R Programming
Lab 01

DATA SCIENTIST'S TOOLBOX OVERVIEW, BASIC R

Lecturer：Professor Pao-Ann Hsiung

Teaching Assistant: Hendrik

Embedded Systems Laboratory
National Chung Cheng University
Chiayi, Taiwan-62102

Learning Objective

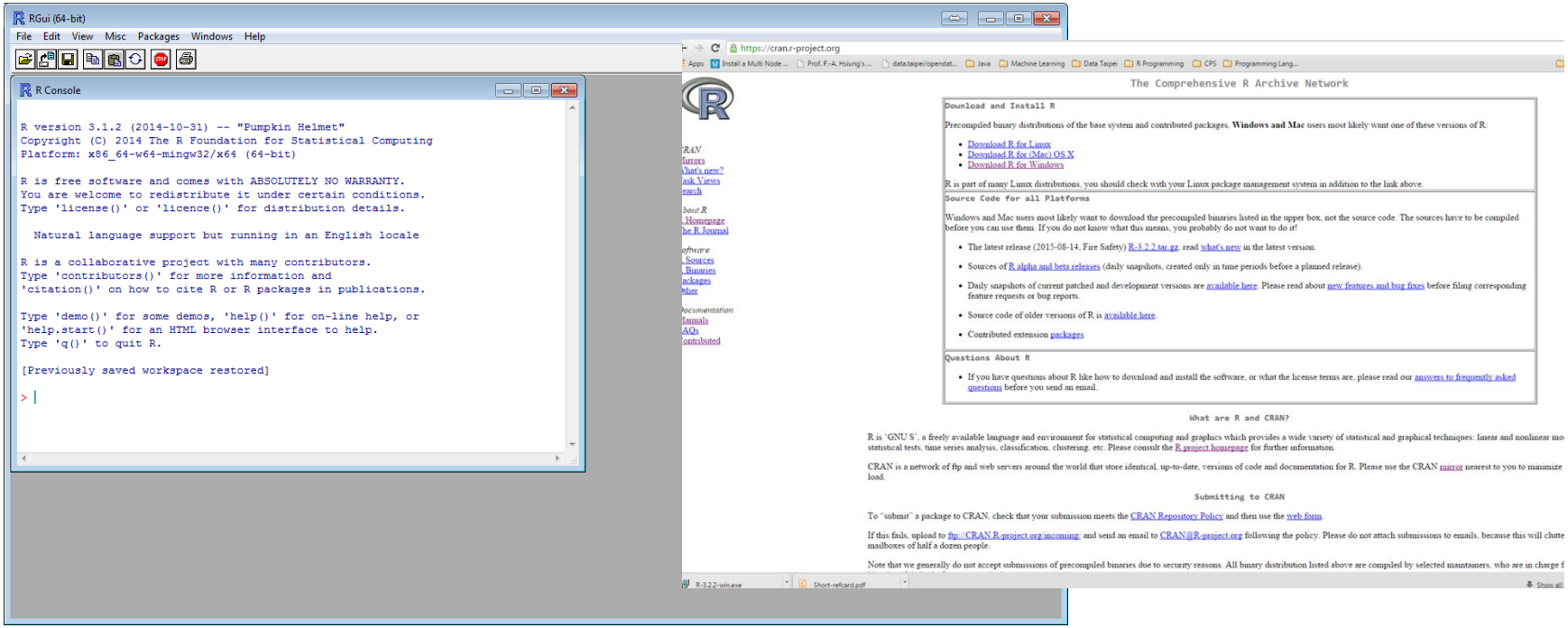
- Install the R and RStudio software packages
- Git and GitHub
- Download and install the swirl package for R
- Execute basic arithmetic operations
- Get to know the sense of R

Outline

- Data Scientist's Toolbox
  - Introduction to basic tools
  - R&Rstudio
  - R tools
  - Git&Github
  - Getting Help
- Basic R
  - Data input
  - First program
  - R packages
- Demo & Assignment

