When you think of “statistics” who are the first 5 people who come up in your mind?

Revisiting the Glass Ceiling: A Study of the Gender Gap in Statistics Academia

Ivy Zhang[1] and Colleen Chan[2]

[1]Department of Statistics, Stanford University
[2]Department of Statistics and Data Science, Yale University

Outline

01
Background

02
Data Collection

03
Results

04
Discussion

Background

Why gender diversity matters

Balanced departments = a balanced society

- **Problem:** gender often disregarded in teaching/research approach, content and analysis
- **Result:** viewpoints, experiences, and needs of half the population risk being overlooked/dismissed

The current landscape

Available Data

Annual Survey of the Mathematics Sciences (ASMS): best gender gap data

Outdated and Limited

Latest report: **Fall 2018; 47 of 60** doctoral granting statistics departments responded

Previous work

(Goldbeck 2016, *Significance.)*

STATISTICALLY SPEAKING

Are women underrepresented in the American statistics profession?

Amanda Golbeck digs into the Annual Survey of the Mathematical Sciences in search of an answer

	Total	Statistics	Biostatistics
Students			
Bachelor's degrees awarded	4.5	4.5	*
Master's degrees awarded	5.1	4.9	5.8
Doctoral degrees awarded	4.6	4.1	5.7
Full-time faculty			
Non-tenure-track	5.4	4.8	5.7
Tenure-track	2.8	2.7	2.8
Untenured, tenure-eligible	3.3	3.4	3.3
Tenured but not full professors	4.0	3.8	4.5
Tenured full professors	2.0	2.1	1.9

*Numbers are too small to produce a reliable estimate.

Expected numbers of women out of every 10 persons, 2014

“There is a **giant hole** in the academic career pipeline” –

Amanda Golbeck, professor of biostatistics and public health administration at the University of Montana

2017-2018 academic year

Among doctoral-granting statistics departments that filled out the ASMS survey, the following **% went to women**

41.0% of **bachelor's degrees**

43.7% of **master's degrees**

37.5% of **Ph.D. degrees**

41.8% of **non tenure-track faculty positions**

24.0% of **tenure-track faculty positions**

These numbers **are slightly lower** than those in 2014!

Gender Barriers in the Academic Funnel

(AFT Higher Education 2011)

Faculty hiring process

- Inequities in compensation
- Lack of standard and transparent processes

Retention

- Hostile or unwelcoming workplace climates
- Inadequate family-friendly policies
- Difficulties in navigating the tenure-track

24.0% of **tenure-track faculty positions**

Why we need better statistics

1. **Transparency**
2. Current data is **outdated**
3. Current data may be **prone to survey nonresponse bias**

Data

Data Collection

Statistics Departments

- **Top 29** (2022 U.S. News Best Statistics Program Rankings)
- **Excluded:** non-tenure track faculty ie: ”Lecturer”, “Adjunct”, “Affiliated” and others ie: “Emeritus” and “Visiting”
- Four dept websites: PhD and master’s students undifferentiated

Journals

- Top journals’ 2021 issues (SRJ Journal rankings in 2021)
  - Annals of Probability
  - Annals of Statistics
  - Journal of the American Statistical Association
  - Journal of Statistical Software

