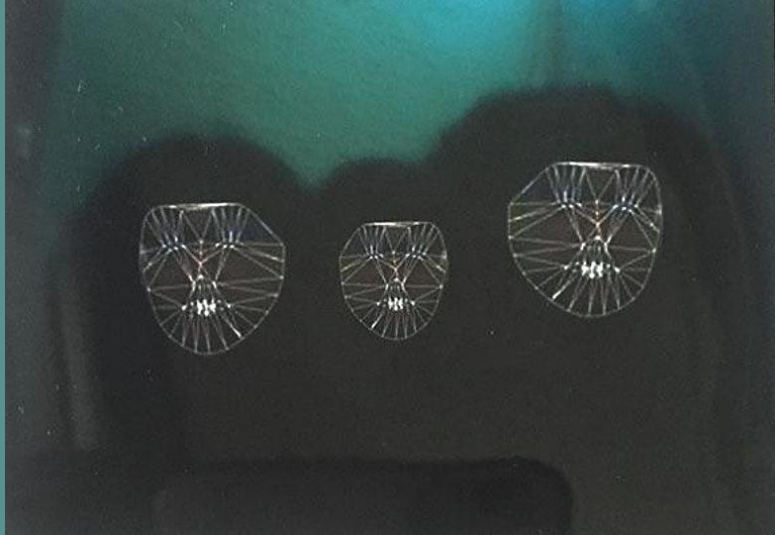


Dataism - Week 1



School of Machines, Making and Make
Believe





Introductions

- What is your background?
- What are your main interests?
- What is your motivation for taking this course?
- What do you hope to get out of it?



Focus Questions for the Course

- **What are some of the key concepts in data science, particularly in the context of activism?**
- **How can a clearer understanding of data and data science allow us to become better citizens and activists?**
- **What do we as citizens need to know in order to avoid being manipulated or subjugated?**
- **Where does the transition between numbers and morals occur?**



Case Study 1 : CA Police Scorecard

Problem: Policing in the United States is plagued by excessive brutality and racial bias.

Response: Use data on arrests, shootings and complaints registered against officers to create a metric grading police departments on:

- Use of excessive force
- Racial bias
- Holding officers accountable
- Approach to policing (ex. “broken windows” policy vs. solving serious crimes)
- **Main Reference:** <https://policescorecard.org/>
- **Video:** https://www.youtube.com/watch?v=isFNFI0O_6k



Case Study 1 : CA Police Scorecard

- **Get Involved:** <https://staywoke.typeform.com/to/jBvCkB>
 - Can join a work group and specify a specialization!
- **Github:** <https://github.com/campaignzero/ca-police-scorecard/>
 - Contains a Jupyter Notebook with methods explained in full detail along with all code
 - Raw and preprocessed data also available
- **Google Drive:**
<https://drive.google.com/drive/folders/1XAT1uFPXj5AsvNTzFeNeeTXGLP09HEIh>
 - More raw data, police handbooks, data for further study



Case Study 2 : CV Dazzle

Problem: Use of facial recognition algorithms has become widespread. This often involves secret surveillance and use of citizen's facial data without their knowledge or permission. Algorithms built on facial recognition often include racial and other types of bias.

Response: Raise public awareness using eye-catching makeup and fashion designs that fool algorithms. Exploit weakness in facial recognition to protect privacy.

- **Main Reference:** <https://cvdazzle.com/>
- **Clearview AI:** <https://www.youtube.com/watch?v=ufAISqXIPB4>
- **News story on the Dazzle Club:**
<https://www.youtube.com/watch?v=ch2ubo7fC84>



Case Study 3 : Social Networks and HIV/Substance Abuse Prevention

Problem: Drug abuse and spread of disease through sharing needles etc. is very prevalent amongst homeless youth.

Response: Both can be curtailed by sharing information on safe practices and providing other forms of support. Information spread and peer intervention groups can be optimized using AI and data on the social circles of homeless youth.

- **Main References:**

- <https://cais.usc.edu/projects/social-network-based-substance-abuse-prevention-for-homeless-youth/>
- <https://cais.usc.edu/projects/hiv-prevention-homeless-youth/>



Course Tools/Resources

- Google Colab (Jupyter Notebook)
- Github (Assignments, Slides, Reading...)
 - <https://github.com/arputtick/dataism>
- Slack (Discussion, sharing resources...)
- Jamboard (Discussion groups, brainstorming)



Why use Colab?

- Don't need to install anything. Runs on a virtual machine with Python and all important libraries (pandas, numpy, tensorflow, pytorch etc.)
- Free access to GPU/more computing power
 - Often needed for large amounts of data
- Easy to share and collaborate
 - Also because runs independent of machine collaborators are using
- (P.S. You can still use ordinary Jupyter Notebooks and run locally if you want.)