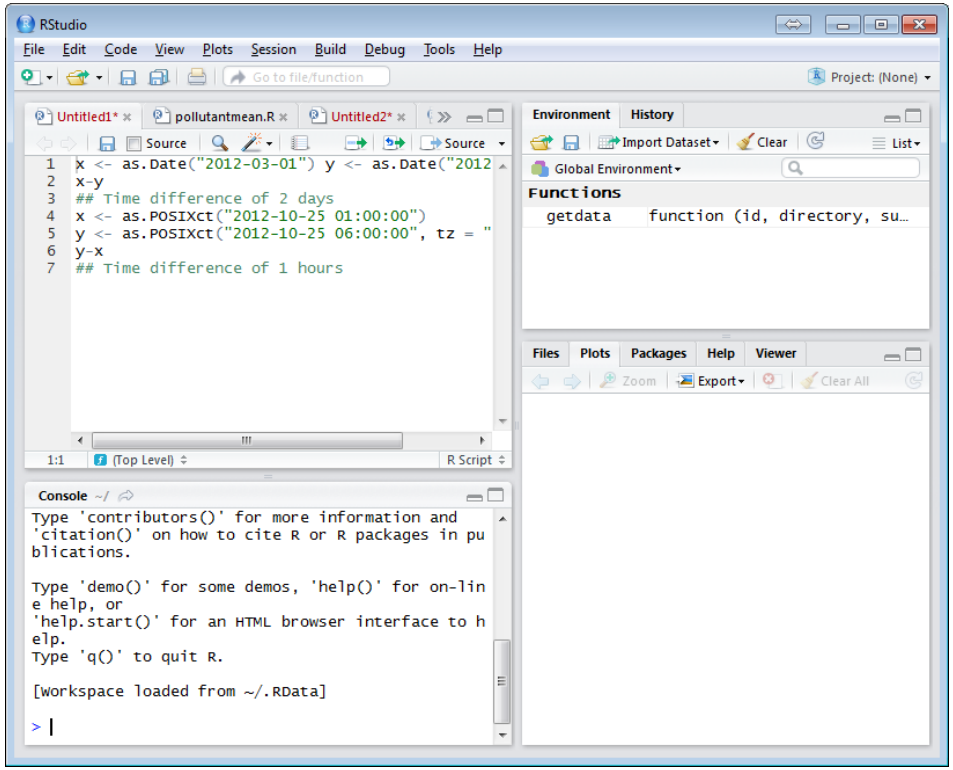
R & RStudio

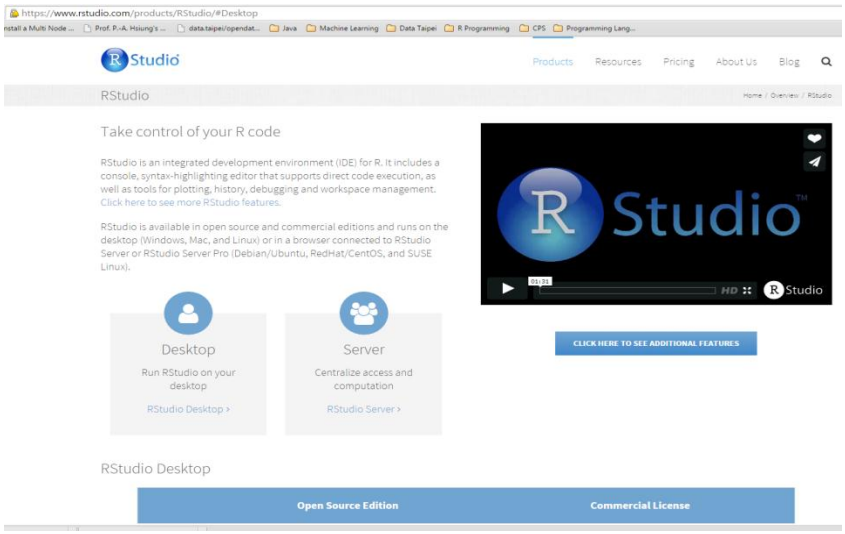
- R distribution available on https://cran.r-project.org/
-



R & Rstudio

- **Rstudio**: IDE for R includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management
- Available on https://www.rstudio.com/products/rstudio/





R Tools

- A collection of tools for building R packages in windows
- Available at https://cran.r-project.org/bin/windows/Rtools/
- For detail refer to that website

* Not necessary for linux or mac

Packages

- See available packages on CRAN

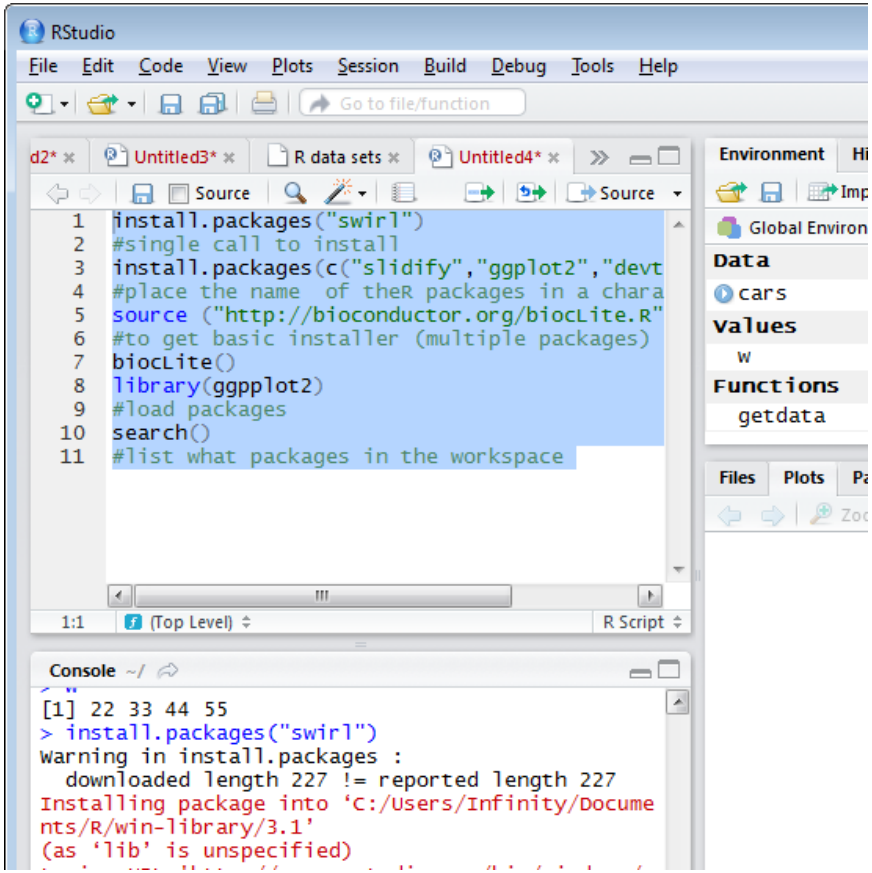
  available.packages()

- Install

  install.packages("namesPackages")

- Load

  library("namesPackages")



