

STAT 20 Redesign

Statistics Department, UC Berkeley

Content

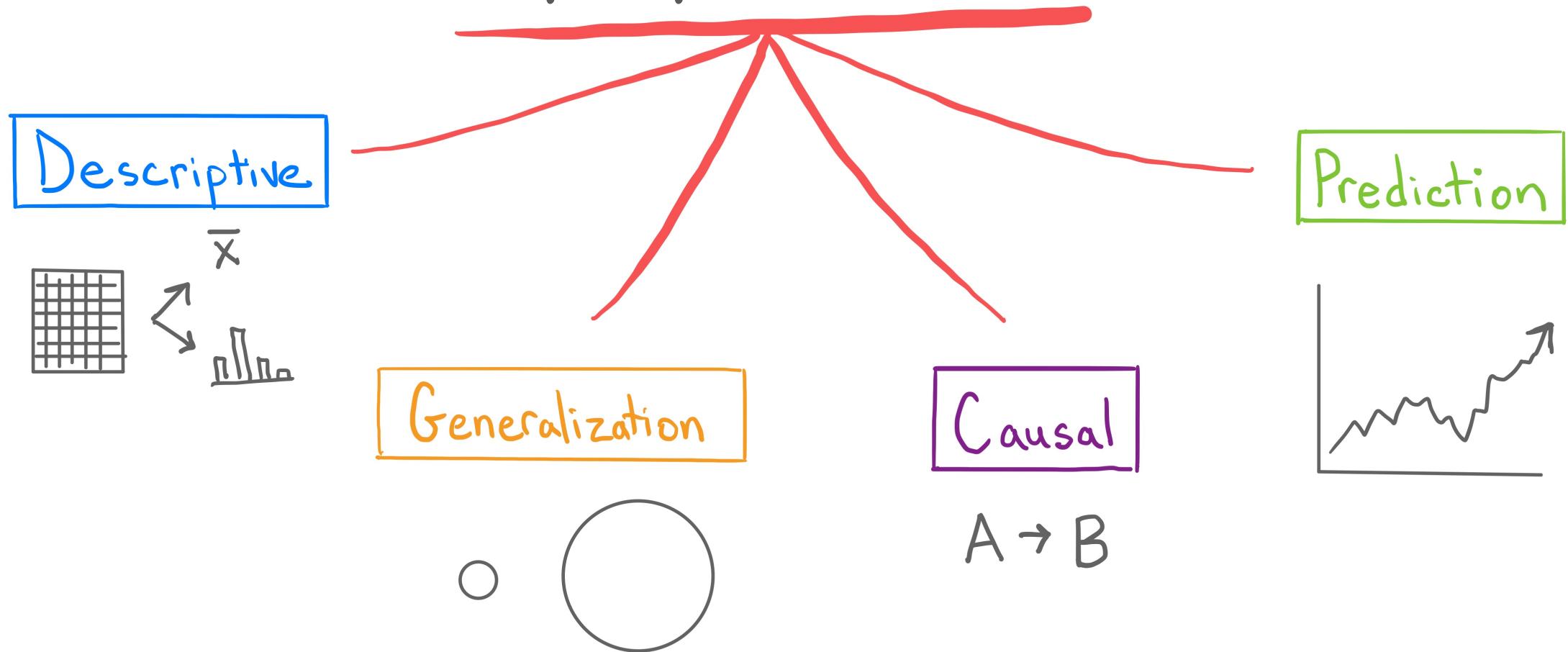
What should a student be able to do after completing their first course in statistics?

STAT 20 2022

- Introductory Statistics course with calculus pre-req
- Fall: 800 students; Spring: 600 students
- Primary audiences: Econ, Business, Statistics

TO LEARN HOW TO CRITIQUE AND CONSTRUCT
CLAIMS MADE USING DATA

TO LEARN HOW TO CRITIQUE AND CONSTRUCT CLAIMS MADE USING DATA



Current focus of development

1. Computing: the Data Frame
2. Data Visualization
3. Multivariate Thinking
4. Tools for Inference: Simulation/resampling
5. Causality

