

Basic stats/plots continued,  
for loops,  
sampling distributions

# Overview

Class logistics

Review and continuation of data frames, statistics and plots in R

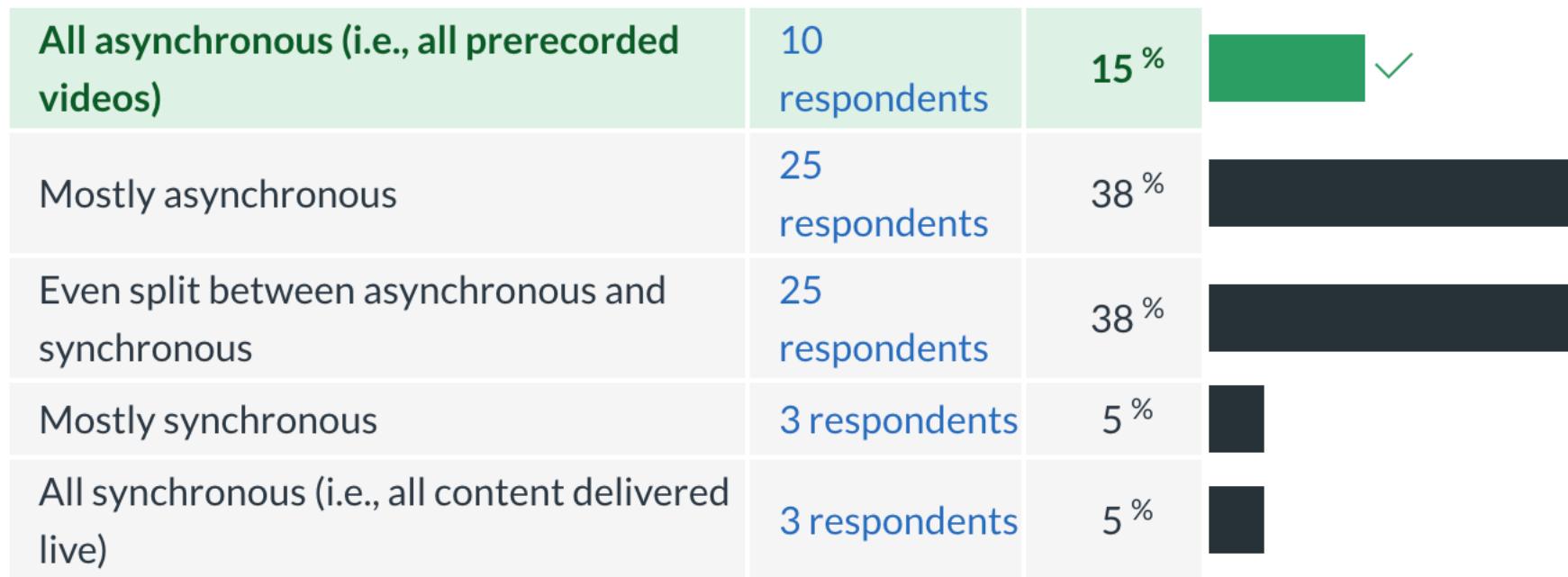
For loops and random numbers

If there is time: sampling distributions

# Logistics: class plan

Today and Thursday we will meet at the regular time (9-10:15)

Going forward, how would you prefer the content in this class is delivered?



# Logistics: class plan

Today and Thursday we will meet at the regular time (9-10:15)

Current plan going forward:

- Monday: asynchronous videos and homework released
- Tuesday class: office hours/live run through - **attendance optional**
- Wednesday **3pm**: Submit survey questions about videos
- Thursday class: synchronous meeting with review and activities