Git & GitHub

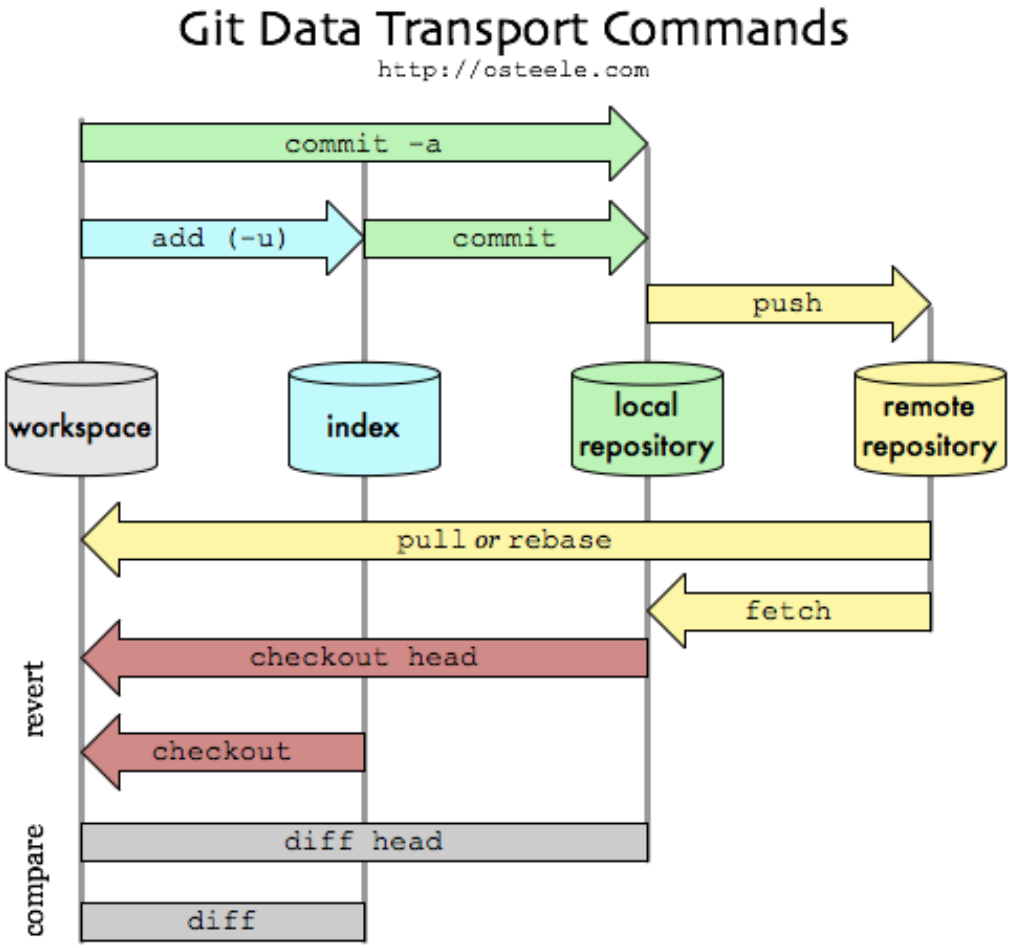
Vs

- You don't need GitHub to use Git
- Git = Local (On your computer); Github=Remote (on the Web)
- GitHub Allow you to:
  1. Share your Repositories with others
  2. Access other user's repositories
  3. Store remote copies of your repositories (on GitHub's Server) in case something happens to your local copies (on your computer)

Go to the following website to download https://git-scm.com/downloads

Git Basic

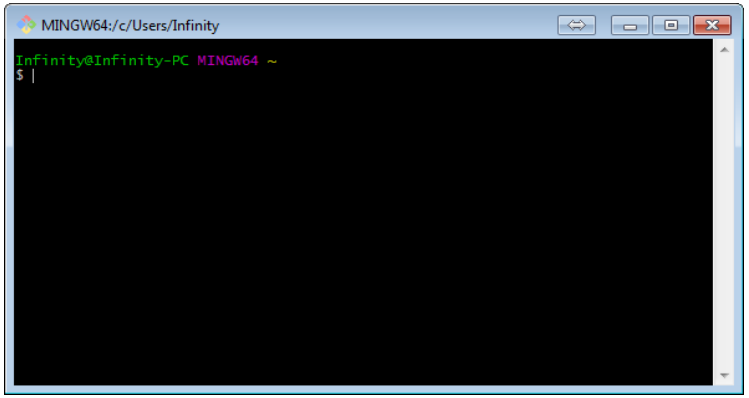
- **init**
- **clone**
- **add**
- **status**
- **commit**
- **log**
- **diff**
- **checkout**

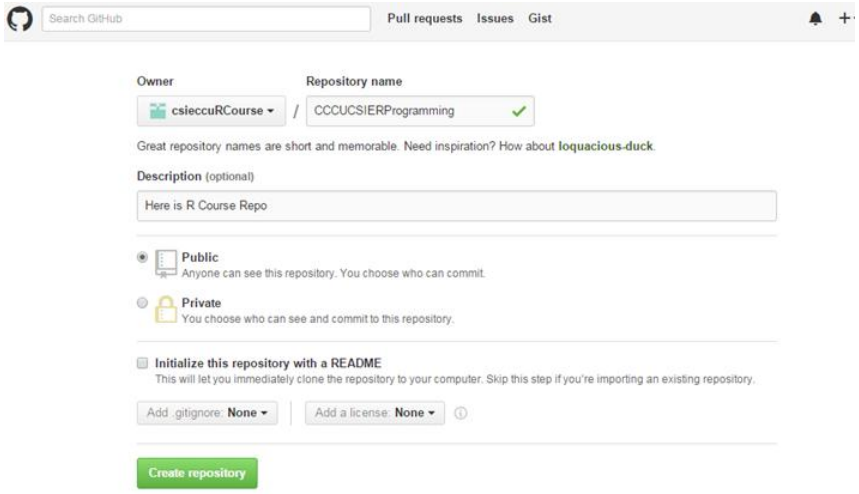


Getting started: Git and Github

- Create a your own github and make a new repository named it *Lab1_studentID*

- Open your git bash, download git here





Creating GitHub Repo

- Two Methods
  1. Start Repository from scratch
  2. Fork another user's repository
- Repo From Scratch
  - Go to your profile page(https://github.com/userNameHere) and click on "Create a new repo" in the righthand corner of the page
  - Or go directly to https://github.com/new
  - Select public(Private repos require a paid or educational account)
  - Check the box next to inialize this repository with a README

Creating a Local Copy

- Now you need to create a copy of this repo on your computer
- Open git bash
- Create a directory on your computer where you will store your copy of the repo

```
$ mkdir ~/test-repo
$ cd ~/test-repo
```

