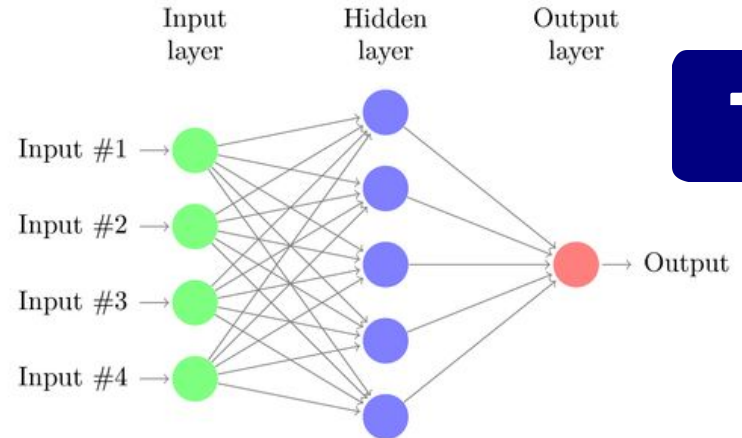
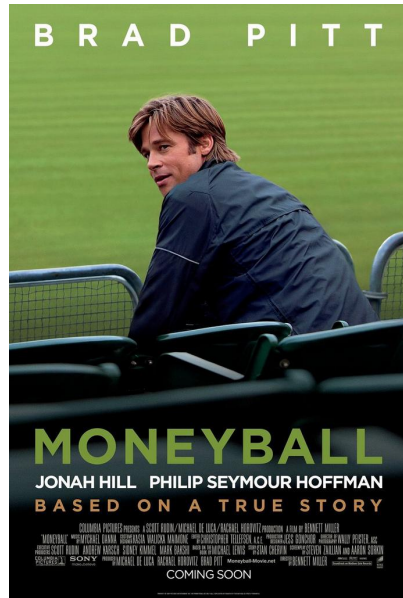
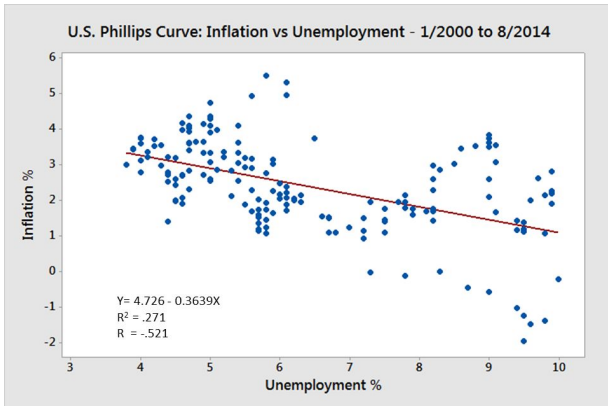
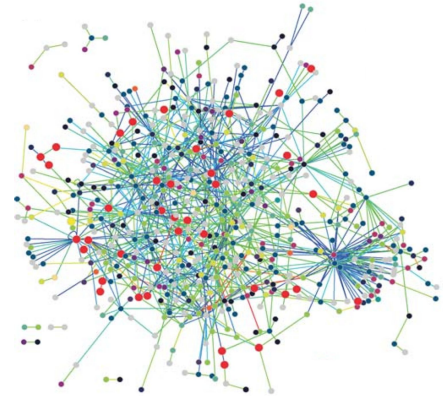
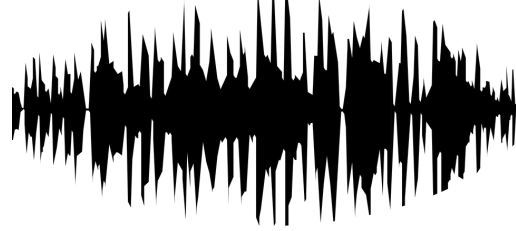


STAT606

Computing for Data Science and Statistics

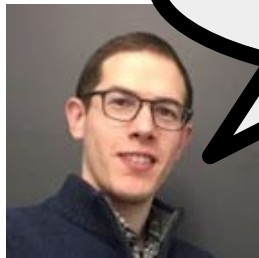
Lecture 0: Introduction and Administrivia

“Data science” has completely changed our world

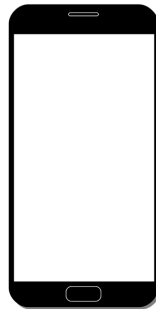


Speech Recognition: then and now

circa 2008

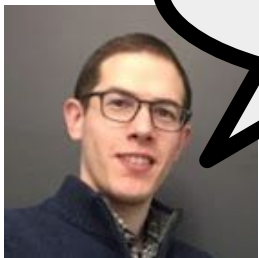


I like the baby green
hippopotamus

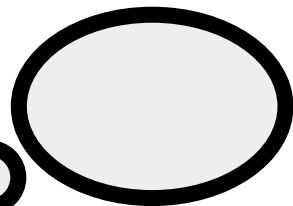


???

circa 2017



啤酒?



