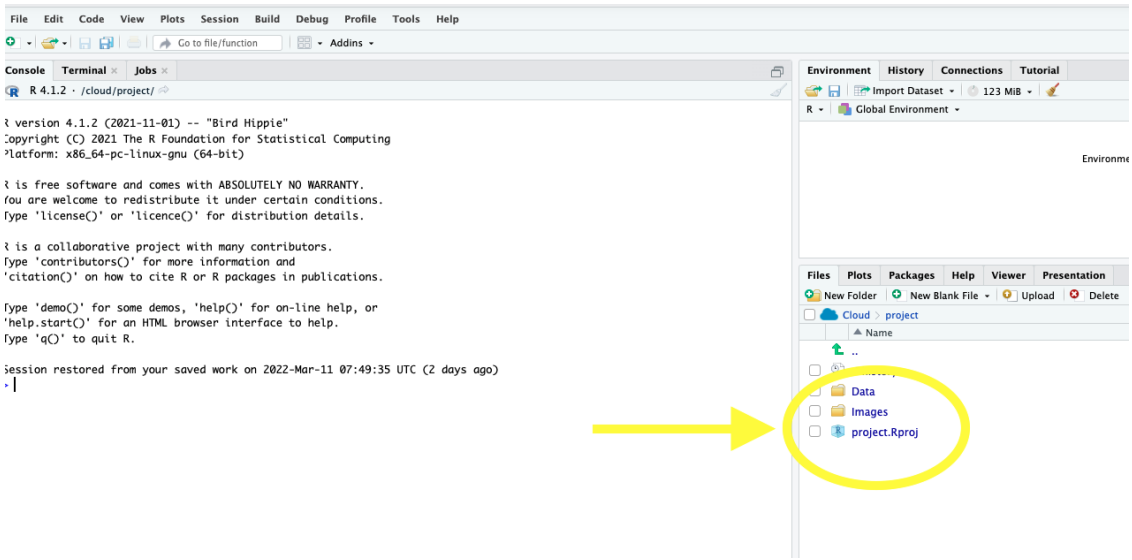
VERY IMPORTANT: The printed final pdf report cannot have more than 5 pages (that does not include the Appendix pages)! You cannot use a data set that has been used for other assignment or in another units. Please read the instructions below carefully.

Assignment description part A

In this assignment, you need to create an Rstudio project that will be able to produce a reproducible html report. For the report, please select a data set that interest you **from this webpage**: Our world in data.

The size of the data set is not relevant, however your data must contain at least 3 numerical variables and a character or factor variable and its size must not be more than 50MB. Your reproducible report should be created using an Rmd file in Rstudio and needs to be knitted into a html file. The report must knit into html without any errors and must have have all the code and code outputs displayed (unless otherwise specified in the instructions below).

The reproducible project files structure should look like this:



Assignment description part B

In this part you will transfer the report you created in Assignment part A into a GitHub repository and add a new section into the report that you created in part A. Please see instructions below.

Instruction for completing the assignment part A

Your assignment should contain the following elements and the R code inside the R code chunks should be visible (unless otherwise specified):

Element 1: R code chunk for loading libraries (1pts)

- You need to load all the libraries that you are going to use in an R code chunk located at the beginning of your Rmd file.

For this part: Make sure you set the necessary R code chunk options so that you do not display the R code or any messages or warnings in the rendered html document. (1pts)

Element 2: Introduction section (5pts)

Using markdown language, provide in 4 sentences **maximum**, the motivation of what you are going to research and why.

Element 3: Research question section (5pts)

- Using markdown language, discuss in 3 sentences **maximum**, the specific question that you are going to investigate/answer in this report using your selected data.

Element 4: Data set introduction section (5pts)

- In this section briefly describe your data (what is the data measuring/recording) in five sentences **maximum** using markdown language. You must provide a hyperlink to the location of the data inserted in the text using markdown language.
- Create a table using the *kable()* function from the *knitr* R package to report the variable names. Add a table caption where you briefly describe the table in no more than 2 sentences. In addition, refer to the table in the report text.