Initialize a local git repo

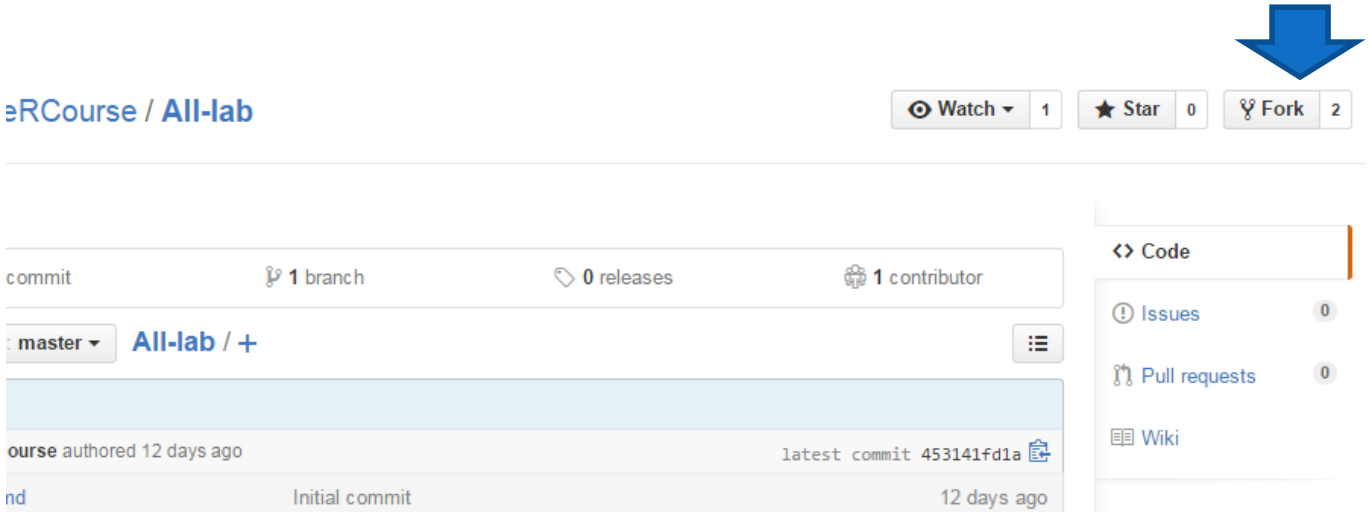
```
$ git init
```

Initialize Point your local repository at the remote repository you just created on the GitHub server

```
$ git remote add origin git@github.com:yourUserNamel/test-repo.git
```

Fork a Another User's Repository

- The second method of creating a repository is to make a copy of someone else's
- Begin by navigating to the desired repository on GitHub website and click the "Fork" button



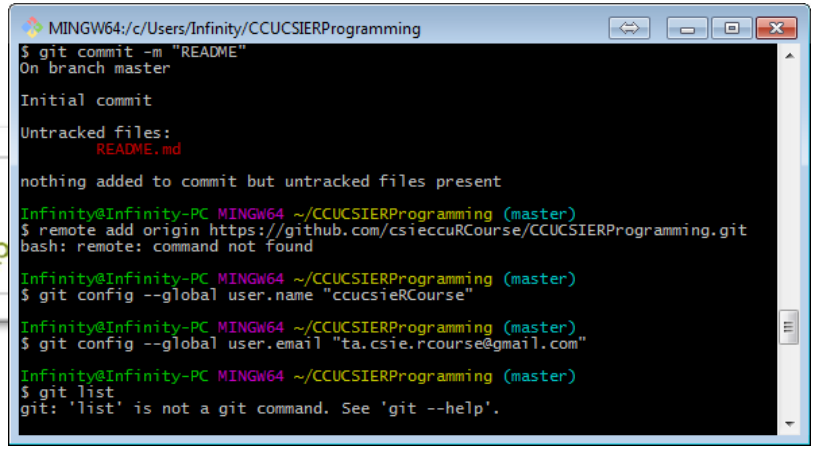
https://help.github.com/articles/fork-a-repo/

Configure username and email

- Each commit to a Git repository will be tagged with the username of the person who made the commit

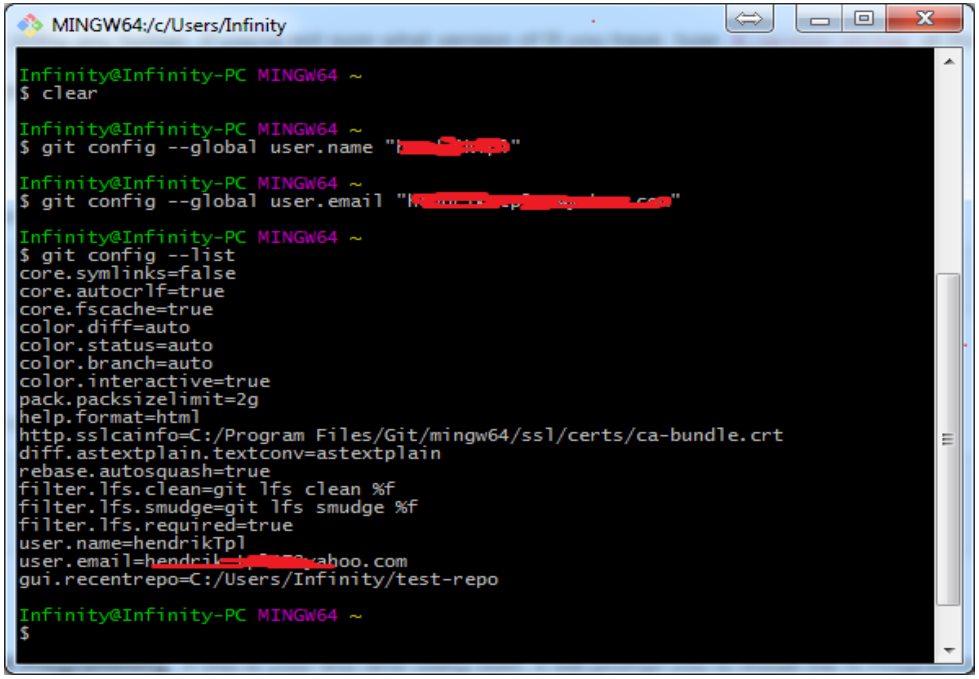
- To set your username and email, Enter the following commands in Git Bash

```
$ git config --global user.name "Your Name Here"
$ git config --global user.email "your_email@examp
```



Cont.

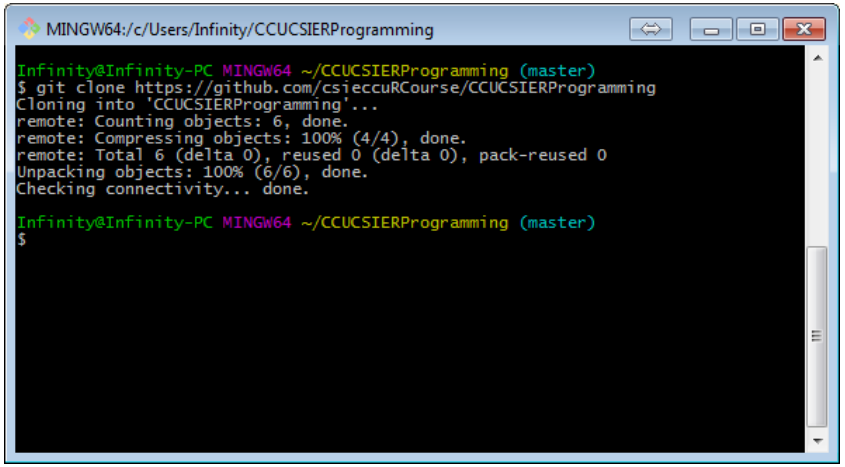
- To confirm your changes, type the following command
  - $ git config –list



