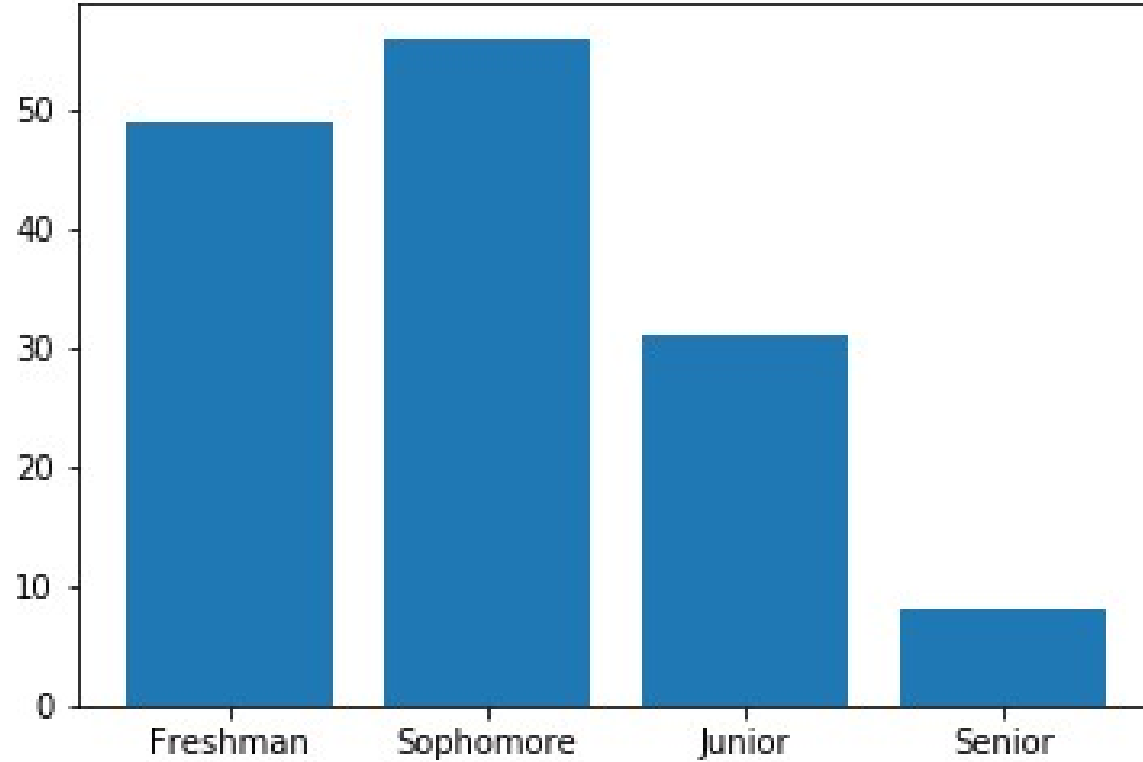


# Welcome to STA101 Data Analysis and Statistical Inference!

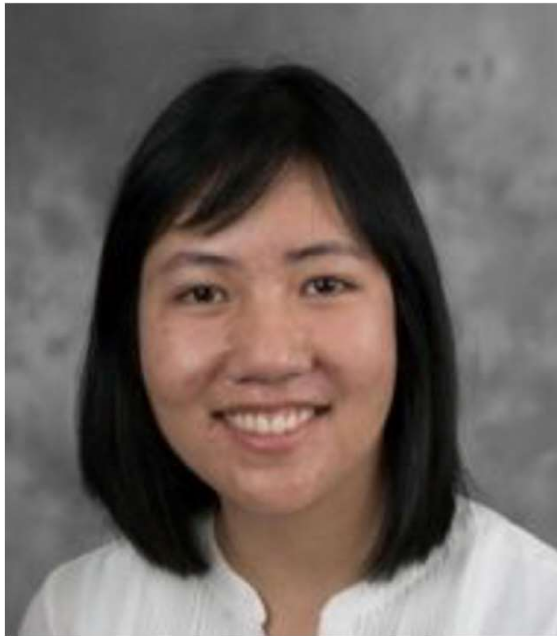


# Getting to Know You!

**OUR STA101 CLASS**



# Getting to know your TAs!



- Lecture TA: **Melanie Lai Wai**

- Melanie is a second-year master's student in statistical science at Duke. Before coming to Duke, she completed her bachelor's degree in mathematics and Chinese at Vassar College in Poughkeepsie, NY. She is particularly interested in interdisciplinary applications of statistics in public health and global health. Aside from statistics, she enjoys reading self-improvement books, learning new languages, and taking long walks.

