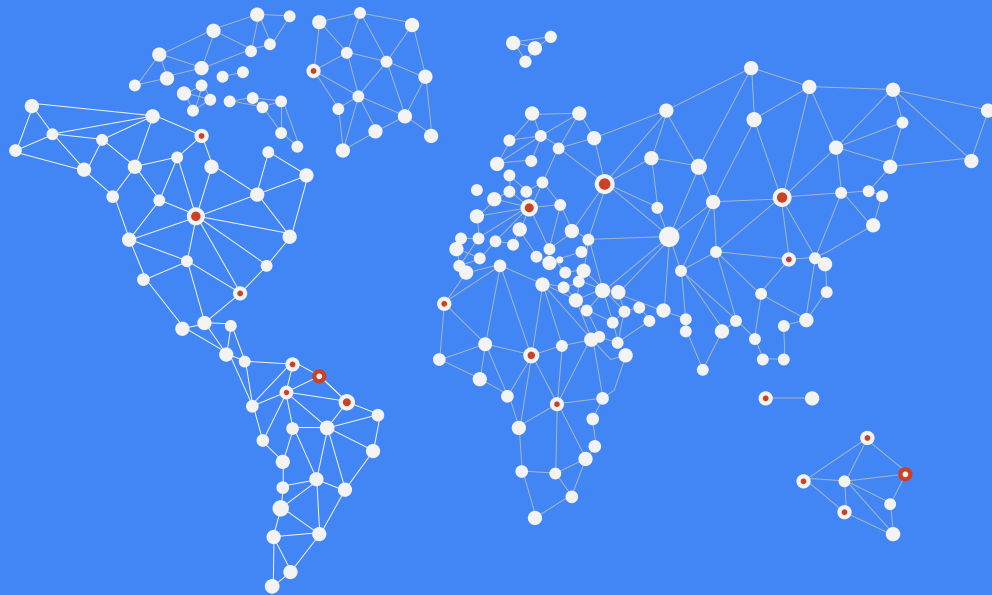


# Intro to Data Science

Week 1 - Introduction  
Lecturer: Eric Hsieh  
6.13.20



# Course Goals

Tech Data Science intuition

Tech how to do Data Science

Help students explore their next data journeys

# Today's Agenda

**1st Hour** : Class/Course/Data Science Intros (50 min)



**Break** (10 min)

**2nd Hour**: Data Science Tech Stack Intro (50 min)

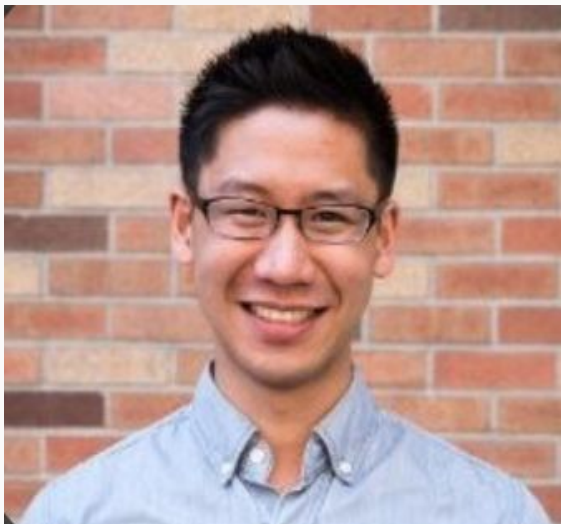


**Break** (10 min)

**3rd Hour**: Statistics as a Data Science Building Block (50 min)

**Conclusion** (10 min)

# Your Instructor



Name: Eric Hsieh

Head of AI & ML Product at AutoGrid

University of Michigan at Ann Arbor

Favorite superhero:



# Go around and introduce yourselves

Name, school, and grade?

Favorite superhero?

Why are you taking this course?

What is Data Science (your understanding)?

