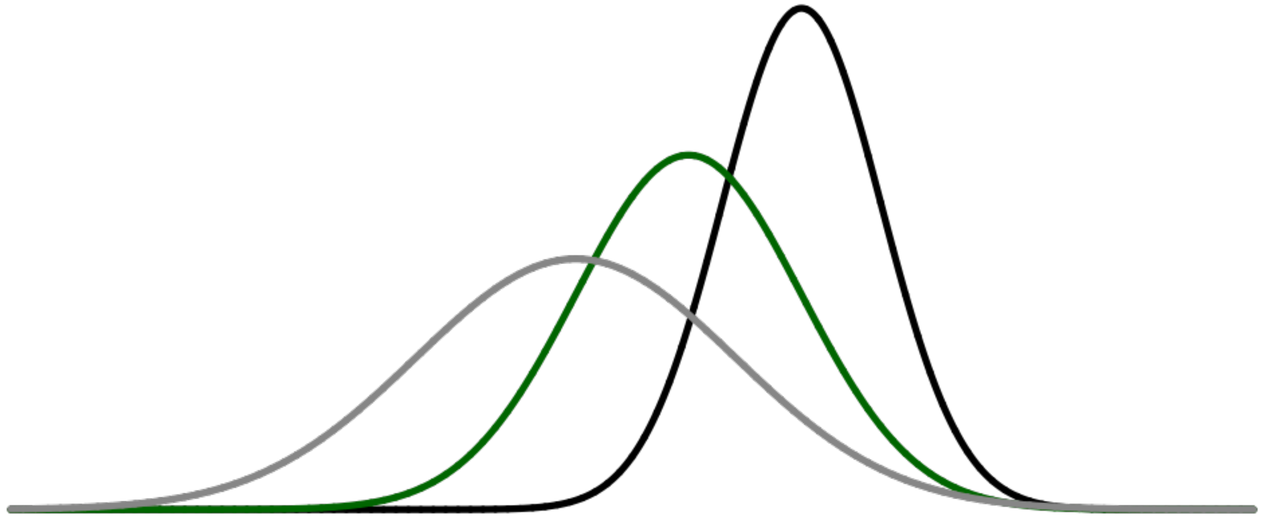


Welcome and Syllabus

STAT 432 | UIUC | Fall 2019 | Dalpiaz

Questions?

Comments? Concerns?

STAT 432

Basics of Statistical Learning

Also ASRM 451….

stat432.org


Hello
my name is
DAVE

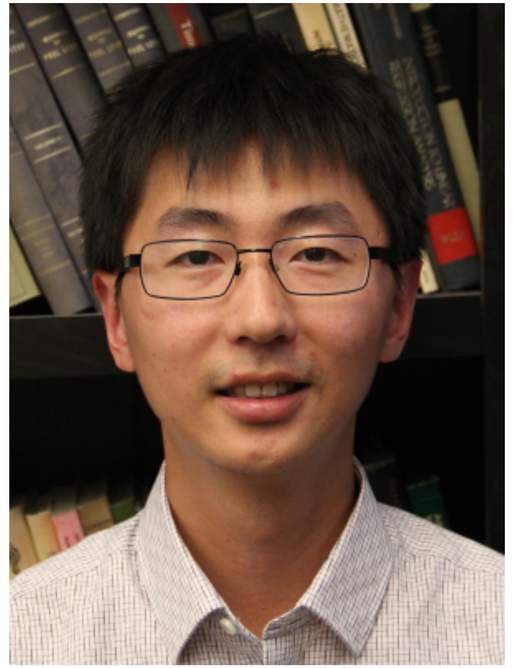
dalpiaz2@illinois.edu

David Dalpiaz

Room 36, 703 S. Wright

David Dalpiaz
Instructor

Mengchen Wang
Teaching Assistant



Zihe Liu
Teaching Assistant

Course Logistics

Prerequisites?

Course Description

Topics in supervised and unsupervised learning are covered, including logistic regression, support vector machines, classification trees and nonparametric regression. Model building and feature selection are discussed for these techniques, with a focus on regularization methods, such as lasso and ridge regression, as well as methods for model selection and assessment using cross validation. Cluster analysis and principal components analysis are introduced as examples of unsupervised learning.