Computing: the Data Frame

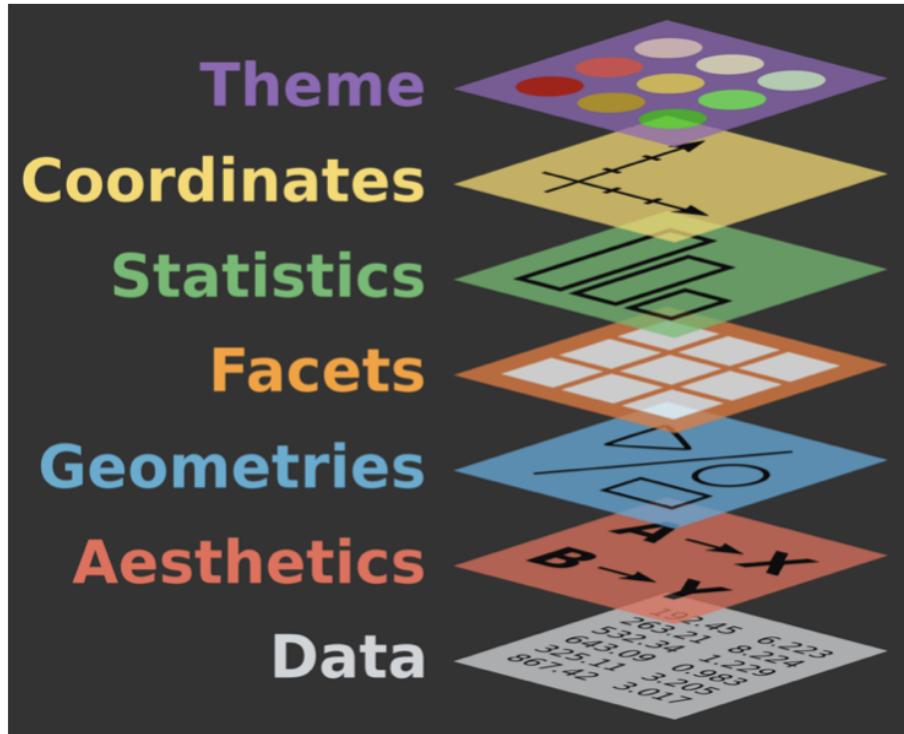
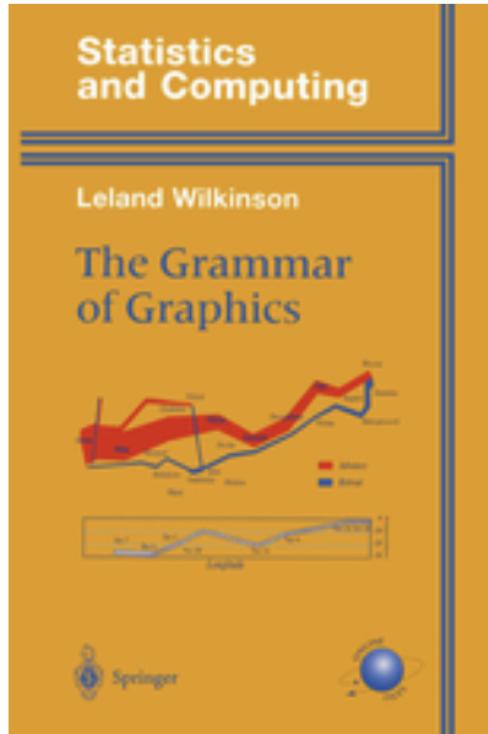
species	island	bill_length_mm	bill_depth_mm
Gentoo	Biscoe	48.4	14.6
Gentoo	Biscoe	44.0	13.6
Adelie	Torgersen	44.1	18.0
Gentoo	Biscoe	49.4	15.8
Adelie	Biscoe	37.9	18.6
Gentoo	Biscoe	44.4	17.3

Goal: Emphasize a recurrent data structure that is useful for understanding the unit of observation and variables.

Data Visualization

Goal: Provide a theory of graphics. More practice.

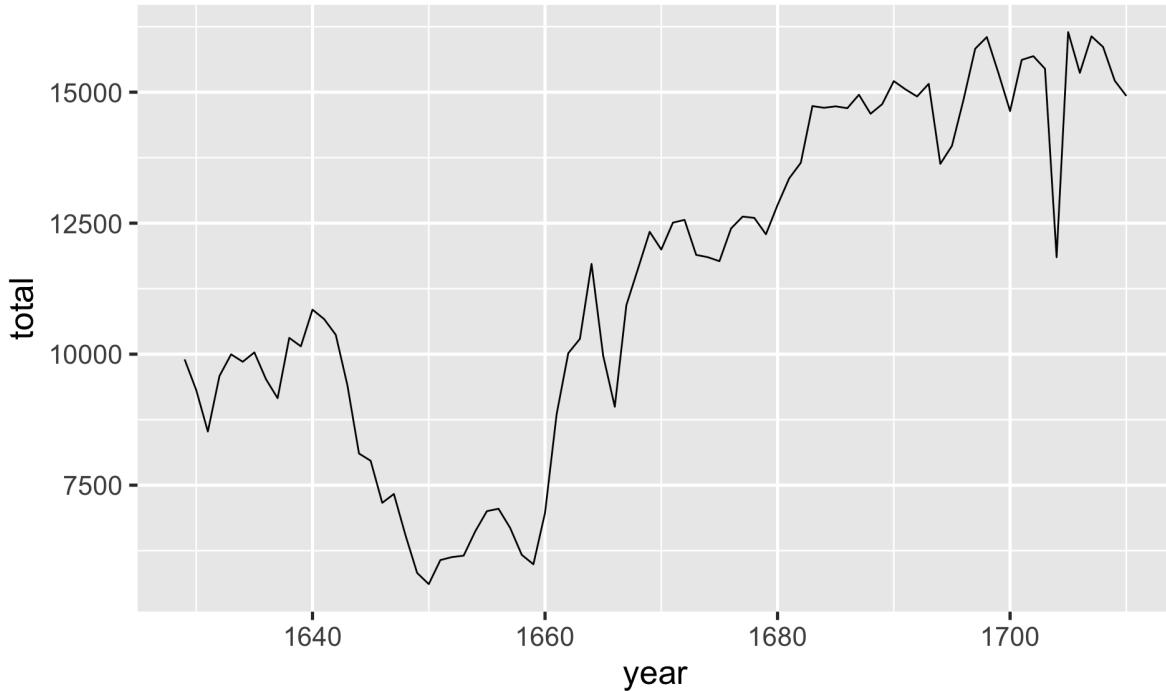
Grammar of Graphics: A statistical graphic is a mapping of *data* variables to *aesthetic* attributes of *geometric* objects.



3 Questions to ask of every plot

1. **Data:** What data frame do I want to depict?
2. **Aesthetics:** How do I want to map the variables of that data frame to the aesthetic attributes of the plot?
3. **Geometry:** How do I want to graphically depict the observations?