Clone

- To make a local repository on your computer
- to get a copy of an existing Git repository on your GitHub Account

git clone https://github.com/yourUserName/repoName

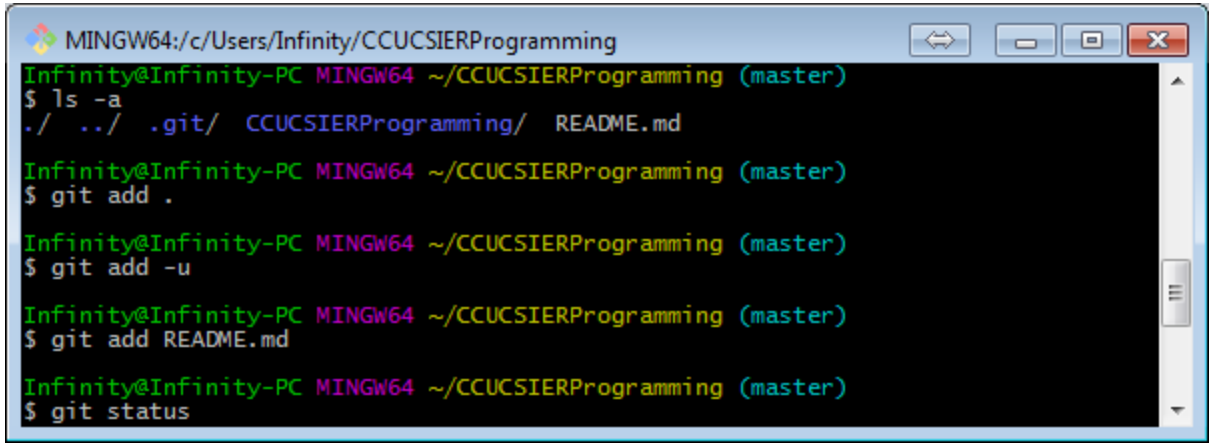


This will clone repo into your current directory

Add

- Suppose you add new files to a local repository under version control
- You need to let git know that they need to be tracked
- Should do this before commiting

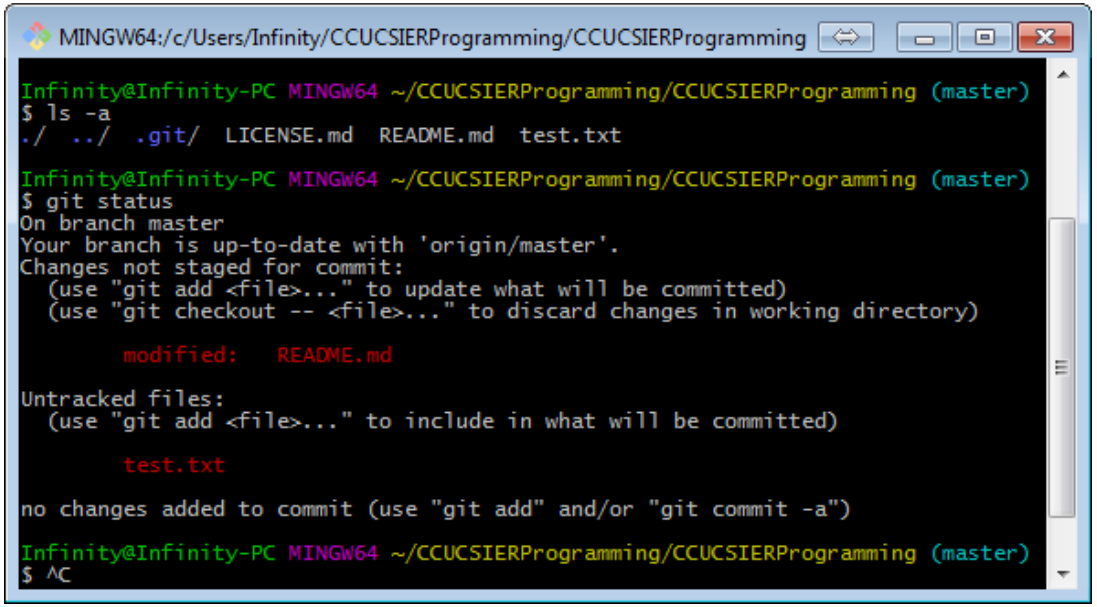
git add . adds all news file
git add –u updates tracking for files that changed names or were deleted
git add –A both



Status

- As you saw in the git add section, in order to see what the status of your staging area is compared to the code in your working directory

```
git status
git status -s
```

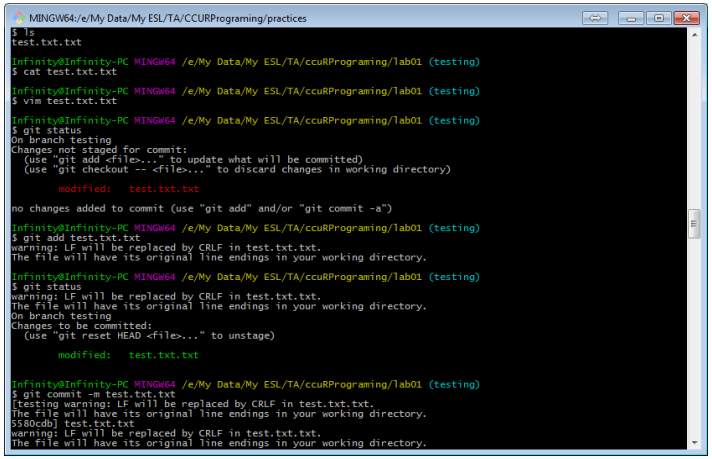


Commit

- Commit the files that you've staged in your local repository

```
$ git commit -m 'First commit'
# Commits the tracked changes and prepares them to be pushed to
a remote repository. To remove this commit and modify the file,
use 'git reset --soft HEAD~1' and commit and add the file again.
```

