Iris dataset:

To do list:

1) Become very familiar with what exactly this dataset includes so that you can answer questions about data collection, the type of data, and the possible analytic choices.

2) Select one of the kernels that include visualizations, and you are interested in reproducing/learning.

3) Copy the kernel in your python code. This *does not mean that you can simply copy and past.*  It will require you to understand the code of someone else’s and EXPLAIN everything that is included in the code in your own words. So, you will include a write-up for every line of code in your GitHub submission (not necessary to include in the Kaggle submission).

4) Once the code is understood and clear, you can start tweaking something – this part is something you will do with whatever you are comfortable doing (i.e., color, axes, 3D, the shape, horizontal/vertical, titles, legends, etc.) For the first part of the project, I will not ask you to do more than 2 (two) of these tweakings, but please feel free to do more.

5) Finally, all the work will need to post on Github (including a detailed readme file an explanation for every line of code), and Kaggle. Please make sure that your Github readme files include a table, italics, headings, bold, or whatever you want – but it needs to use Github markdowns and looks visually presentable.

*This first time, I am not assigning points to each of these headings because I want you to learn the process very well, but regarding importance here is the ranking I have in mind:*

*Section 3 – explaining everything*

*Section 5 – successfully posting it on Github and Kaggle and including a detailed readme*

*Section 1 – familiarity with the dataset*

*Section 4 – tweaking something*

*Section 2 – selecting a kernel*