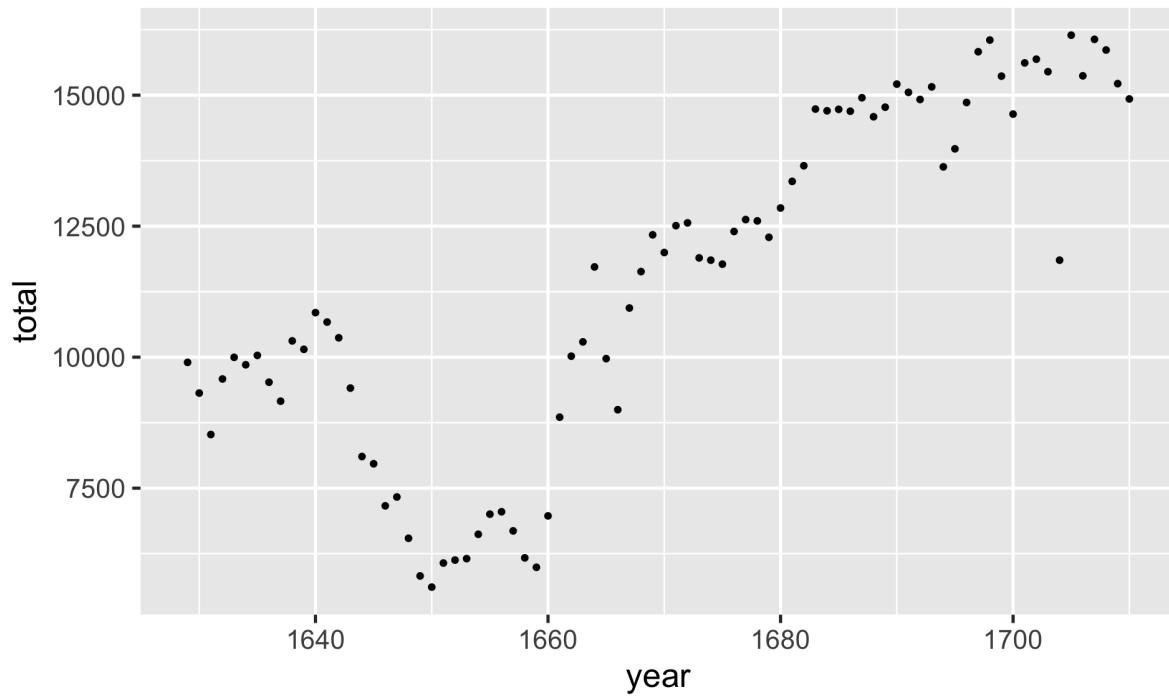
John Arbuthnot's Christening Records



- **Data:** `arbuthnot`
- **Aesthetics:** map `year` to the x-axis, `total` count to the y-axis
- **Geometry:** express observations as a line

```
1 ggplot(arbuthnot,  
2       aes(x = year,  
3              y = total)) +  
4   geom_line()
```

John Arbuthnot's Christening Records



- Data: `arbuthnot`
- Aesthetics: map `year` to the x-axis, `total` count to the y-axis
- Geometry: express observations as `points`

```
1 ggplot(arbuthnot,  
2       aes(x = year,  
3              y = total)) +  
4   geom_point()
```

Multivariate Thinking

Goal: Become accustomed to thinking about the relationships between more than two variables.

Where does this come up?

- Higher dimensional data visualization
- Multiple regression

Tools for Inference

Computational

- Simulation
- Bootstrapping
- Permutation tests

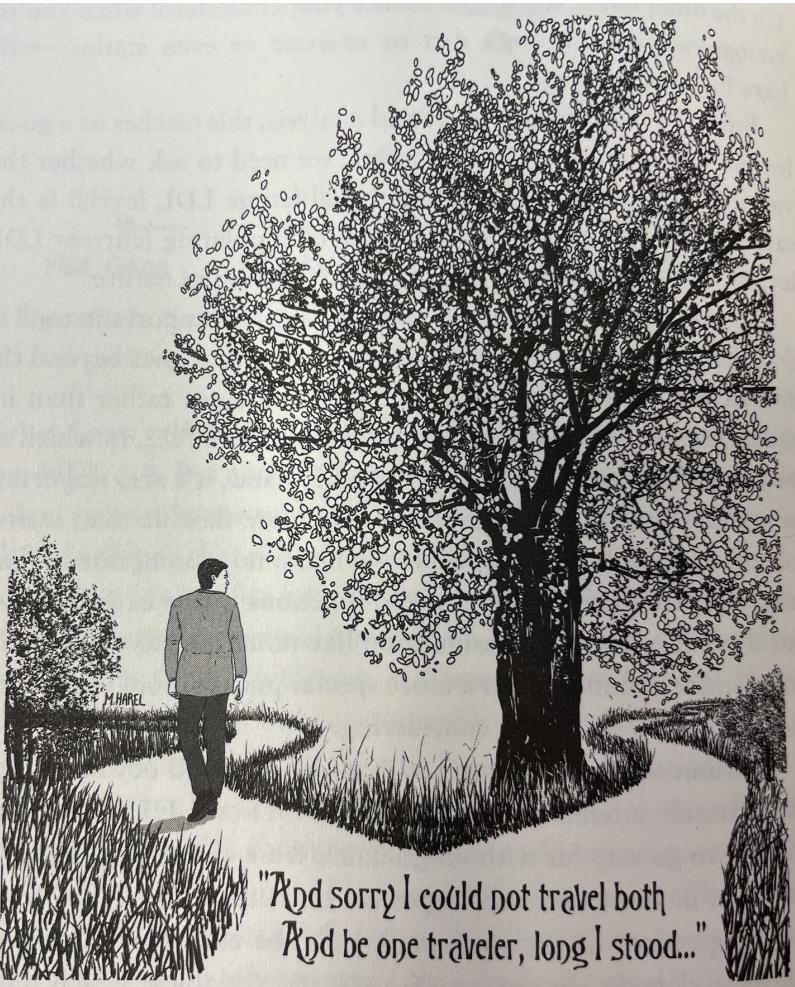
Mathematical

- Binomial
- Normal approximation
- t -test
- χ^2
- F

Goal: Simplify, build intuition, build foundation

Causality

Goal: Approach causality constructively with a definition and different approaches to estimating causal effects.