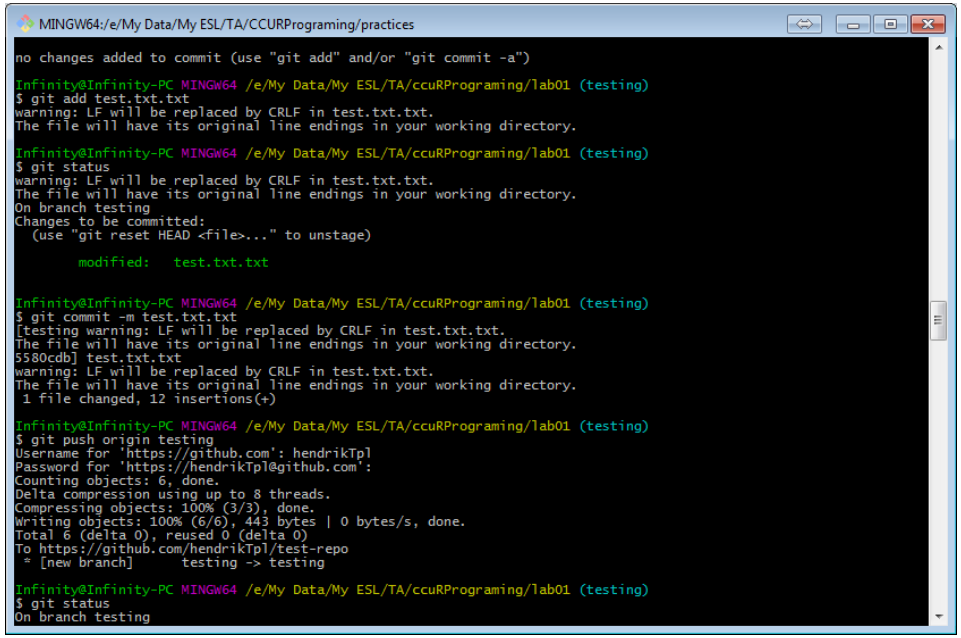
- This only update your local repo , not the remote on github



Push

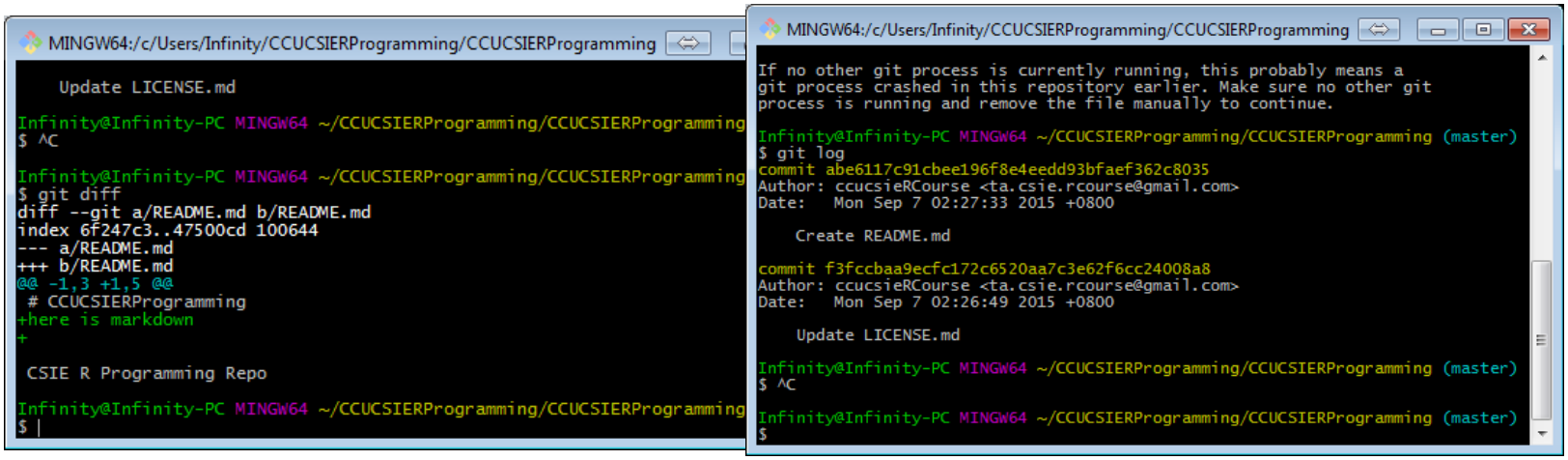
- Update local commits on the remote (Github)

```
$ git push origin master
# Pushes the changes in your local repository up to the remote repository you specified as the origin
```



log & diff

- git log to find specific commits in your project history - by author, date,
- git diff to compare two different points in your history - generally to see how two branches differ or what has changed from one version of your software to another content or history.



Branches

- Sometimes you are working on a project with a version being used by many people. You may not want to edit that version
- So you can create a branch with the command

