Tools for Data Science

▼ Week 5

Introduction to R and RStudio

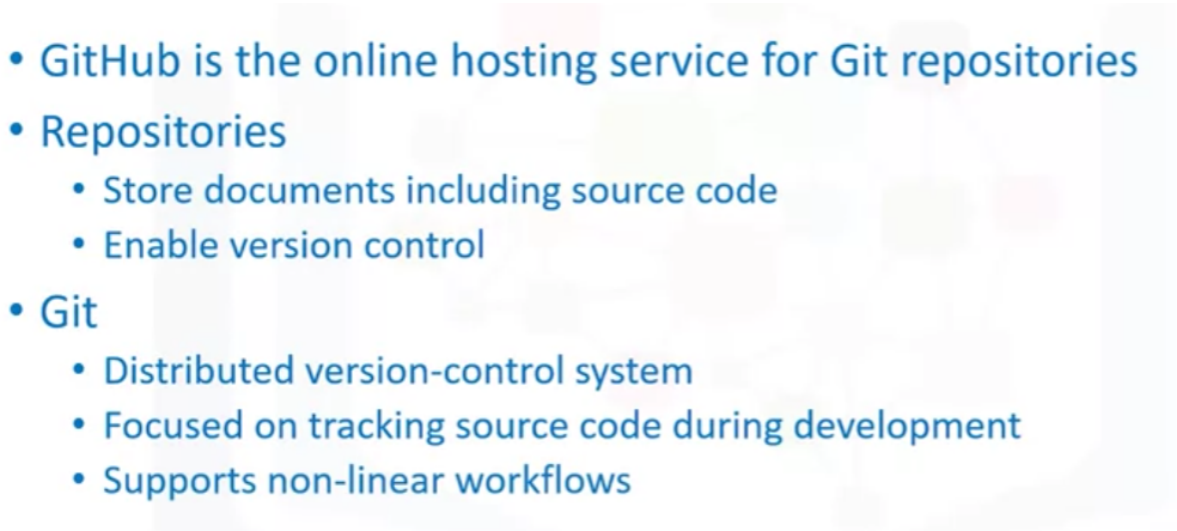
Download & Install R and RStudio: **downloadandinstallrandrstudio.pdf**

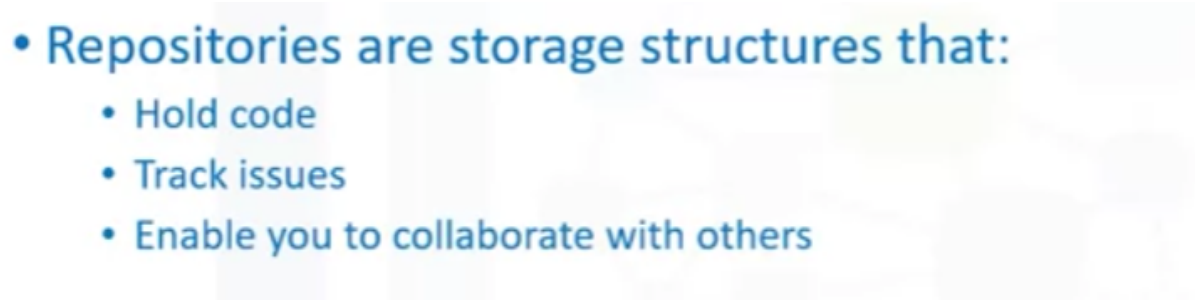
Getting started with RStudio and Installing packages: **GettingStartedwithR.pdf**

Plotting in R:

Introduction to Git & Github:

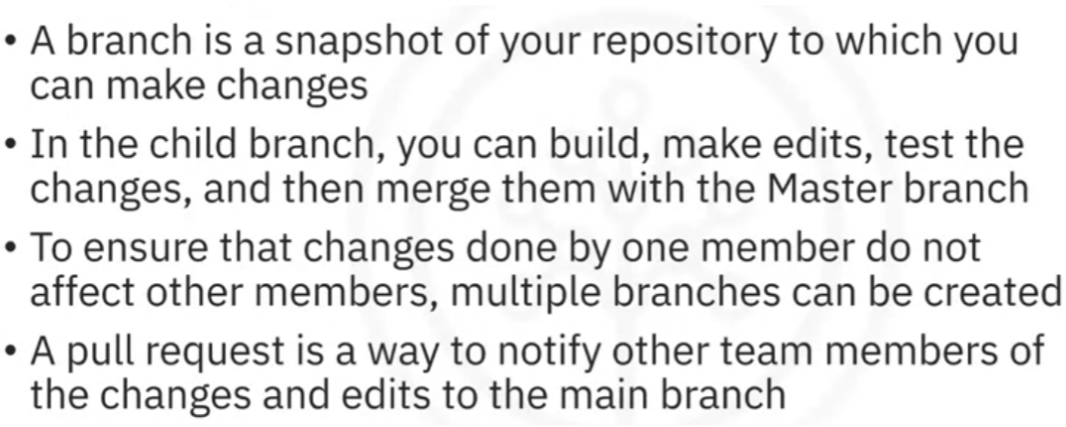
Tutorial: https://youtube.com/playlist?list=PLEPye7A7EcQZrT3VSBb7jtxnxIfY3yyG6

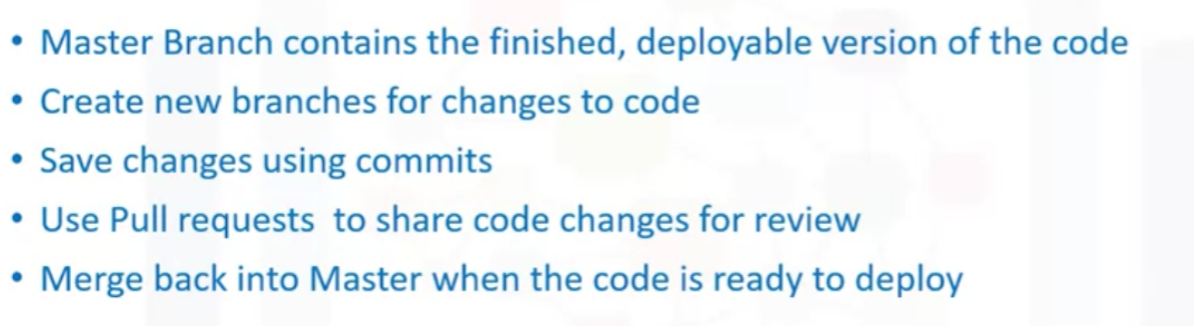




Getting Started with Github: **GettingStartedwithGitHub.pdf**

Git Branches:





Summary:

- The capabilities of R and its uses in Data Science.

- The RStudio interface for running R codes.
- Popular R packages for Data Science.
- Popular data visualization packages in R.
- Plotting with the inbuilt R plot function.
- Plotting with ggplot.
- Adding titles and changing the axis names using the ggtitle and lab’s function.
- A Distributed Version Control System (DVCS) keeps track of changes to code, regardless of where it is stored.
- Version control allows multiple users to work on the same codebase or repository, mirroring the codebase on their own computers if needed, while the distributed version control software helps manage synchronization amongst the various codebase mirrors.
- Repositories are storage structures that:
    - Store the code
    - Track issues and changes
    - Enable you to collaborate with others
- Git is one of the most popular distributed version control systems.
- GitHub, GitLab and Bitbucket are examples of hosted version control systems.
- Branches are used to isolate changes to code. When the changes are complete, they can be merged back into the main branch.
- Repositories can be cloned to make it possible to work locally, then sync changes back to the original.

Glossary: **Glossary.pdf**