Counterfactual: relating to or expressing what has not happened or is not the case.

- The outcome of taking the left path.
- The outcome of taking the right path.

Only one of these are observed.

Lab: A Matter of Taste

Students devise an experiment to test a claim *before* being taught experimental design.



- Experiments are assessed based on their clarity, reproducibility, and internal consistency, not on correctness.

Claim and Null Hypothesis

In one sentence, write the *claim* you intend to test. Then write down the corresponding *null hypothesis* that you will gather evidence against and identify the test statistic that you will use.

Claim: By using a straw, the taster's ability to differentiate between flavors would be diminished.

Null: There would be no difference in differentiating the taste between flavors when using a straw.

Protocol Test Stat: The difference in proportion (^{correct guesses}_{to fail})

Record below the step-by-step protocol that you will use to collect data to bear on the claim above. Be precise. You should be able to hand this off to another group and have them conduct your experiment.

- ① Label 4 cups according to flavor in such a way that only the facilitator knows and the taster doesn't.
- ② Fill 4 cups with the 2 different flavors of La Croix (2 cups per flavor).
- ③ Have the taster taste a sample from each of the 4 cups (without a straw). Subject must also eat a saltine cracker between each taste.
- ④ Have subject blindly guess what they believe each flavor is and facilitators will record data.
- ⑤ Repeat steps 3 and 4 but this time using straws.

Data

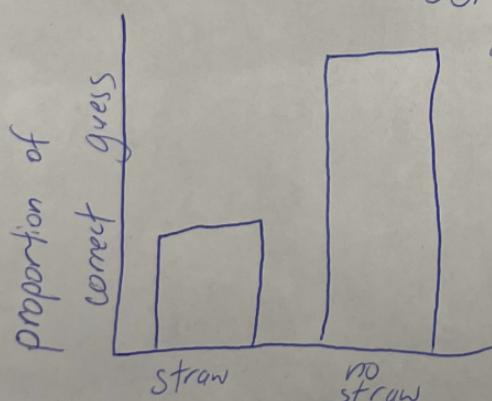
In the space below, sketch out an empty data frame that identifies the important characteristics of your data. Create the data frame such that the number of rows reflect your sample size, n , and the number of columns represent the variables that you intend to collect. Label the columns with the names of those variables and note how you are defining your unit of observation.

Trial	Correct_guess	using_straw	Unit of observation: individual sips

Exploratory Data Analysis

Sketch a plot that you will use to use to determine how consistent your data are with your original claim. Please label the axes. Since you do not yet have data, create two sketch of what the plot *might* look like in two different scenarios. Below each plot, write out the ggplot2 code that will create that graphic.

