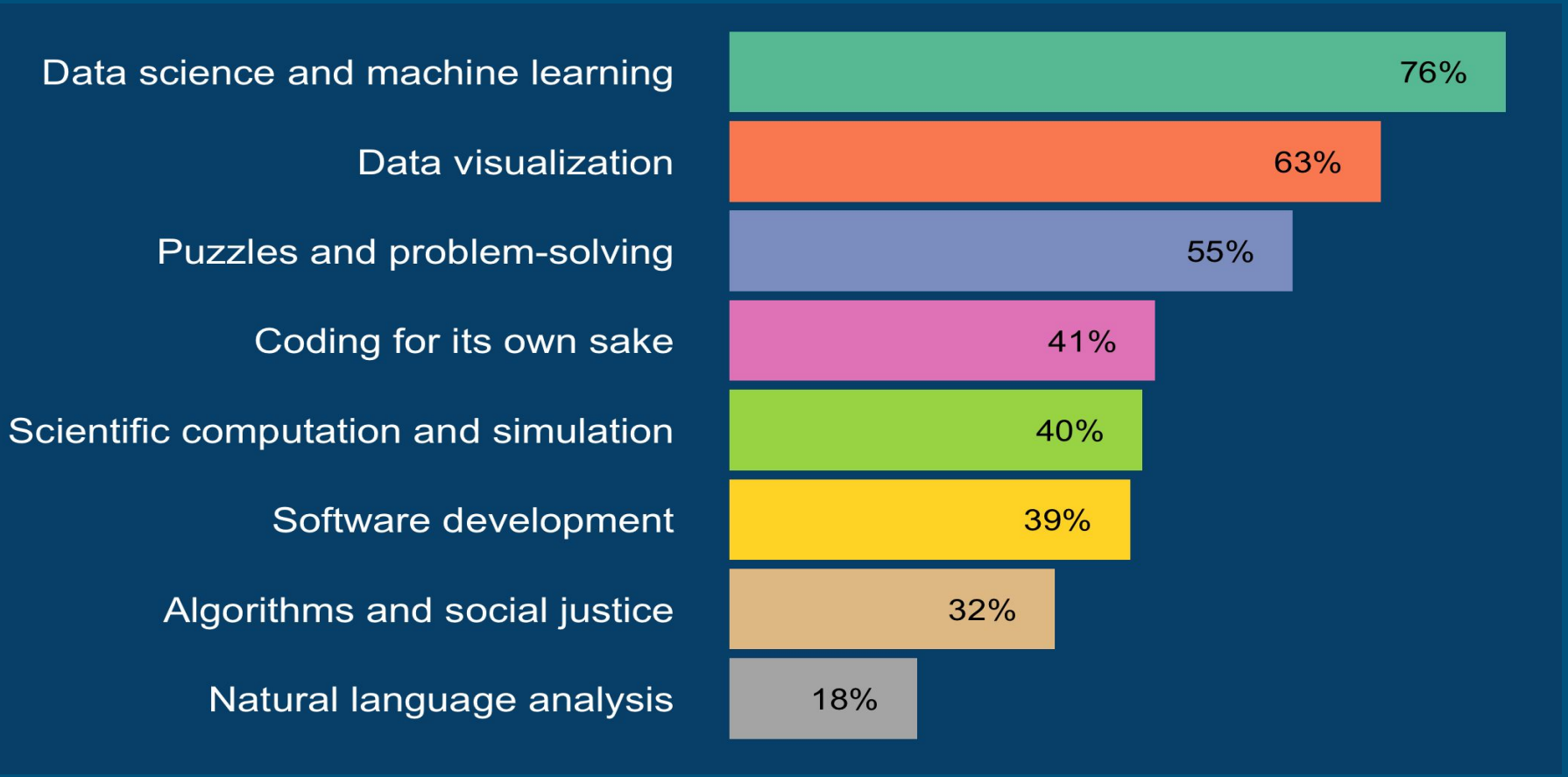
PIC16A Project Overview

Prof. Michael Perlmutter
UCLA Department of Mathematics

How we're feeling in general

How we're feeling about PIC16A

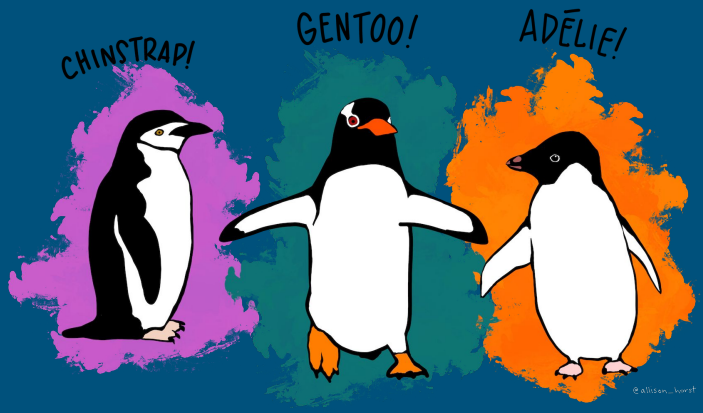
Your Interests



Learning Objectives

1. **Design** and **conduct** a real data analysis in Python.
2. Clearly and concisely **communicate** your findings at each stage.
3. **Make collaborative decisions** about approaches, interpretations, and significance.

Project Data Set



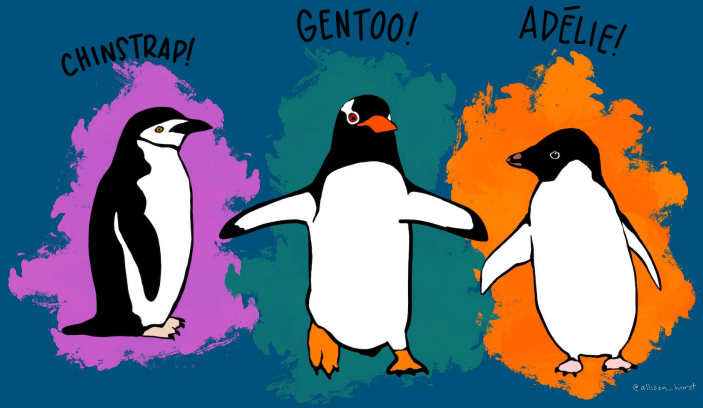
Artwork by @allisonhorst

Palmer Penguins

Physiological measurements on 3 species of penguins collected by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER.

We will be using this data in many lectures and discussions in the coming weeks.