  git checkout –b branchname
- To see what branch you are on type

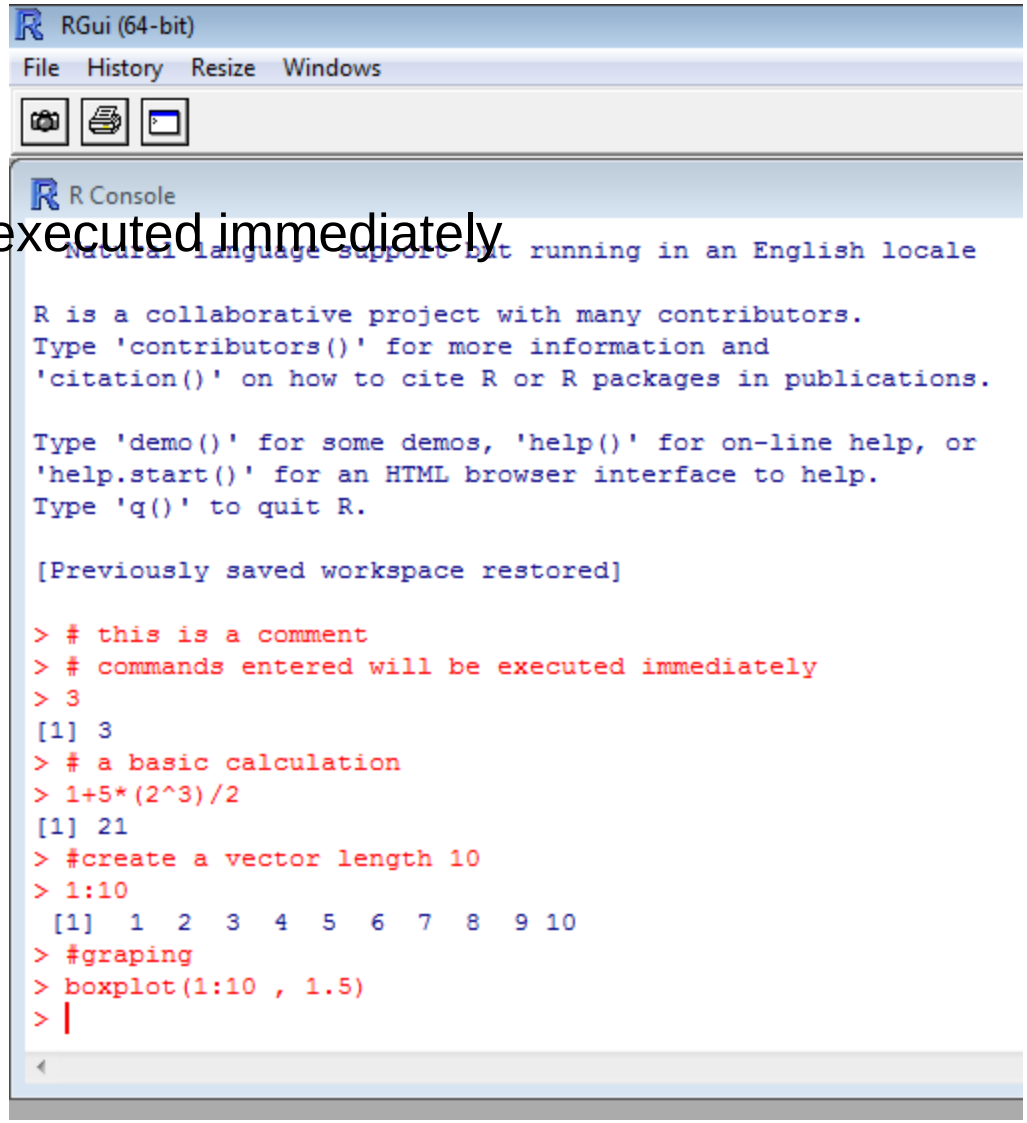
  git branch
- To switch back to master branch type

  git checkout master

More about git, please kindly go to https://git-scm.com/doc or http://gitref.org/

Getting Started

- Open R console
  - Commands entered will be executed immediately
- Demo some simple tasks
- Calculation
  - E.g. 1+5*(2^3)/2
- Statistical analysis
  - E.g. linear regression
- Graphing
  - E.g. boxplot



Entering Input & Evaluation

- <- symbol is the assignment operator
- # character indicates a comment

VarNames <- c("variables")

```
> x <- 1
> print(x)
   [1] 1
> x
  [1] 1
> msg <- "hello"
> x <- ## Incomplete expression
```

When a complete expression is entered at the prompt, it is evaluated and the result of the evaluated expression is returned

```
> x <- 5 ## nothing printed
> x ## auto-printing occurs
    [1] 5
> print(x) ## explicit printing
    [1] 5
```

The [1] indicates that x is a vector and 5 is the first element

First R

- X <-1:10 #create x with 10 numbers
- Print(x) #print numbers in x
- Y <-2*x-1 #create y based on x (also 10 num)
- Print(y)
- Mean(y) #calculate average of numbers in y
- Model <- lm(y~x) #fit a linear regression model
  # y=B0+(B1*X)

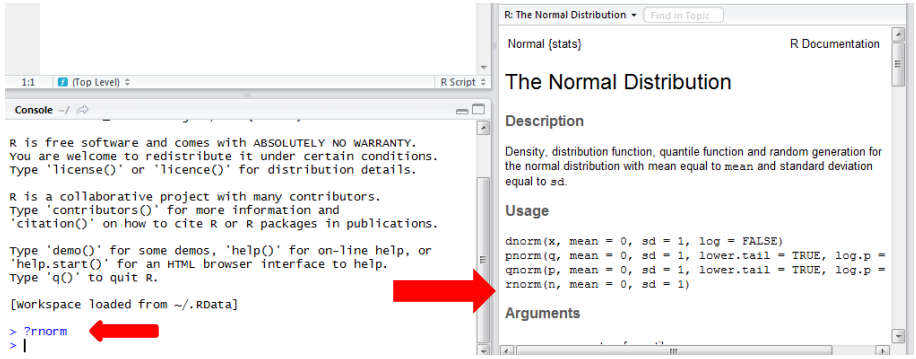
Print(Model) # print the linear model
#B0 = -1 , B1=2

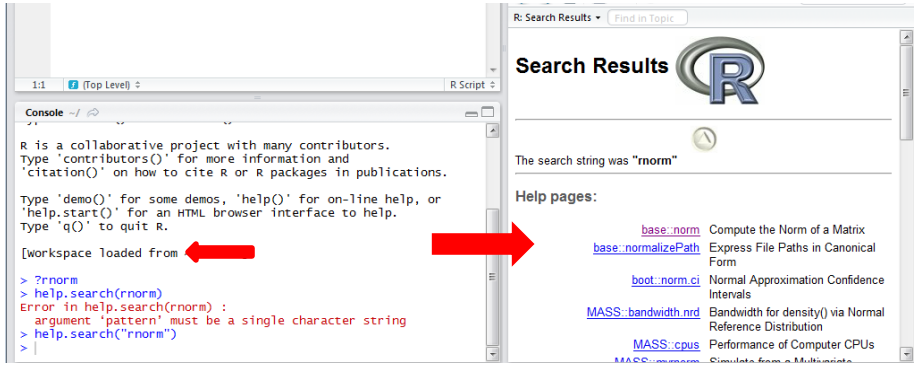
Boxplot(x,y) #plot distribution of x and y

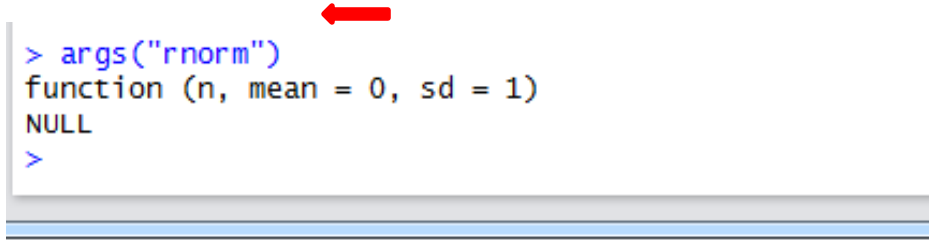
Getting help

- Access help file
  - ?topic
    - E.g. ?rnorm

      for Normal Distribution
- Search help files
  - help.search("topic")
- Get arguments
  - args("topic")