Element 5: Data set description subsection (5pts)

Here you need to report details about your data. Specifically, the size of our data set such as number of observations, variables and variable types. Below I give you the list of things that you need to include:

- Using *inline R code* write a sentence describing the number of variables and observations in your data set.
- In addition, create a screenshot image of your inline R code and save it as a png file. Upload this png file inside the "Image" folder in your Rstudio project.
- Then create an R code chunk to display the figure that you created with the inline code and add below a sentence explaining the code you use. Hint: Remember that for inserting figures you can create an R code chunk and use the R command *knitr::include_graphics()*. Make sure you have the correct relative path to the Image folder.

- Using the function *head()* display the first 2 rows of the data so that you can show the type of variables that are in the data set (numeric, character/factor etc).

Element 6: Results section (5pts)

Using visualizations of the data you selected discuss the answer to your research question. For that:

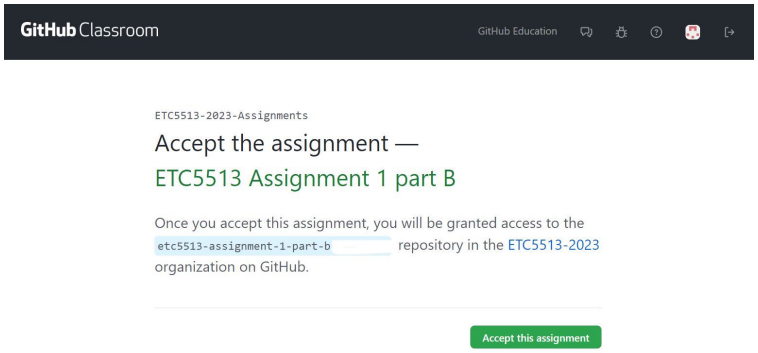
- Create 2 figures maximum of your data that will help you answer your research question and add a caption in each of them using the relevant options inside the R code chunk. Create the figures using the *ggplot2* R package.
- Using markdown language, add a list with two bullet points describing what you see in the each of the figures using italic font and how that can help you answer your research question. Make sure you refer to these figures in the text.

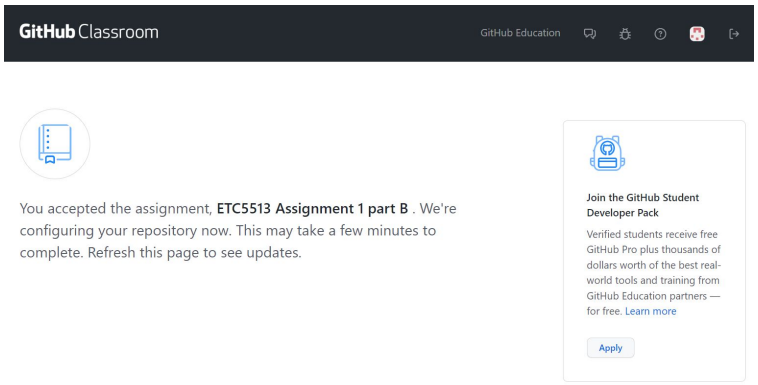
Instruction for completing the assignment part B

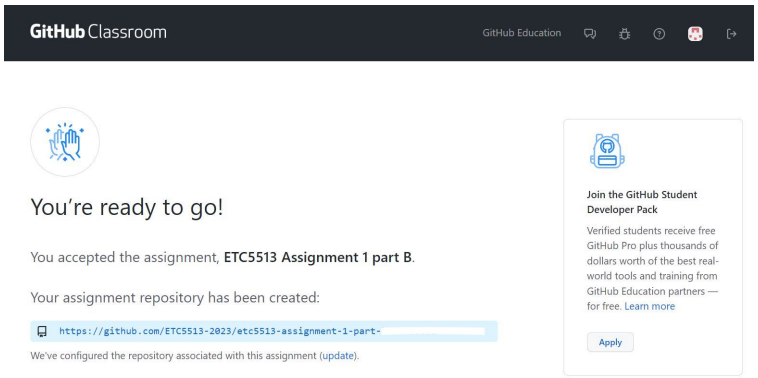
I have created a GitHub repository for you. The assignment is inside GitHub classroom in the link below (**please log in into your GitHub account before accessing the link below**):