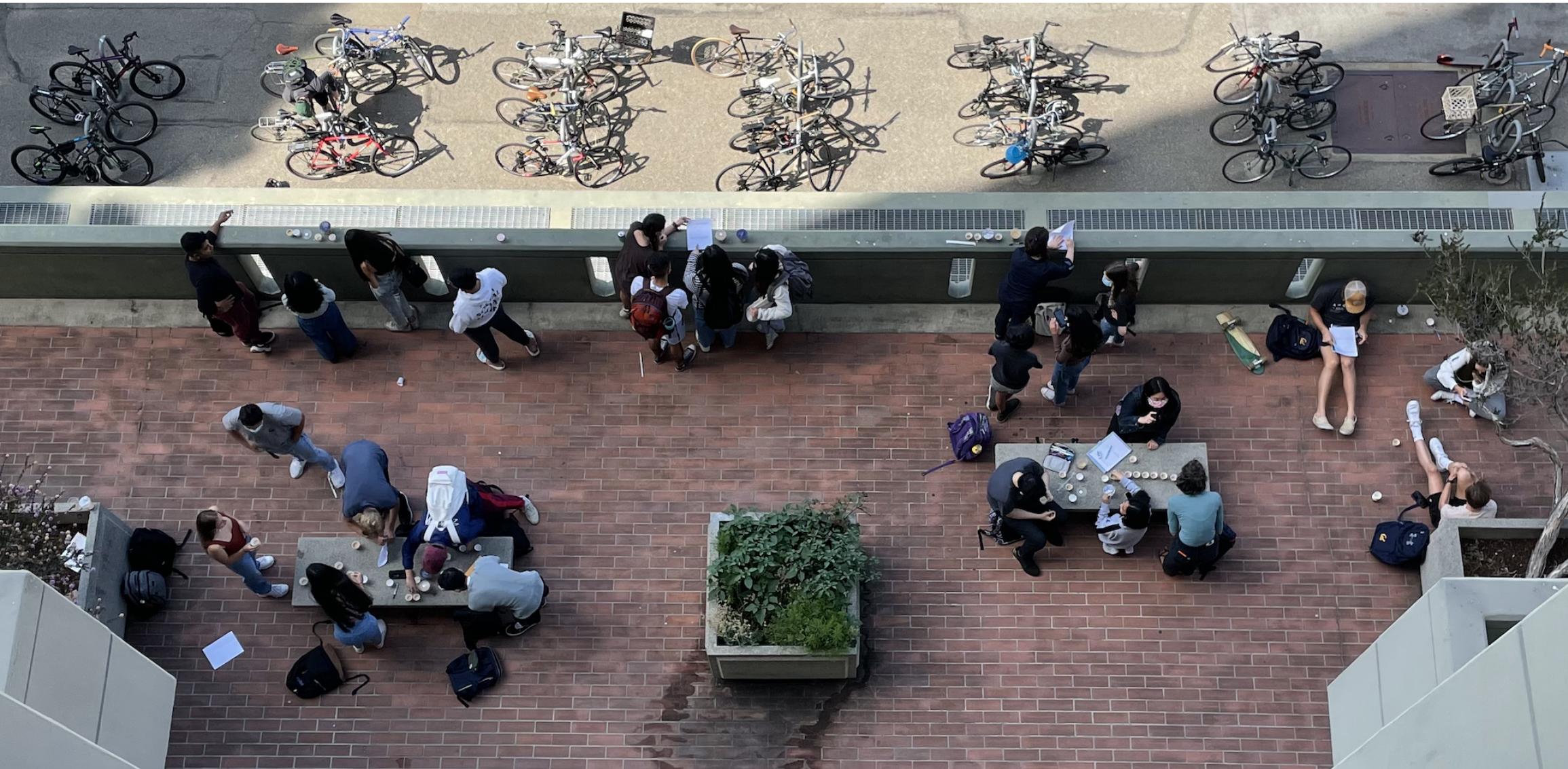
Scenario 1: Data supports claim



Scenario 2: Data does not support claim

```
df %>%  
  group_by(using_straw) %>%  
  summarise(ave_g = mean(correct_guess)) %>%  
  ggplot(aes(x = using_straw, y = ave_g)) +  
  geom_box()
```





treatments.

Why do an Experiments?

- More control
- Establish causation



Why do an Obs. Study?

- Ethical
- Studies "natural" behavior
- Can study historical data
- Easier/cheaper

cup	true flavor	guess	correct
1	R	R	T
2	D	P	F
3	D	P	T
-	:	:	:
1	R	R	T
2	D	P	F
3	D	P	T
-	:	:	:
1	R	R	T
2	D	P	F
3	D	P	T
-	:	:	:

trial	flavor	guess	T or F
1	1	1	T
2	2	1	F
3	2	2	T
4	2	1	F
5	1	2	F
-	1	2	F
20			

Reinforcement Data Analysis

Below is a graph you can use to determine how accurate your rats are with their predictions.

Below is a graph you can use to determine how accurate your rats are with their predictions.

Source: J. Dohm response time

Source: J. Dohm response time



Current focus of development

1. Computing: the Data Frame
2. Data Visualization
3. Multivariate Thinking
4. Tools for Inference: Simulation/resampling
5. Causality

Course Design

Alternative Design

Split the massive lecture course into many parallel courses in a flipped structure with a shared curriculum.

Motivation

1. Construct an effective learning environments for students
2. Allow curriculum to innovate while being robust to staffing changes
3. Keep the instructor focus on teaching

Attendance

Average lecture attendance across Statistics undergrad courses:

27%

- The proportion of enrolled student attending live lecture is lower (large classes have lower attendance).
- Discussion attendance is lower.

