Data Analytics
CS301
Tools for Working with Data

Week 1: 6th July
Summer 2021

Oliver BONHAM-CARTER

To install for this class

- `ClassDocs` – all class material
- `Git` - to work with GitHub
- `Atom` – an editor
- `Docker` – run programs in an environment on your computer
- `RStudio` – Used for programming in R
  - Two ways to install this:
    - Locally
    - Using Docker

Notes to install each software are below

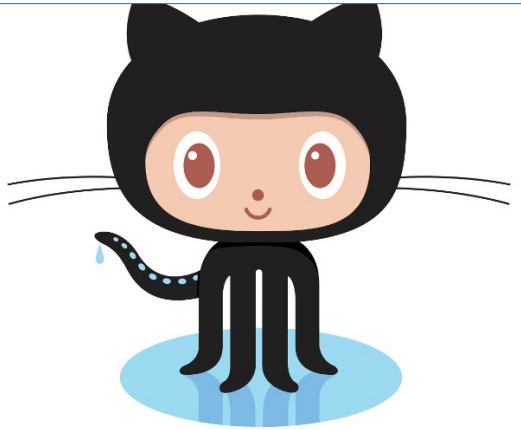
ClassDocs: All Class Materials

- We will be using GitHub to manage all class material. The links below are used to *pull* over your to `classDocs` repository to get slides and labs.

- **HTTP based repository pull**: works in absence of installed ssh keys.
  - `https://github.com/cs301summer2021/classDocs.git`

- **SSH based repository pull**: uses installed ssh keys.
  - `git clone git@github.com:cs301summer2021/classDocs.git`

GitHub



git

Installing Git

- **MacOS**: go to your *Terminal*, type in "`git`" and if not installed, MacOS will offer to install the free *Xcode* software development suit from Apple that contains `git`.
- **Ubuntu**: `Git` may already be installed. If not, use the command, `sudo apt install git` to install `git`. You will need your password.
  - Good ref: https://www.digitalocean.com/community/tutorials/how-to-install-git-on-ubuntu-20-04
- **Windows**: Git does not come with the Windows OS and so it must be installed. Please visit https://gitforwindows.org/ to install and learn more.

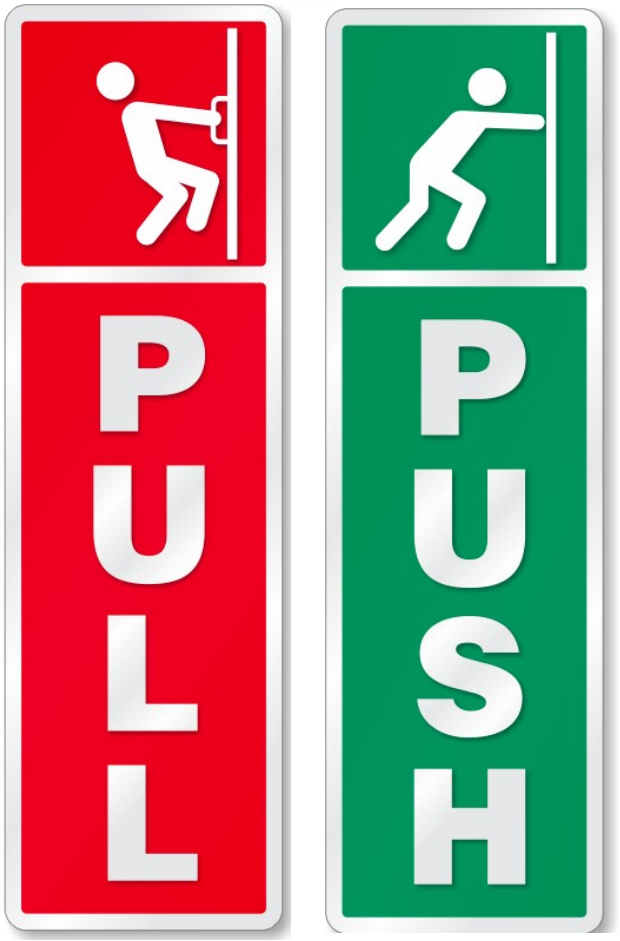
Git and Your Class Repositories

- **PULL** your classDocs before class (cloud data sent to you).

```
git pull
```

- **PUSH** assignment repos to submit homework (your data sent to the cloud)

```
git add -A
git commit -m “My commit mesg”
git push
```