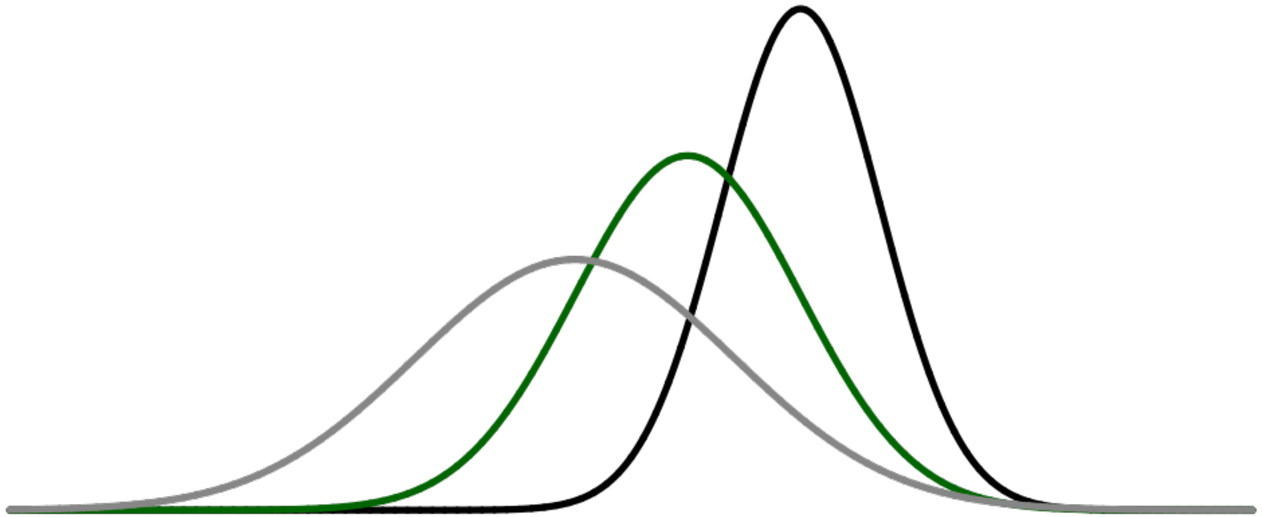
Course Description

Machine learning form the perspective of a statistician who uses R.

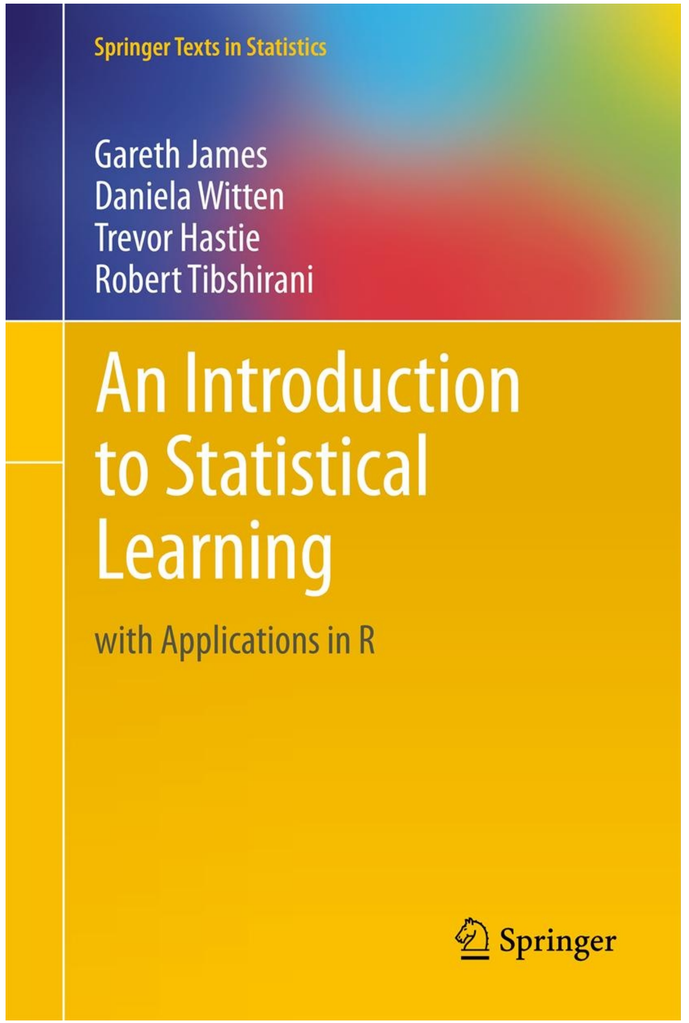
Learning Objectives

After this course, students should be expected to be able to …

- ***identify*** supervised (regression and classification) and unsupervised (clustering) learning problems.
- ***understand*** the fundamental theory behind statistical learning methods.
- ***implement*** learning methods using a statistical computing environment.
- ***formulate*** practical, real-world, problems as statistical learning problems.
- ***evaluate*** effectiveness of learning methods when used as a tool for data analysis.



