Introduction to R and RStudio

R is statistica programming language used for data processing and manipulation, statistical inference, data analysis, and ML.

R is used most by academics, healthcare and the government.

R supports importing data from different sources. Flat files, databases, web, statistical software such as SPSS or stata

R capabilities

Easy to use

Great for Visualization

Basic data analysis doesnt need installing packages

What is RStudio?

popular IDE for R.

Increases productivity in running R

includes syntax highlighting editor.

Console

Workspace and history tab

Panel that contains files, plots, packages , Help tabs

Tabs

Files tab shows the files we have

plots tab shows the plots generated and has the abiluty to export them.

packages tab shows the external packages that are available to install.

Help tab has links that R resources, support and packages

Popular R Libraries:

dplyr-- Data manipulation

stringr-- string manipulation

ggplot-- Data Visualization

caret-- ML

Reading:Download and install Rstudio

<https://author-ide.skills.network/render?token=eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJtZF9pbnN0cnVjdGlvbnNfdXJsIjoiaHR0cHM6Ly9jZi1jb3Vyc2VzLWRhdGEuczMudXMuY2xvdWQtb2JqZWN0LXN0b3JhZ2UuYXBwZG9tYWluLmNsb3VkL0lCTURldmVsb3BlclNraWxsc05ldHdvcmstRFMwMTA1RU4tU2tpbGxzTmV0d29yay9sYWJzL01vZHVsZTIvSW5zdGFsbGF0aW9uJTIwbWFya2Rvd25zL1IlMjBhbmQlMjBSc3R1ZGlvJTIwSW5zdGFsbGF0aW9uL1JfUlN0dWRpb19JbnN0YWxsYXRpb24ubWQiLCJ0b29sX3R5cGUiOiJpbnN0cnVjdGlvbmFsLWxhYiIsImFkbWluIjpmYWxzZSwiaWF0IjoxNjc3NzYyMjA1fQ.oBbRhZtDRuZDvQ7-zvZcDkuXPyzXXK2Zwf6m839ezxc>

RStudio basics:

<https://author-ide.skills.network/render?token=eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJtZF9pbnN0cnVjdGlvbnNfdXJsIjoiaHR0cHM6Ly9jZi1jb3Vyc2VzLWRhdGEuczMudXMuY2xvdWQtb2JqZWN0LXN0b3JhZ2UuYXBwZG9tYWluLmNsb3VkL0lCTURldmVsb3BlclNraWxsc05ldHdvcmstRFMwMTA1RU4tU2tpbGxzTmV0d29yay9sYWJzL0xhYnNfVjQvUl9iYXNpY3NfdXNpbmdfUlN0dWRpby5tZCIsInRvb2xfdHlwZSI6InRoZWlhIiwiYWRtaW4iOmZhbHNlLCJpYXQiOjE2NzE1NzAxODF9.CNdjbbSdLmRDF0qcGejWb6PfilQ247WrKPaQo5_SoFw>

Week5\_Program1.R

Week5\_Program2.R

Subtract.R

Plotting in RStudio

Command to install packages:

install.packages <package name>

Packages for plotting are:

ggplot : histo, bar, scatter

plotly : Web based data vis . Can be saved as html

Lattice: Complex, milti-variable data sets . high level

Leaflet: Interactibe plots

Example Program

# Define the cars vector with 5 values

cars <- c(1,4,6,5,10)

# Graph the cars vector with all defaults

plot(cars, type= ‘o’)

# Create a title

title(main=”cars vs Index”)

Using ggplot

adds layers of functions and arguments

library(ggplot)

# mtcars is the dataframe

#x-axis - mps - miles per gallon

# y-axis - weight

# geom\_point() - scatter plot. Return an empty plot if not used.

ggplot(mtcars,aes(x=mpg, y= wt))+geom\_point()

# adding titles and axis names

ggplot(mtcars,aes(x=mpg, y= wt))+geom\_point()+ggtitle(“Miles per gallon vs weight”)+labs(y=”weight”, x=”Miles per gallon”)

ggalley- built on top of ggplot.

Installing Packages

<https://author-ide.skills.network/render?token=eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJtZF9pbnN0cnVjdGlvbnNfdXJsIjoiaHR0cHM6Ly9jZi1jb3Vyc2VzLWRhdGEuczMudXMuY2xvdWQtb2JqZWN0LXN0b3JhZ2UuYXBwZG9tYWluLmNsb3VkL0lCTURldmVsb3BlclNraWxsc05ldHdvcmstRFMwMTA1RU4tU2tpbGxzTmV0d29yay9sYWJzL01vZHVsZTIvRFMwMTA1RU4tMi1MYWItUlN0dWRpbyUyMCVFMiU4MCU5MyUyMFRoZSUyMEJhc2ljcy5tZCIsInRvb2xfdHlwZSI6InRoZWlhIiwiYWRtaW4iOmZhbHNlLCJpYXQiOjE2NzIzMDYxMjh9.lz-CLk299yFHmOsNn1z7v_5LP1D7LTFqVBQZ2y0tLbg>