# Logistics: class participation

## Class participation grade (6%)

1. Asking and answering questions on Piazza
2. Filling out lecture video surveys

**X** Questions during class are highly encouraged but will **not** count toward your class participation grade



# Logistics: Homework 1

Homework 1 has been posted

- We have already covered everything on the homework except for loops

> `SDS230::download_homework(1)`

Due on Gradescope by 11:59pm on Sunday September 13<sup>th</sup>

- Instructions for how to submit homework on Gradescope are on Canvas

# Questions about the logistics?





# Review and continuation of...

Data frames and structured data

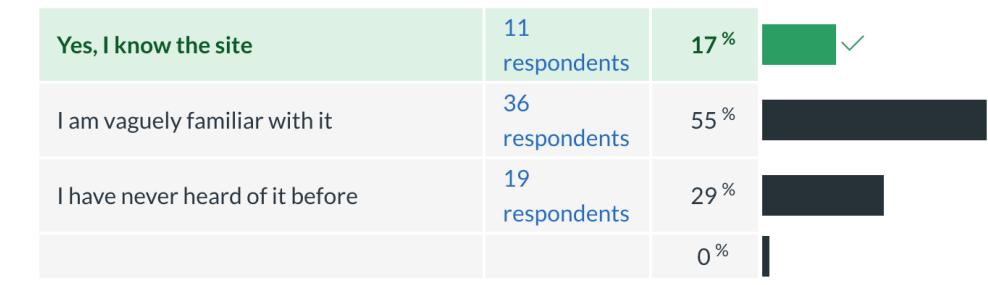
Statistics and plotting of categorical and quantitative data

# OK Cupid data

The screenshot shows the OkCupid website interface. At the top, there's a blue header bar with the OkCupid logo, a 'Welcome, BigDaddyC\_taco' message, and a 'Sign out' button. Below the header is a navigation menu with links for 'Messages', 'Matches', 'Connections', 'Treasures', and a search bar. The main content area displays a user profile for 'BigDaddyC\_taco'. The profile includes a large photo of a smiling man, his username 'BigDaddyC\_taco', his status '21 / M / Straight / Single Chicago, Illinois', and an 'Online Now' indicator. Below the profile are tabs for 'About', 'Photos', 'Questions', and 'Personality'. The 'About' tab is selected. Under 'My self-summary', the user writes: "I'm a young, ambitious and outgoing individual. I love traveling, having recently been to South America and through the southern states on a road trip with friends. I'm a very caring/emotional person. I enjoy anything artistic and always up for new activities. Also, I've been told I'm too perfect." The 'What I'm doing with my life' section lists: "Working two marketing jobs in downtown and Lincoln Park areas of Chicago.", "Full-time student at DePaul University studying Marketing/Sales.", "Volunteer on South Side of Chicago (Pilsen, Little Village & Englewood).", and "Writer for my blog, The Plaid Tie". To the right of the profile is a 'My Details' sidebar with information like 'Last Online: Online now!', 'Ethnicity: Hispanic / Latin', 'Height: 6' 0" (1.83m)', 'Body Type: Fit', 'Diet: Mostly anything', 'Smokes: No', 'Drinks: Rarely', and 'Drugs: Never'. On the left side of the page, there are sidebar sections for 'You might like...', 'Favorites' (empty), and 'Profile Completion' (65%). A call-to-action 'Contact 5 new people to get to 70%' is also present.

**Survey question 1:** Are you familiar with the website OkCupid?

Are you familiar with website OkCupid?



# Data Frames

Data frames contain structured data

	age	body_type	diet	drinks	drugs	education
1	22	a little extra	strictly anything	socially	never	working on college/university
2	35	average	mostly other	often	sometimes	working on space camp
3	38	thin	anything	socially	NA	graduated from masters program
4	23	thin	vegetarian	socially	NA	working on college/university
5	29	athletic	NA	socially	never	graduated from college/university
6	29	average	mostly anything	socially	NA	graduated from college/university

# Data Frames

## Variables

Cases

	age	body_type	diet	drinks	drugs	education
1	22	a little extra	strictly anything	socially	never	working on college/university
2	35	average	mostly other	often	sometimes	working on space camp
3	38	thin	anything	socially	NA	graduated from masters program
4	23	thin	vegetarian	socially	NA	working on college/university
5	29	athletic	NA	socially	never	graduated from college/university
6	29	average	mostly anything	socially	NA	graduated from college/university