1801 PhD students, 875 faculty, 1382 authors

Webscraper

Python

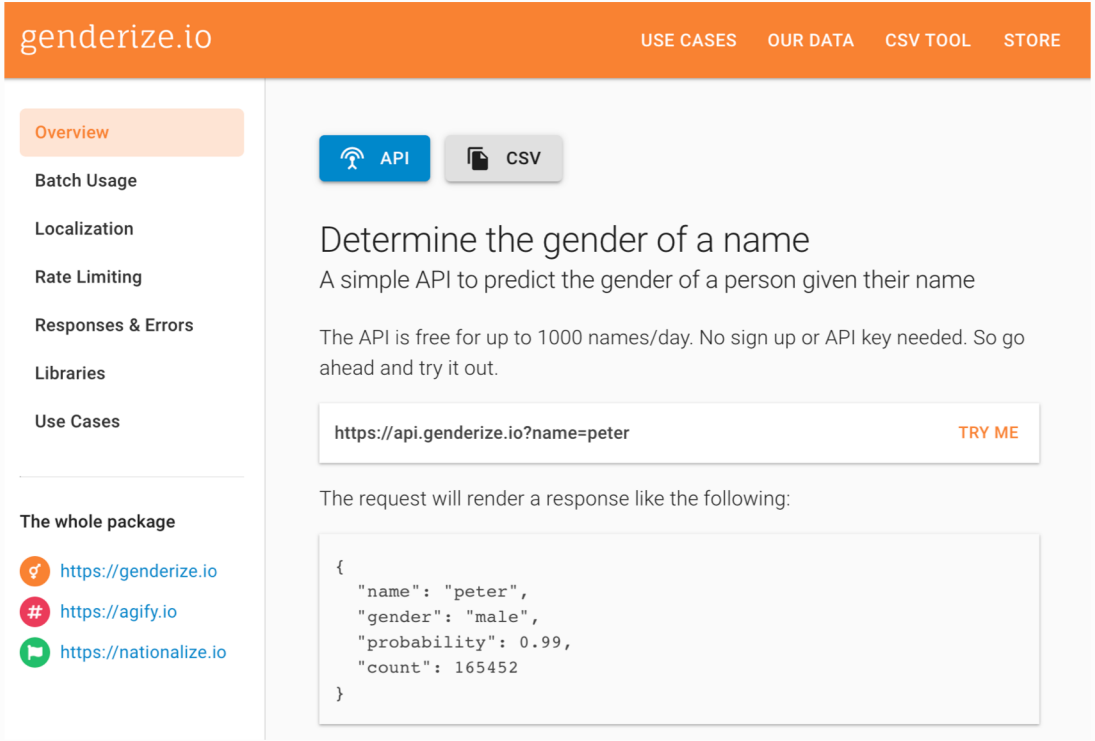
BeautifulSoup html parsing library

NameScraperPipeline

ExtractSchoolNames and **ExtractAuthorNames** classes: pulls and cleans names from department and journal websites

Gender Classification

We used the **genderize.io** package to classify first names



Validation of Gender Classifier

Manual labels

Classify male vs female from available photos (n = 206) from Michigan, Rice, Texas A&M

Comparison

genderize.io classified **93.4%** of names with **93.8% accuracy**

Closer look

Of the12 incorrectly classified, 75% of the women labeled men: **classifier provides an upper bound**