# Getting to know your TAs!

- Lecture TA: **Lavonne Hoang**



- I am a sophomore majoring in Computer Science and Statistics and minoring in Spanish. I am really interested in pursuing either a career in the financial industry or the tech industry (specifically data science). I love to travel, read, and dance in my free time.

# Getting to know your TAs!

- Head TA + Lab TA: Tessa Johnson



- Hello! My name is Tessa, and I'm a second year PhD student in the statistics department. I did my undergraduate degree at Texas A&M in two years with a double major in Mathematics and Statistics. Outside of school I spend most of my time reading or playing with my puppy, Rory. I was one of the head TAs for STAT 101 in Fall 2017, and I was an instructor of record (i.e. the professor) for STAT 101 during Summer Session I 2018. Feel free to reach out to me at any time via email or my office hours!

# Getting to know your TAs!

- Lab TA: **Vanessa Alwan**



- My name is Vanessa Alwan and I am a 1st year masters student in the economics department. I majored in statistics as an undergraduate and have always loved statistics and data analysis. I hope to pursue a PhD one day in an area that allows me to utilize my skills in statistical methods.

# Getting to know your TAs!



- Lab TA: **Morgan Bird**

- I am a junior majoring in Statistical Science and minoring in Sociology. This past summer I completed an independent DukeEngage in Columbia, SC, where I created web pages highlighting the nonprofit's educational resources. During the summer of 2019, I will be interning in Charlotte, NC, on an alternative investments team at a financial services firm. Outside of class, you can find me with friends, listening to Spotify playlists, at the gym, or trying new restaurants in Durham.

# Getting to know your TAs!



- Lab TA: **Tess Chandler**

- Tess is a junior majoring in Statistics and minoring in Spanish. She recently returned from studying abroad in Madrid. Last summer, she was a research assistant in the School of Health and Human Services at the University of New Hampshire where she assisted in running data analysis for several neuroscience research initiatives. In addition, Tess has worked as a teaching assistant for calculus 111 and is a member of Duke's Women's Club Lacrosse team.

# TO-DO:



1. Get clickers if you haven't already and bring to class on Monday.
2. Download or purchase book.
3. Complete the Pre-test and Getting to Know you Survey (instructions emailed).
4. Watch Unit 1 videos by next Monday (course website).

# Coming up...

## Thursday:

1. Go to lab and work on **Lab Assignment 0**.

## Next Monday:

1. Begin registering clickers (roll-call).
2. Take our **Unit 1 Readiness Assessment** (not graded this time.)
3. Lesson 1.2 and **Application Exercises**.



# Check out the Course Website



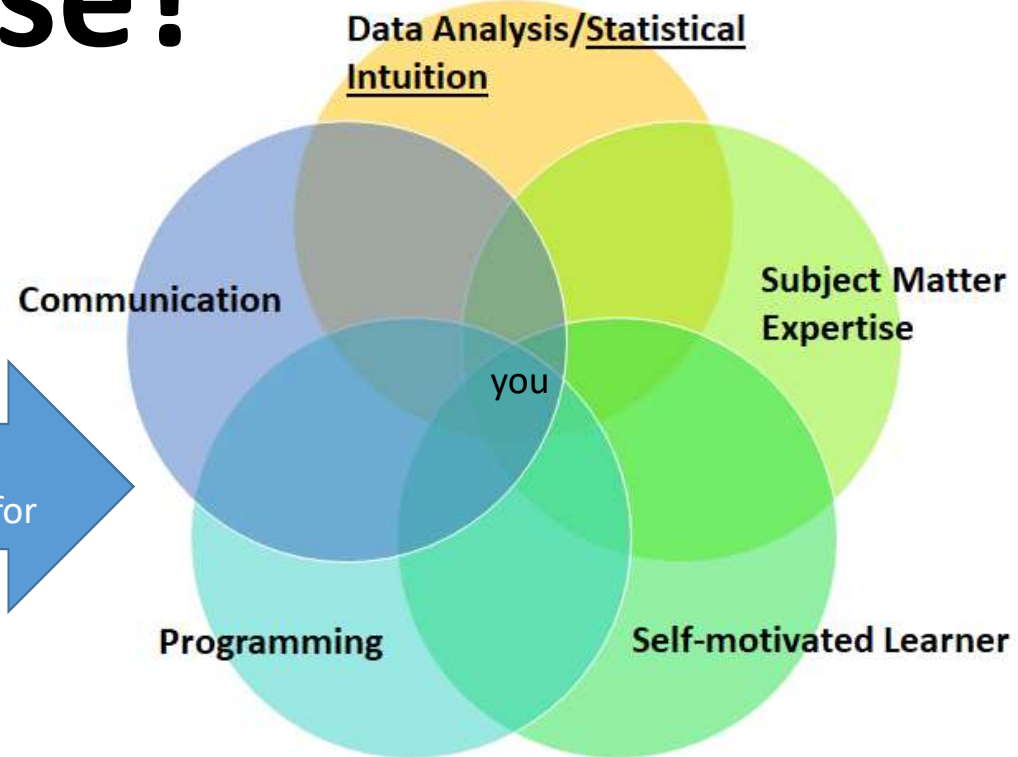
<https://www2.stat.duke.edu/courses/Spring19/sta101.001/>

