

Data Analytics

CS301

Introduction to Data Analytics

Week 1: 15th Jan

Spring 2020

Oliver BONHAM-CARTER

ClassDocs: All Class Materials

- We will be using GitHub to manage all class material. Clone your repository with using the following:

- Ssh: `git@github.com:Allegheny-Computer-Science-301-S2020/classDocs.git`
- HTTPS: `https://github.com/Allegheny-Computer-Science-301-S2020/classDocs.git`

GitHub



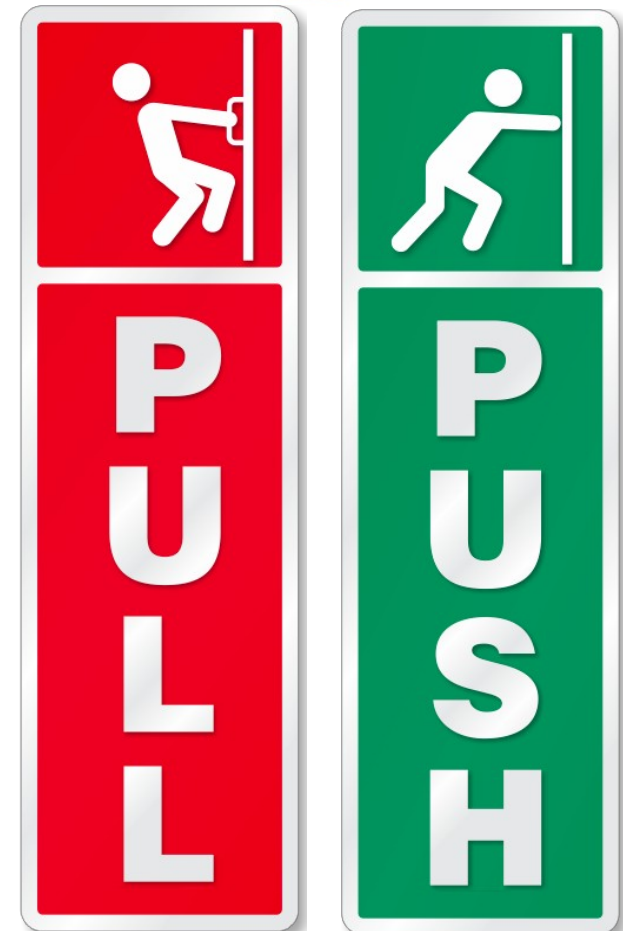
Class Repositories

- **PULL** your classDocs before class.

```
git pull
```

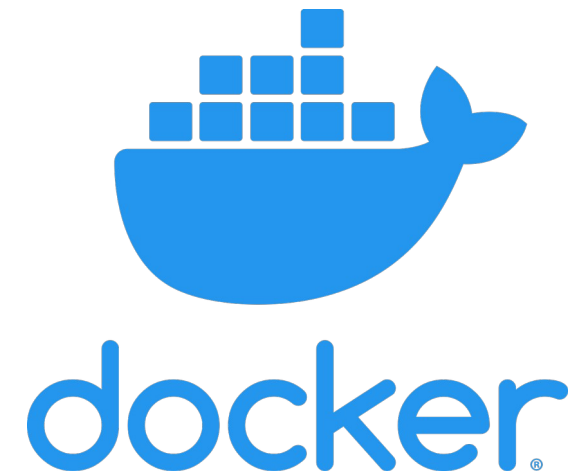
- **PUSH** assignment repos to submit homework

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



Docker for Running Software

- We will be using Docker in lab and class
- Please be sure that your machine will work with the regular Docker, **not** Docker ToolBox.
- Verify: www.cs.alleggheny.edu/canirundocker



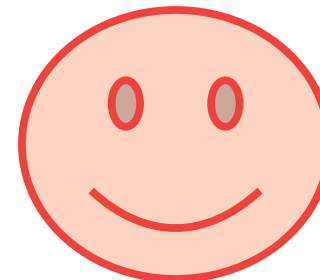
Yes!

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

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.



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



Group Discussion Questions

- According to the article, who uses (or could use) the collected data?
- What are three potential benefits of using this data or similar types of data?
- What are three potentially unethical consequences of using this data or similar types of data)?
- How can we prevent these unethical consequences?

Opinion | THE PRIVACY PROJECT

Twelve Million Phones, One Dataset, Zero Privacy

By Stuart A. Thompson and Charlie Warzel

DEC. 19, 2019





Consider This ...

- You are given the lists of words from several main stream-news articles.
- Pick a list to work on with a group of your peers.
- Although the article text cannot be read directly, can you determine the general sense of the article from a list of its words?
- What is the general subject of your article?
 - Are there names of people you recognize in your list? What can you infer about the article from the name(s)?
 - Do the listed nouns support your conclusions?
 - What type of media source would contain such a story?

THINK