Basics of Statistical Learning



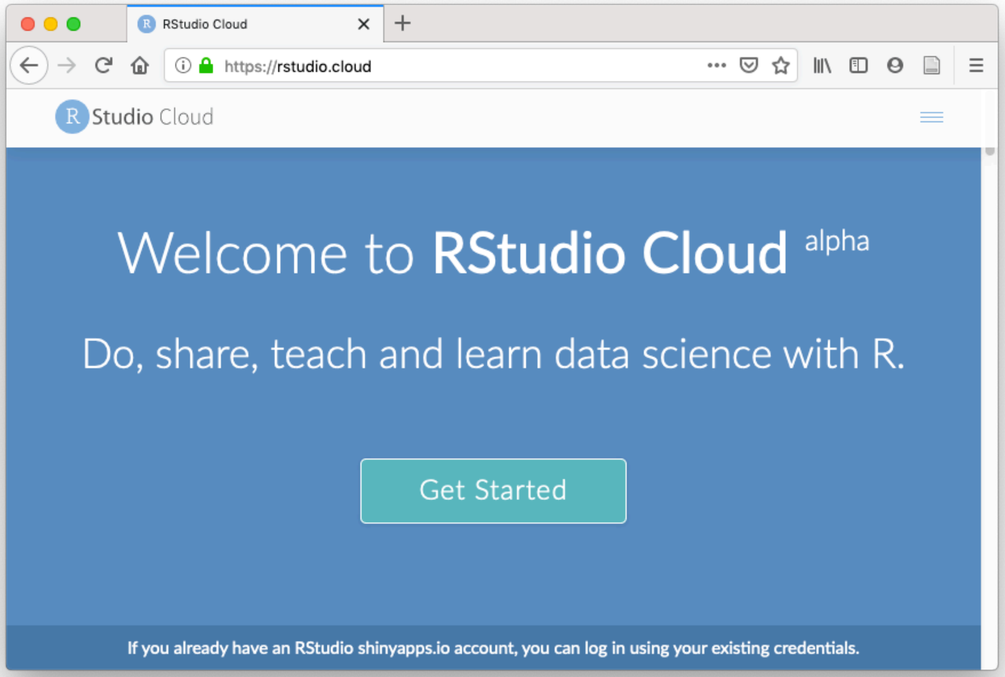
Course Format

- Three lectures per week. (Unimportant?)
    - Sometimes slides, sometimes board notes, sometimes computing.
- (Important!) Things you will do:
        - (Practice) Quizzes on PrairieLearn
        - Exams at the CBTF
        - Data Analyses
        - Projects

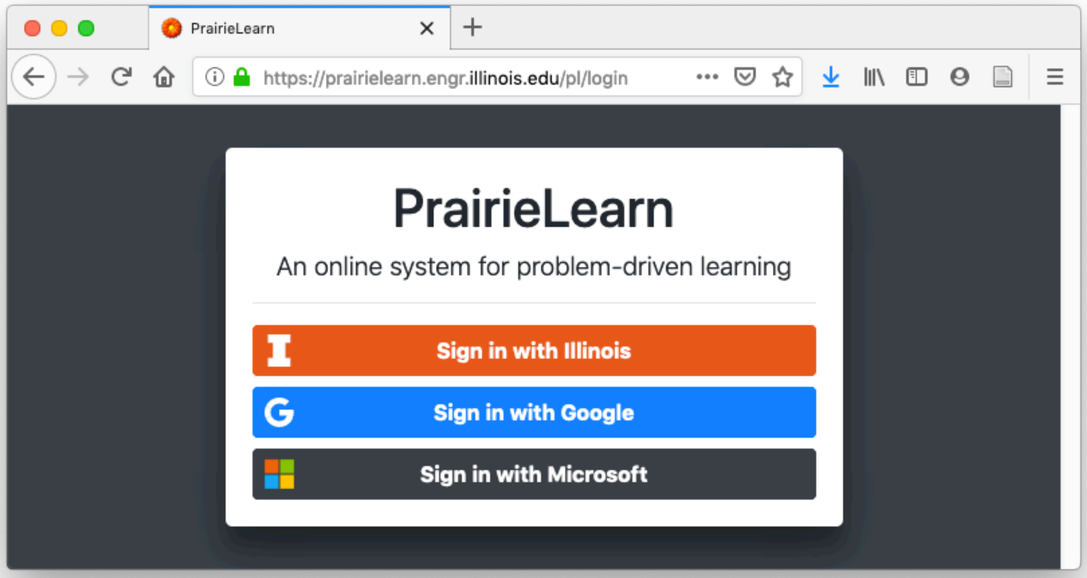
Assessment	Percentage
PrairieLearn Quizzes	20
CBTF Exam I	10
CBTF Exam II	10
CBTF Exam III	20
Practice Data Analyses	10
Data Analyses	10
Group Final Project	15
Graduate Project	5

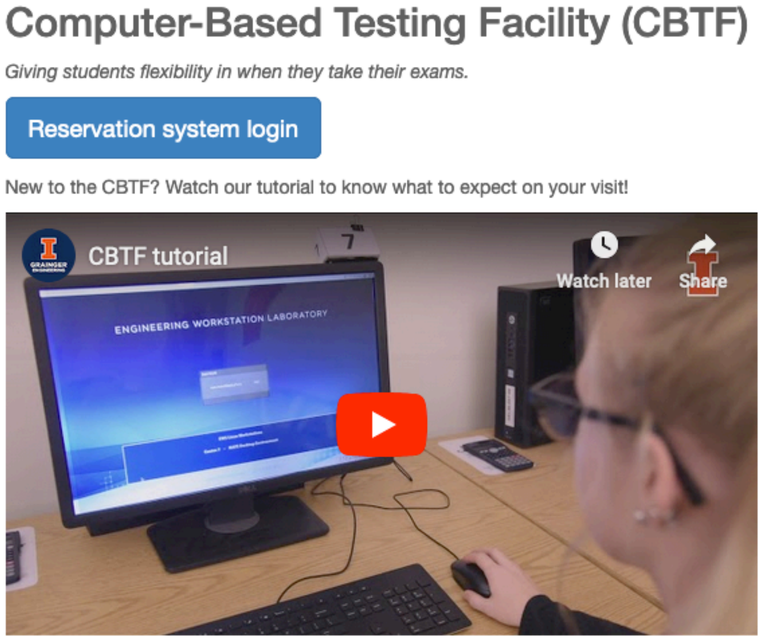
A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-
TBD	93%	90%	87%	83%	80%	77%	73%	70%	67%	63%	60%