# What do you get out of this course?

Data is everywhere!



What industry +  
academia is looking for

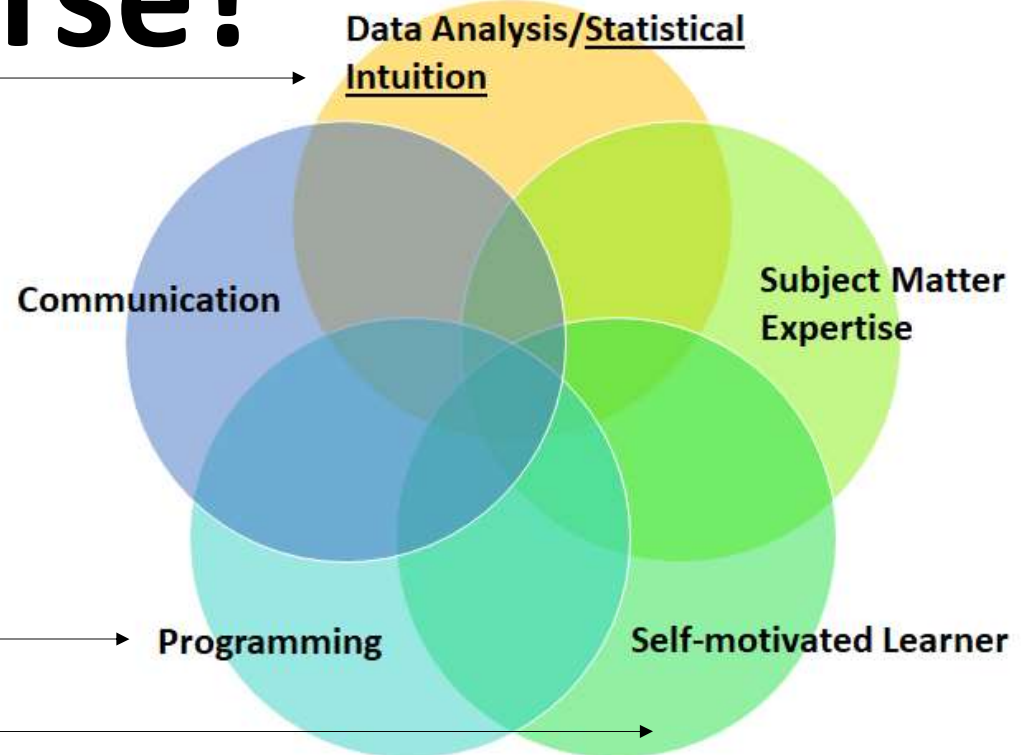


<https://www.amstat.org/asa/education/Curriculum-Guidelines-for-Undergraduate-Programs-in-Statistical-Science.aspx>

# What do you get out of this course?

## Current State of Data Analysis/Science:

1. Field of data analysis is broad! (Impossible to know *everything*.)
2. New and useful statistical analyses and algorithms are developed everyday!
3. Datasets are larger!



# Taken a flipped class before? Want to post tips for your fellow classmates on Piazza?

- Overview
- Announcements
- Syllabus
- Calendar
- Email
- Drop Box
- Resources
- Assignments
- Gradebook
- Statistics
- Site Info
- Tests & Quizzes
- Piazza
- Help

The screenshot shows the Piazza Q&A interface for a course (STA. 101.001.SP19). The left sidebar contains navigation options. The main content area shows a draft post titled "Flipped Class Tip" with the text: "When going through the videos, make a list of questions you don't understand. Bring these questions to class, lab, or office hours and&nbsp;". Below the draft is a list of pinned posts, including "Search for Teammates!", "Introduce Piazza to your stu...", "Get familiar with Piazza", "Tips & Tricks for a successf...", and "Welcome to Piazza!". The right side of the interface shows the "New Post" form, where the "Note" post type is selected. The "Post to" option is set to "Entire Class". The "Summary" field contains "Flipped Class Tip". The "Details" section shows a rich text editor with the text: "When going through the videos, make a list of questions you don't understand. Bring these questions to class, lab, or office hours and reach out to Dr. Ellison, one of the TAs, or your classmates to get them answered!". A black arrow points from the main heading to the rich text editor.