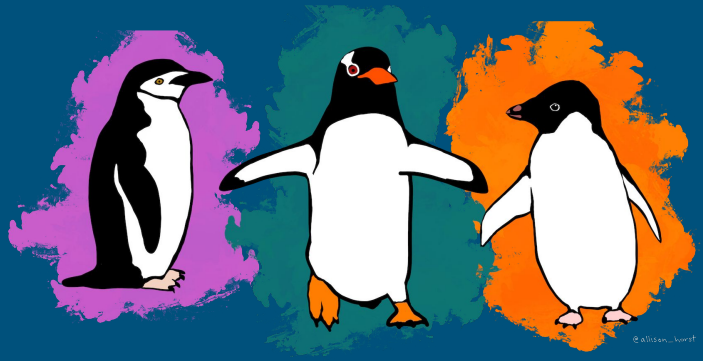
Project Data Set



Artwork by @allisonhorst

Species	Region	Island	Stage	Individual ID	Clutch Completion	Date Egg	Culmen Length (mm)	Culmen Depth (mm)	Flipper Length (mm)	Body Mass (g)	Sex	Delta 15 N (o/oo)	Delta 13 C (o/oo)	Comments
Adelie Penguin (Pygoscelis adeliae)	Anvers	Torgersen	Adult, 1 Egg Stage	N1A1	Yes	11/11/07	39.1	18.7	181.0	3750.0	MALE	NaN	NaN	Not enough blood for isotopes.
Adelie Penguin (Pygoscelis adeliae)	Anvers	Torgersen	Adult, 1 Egg Stage	N1A2	Yes	11/11/07	39.5	17.4	186.0	3800.0	FEMALE	8.94956	-24.69454	NaN
Adelie Penguin (Pygoscelis adeliae)	Anvers	Torgersen	Adult, 1 Egg Stage	N2A1	Yes	11/16/07	40.3	18.0	195.0	3250.0	FEMALE	8.36821	-25.33302	NaN
Adelie Penguin (Pygoscelis adeliae)	Anvers	Torgersen	Adult, 1 Egg Stage	N2A2	Yes	11/16/07	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Adult not sampled.
Adelie Penguin (Pygoscelis adeliae)	Anvers	Torgersen	Adult, 1 Egg Stage	N3A1	Yes	11/16/07	36.7	19.3	193.0	3450.0	FEMALE	8.76651	-25.32426	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Gentoo penguin (Pygoscelis papua)	Anvers	Biscoe	Adult, 1 Egg Stage	N38A2	No	12/1/09	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Gentoo penguin (Pygoscelis papua)	Anvers	Biscoe	Adult, 1 Egg Stage	N39A1	Yes	11/22/09	46.8	14.3	215.0	4850.0	FEMALE	8.41151	-26.13832	NaN
Gentoo penguin (Pygoscelis papua)	Anvers	Biscoe	Adult, 1 Egg Stage	N39A2	Yes	11/22/09	50.4	15.7	222.0	5750.0	MALE	8.30166	-26.04117	NaN
Gentoo penguin (Pygoscelis papua)	Anvers	Biscoe	Adult, 1 Egg Stage	N43A1	Yes	11/22/09	45.2	14.8	212.0	5200.0	FEMALE	8.24246	-26.11969	NaN
Gentoo penguin (Pygoscelis papua)	Anvers	Biscoe	Adult, 1 Egg Stage	N43A2	Yes	11/22/09	49.9	16.1	213.0	5400.0	MALE	8.36390	-26.15531	NaN

Your Task



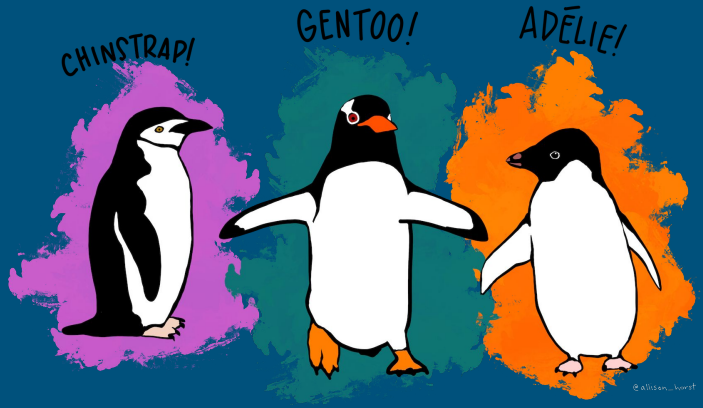
Artwork by @allisonhorst

Identifying the species of a penguin usually requires biological expertise. Taking measurements is difficult and can be stressful to the penguin.

*Can you build a machine learning model that can accurately **predict the species of a penguin** based on a **small number of measurements**?*

Project Outline + Timeline

Required Components



Artwork by @allisonhorst

Exploratory data analysis

- At least 3 figures.
- At least 1 displayed table.

Modeling

- Multinomial logistic regression
- Decision trees
- One more (your choice)

Writing

- Explanation + discussion
- Comments + docstrings

Rough Timeline

W6	W7	W8	W9	W10
Introduce data visualization Introduce data frames (pandas)	Start machine learning	More machine learning	⇦ Needed lecture topics *(note: there are still new topics in W9-10!)*	
Group Work ⇨	**Discussion activities to get you started.** **Homework problem(s).**	**Discussion activity.** **Homework problem(s).**	Work with group to build out project.	Finish project with group. **Project due Friday, 3/12 (group submission).**

Group Work

Project is graded as a group.

Include a short *Group Contributions Statement* describing how each group member supported the project.

Your group is responsible for finding time outside of class to work together.

Equitable delegation is fine.