# Hour #1:

## Class/Course/Data Science Intros

# Course Plan

Week	Subject	Assignment
1	Introduction	#1 - Help the Poor Data Scientists (Data Sampling)
2	Statistics	#2 - Help Your Poor Profs (Stats Via Python)
3	Exploratory Data Analysis (EDA)	#3 - All of the Plots (EDA Exercise)
4	Data Cleaning / Feature Engineering	#4 - Not Your Typical Cleaner (Data Manipulation Exercise)
5	Scoring/Testing Models	#5 - Data Referee (Train/Score/Test First ML Model)
6	ML (Regression)	#6 - Guess a Number! (Regression Modeling)
7	ML (Classification)	#7 - Be Better than a Coin (Classification Modeling)
8	ML (Ensemble + KNN)	#8 - End Game (Student Case Studies)
9	ML (Other)	#8 - End Game (continue)
10	Case Studies Presentations	N/A

# What is Data Science?

The art of using data to help the world **better decide** via **different granularity and frequency**.

Data Science = Classical **Math/Stats** Theories + Modern **Engineering** Tools



# What are some Data Science Use Cases?

Predict if/when an airline engine will fail

Predict weather (i.e., temperature)

Image recognition of whether or not a mole is cancerous

Recommending your next movie



NLP Bot that chats with you as a customer service representative

NLP Bot that talks to you as a friend

# Different Types of Data Scientists



Data Engineer



ML Researcher



ML Engineer



Data Scientist (core)



Data Scientist (analytics)



Visualization Guru

# Typical Data Science Workflow



# Setting Up: JupyterHub + Pull Course Repo

**Step #1:** Log in to your respective  
JupyterHub account



Sign in

Warning: JupyterHub seems to be served over an unsecured HTTP connection. We strongly recommend enabling HTTPS for JupyterHub.

**Username:**

**Password:**

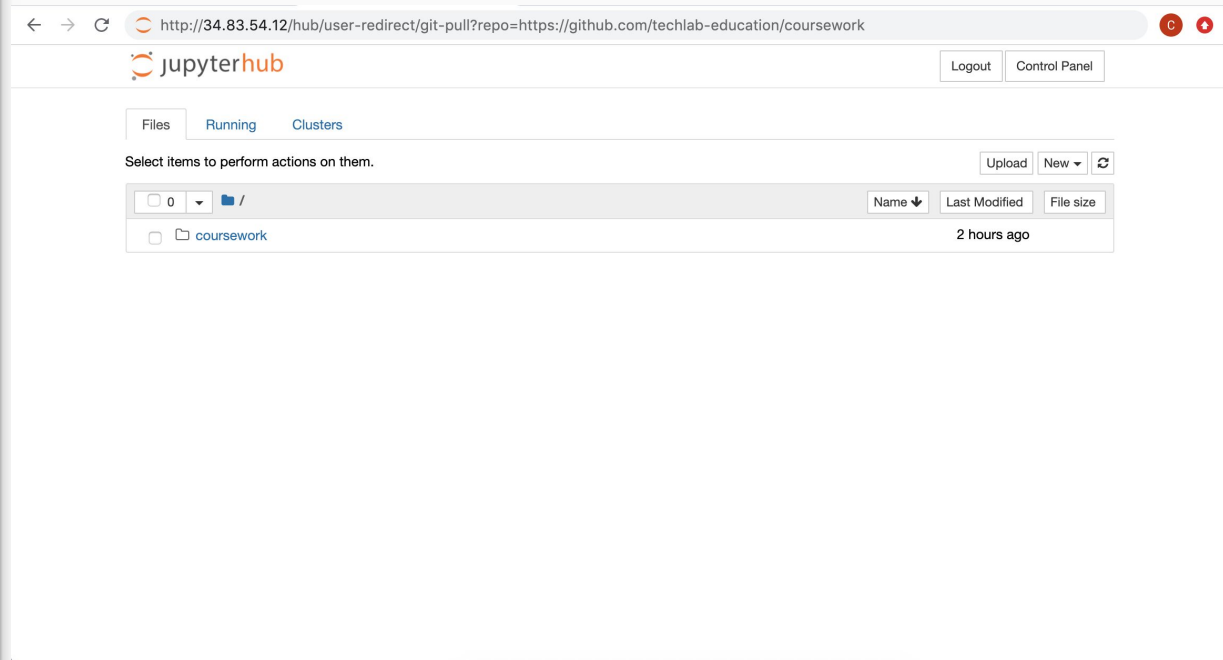
Sign In

# Setting Up: JupyterHub + Pull Course Repo (continue)

**Step #2:** “Pull” the course folder (repo) from online into the server on your JupyterHub account by putting the below address in the search bar of the same browser where JupyterHub is opened.