# Have questions about content? Ask on Piazza!

- Overview
- Announcements
- Syllabus
- Calendar
- Email
- Drop Box
- Resources
- Assignments
- Gradebook
- Statistics
- Site Info
- Tests & Quizzes
- Piazza**
- Help

The screenshot displays the Piazza Q&A interface for the course STA. 101.001.SP19. The interface is divided into several sections:

- Navigation Sidebar:** Includes Overview, Announcements, Syllabus, Calendar, Email, Drop Box, Resources, Assignments, Gradebook, Statistics, Site Info, Tests & Quizzes, **Piazza**, and Help.
- Course Header:** Shows the course name, navigation tabs (polls, hw1, hw2, hw3, hw4), and the user's name (Victoria Ellison).
- Post Creation Area:** Features a 'New Post' button, a search bar, and a 'DRAFTS' section. A pinned post titled 'Flipped Class Tip' is visible, with the text: '<p>When going through the videos, make a list of questions you don't understand. Bring these questions to class, lab, or office hours and&nbsp;'. Below this are several other pinned posts, including 'Search for Teammates!', 'Introduce Piazza to your stu...', 'Get familiar with Piazza', 'Tips & Tricks for a successf...', and 'Welcome to Piazza!'.
- Post Editor:** Contains a 'Post Type' selector (Question, **Note**, Poll/In-Class Response), a 'Post to' selector (Entire Class, Individual Student(s) / Instructor(s)), and a 'Select Folder(s)' dropdown menu. The 'Summary' field contains the text 'Flipped Class Tip'. The 'Details' section includes a rich text editor with a toolbar containing options like Edit, Insert, View, Format, Table, Bold, Italic, Bulleted List, Numbered List, Link, Image, Video, Facebook, Code, Text, Markdown, and Help. A large black arrow points from the top right towards the 'Help' button in the editor toolbar.

# Questions we can answer in this class:

“Is there an association between **age** and **political affiliation**?”

Sample Data

Gender (Categorical with 2 Levels)	Political Affiliation (Categorical with 3 Levels)	Voted in 2018? (Categorical with 2 Levels)	Age (Numerical - Continuous)	Income (Numerical - Continuous)	Number of Dependents (Numerical - Discrete)
Male	Democrat	Yes	22	\$20,000	0
Female	Republican	Yes	39	\$40,000	3
Male	Independent	Yes	47	\$400,000	0
Male	Republican	Yes	18	\$129,000	0
Female	Republican	No	29	\$85,000	2
Female	Democrat	No	80	\$72,000	1
Female	Democrat	Yes	56	\$55,000	0
Male	Independent	Yes	72	\$34,000	0
...	...	...	...	...	...

# Questions we can answer in this class:

“Is there a linear association between **age** and **income**?”

Sample Data

Gender <i>(Categorical with 2 Levels)</i>	Political Affiliation <i>(Categorical with 3 Levels)</i>	Voted in 2018? <i>(Categorical with 2 Levels)</i>	Age <i>(Numerical - Continuous)</i>	Income <i>(Numerical - Continuous)</i>	Number of Dependents <i>(Numerical - Discrete)</i>
Male	Democrat	Yes	22	\$20,000	0
Female	Republican	Yes	39	\$40,000	3
Male	Independent	Yes	47	\$400,000	0
Male	Republican	Yes	18	\$129,000	0
Female	Republican	No	29	\$85,000	2
Female	Democrat	No	80	\$72,000	1
Female	Democrat	Yes	56	\$55,000	0
Male	Independent	Yes	72	\$34,000	0
...	...	...	...	...	...