Computing Resources

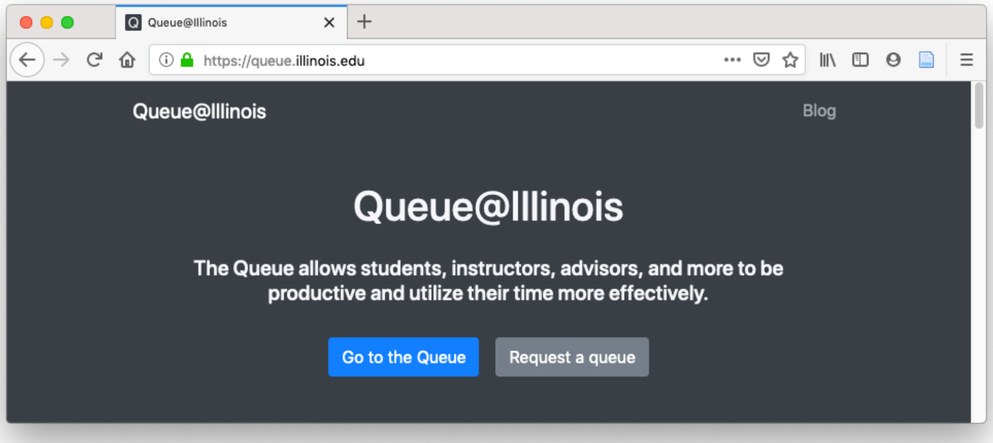


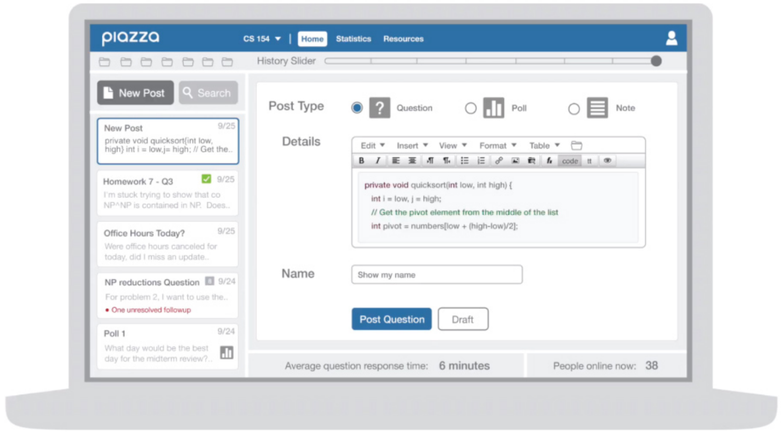
PL and CBTF


PrairieLearn
https://prairielearn.engr.illinois.edu/pl/login
PrairieLearn
An online system for problem-driven learning
Sign in with Illinois
Sign in with Google
Sign in with Microsoft


Computer-Based Testing Facility (CBTF)
Giving students flexibility in when they take their exams.
Reservation system login
New to the CBTF? Watch our tutorial to know what to expect on your visit!
CBTF tutorial
Watch later
Share
ENGINEERING WORKSTATION LABORATORY

Additional Class Technology





Blackboard

- Use @illinois.edu email
- Begin subject with **[STAT 432]**
- Get to the point!
- Probably just use **Piazza**…

Office Hours

Wednesday 4:00 - 7:00

“I don't know who you are. I don't know what you want. If you're looking for ransom, I can tell you I don't have money... but what I do have are a very particular set of skills. Skills I have acquired over a very long career. Skills that make me a nightmare for people like you…”

Not registered?

“I am altering the deal, pray I don’t alter it any further.”

Questions?