Using Colab

1. Create a Google Account if you don't already have one:
 - a. <https://accounts.google.com/signup/v2/webcreateaccount?hl=en&flowName=GlifWebSignIn&flowEntry=SignUp>
2. Go to Colab: <https://colab.research.google.com/>
3. Select 'Github' on welcome menu.
4. Search for 'arputtick/dataism' to see the notebooks available in the course repository.

You should see something like this:

Examples

Recent

Google Drive

GitHub

Upload

Enter a GitHub URL or search by organization or user

☐ Include private repos

arputtick/dataism


Repository: [🔗](#)



Branch: [🔗](#)

arputtick/dataism

master

Path

 assignments/Assignment_0.ipynb

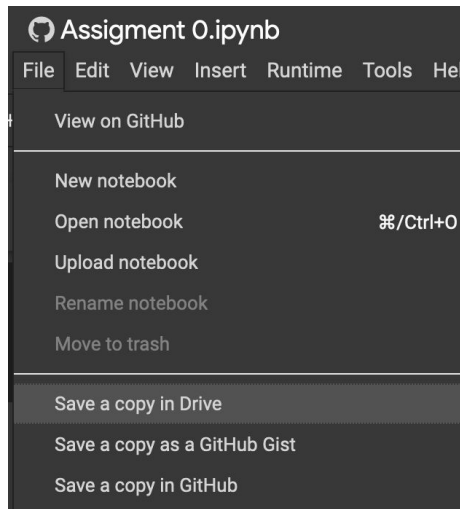
 

Click on a notebook to open it in Colab.



Working in Colab

- Once you open a Notebook, you can work on it freely. To save your work:



ANY CHANGES YOU MAKE WON'T BE SAVED UNLESS YOU SAVE A COPY IN DRIVE OR GITHUB!!!

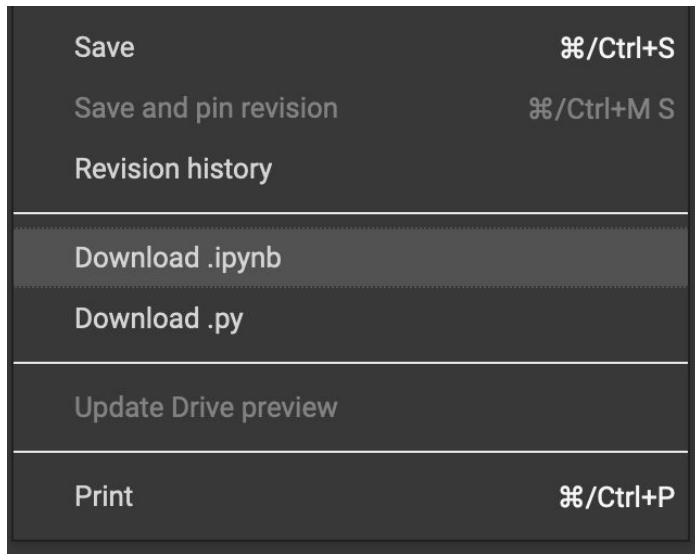


Working in Colab

You can also download a copy of the notebook:

‘File’ -> ‘Download .ipynb’ or

‘Download .py’





Course Structure

- ~60 min. Lecture
- ~30 min. Interview
- ~30 min. Discussion/Workgroup

With breaks ;)



Homework:

- Exercises with simplified tasks to illustrate concepts
- Simultaneously familiarize students with tools to the point that they can get started on their own projects
- Students should also investigate each case study on their own BEFORE we discuss in class (I'll give specific recommendations)



Final Project

We'll make this more precise as the course goes on, but I'm thinking that students (perhaps in groups) select an issue that is important to them and either:

1. Find an existing case study, investigate it and present it to the group, or
2. Start conceptualizing their own project:
 - a. Translating from qualitative ideas to measurable data
 - b. Searching for available data
 - c. Methods/Planning how to use it
 - d. Start prototyping



Supplementary Material

- **Books:** Weapons of Math Destruction, Donut Economics
 - Both give a great idea of the power data models and algorithms have to shape society once they become widespread. They also present alternatives to the current norm.
- **Online Courses:**
 - **Coursera:** deeplearning.ai, statistics for python, tensorflow in practice (for later)
 - **Fast.ai:** deep learning course with the goal of getting personal projects up and running AFAP
 - **Stanford CS231:** if you feel like going deeper



Note on Learning

I've always found it best from a motivational standpoint to learn “top-down” and “bottom up” simultaneously:

- **Bottom up**, e.g. this course. Minimal theory. Goal of doing/understanding something practical as quickly as possible. Can also be a personal project, but starting with workshop or course helps
- **Top down**, more theory/techniques/tools, like a book on deep learning or a class focused statistics and python

Regular interaction with humans with similar interests may also help.





Go to Github (<https://github.com/arputtick/dataism>) for the homework for next week :)