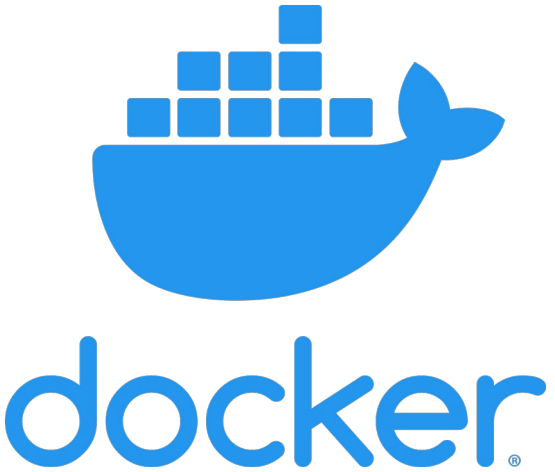


The Atom Editor: Suggested for Programming

- We will be programming and Atom facilitates this task
- If you do not already have it, please download it from: https://atom.io/

Docker for Running Software

- A container in which to run programs in isolation.
- Please be sure that you machine will work with the regular Docker, **not** Docker ToolBox.
- Verify: www.cs.allegheny.edu/canirundocker



Yes!

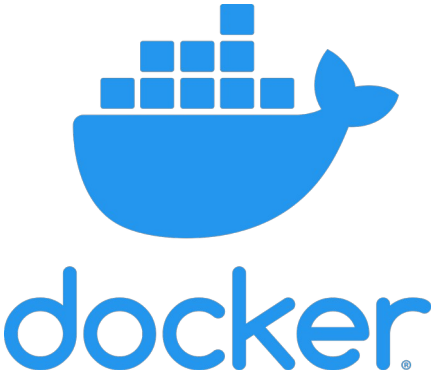
Check the docker docs for more information about the Linux system requirements and installation procedure.

No / Maybe

- Windows: Purchase a Windows Enterprise activation key
- Dual boot: Linux and Windows
- Use another computer

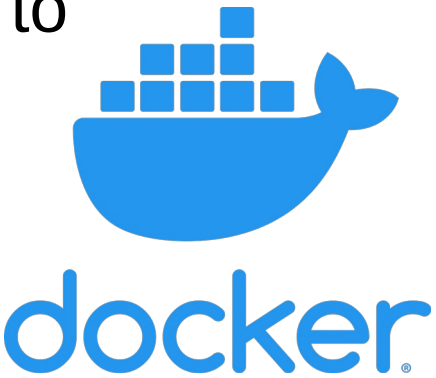
Get Started With Docker

- Running and Testing Programs with Docker and GatorGrader (Dr. Jumadinova):
  - https://www.youtube.com/watch?v=iceAgNEORCA
- Main site
  - https://www.docker.com/
- Downloads
  - https://www.docker.com/get-started
- Tutorial
  - https://www.docker.com/101-tutorial

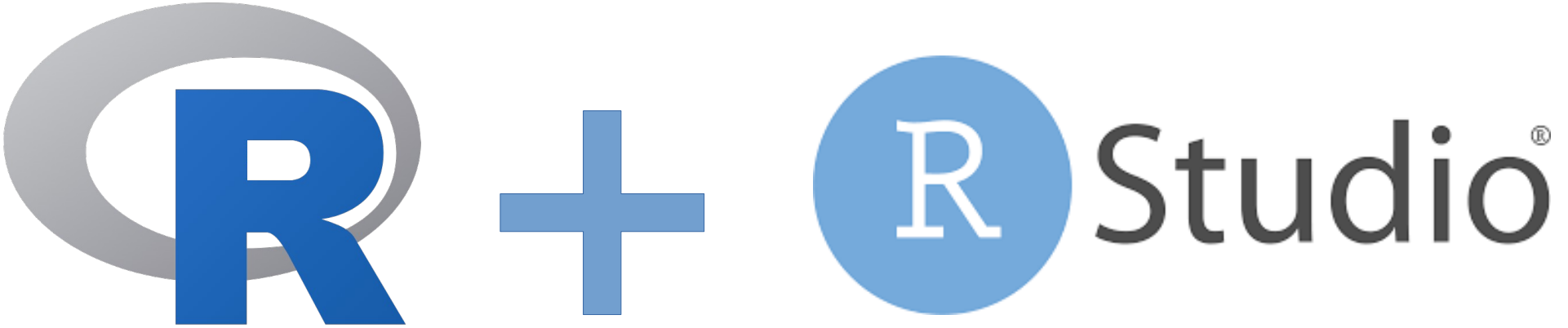


Learning About Docker

- Play-with-Docker
  - https://www.docker.com/play-with-docker
- Once Docker has been installed, you can play with it.
- First, build a work container:
  - `docker run -dp 80:80 docker/getting-started`
- Then, to learn more use your browser to go to the url:
  - http://localhost/