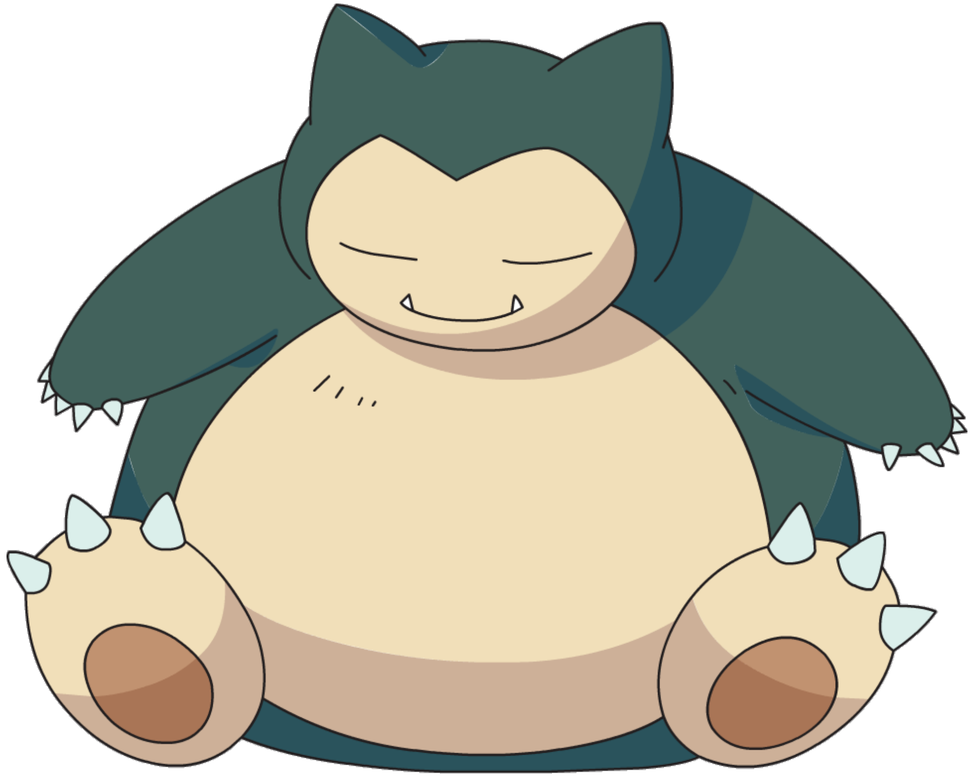
Comments? Concerns?