Course components

- Staffing
- Scheduling
- Spaces

Model 1

Class size: 600 students

- 1 Instructor
 - Lecture to 600



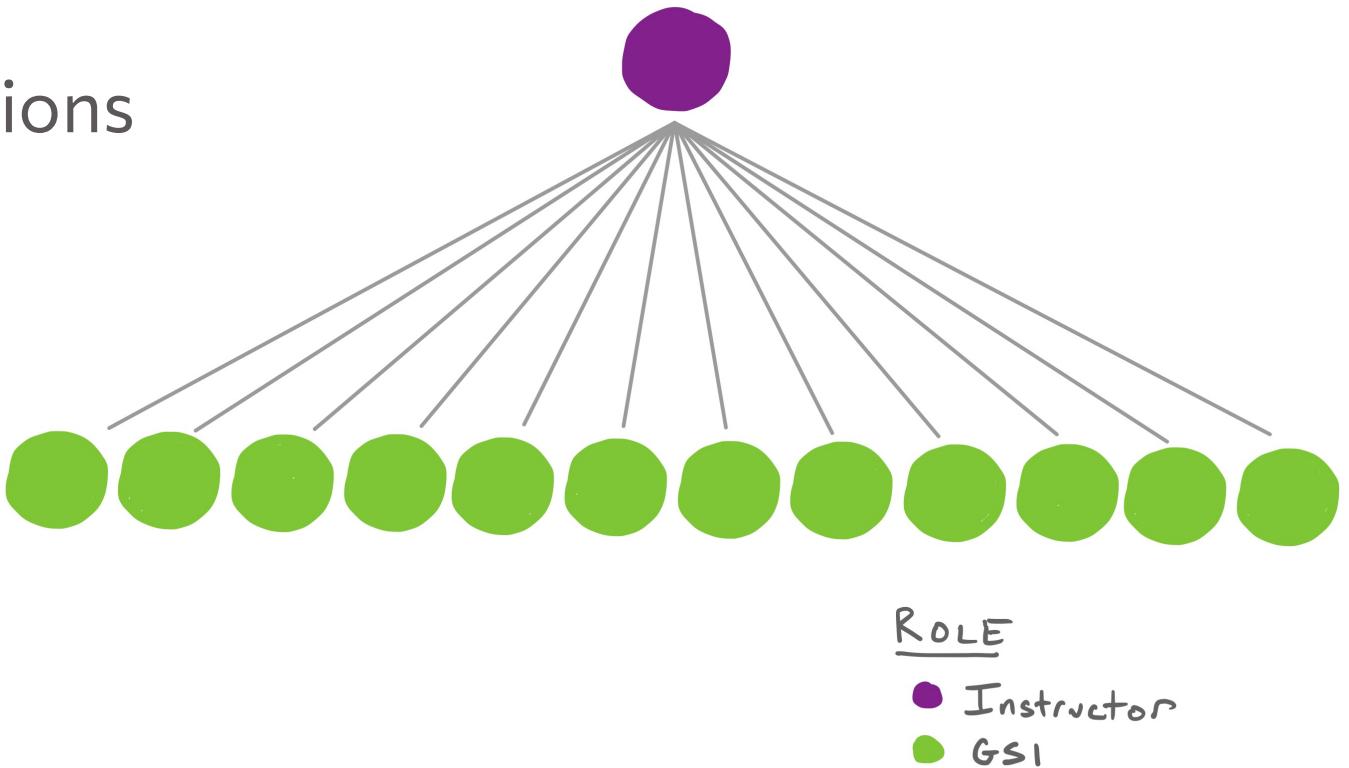
ROLE

- Instructor

Model 1

Class size: 600 students

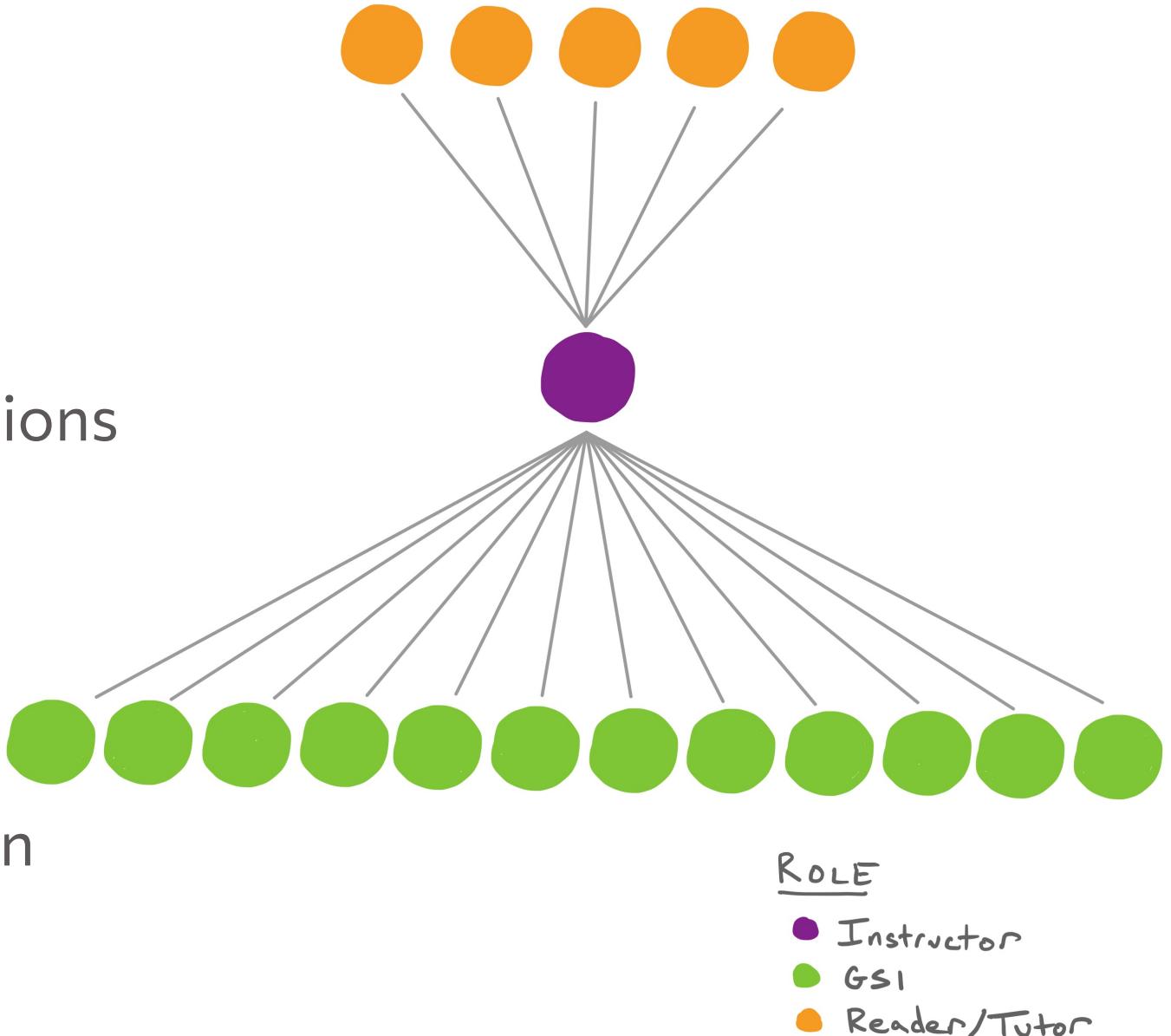
- 1 Instructor
 - Lecture to 600
- 12 GSIs
 - Teach labs to 2 sections



Model 1

Class size: 600 students

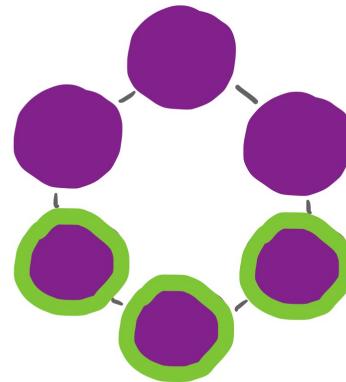
- 1 Instructor
 - Lecture to 600
- 12 GSIs
 - Teach labs to 2 sections
 - Grade
- 5 Tutors
 - Grade
 - Run tutoring session