# Data Frames

Quantitative Variable

Categorical Variable

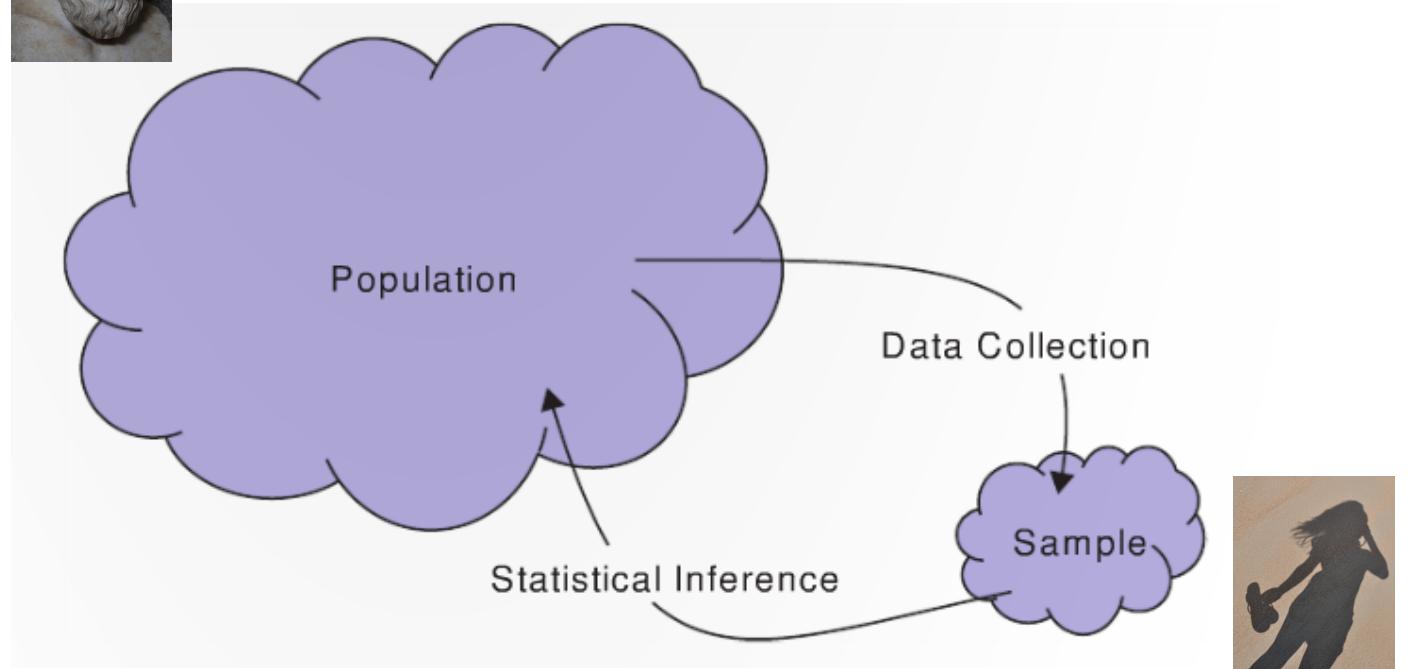
Cases  
(observational units)

	age	body_type	diet	drinks	drugs	education
1	22	a little extra	strictly anything	socially	never	working on college/university
2	35	average	mostly other	often	sometimes	working on space camp
3	38	thin	anything	socially	NA	graduated from masters program
4	23	thin	vegetarian	socially	NA	working on college/university
5	29	athletic	NA	socially	never	graduated from college/university
6	29	average	mostly anything	socially	NA	graduated from college/university

