

Data Analytics

CS301

Tools for Working with Data

Week 1: 3rd Sept
Fall 2020

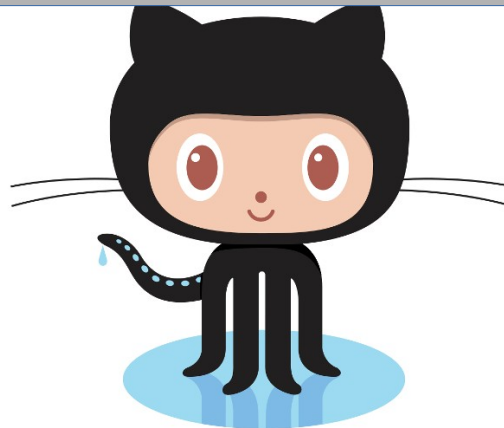
Oliver BONHAM-CARTER

ClassDocs: All Class Materials

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

- **HTTP based repository pull:** works in absence of installed ssh keys.
 - `https://github.com/Allegheny-Computer-Science-301-F2020/classDocs.git`
- **SSH based repository pull:** uses installed ssh keys.
 - `git clone`
`git@github.com:Allegheny-Computer-Science-301-F2020/classDocs.git`

GitHub



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.



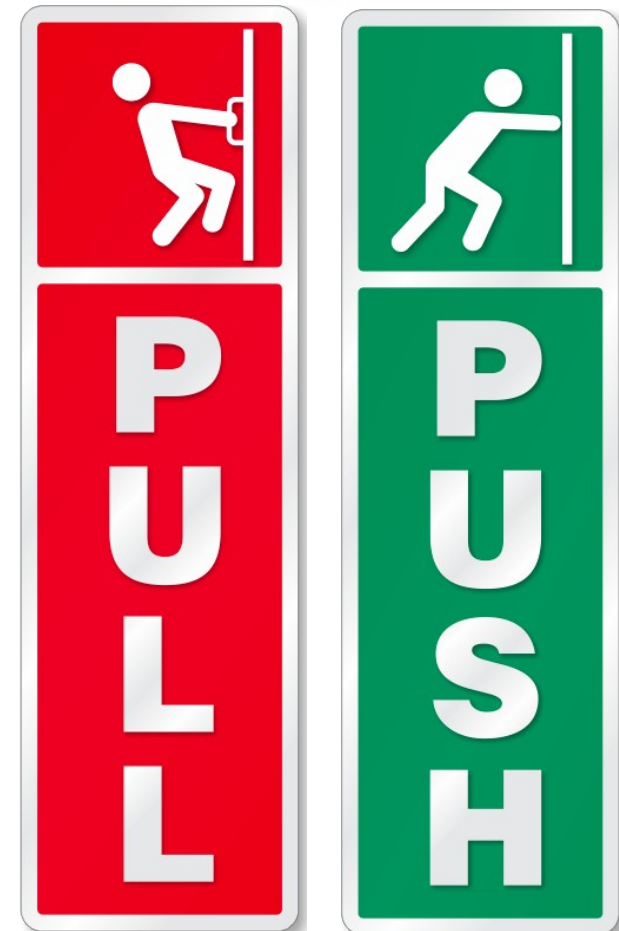
Git and Your Class Repositories

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

```
git pull
```

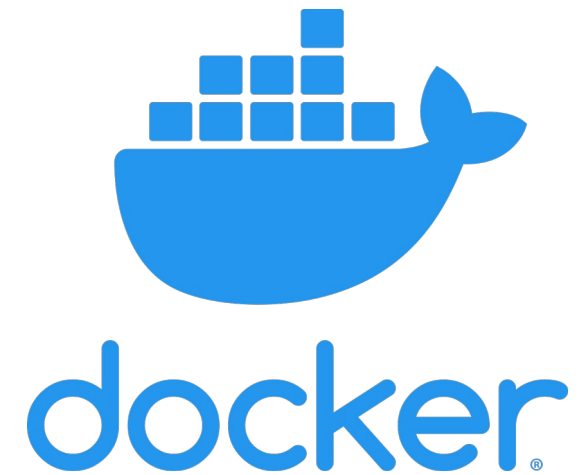
- **PUSH** assignment repos to submit homework (your data sent to the cloud)

```
git add -A  
git commit -m "My mesg"  
git push
```



Docker for Running Software

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



Yes!

Check the [docker docs](#) for more information about the Linux system requirements and installation procedure.

No / Maybe



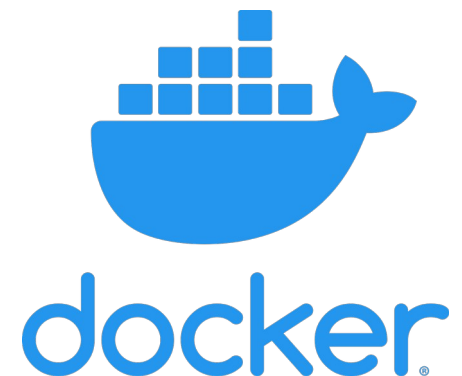
All Set!



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

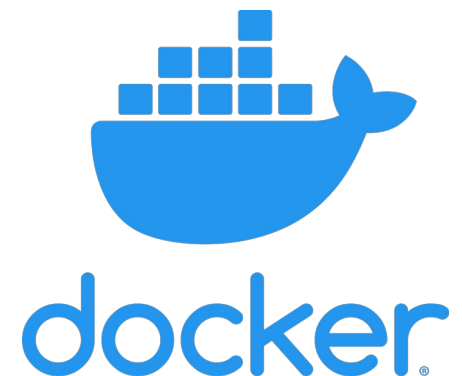
Get Started With Docker

- Dept of CS video (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>



Exploring Docker

- Play-with-Docker
 - <https://www.docker.com/play-with-docker>
- Once Docker has been installed, you can play with it.
- Build the container:
 - `docker run -dp 80:80
docker/getting-started:pwd`
- Then browser url: `http://localhost/`



Atom:

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/>



Please Install Your Software

- We will be using Git and GitHub. Please setup your account **by next class** at <https://github.com/> and also download a Git client software from <https://git-scm.com/downloads>
- We will also be using the Atom editor to write code. Please download and install your editor from <https://atom.io/>
- For most labs, we will be using Docker. Please download and install your Docker Desktop installation (note: not the Docker ToolBox) from <https://www.docker.com/>. Help: <https://hub.docker.com/>
- If necessary, please help each other to install this software. Or see the department's Technical Leaders with questions.

Links to download sites are above!



Article For Class

- *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>

Opinion | THE PRIVACY PROJECT

Twelve Million Phones, One Dataset, Zero Privacy

By Stuart A. Thompson and Charlie Warzel

DEC. 19, 2019

THINK



Discuss With Your Group

- 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 do now have data that you did not have prior?

Opinion | THE PRIVACY PROJECT

Twelve Million Phones, One Dataset, Zero Privacy

By Stuart A. Thompson and Charlie Warzel

DEC. 19, 2019





Next Week's Focus



- Web site analysis
 - Yandex Metrica
 - Matomo
 - Google analysis



Looking at Websites and Data

- The Internet houses websites that are used to give product information to *consumers* and potential *customers*.
- Q: How do owners of these sites know that they are fulfilling their roles to generate business?
- Web Analytics to study web traffic to and from a site
- Yandex Metrics dashboard demo:
 - <https://metrika.yandex.com/dashboard?group=day&period=week&id=44147844>
- Matomo dashboard demo
 - Demo: <https://demo.matomo.org/>