Model 2

Class size: 600 students

- 6 Instructors
 - 100 students each
 - Faculty + GSI



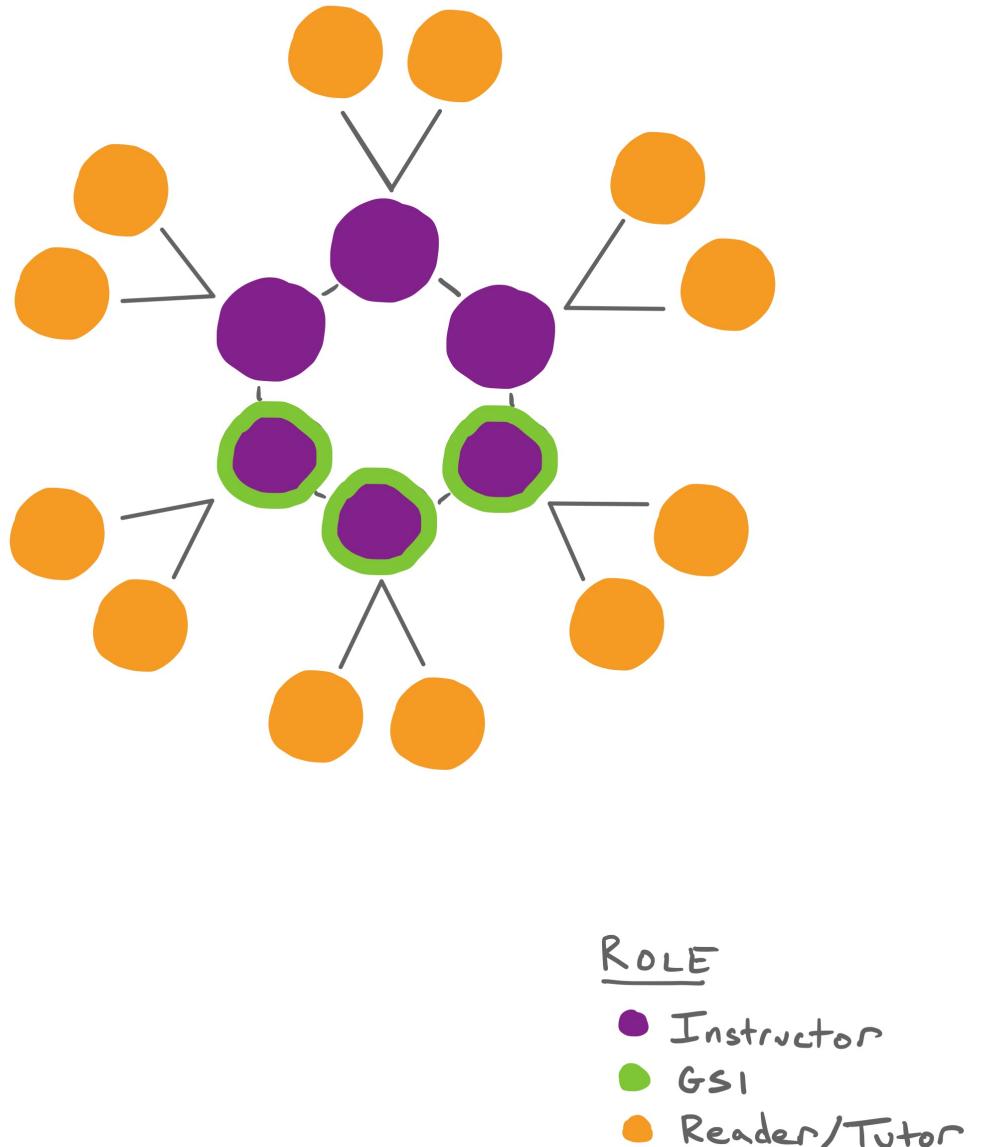
ROLE

- Instructor
- GSI
- Reader/Tutor

Model 2

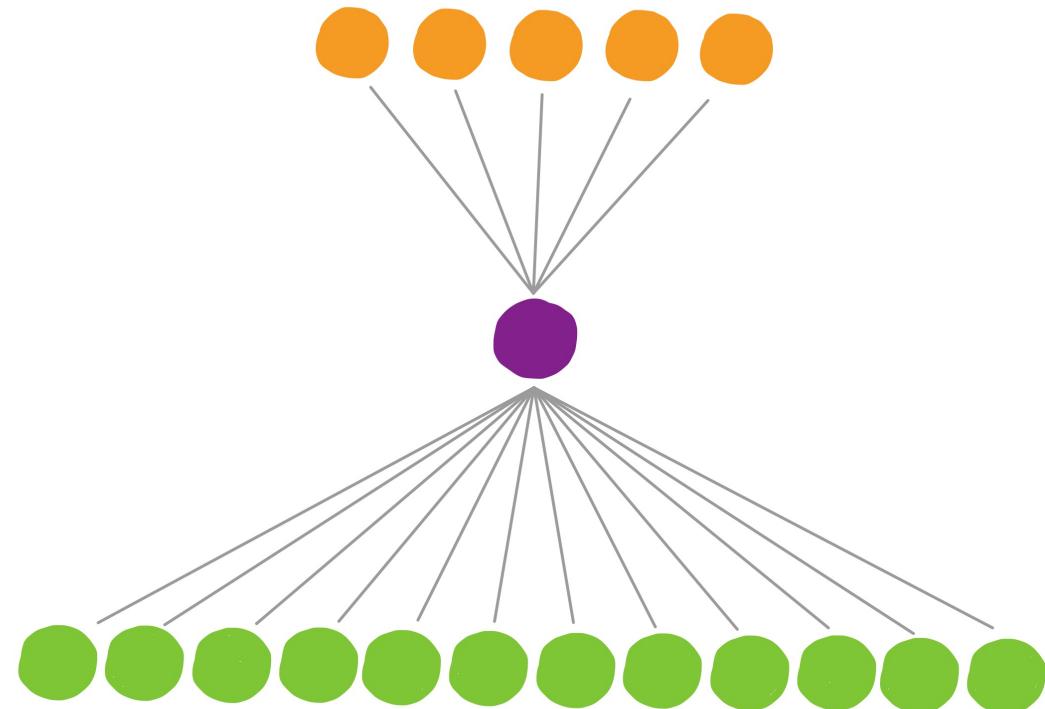
Class size: 600 students

- 6 Instructors
 - 100 students each
 - Faculty + GSI
- 12 Assistants
 - Grade
 - Assist during class