# Where does data come from?



**Population:** all individuals/objects of interest



**Sample:** A subset of the population

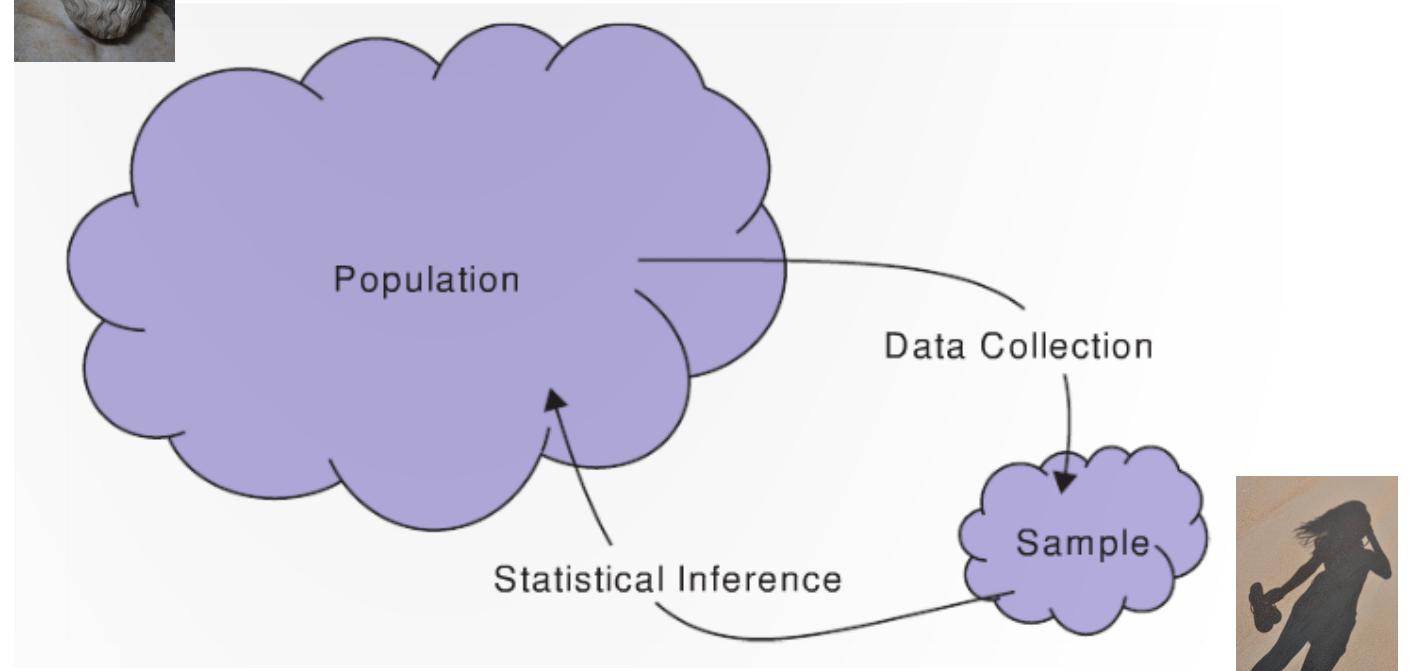
# Where does data come from?

**Question:** Is the okcupid profiles data frame a population or a sample?

**Question:** If the OkCupid profiles data frame is a sample, what is the population?



**Parameters:**  $\pi, \mu, \sigma, \rho, \beta$



**Statistics:**  $\hat{p}, \bar{x}, s, r, b$

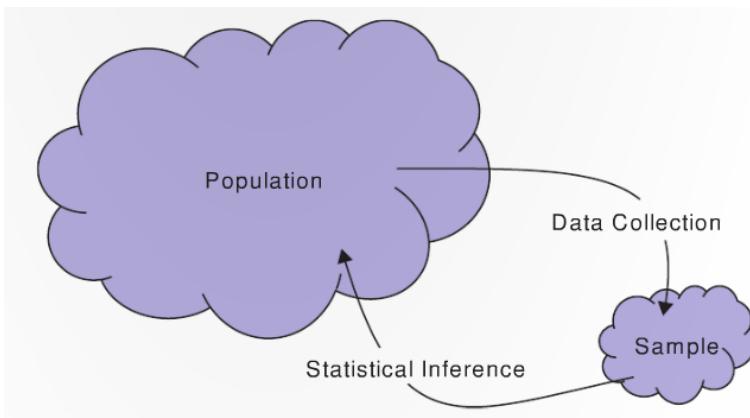
# How do we get sample of data?

**Simple random sample:** each member in the population is equally likely to be in the sample  
“Random selection”

*Soup analogy!*

**Q:** Why is this good?

**A:** Allows for generalizations to the population! (no bias)



**Question:** Is the OkCupid profiles data a simple random sample?

# Observational and experimental studies

An **experiment** is a study in which the researcher actively controls one or more of the explanatory variables.

An **observational study** is a study in which the researcher does not actively control the value of any variable but simply observes the values as they naturally exist

**Question:** Is the OkCupid profiles data from an experiment or observational study?



# Observational and experimental studies

**Question:** Would it be ethical to run an experiment on OkCupid users?

- [OkCupid experiments on humans](#)

OkCupid deleted all the text on everyone's profiles and measured how star ratings changed

- i.e., only the picture was available

5% Match  
3% Friend  
98% Enemy

H<3artbreaker  
19 / M

Rate this profile

Message

Save to Favorites

Hide Report

About Photos The Two of Us

**What I'm doing with my life**  
I'm a singing sensation, pretty much the biggest there ever was. some might say Elvis, you know, or Michael Jackson, and I mean no disrespect but baby, baby, baby NO.