Plotting with ggplot

<https://author-ide.skills.network/render?token=eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJtZF9pbnN0cnVjdGlvbnNfdXJsIjoiaHR0cHM6Ly9jZi1jb3Vyc2VzLWRhdGEuczMudXMuY2xvdWQtb2JqZWN0LXN0b3JhZ2UuYXBwZG9tYWluLmNsb3VkL0lCTURldmVsb3BlclNraWxsc05ldHdvcmstRFMwMTA1RU4tU2tpbGxzTmV0d29yay9sYWJzL01vZHVsZTIvcGxvdHRpbmdfd2l0aF9SLm1kIiwidG9vbF90eXBlIjoidGhlaWEiLCJhZG1pbiI6ZmFsc2UsImlhdCI6MTY4MzgxNjM3Mn0.2BgXGZYX0gsCtTCjuOufr1-t_nkaBoL2bfic53Zu6pE>

Program:

Week5\_plotting with ggplot

Plotting with GGally

<https://author-ide.skills.network/render?token=eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJtZF9pbnN0cnVjdGlvbnNfdXJsIjoiaHR0cHM6Ly9jZi1jb3Vyc2VzLWRhdGEuczMudXMuY2xvdWQtb2JqZWN0LXN0b3JhZ2UuYXBwZG9tYWluLmNsb3VkL0lCTURldmVsb3BlclNraWxsc05ldHdvcmstRFMwMTA1RU4tU2tpbGxzTmV0d29yay9sYWJzL01vZHVsZTIvUGxvdHRpbmclMjB3aXRoJTIwZ2dwbG90Lm1kIiwidG9vbF90eXBlIjoidGhlaWEiLCJhZG1pbiI6ZmFsc2UsImlhdCI6MTY3MjQyNDA1NH0.1_csippFJBrd38raMrUuz7b9dYvv3BocWRMWacJKHYg>

Program:

Week5\_plotting with GGally

Overview of Git/Github

Git and Github are popular among developer for versioning the source code and collaboration.

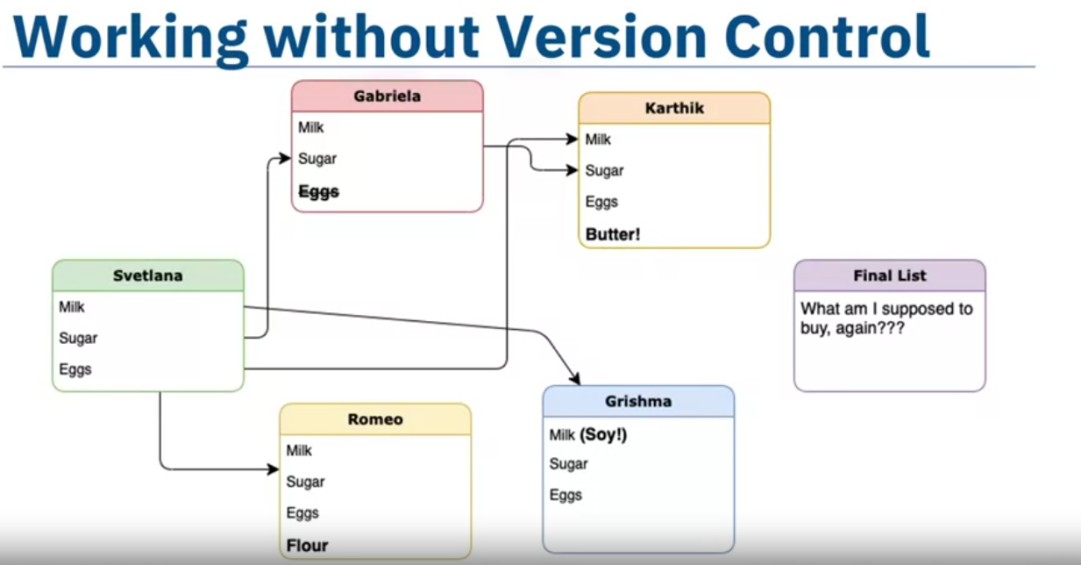
Version Control

Allows to keep track of changes done in the code.