Reference: https://cran.r-project.org/doc/contrib/Short-refcard.pdf

DEMO & ASSIGNMENT

Part1(setup your own environment)

You are asked to setup your R & Git

- R
  - Install R
  - Install Rstudio
  - Open R &Rstudio and take a screenshot

- Github
  - Set up a Github account (you may use your own name or a pseudonym).
  - Create a repo called ***Lab1_studentID (***scratch***)***
  - Clone your github into local repo
  - Create a text file called HelloWorld.md
  - Add the line

    ##This is a secondary heading

    ### This is a Tertiary heading

    * first item list

    * second item list
  - Commit every changes
  - Push the document to the ***Lab1_studentID*** repo you created on Github
  - Fork the data sharing repository here https://github.com/ccucsieRCourse/sharing

Part2 (My first R)

write your first R

- Do some calculations (e.g. 1*1+2*2+3*3+…+100*100)
- Learn some datasets (e.g. ?inspect)
- Learn some function (e.g. ?seq, ?mean)
- Learn some package (e.g. ?stats, ?swirl)
- Try example in the help file

Part 3 (Simple Data Analysis)

- Table morphometric measurements of eight birds.

Wingcrd	Tarsus	Head	Wt
59	22.3	31.2	9.5
55	19.7	30.4	13.8
53.5	20.8	30.6	14.8
55	20.3	30.3	15.2
52.5	20.8	30.3	15.5
57.5	21.5	30.8	15.6
53	20.6	32.5	15.6
55	21.5	NA	15.7

- Use the c function, create a variable wingcrd to hold the wing lengths for the 8 birds
  - Similary, create Tarsus , head and wt
  - Compute basic statistics of the data (mean, median , min , max)

Swirl exercise(home)

- 1) Make sure you have a recent version version of swirl:

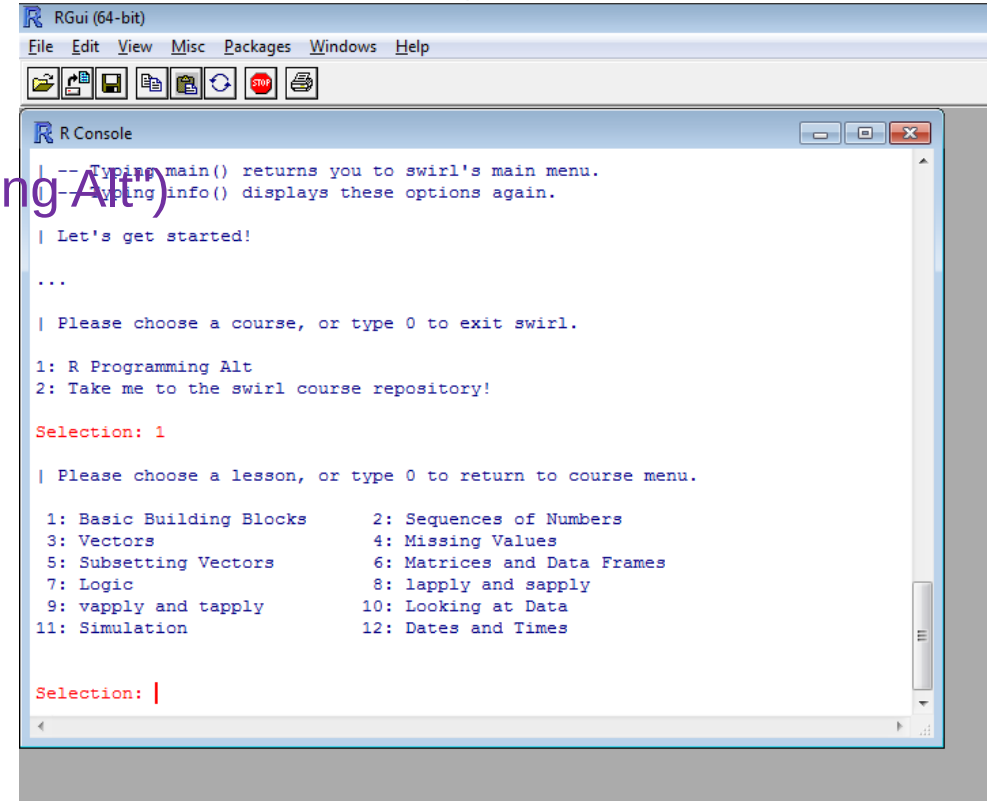
  install.packages("swirl")

- 2) Enter the following from the R console, substituting the name of the course that you wish to install:

  library(swirl)

  install_from_swirl("R Programming Alt")

  swirl()



Hand in

- To submit this lab
  - Use your own repo that you created on your Github
  - R scripts are called regards to part number (e.g. part2.R)
  - Zip Lab1_StudentID that contain R&Rstudio screenshot, part2.R , part3.R
  - Commit each file you have modified
  - Submit the link to the forked repository on your Github account

Hand in

- In order to record to assignment you still need to upload to
- The E-course System
  - https://ecourse.ccu.edu.tw/
- Upload StudentID.zip into “Lab1”

GRADING POLICIES

Grading Policies

- There will be 8 Labs =24%
  - Lab1 =2%, Lab2=2%, Lab3=3%, Lab4=3%, Lab5=3%, Lab6=3%, LAb7=4%, Lab8=4%
- Finish All Lab1 in order to get full credit (part1-part3)
  - In case you haven't finished by the end of class, you can still send it by reducing 20% from full score for each day..
  - e.g. Total score lab1 100, it will 80, 60, 40, ….

References

- Roger, D. Peng.(2015).*R Programming for Data Science*. Victoria, British Columbia : Leanpub.
- Joseph, A. (2012). *R IN A NUTSHELL: A Desktop Quick Reference, 2nd Edition*. Sebastopol, CA : O'Reilly.
- Alain, F. Z. ,Elena N. L., Erik H.W.G. (2009). *A Beginner's Guide to R.* Newburgh,United Kingdom : Springer
- https://www.r-project.org/
- http://gitref.org/
- https://git-scm.com/book/en/v2/Git-Basics-Getting-a-Git-Repository
- https://www.coursera.org/jhu