**I'm really good at**  
singing, dancing, looking in the mirror, picking my own outfits, taking off my shirt, taking selfies, taking Instagram videos, walking smiling, breathing, being #nohate

**The most private thing I'm willing to admit**  
Im crazy, Im nuts... that's just the way my brain works

His Details	
Height	5' 7"
Diet	Mostly anything
Drinks	--
Drugs	--
Sign	Pisces
Job	Artist/musical/writer
Income	More than \$1,000,000
Pets	Likes dogs

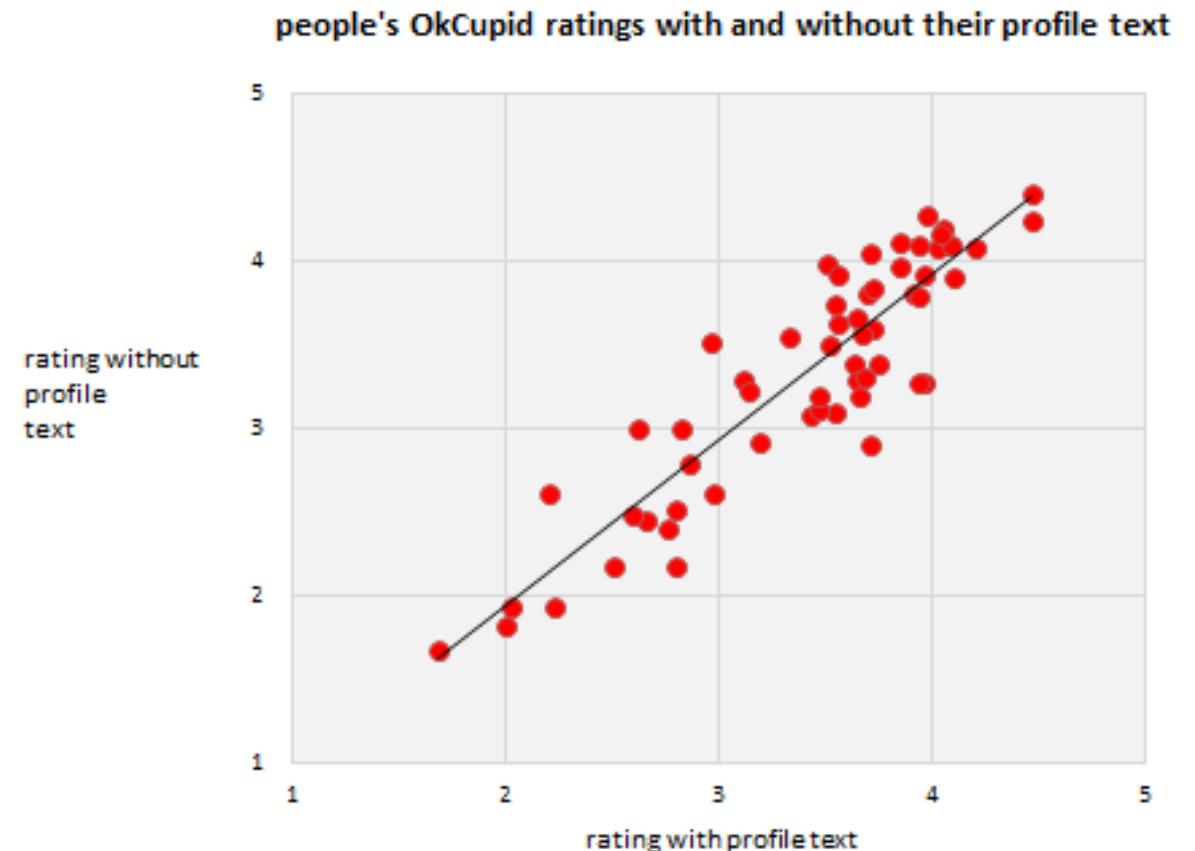
# Observational and experimental studies

**Question:** Would it be ethical to run an experiment on OkCupid users?

- OkCupid experiments on humans

OkCupid deleted all the text on everyone's profiles and measured how star ratings changed

- i.e., only the picture was available



# Questions?



Let's quickly review plots and statistics in R