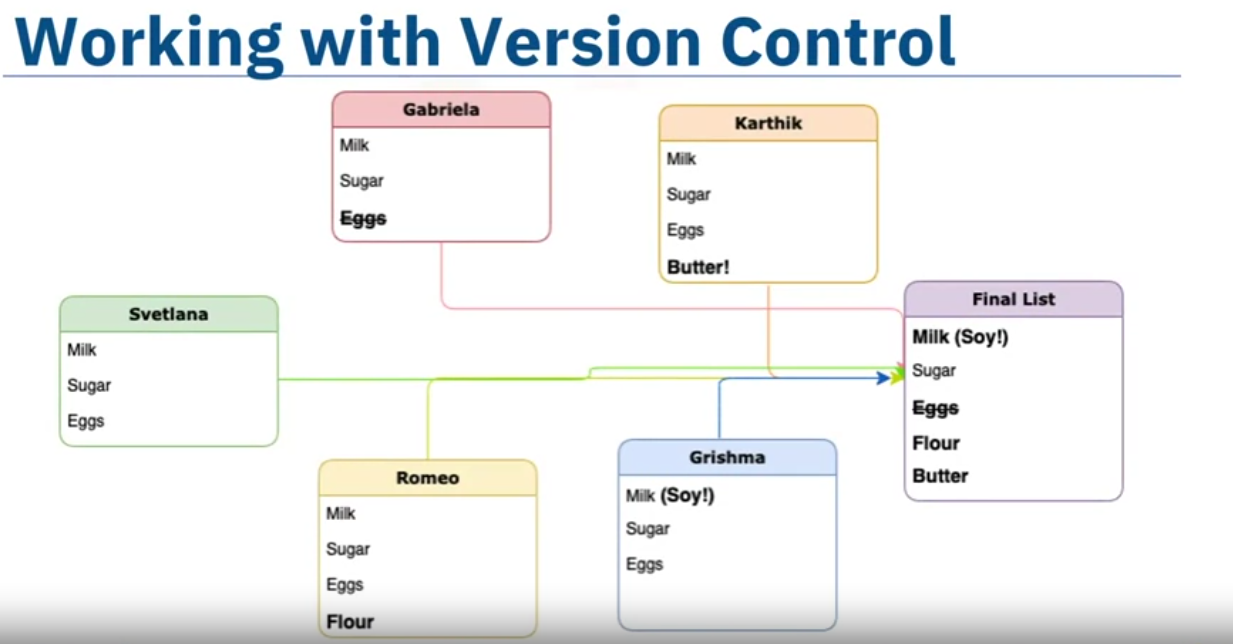
Easy to get back older version if any mistake is made.

Working without version control

Example: Friends providing list of items to shop



With Version Control:



Lot more cleaner than without version control.

Git:

Free and open source version controlling software

Distributed version control

Accessible anywhere in the world.

most commonly used because of the distributed nature.

Can also version control images, documents etc.,

Other version control softwares exist

Can be use with command line interface

but Git +Github gives lot more features.

Github alternatives

gitlab, bitbucket, beanstalk

Term to know:

ssh :secure shell : A method for secure remote login from one computer to other.

Repository: A project folder with version control

Fork: repo copy

pull req: A process used to request that someone revies and approves your changes before they become final.

Working DirL A directory on your file systerm , including its files, sib dirs that are associated with a git repo

basic Git command

init - setup git repo locally

add- adds all changes that is made in the repo into the staging area.

status - displays the status of the git repo and stage snapshot of our changes

commit - commits stage snapshot and commits the changes,

reset - resets a ny changes,

log - browse previous changes to the process by displaying all the commits,

branch - creates a new isolated env called a branch so a new feature can be work on witout modifying the working code.

checkout - switch to a specific branch,

merge : merges the branches,

Introduction to Github

In 2003 Linux was development was managed by bitkeeper which was free to use system.

In 2005 Bitkeeper changed to for free system problematic

Linus Tolrvalds lead a team to develop a replacement source control.

Character

Strong support for non-linear dev- linux patches where then arriving at 6.7 per sec.

Distbuted development - Each dev has a local copy of full dev history.

Compatibility with existing systems and prtotocols

Efficient handling of large projects

Cryptographic authentication of history,

Pluggable merge strategies.

Git Repository Model:

Distributed version-control system.

Tracks source code

Coordinates among programmers

Tracks changes

supports non-linear workflows

Created in 2005 by Linus Torvalds

What is git?

Distributed version-control system.

Tracks changes

Provides central point for collaboration.

Allows for centralized administration

Teams have controlled access scope

The main branch should always correspond to deployable code

IBM Cloud is built around open-source tools including Git repositories.

What is GitHub

Online Hosting service for Git repositories

by subsidiary of Microsoft

Offers free, professional and enterprise accounts

As of August 2018, GitHub had over 100M repos

What is a Repo:

A data structure forstoring documents including app source-code

A repo can track and maintain version-control.

What is GitLab?

A Devops platform, delivered as a singlr app

Provides access to Git Repos

Provides source code management

GitLab enables developers to:

Collaborate

Work from a local copy

Branch and merge code

Streamline testing and delivery with CI/CD