Expected Effort

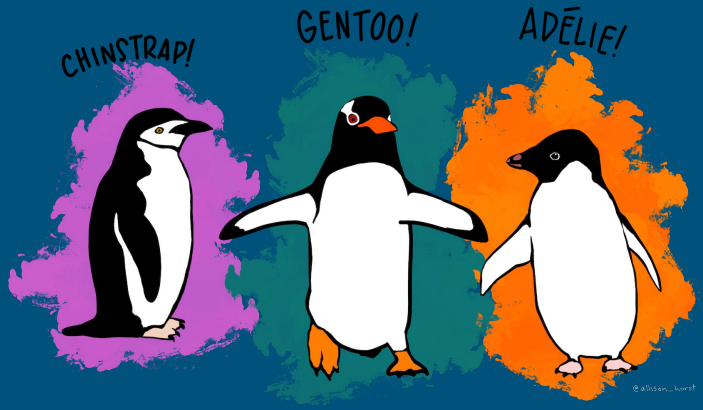
Roughly 10 hours per group member:

- 3 hours in scheduled Discussion.
- ~3-4 hours in homework problems (solo or with group).
- ~3-4 additional hours (coordinate with group).

Grading will account for how many people are in your group.

Project Expectations

Exploratory Data Analysis



Artwork by @allisonhorst

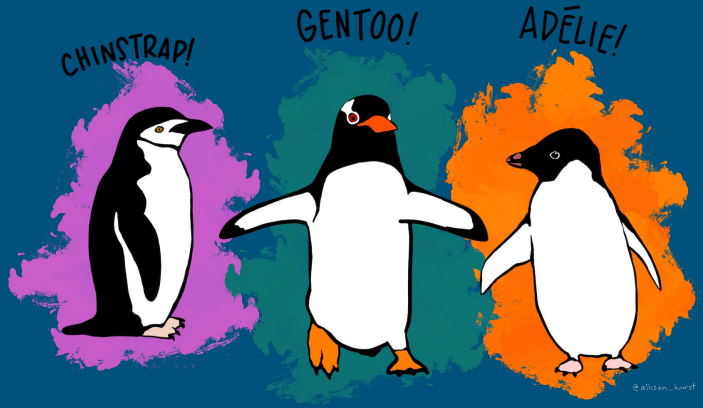
3 figures, 1 table

Your figures and tables should support future decisions made in your modeling.

Your **writing** should make clear what you have learned from your figures and tables, and how it informs your modeling choices down the line.

Make it clear: what's "the point" of each figure?

Modeling Goals



Artwork by @allisonhorst

Find a model and a small set of measurements with good predictive performance.

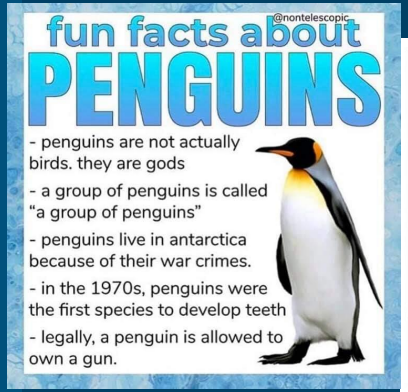
- Good prediction with **three or fewer** features → full credit.
- Perfect prediction with the complete data set → partial credit.

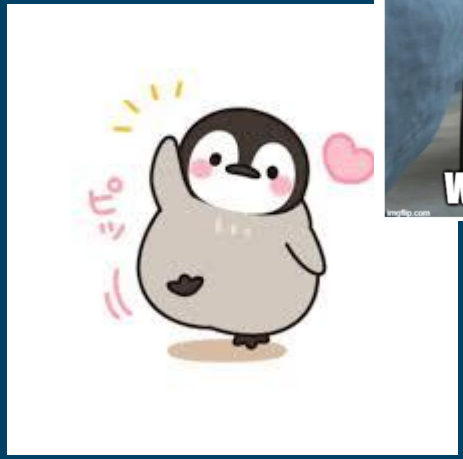
Point Values

- (20 points): Clear and equitable Group Contributions Statement.
- (20 points): Writing, documentation, and style.
- (20 points): Exploratory data analysis -- **relevant** figures and tables.
- (20 points): Modeling -- 3 models, evaluation, recommendation.
- (20 points): 20x model accuracy on test data w/3 or fewer features.

PIC16A Penguins Gallery

Submissions welcome!





Thanks!

Questions?