ML in 5 Minutes

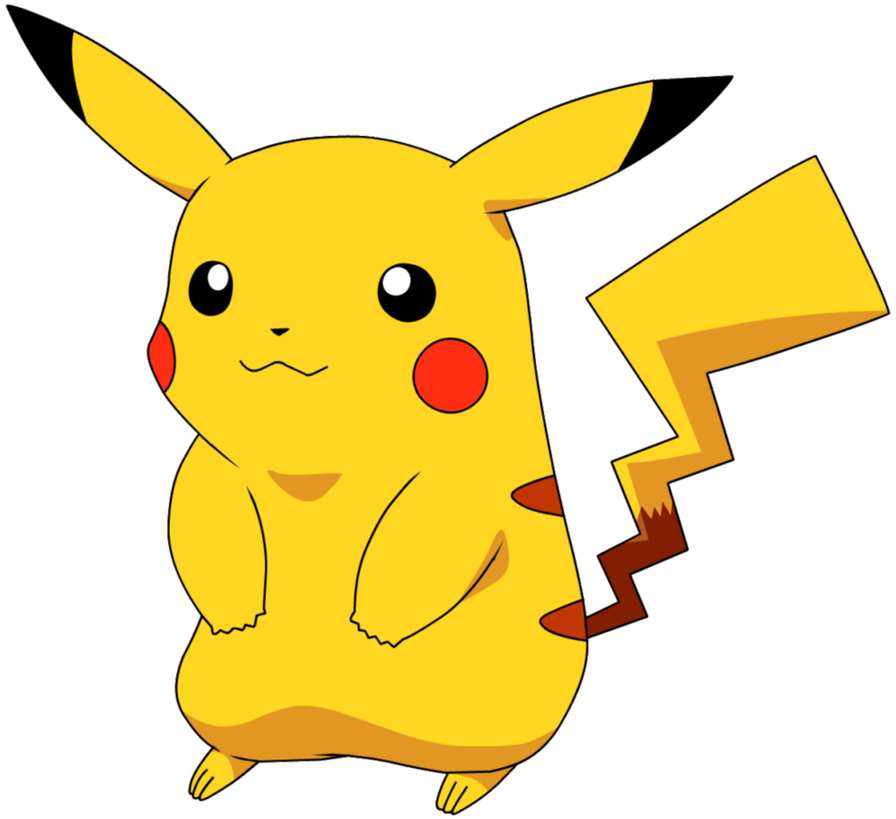
Supervised Learning

Classification

Let’s train *you* to be a classifier…



This is a **Snorlax.**



This is a **Pikachu.**

This is a **Raichu.**

This is a **Snorlax.**

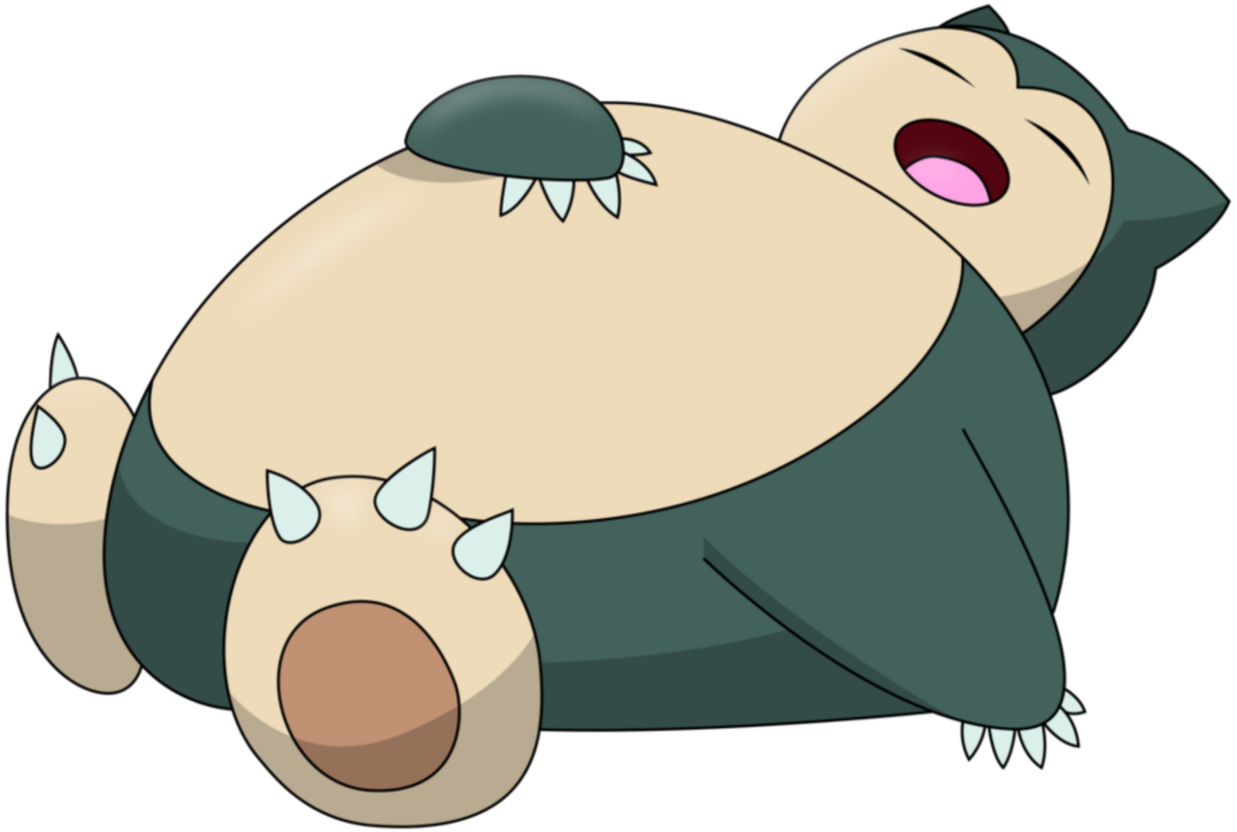
This is a **Raichu.**

This is a **Pikachu.**

Now that you are a classifier,
let's make some **predictions**...

What Pokémon is this?

What Pokémon is this?



What Pokémon is this?

What might the “data” look like?

Class (y)	Color (x1)	Height (x2)	Weight (x3)	Type (x4)
Pikachu	Yellow	0.4 m	6.0 kg	Electric
Snorlax	Blue	2.1 m	460.0 kg	Normal
Raichu	Orange	0.8 m	30.0 kg	Electric
…	…	…	…	…

A non-exhaustive list of questions…

- How would you go from an image to a data frame?
- Which predictors should we use in our model?
- How do we model the response as a function of the predictors?
- How to we use our model to make predictions?
- How do we know if our model is working well?
- Who cares?

Supervised Learning

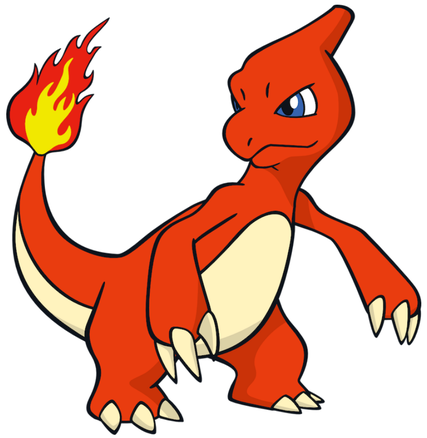
Regression

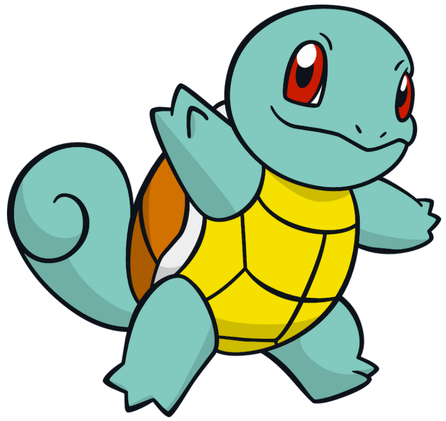
It’s pretty much the same as classification except you’re predicting a **number** instead of a category.