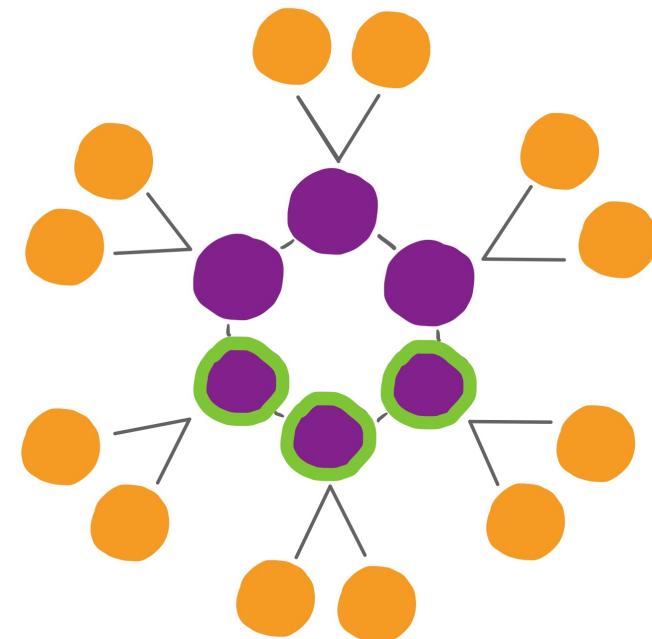
Who are the Instructors?

- Senate faculty? Lecturers?
- Our Ph.D.s? Others?



Is this financially doable?

- 20 hr GSI: \$22k
- 10 hr Reader/Tutor: \$2.2K



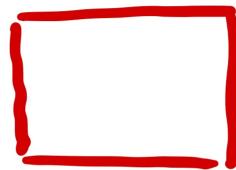
ROLE

- Instructor
- GSI
- Reader/Tutor

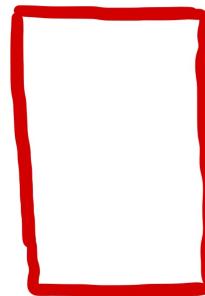
Schedule

Week 1

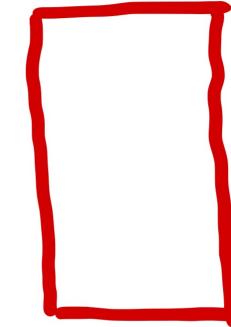
M



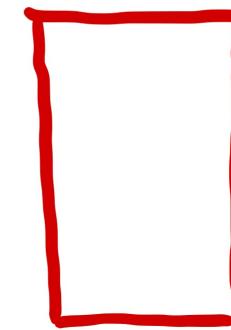
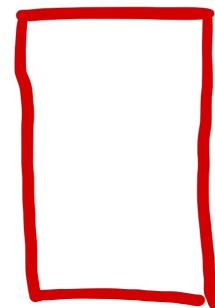
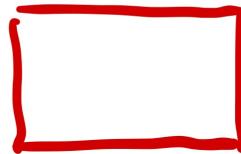
W



F

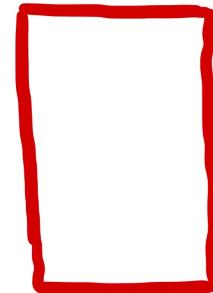


Week 2

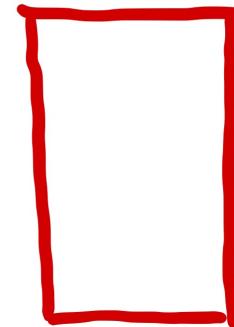
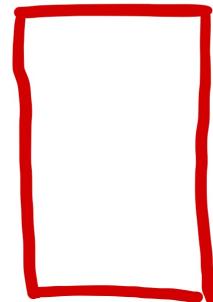
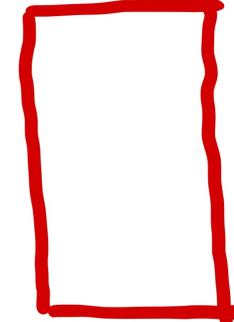


- 2 hrs
- ~100 students
- Instructor + Assistants
- Blended format
 - Worksheets / notebooks
 - Activities
 - Mini-lectures

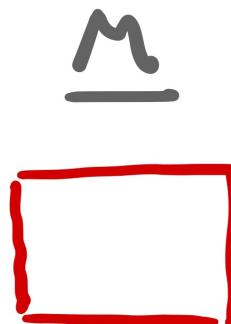
w



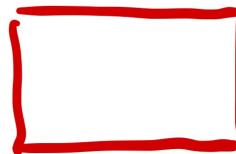
F



Week 1



Week 2



- 1 hr

Week 1

- ~100 students
- Assistants
- In-class quiz

Week 2

- ~600 students
- Instructor(s)
- Plenary lecture in evening

Spaces



Models

- TEAL classes at MIT
- PS 3, UC Berkeley, David Broockman

Feedback

What characteristics in a course would make it more appealing to teach?