Results

% Women

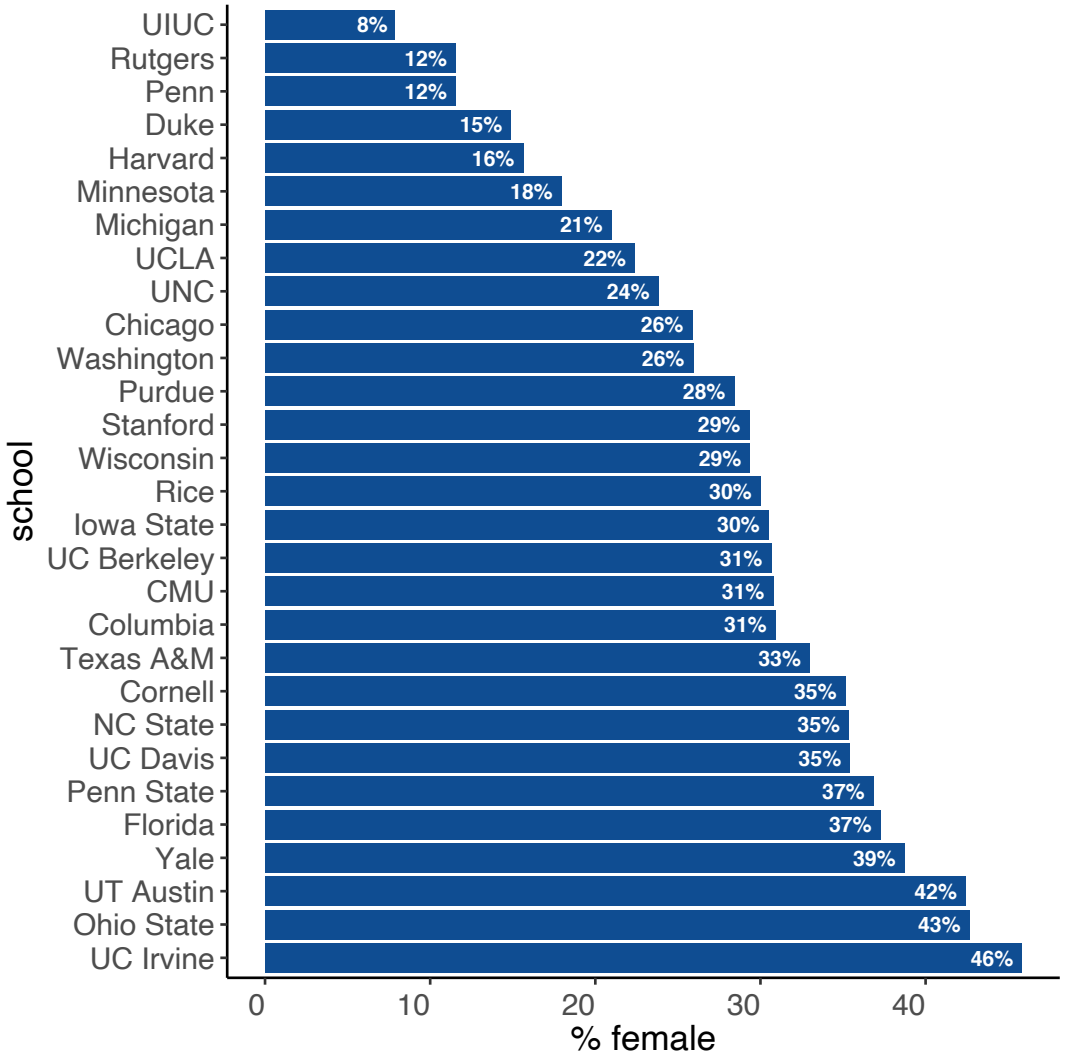
28.9% of PhD students

28.7% of tenure-track faculty

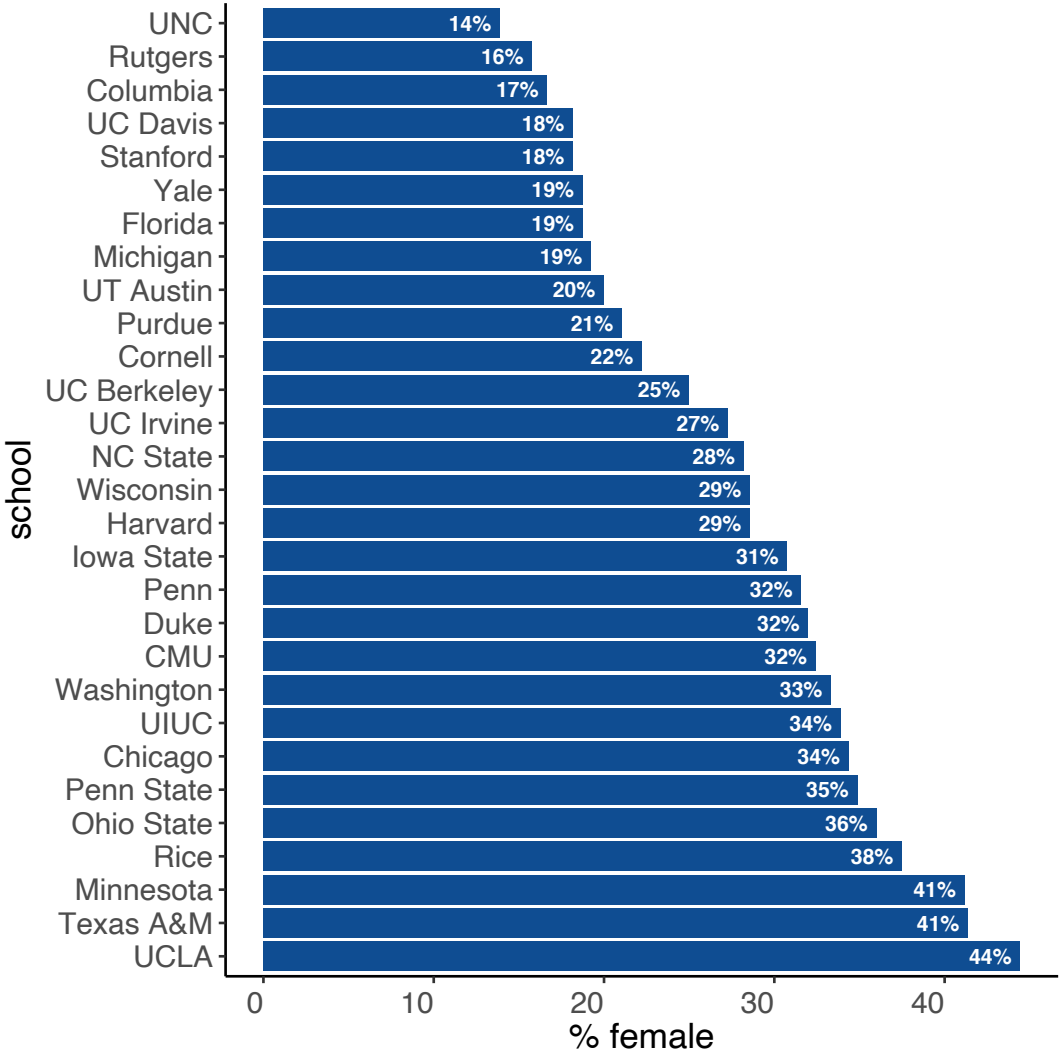
20.2% of stats journals authors

based on 96.7% of names that the classifier could predict

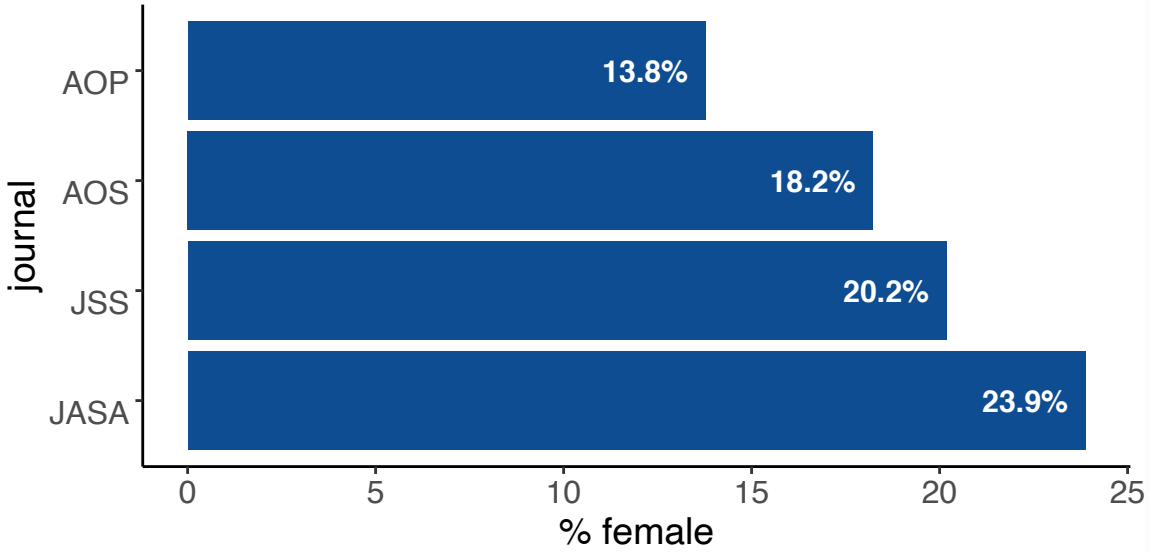
By PhD Students



By Faculty

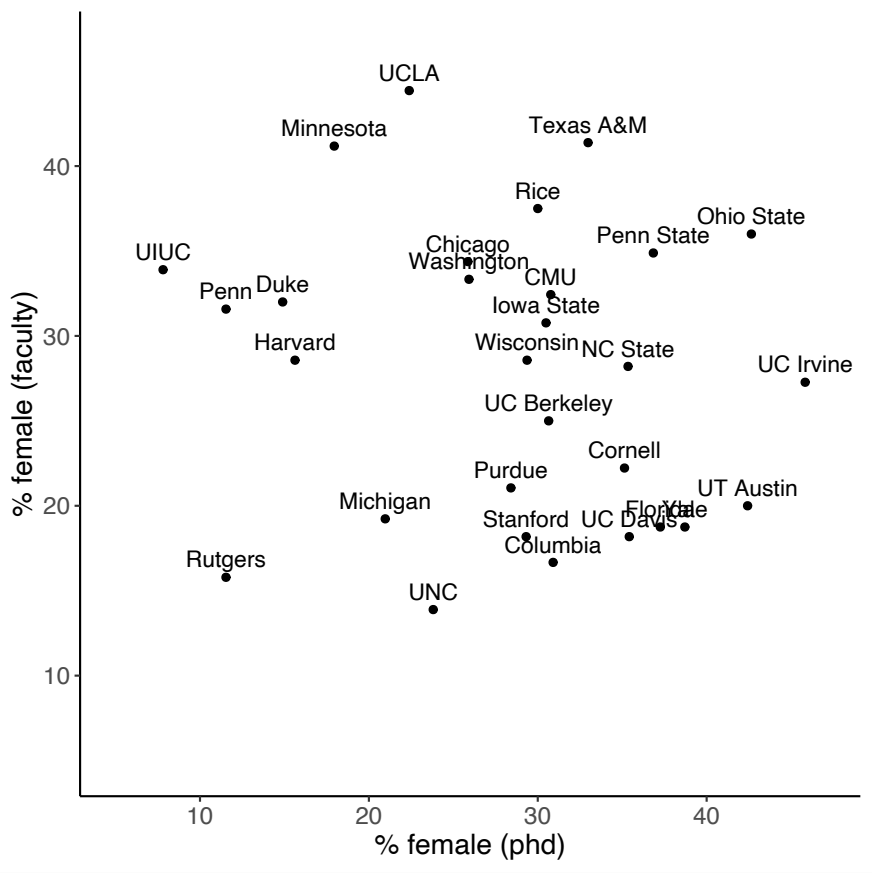

school
UNC
Rutgers
Columbia
UC Davis
Stanford
Yale
Florida
Michigan
UT Austin
Purdue
Cornell
UC Berkeley
UC Irvine
NC State
Wisconsin
Harvard
Iowa State
Penn
Duke
CMU