Natural Language Processing: then and now

1966: ELIZA (<https://en.wikipedia.org/wiki/ELIZA>)

- Rudimentary “chat bot”; no learning; no statistical model

2010s: Statistical Machine Translation → Neural Machine Translation

- Until the early 2010s, machine translation based on statistical modeling
- In the 2010s [neural sequence-to-sequence](#) models emerged

Circa 2020: GPT-2 (<https://en.wikipedia.org/wiki/GPT-2>)

- Increasingly “human-like” text generation, but still made weird mistakes

2022: ChatGPT (<https://en.wikipedia.org/wiki/ChatGPT>)

- Capable of producing long-form text, often indistinguishable from human

Course goals

- Establish a broad background in Python programming
- Introduce basic concepts in computer science
- Survey popular tools in academia/industry for data analysis and exploration
- Learn basic distributed computing frameworks
- Learn how to read documentation and quickly get new tools up and running

These tools will be obsolete some day...

...but not your ability to learn new frameworks and solve problems!

Course structure

Unit 1: Introduction to Python

Data types, functions, Jupyter, classes, objects, functional programming

Unit 2: Numerical Computing and Data Visualization

numpy, scipy, scikit-learn, matplotlib

Unit 3: Dealing with structured data

Python pandas, retrieving web data, SQL, APIs

Unit 4: Big data and distributed computing

Hadoop, Spark, TensorFlow

Schedule (tentative) and other information on Canvas or course webpage:

<http://pages.stat.wisc.edu/~kdlevin/teaching/Spring2023/STAT606/>

Prerequisites

This course has no formal prerequisites, but...

I am, for the most part, assuming that every student in the class took STAT605

- You can review that material on the webpage for STAT605 (Fall 2020)
<http://pages.stat.wisc.edu/~kdlevin/teaching/Fall2020/STAT605/index.html>

This course is probably not for you if:

- you have no programming background
- you have never used the UNIX/Linux command line

Important note: STAT606 was offered previously as STAT679.

- If you took this course previously as STAT679, you **may not** take it again, and should drop the course.

Course information

Instructor: Keith Levin

- Email: kdlevin@wisc.edu
- Office: 6170 MSC
- OH: TBA or by appointment

We have a grader, but no TA

- Please be patient with grading!

Textbook: None

- Readings posted to the website

Grading: ~13 HWs,

- Weighted approximately equally
- No midterm, no final
- No class project
- Late days (see syllabus)

See syllabus on Canvas or at

<http://pages.stat.wisc.edu/~kdlevin/teaching/Spring2023/STAT606/syllabus.pdf>

Before we continue...

Readings:

For the first part of the course, readings will be given in both
Allen B. Downey's [*Think Python*](#) and
Charles Severance's [*Python for Everybody*](#)

You can do the readings out of either one, whichever you prefer!

Later, we'll make exclusive use of Severance

In the last part of the course, your textbook will be documentation!

A Note on Readings

I will post weekly readings throughout the course

Some of the readings consist of technical documentation

It is a goal of this course to get you comfortable reading docs!

Read and understand what you can, google terms you don't understand...

...and it's okay to set things aside to come back to later!

Course format

Originally, STAT606 was a lecture-based course

in Spring 2022, I tried an experiment:

- Lectures (with slides, small programming examples, etc) posted online

- In-person meetings dedicated to hands-on programming exercises

- Students were nearly unanimous in preferring this workshop-style format

So, we will continue it this semester. That means:

- Do the readings and/or watch video lectures *before* our meetings!

- Bring your laptop to class meetings, if you plan to do the in-class exercises!

Policies

Don't plagiarize!

- You may discuss homeworks with your fellow students...
- ...but you must submit your own work
- Disclose in your homework whom (if anyone) you worked with

Late homeworks are not allowed!

- Instead, we have “late days”, of which you get 7
- One late day extends HW deadline by 24 hours

Refer to the syllabus for details.

Survey: background and availability

I have posted a survey to Canvas to:

- 1) Assess your background
- 2) Decide what topics to cover in more/less detail
- 3) Figure out when to schedule office hours

Please fill it out promptly!

Things to do very soon:

Install Python 3 and pip/conda

Download Python here: <https://www.python.org/downloads/>
and make sure you are comfortable running it from the command line

Pick an editor/IDE for python (e.g., IDLE)

or just use a text editor, or just write directly in jupyter

Install and familiarize yourself with jupyter:

<https://jupyter-notebook.readthedocs.io/en/latest/>

Note: we will use only Python 3 in this course. To check that you have Python 3 installed, call `python --version` on the command line.

Other things

HW0 is posted to canvas and the website. **Get started now!**

If you run into trouble, come to office hours for help

- But also please post to the discussion board on Canvas
- If you're having trouble, at least one of your classmates is, too
- You'll learn more by explaining things to each other than by reading stackexchange posts!

Email policy:

I will **not** provide tech support over email!

If you are having trouble, post to the discussion board and/or come to OHs!

A parting note...

Homework:

Start your homework early!

If you run into technical issues, you'll want to have time to come get help!

A note on pace and difficulty

I aim to teach Python from scratch in this course, but...

...time spent on Python is time not spent on the stuff you're really here for

So, I expect that you are willing to work hard to keep up

If lectures are moving too fast, or you don't understand something, speak to me promptly!