Unsupervised Learning

Clustering

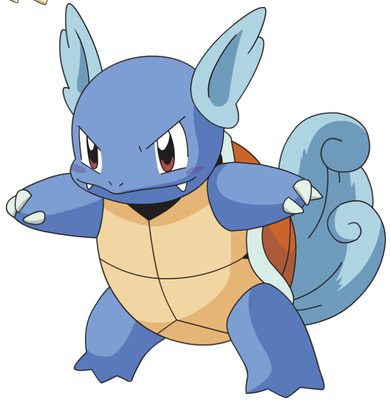
Can you “group” these Pokémon?

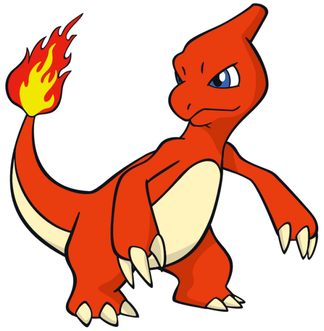


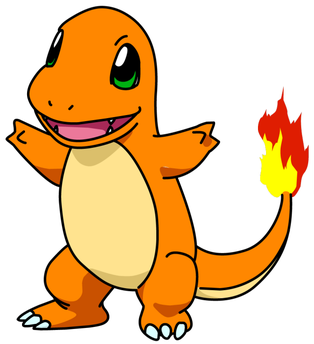


Maybe like this?

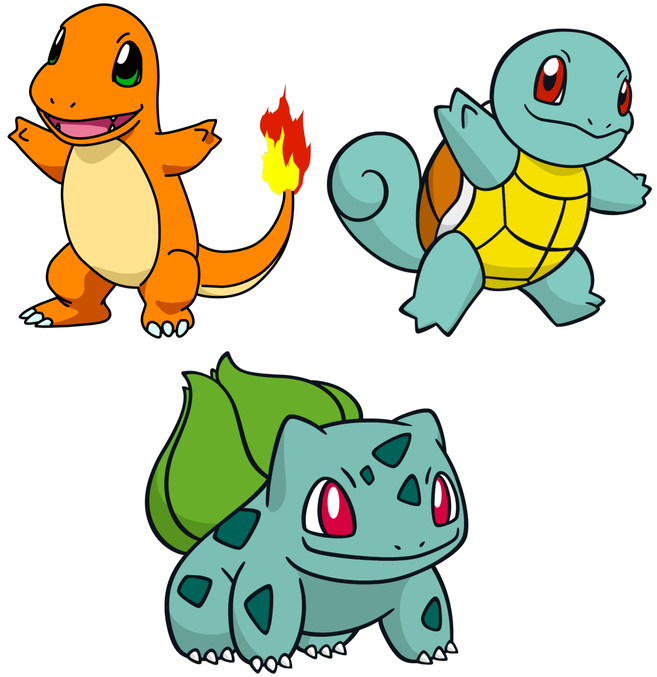
How about like this?







Why not like this?



An non-exhaustive list of questions…

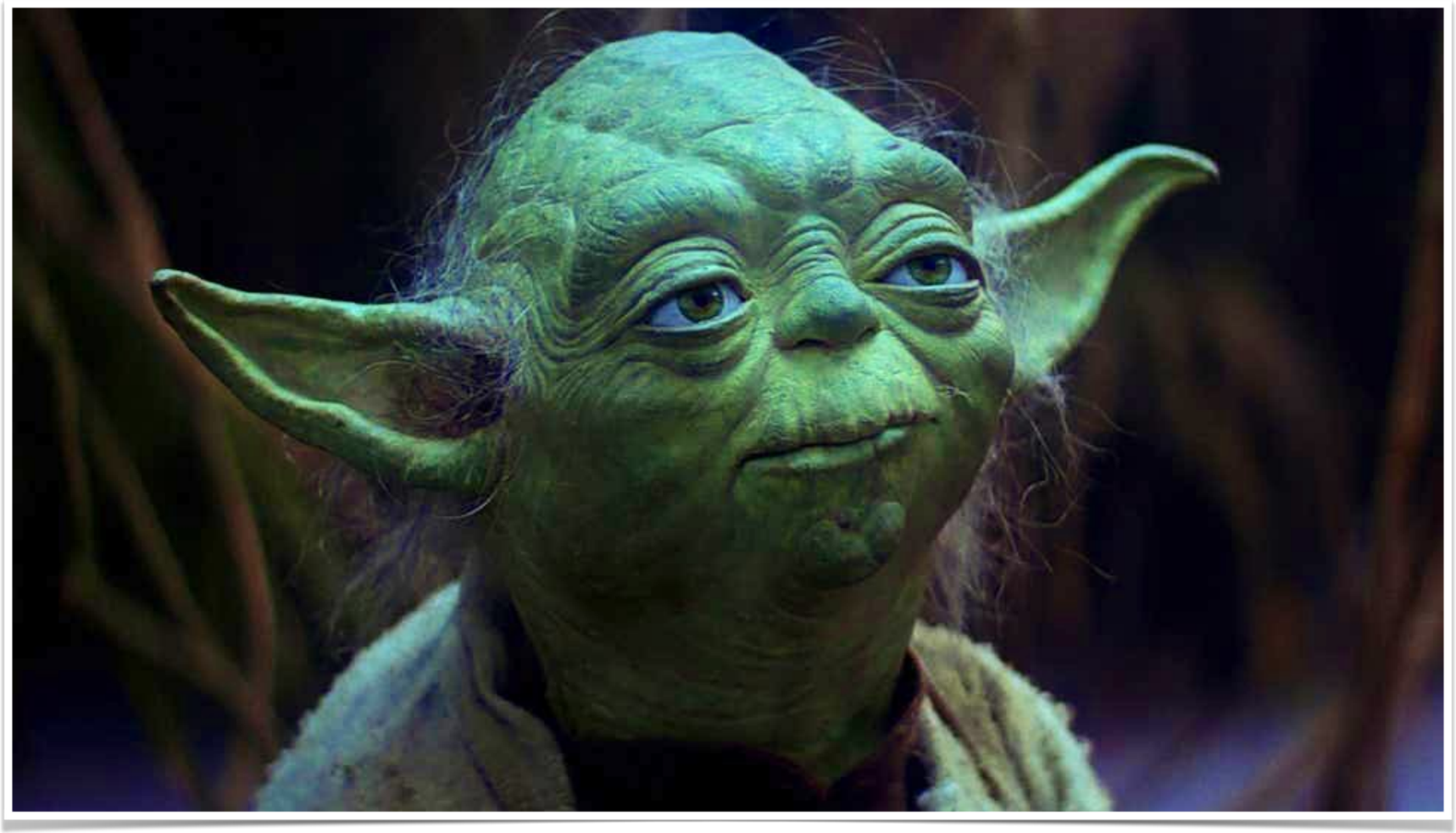
- How do you measure the similarity between observations?
- How many groups should there be?
- How do you assign observations to groups?
- Who cares?