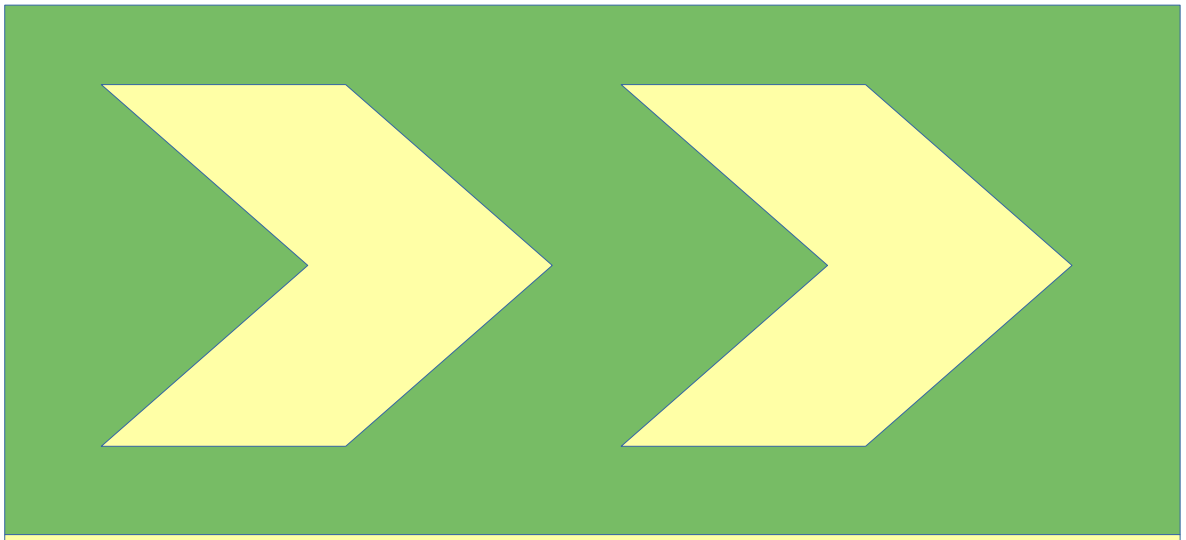
A Local Install of rStudio



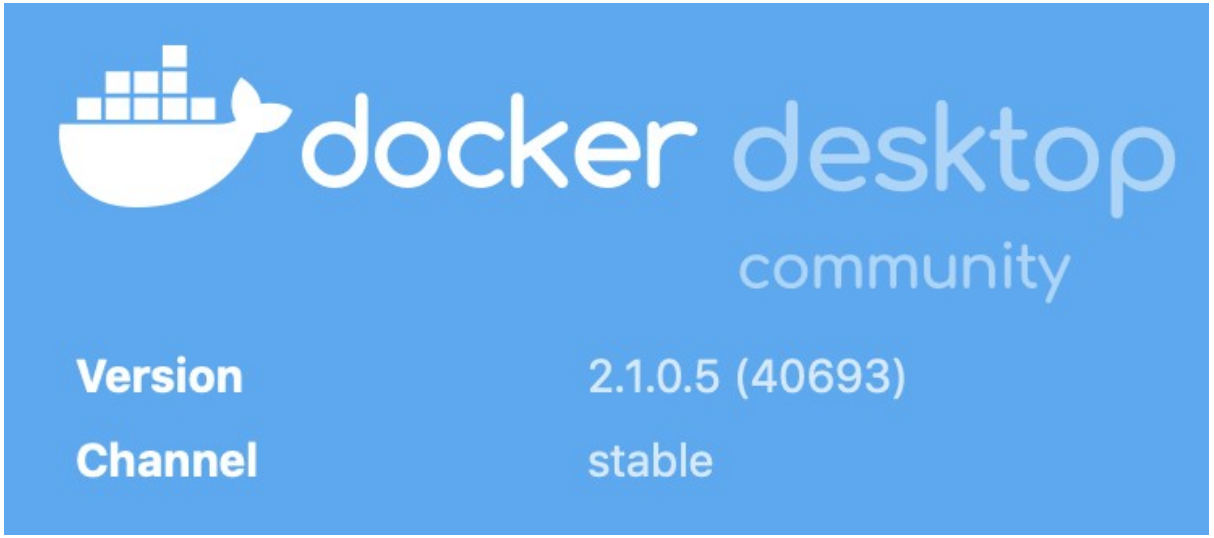
- **You must first install R and then rStudio**
  - The R programming language
    - https://cran.rstudio.com/
  - Rstudio
    - https://rstudio.com/products/rstudio/download/