Washington
UIUC
Chicago
Penn State
Ohio State
Rice
Minnesota
Texas A&M
UCLA
14%
16%
17%
18%
18%
19%
19%
19%
20%
21%
22%
25%
27%
28%
29%
29%
31%
32%
32%
32%
33%
34%
34%
35%
36%
38%
41%
41%
44%
0
10
20
30
40
% female

By Journal



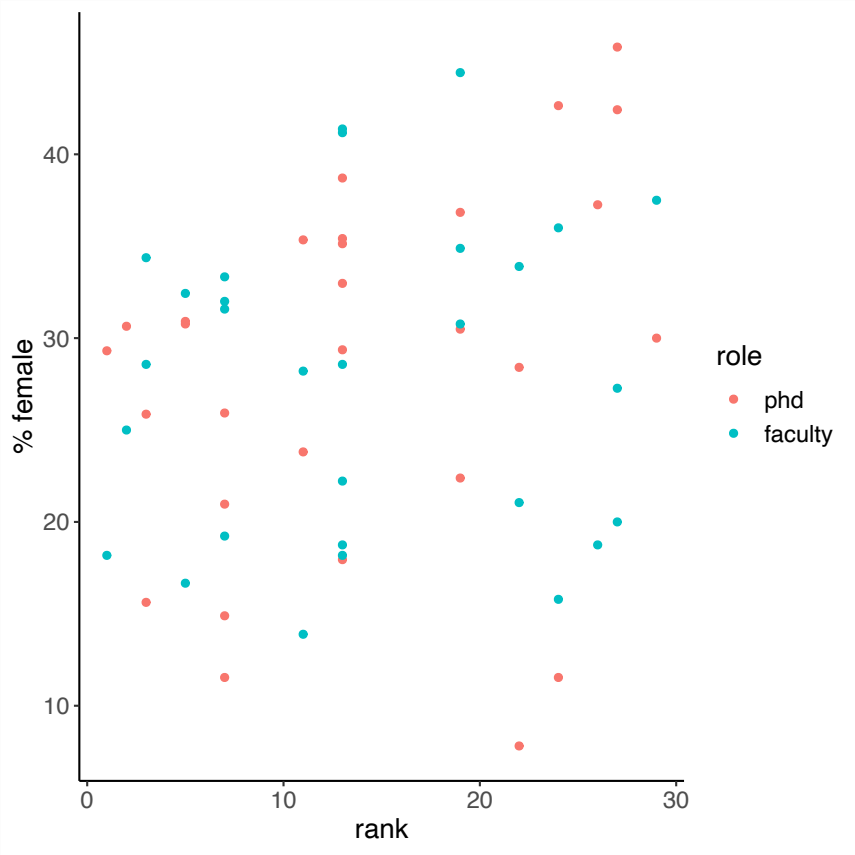
PhD Students vs Faculty

No relationship in female representation between PhD students and faculty



Rank vs Female Representation

Higher ranking departments seem to have **less female representation**



Discussion

Summary

Limitations

- Only as accurate as websites' data
- Classifier not 100% accurate and limited to gender binary

Future work

- Track over time
- Expanding the number of departments and journals

We hope that our results will better inform department hiring and admissions committees when considering diversity

Github Replication

https://github.com/yzhangi96/StatNameScraper

Thank you!

Donna Spiegelman, Susan Dwight Bliss Professor of Biostatistics, Yale School of Public Health

Women in Statistics and Data Science

WSDS ASA Student Travel Award