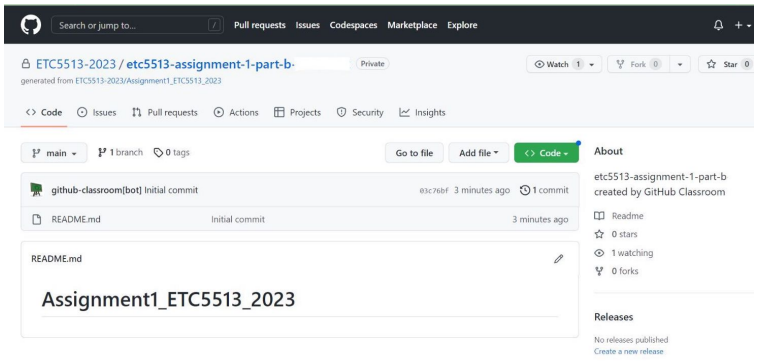
- Once you access the link below, please follow the instructions (as in the pictures below) so that you can access the GitHub repo that contains the assignment repo. Remember that you first need to log in into your GitHub account!

Assignment GitHub classroom link: https://classroom.github.com/a/KQoidX0m









Once you have completed the steps above the GitHub repository will appear as a repo in your GitHub account and you will be working with it (clone, commit, push and pull) as we have done with Github repositories during the lectures and tutorials (1pt).

- The GitHub repo will appear in your Github classroom space and it is essential that you work on that repository. That repository will be the one that you will clone into your computer. **Do not work or create a different repository!**
- Once you have clone that repository in your local computer please add all the files for the reproducible report that you created in the assignment part A and push all those changes into the remote GitHub repository with a relevant commit clear message (3pts).

- Working in the same Git repo in your local machine, now please add a final section called "Conclusions" in your report and add a max of 4 sentences to discuss the findings/conclusions of your research report (3pts) and push those changes into the same GitHub repository with a clear message (2pts).

Marking rubric

In addition to the points for each section as described above, you will be graded on your:

- Report template (5pts). For example the default Rmd template will be marked with 1pt, if you use more sophisticated templates that will add points into this component of the assignment (to a max of 5pts)
- Issues with spelling and grammar (up to -5pts)
- R code style (5pts)
- Report quality: Sections in the report are connected and aligned with the research question in a coherent way. (10pts)

The usage of AI (chatGPT for example)

- You may use chatGPT to **correct your English** or to help you with your R code (you can use chatGPT to query **your code*** and for example to find code bugs or to ask for hints to enhance **your existing code**). However, if you use chatGPT you must declare it by adding a section in your Rmd report called **Appendix** where you display screen shots of your chatGPT queries and all the interactions that you have had with chatGPT related to this assignment.
- You **cannot** use chatGPT to generate content for the assignment from scratch or to create code for you from scratch. You must always provide chatGPT with an input and you must screenshot all the those interactions, and report that usage in the appendix of your reproducible report.
- Monash University supports the responsible and ethical use of generative AI. For more info please refer to Monash Policy and practise guidance around acceptable and responsible use of AI technologies.

Remember you can be better than chatGPT. If you just use chatGPT to create content for you, where is your value and why are your skills special?!

Assignment submission: Moodle

The report must be rendered into html and all the code and code outputs must be visible.

Items to be submitted in Moodle:

1. Please, print the html version of the report into a pdf file using a browser and upload this file into Moodle. (Open your html in a browser for example Google Chrome, and print the file into a pdf file.)
2. Paste the http of your GitHub repo in the text box.
3. A zip file of your entire repository.

VERY IMPORTANT: The printed pdf report cannot have more than 5 pages (that does not include the Appendix pages)! You cannot use a data set that has been used for other assignment or in another units.

Submission deadline

Thursday March 30, 11:55pm (Melbourne time). **Late submissions will not be accepted.**

Plagiarism

Monash University is committed to honesty and academic integrity. There are serious consequences for plagiarism and collusion. If plagiarism and/or collusion is detected further actions will be taken according to Monash University policy and procedures. More info here:

https://www.monash.edu/students/admin/policies/academic-integrity

- You cannot re-use assignments that have been submitted or used in other units!