If you install these, you may not need to use Docker containers for your R programming.

RStudio With Docker



FYI: Using containers



Docker Alternative of: R Programming at Bash

Note: the directory where you run this becomes your local directory in the container.

```
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

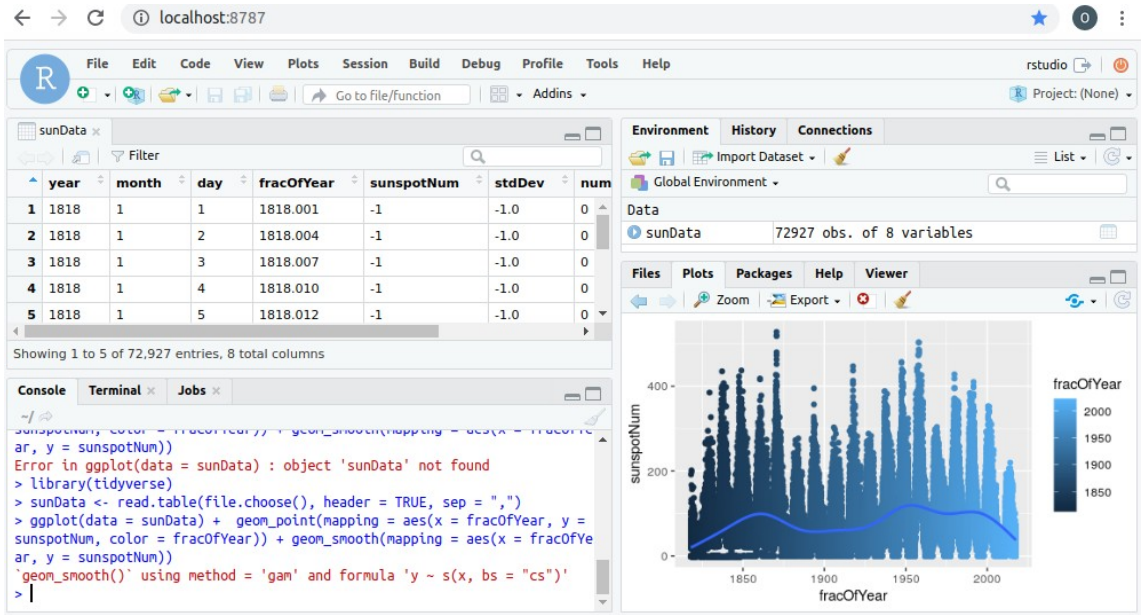
- Build and run container:
  - docker run -ti --rm r-base

- Build, mount local drive and run container :
  - docker run -ti --rm -v "$PWD":/home/docker -w /home/docker -u docker r-base

Docker Container Setup: rStudio

Note: the directory where you run this becomes your local directory in the container.

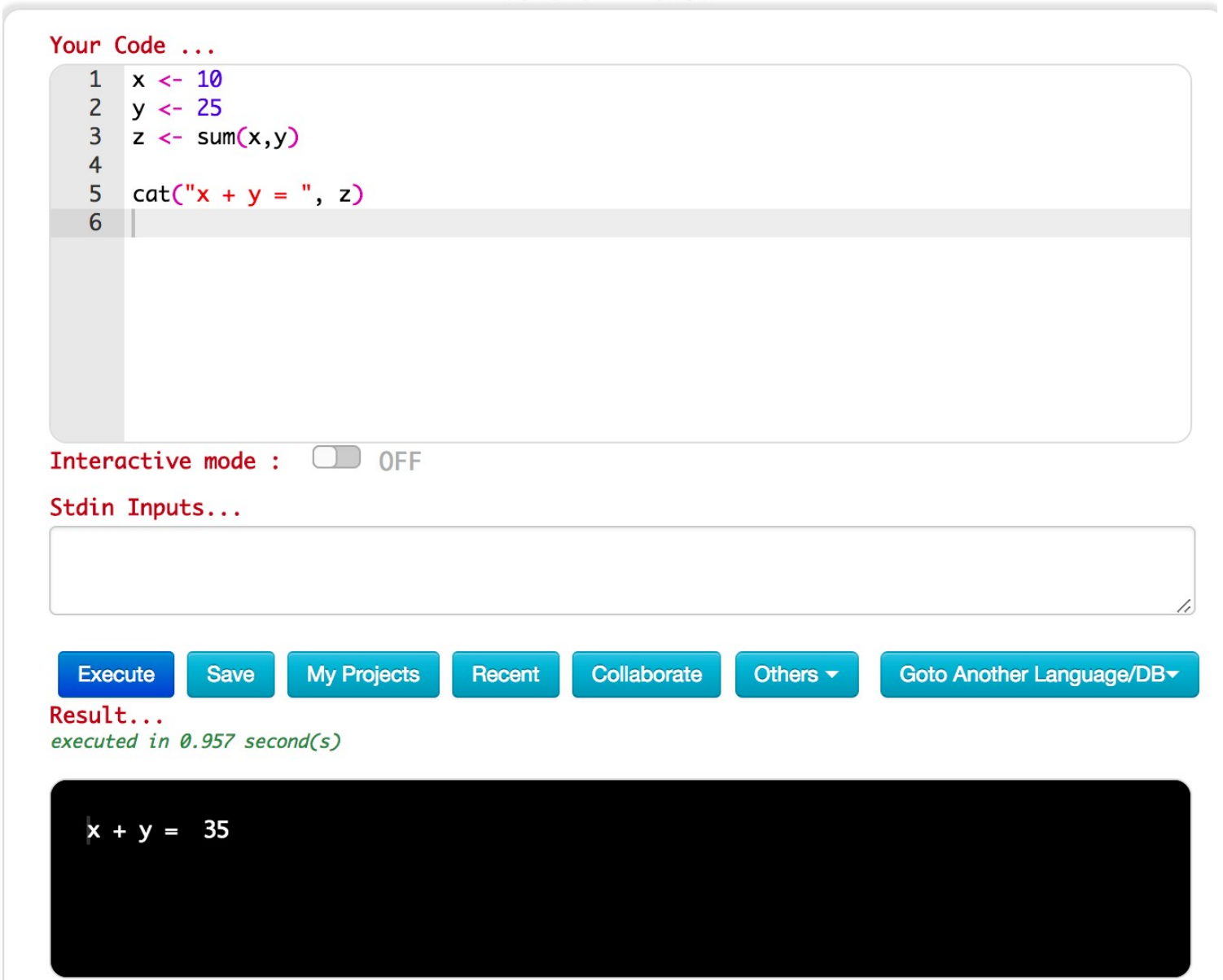
Username: *rstudio*
Password: *letmein*



- Build, mount local drive and run container:
  `sudo docker run --rm -e PASSWORD=letmein -p 8787:8787 -v $PWD:/home/rstudio/ rocker/verse`
- Browser:
  - URL: Use Browser address: http://localhost:8787/

R by Jdoodle

- https://www.jdoodle.com/execute-r-online



Please Read for Next Class

- Come prepared to discuss
- *Twelve Million Phones, One Dataset, Zero Privacy,* A New York Times opinion piece
- Link: https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html

Consider for Discussion

- Why is smart-phone location data considered to be *sensitive and confidential*?
- Discuss any two issues of personal privacy which may likely be discovered when this data is analyzed.
- How could the found trends in the data be used in unethical ways? Who would gain/lose something?
- After reading this article, what concerns you about data handling that did not concern you prior to reading?

Opinion | THE PRIVACY PROJECT

Twelve Million Phones, One Dataset, Zero Privacy

By Stuart A. Thompson and Charlie Warzel

DEC. 19, 2019

Activity 01

- Read the article
- Accept and Pull your activity from GitHub to be completed by next class
- Read the `README.md` file in the repository for more details.

GitHub Activity Repository:
https://classroom.github.com/a/QEJKluQX
Due at 9:30AM (EST) on 7 July 2021
Work file: **reflection.md**