# Web Exercise 1: Introduction of GitHub and Online Data Science Resources

1. What are the key functions of GitHub? Who are the users? ( Please use your own words to describe them in 200 words – 300 words).

Github expands the use of the version control system Git. Its functions are community-based. The platform allows you to share your repositories (repos) between your local (personal computer) and your remote (GitHub account) machines. This requires a set of “push” and “pull” commands to retrieve or send commit repos.

Github is an important tool for collaboration work. If a repo is public, someone from the internet can pull the master branch (or other branches) into their own computer and edit the code. The branch can be pushed back into Github as a pull request. If the owner of that branch approves the changes, they can merge the pull request. Here on out, the approved code changes are now applied to the branch. In some cases, merging pull requests may not be so easy. There may be merge conflicts where the owner of the branch needs to edit the code properly and commit the changes. Git helps with that process in separating the merge conflict code from the branch and code from the pull request.

Regardless, Github is an important tool for agile project management. It helps teams to finish projects in a timely manner and guarantees high-quality products from the increased availability of help. Thus, the users of Github tend to be teams working on a variety of projects. Many companies use Github for software development. Also, Github users can be individuals with their own personal projects. Someone they don’t know can essentially help edit their repo. But as long as you are willing to collaborate and have some background with coding, Github can be a sandbox of possibilities.

1. What are the differences between “clone” and “pull request” in GitHub?

Git clone creates a copy of the repo from an existing repo. It also tracks the remote branches they were copied from. For example, after a Github repo is cloned to your local computer, git status can let you know if there’s discrepancies between your repo and the GitHub repo or if everything is up to date. A pull request is when someone pushes a branch into the Github repo. Usually the person modified the branch and wants to see if their changes can be approved to merge into the main Github branch. So essentially cloning is copying a repo and pull request merges the original repo with another similar, modified repo.

1. Introduce the selected “Data Set” from the [awesome-datascience] GitHub (with the data source URL) and describe its potential applications and values.

URL: <https://catalog.data.gov/dataset>

Data.gov was launched in 2009 by U.S. General Services Administration, Technology Transformation Service (<https://www.data.gov/about>). Essentially, research and work done under the U.S. government that can be made public are accessed in Data.gov. Data can be filtered from the following categories:

* Location
* Topics
* Topic Category
* Dataset Type (Geospatial or Non-Geospatial)
* Tags
* Formats (CSV, HTML, etc.)
* Organization Types (Federal, State, City, etc.)
* Organizations
* Publishers
* Bureaus

If you don’t know what data you want, the filters are helpful in finding a data set you may like. Potential applications can range from data synthesizing results from different organizations, publishers, etc. to data analysis of a particular subject in one data set. Another very useful application is finding relevant geospatial data that is available to practice with.

1. Introduce ONE selected “free-data-science-book” with URL and the author/institute information. Explain briefly about why you like this book.

URL: <http://cran.r-project.org/doc/manuals/R-intro.pdf>

Author/Institution: W. N. Venables, D. M. Smith, and the R Core Team

This book caught my interest because I have no experience with the R language and would like to learn more about it. Since I have some experience with Python, I want to see the difference between the two languages. The book title caught my eye because it is self explanatory that the book will provide a good introduction. The “Beginner” tag was also helpful in making sure I didn’t end up reading a book above my experience level.

1. Print out the README.md (in a regular display mode) as a PDF file for attachment #1.
2. Print out your GitHub Website from a Web Browser (Chrome or others) as a PDF file for attachment#2.

URL: <https://ohkaaaaay.github.io/Fabio.Github.io/>