# Questions we can answer in this class: “Is there an association between gender and political affiliation?”

Sample Data

Gender <i>(Categorical with 2 Levels)</i>	Political Affiliation <i>(Categorical with 3 Levels)</i>	Voted in 2018? <i>(Categorical with 2 Levels)</i>	Age <i>(Numerical - Continuous)</i>	Income <i>(Numerical - Continuous)</i>	Number of Dependents <i>(Numerical - Discrete)</i>
Male	Democrat	Yes	22	\$20,000	0
Female	Republican	Yes	39	\$40,000	3
Male	Independent	Yes	47	\$400,000	0
Male	Republican	Yes	18	\$129,000	0
Female	Republican	No	29	\$85,000	2
Female	Democrat	No	80	\$72,000	1
Female	Democrat	Yes	56	\$55,000	0
Male	Independent	Yes	72	\$34,000	0
...	...	...	...	...	...

# Questions we can answer in this class: “Are gender, political affiliation, and age predictors of income ?”

Sample Data

Gender (Categorical with 2 Levels)	Political Affiliation (Categorical with 3 Levels)	Voted in 2018? (Categorical with 2 Levels)	Age (Numerical - Continuous)	Income (Numerical - Continuous)	Number of Dependents (Numerical - Discrete)
Male	Democrat	Yes	22	\$20,000	0
Female	Republican	Yes	39	\$40,000	3
Male	Independent	Yes	47	\$400,000	0
Male	Republican	Yes	18	\$129,000	0
Female	Republican	No	29	\$85,000	2
Female	Democrat	No	80	\$72,000	1
Female	Democrat	Yes	56	\$55,000	0
Male	Independent	Yes	72	\$34,000	0
...	...	...	...	...	...

# Getting to Know You!

What's a large population that *you* would like to learn more about? What would you like to learn about this population?

What questions can we answer by the end of this lecture?

- How does a flipped class work?



What questions can we answer by the end of this lecture?

- What does it mean to make an inference about a population?



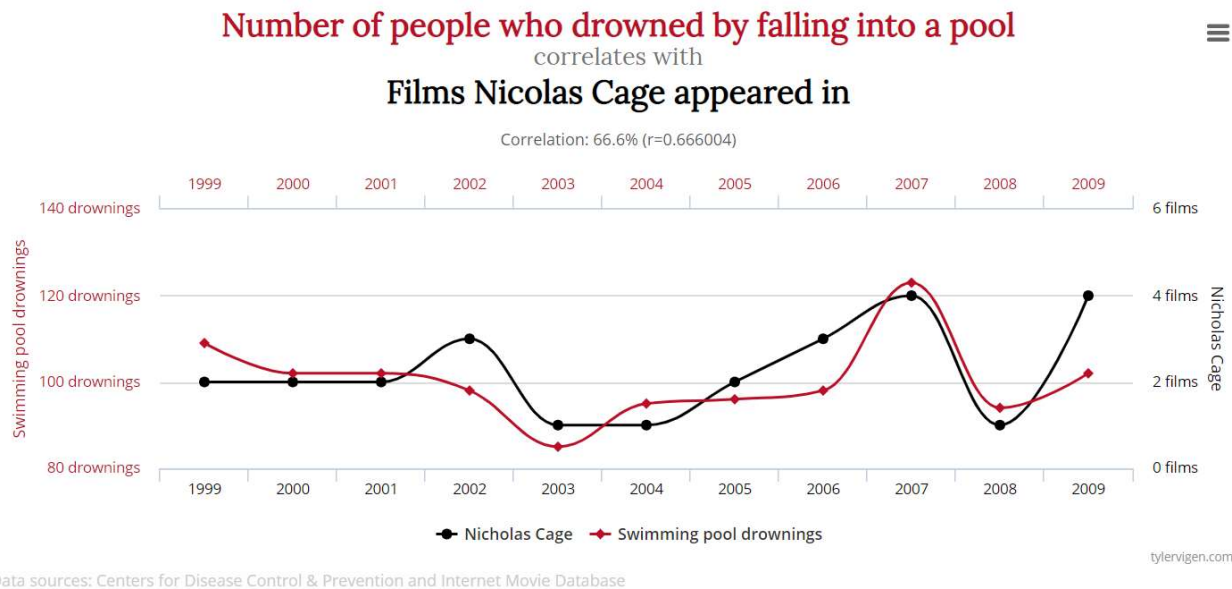
What questions can we answer by the end of this lecture?

- When can we *generalize* a set of results about a small data set to a larger population?



# What questions can we answer by the end of this lecture?

- When can we assess if one variable *causes/affects* another? When can we not?



What questions can we answer by the end of this lecture?

- What are four ways to collect a random sample? Why would we use each one?
- What are three biases a sample can suffer from?
- How do we create a good experimental design?