The Extended Syllabus

At the end of the course, I hope that students feel they are…

- A better *statistician*.
- A better *programmer*.
- A better *learner*.

$$\textbf{grade} = f(\textbf{prior knowledge}, \textbf{effort}, \textbf{luck})$$



“You must unlearn what you have learned.”

Things I sort of wish you didn’t know about:

- R-Squared
- Leverage
- Cook’s Distance
- Variance Inflation Factors
- P-Values???

Things I would be happy to never see or talk about in this course.

- MSE as a model metric. (Hint: use RMSE. MSE is appropriate in theoretical discussions.)
- Removing outliers based on leverage or Cook's distance.
- Removing predictors to reduce variance inflation factors.
- Calling a standard error a standard deviation or vice versa.
- Model selection based on p-values of individual coefficients.
- R-Squared.
- Causality. (Unless you're really sure you should. Hint: you shouldn't.)
- SAS. (Feel free to bug me about Python though…)
- Mixing assignment operators. (Or poorly styled code in general.)
- Using ASRM instead of STAT. (There are eight sections of this course because of this…)

Facts versus Opinions

Data Science
Big Data
Deep Learning
Predictive Analytics
Artificial Intelligence

Machine Learning

“Won’t you be my neighbor?”

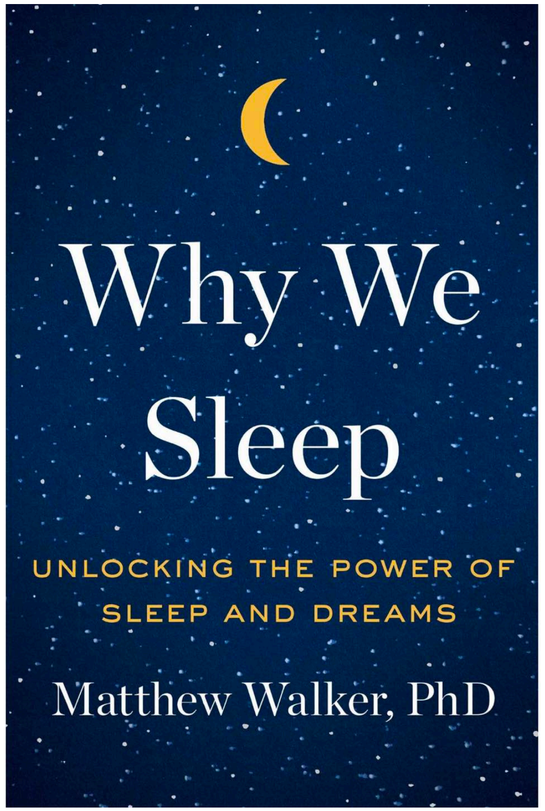
“There are known, knowns…”

“Show up, don’t quit, ask questions.”

–Dan John

Student Health

- Diet
- Exercise
- Sleep



Expectations?

Feedback?

Questions?

Comments? Concerns?

Homework

- Bookmark the course website.
- **Read the full syllabus!!!**
- Read the extended syllabus.
- Register for course on **PrairieLearn**.
- Register for course on **Piazza**.
- Register for the **CBTF** Syllabus Exam.
- Register for course on **RStudio Cloud**?
    - We'll walk through this next time.


That's all Folks!