Web address to copy paste:

<http://34.83.54.12/hub/user-redirect/git-pull?repo=https://github.com/techlab-education/coursework>

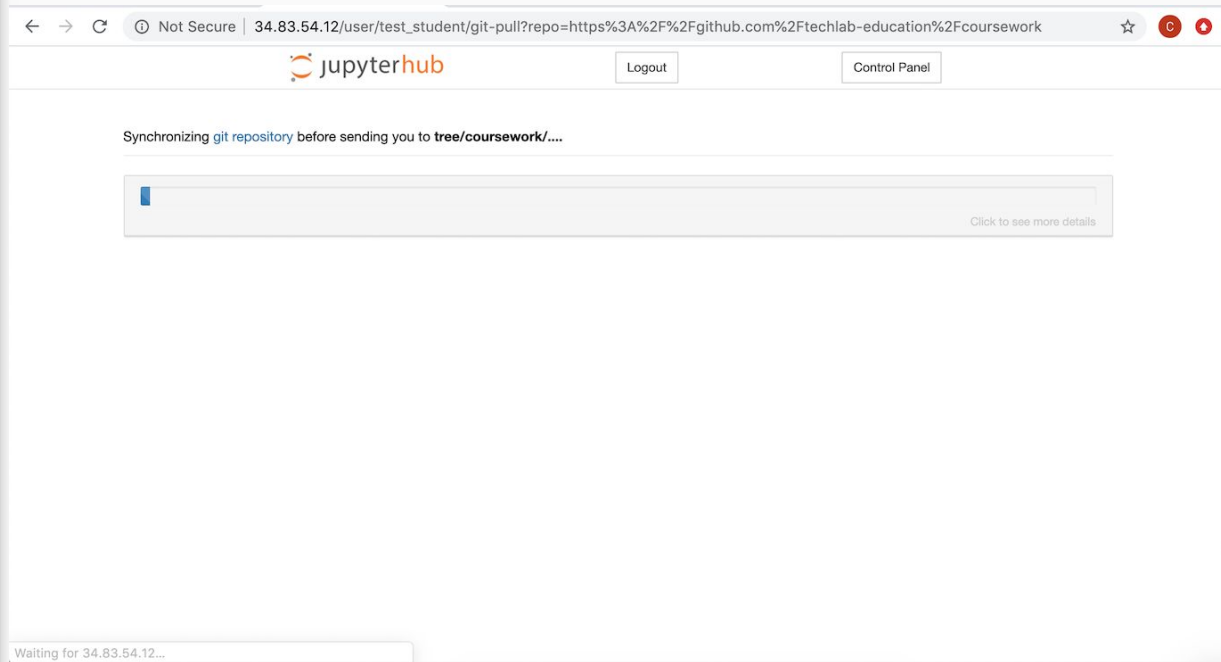


# Setting Up: JupyterHub + Pull Course Repo (continue)

Page of the left will appear (JupyterHub pulling the repo).

Once done, you will have the most updated course repo/materials!

**Step #3:** Remember to repeat this step every week before class to get the most updated course materials (we will release homework/lecture materials as the summer unfolds)



# Course Materials Structure

<input type="checkbox"/>	0	▼	📁 / coursework
	📁	..	
<input type="checkbox"/>	📁	Cheat_Sheets	
<input type="checkbox"/>	📁	Data	
<input type="checkbox"/>	📁	Homeworks	
<input type="checkbox"/>	📁	Intro to Python	
<input type="checkbox"/>	📁	Lectures	
<input type="checkbox"/>	📄	README.md	

# Break (10 min)

For the rest of the lecture, we will use Jupyter Notebook (**Week 1 - Lecture #1 - Notebook.ipynb**) under folder **coursework/Lectures** on JupyterHub.

Please open the notebook to follow along. 