Patching the Leaky Pipeline

Do female PIs at ISTA have more female scientists in their groups?

Vasudha Kulkarni

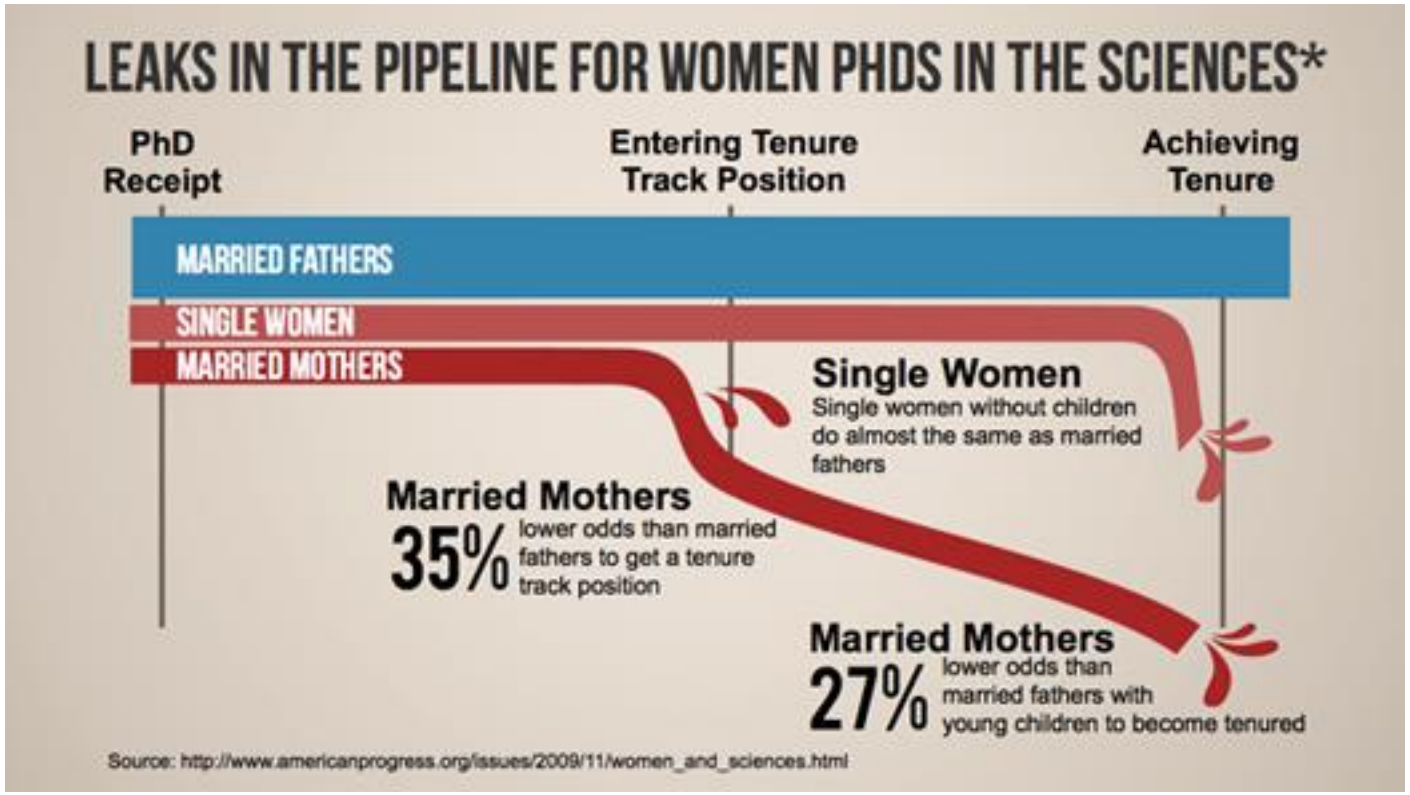
Experimental Design and Statistical Analysis Presentation

18.11.2024

Picture A Scientist

Sharon Shattuck and Ian Cheney, 2020 (https://www.pictureascientist.com/)

Leaky Pipeline in STEM



1. Organizational problems
2. Societal expectations
3. Bias and discrimination

Why care about diversity in science?

- Idealist argument

  "Everyone should get equal opportunities to pursue their goals."

- Scientific argument

  "If women were included, twice as many people would have worked on problems."

- Selfish argument

  "Accessibility for certain groups makes everyone's life easier."

The Harmful Privilege: Why men should care about diversity in science by Dr. Carl Goodrich (2023); Omland, Rose and Odom (2020)

Impact of female professors on students*

Role model effect

- Female and minority students' choice of major is positively affected by number of classes taught by professors “like them” [1]
- Female students in quantitative disciplines took additional courses if the introductory classes were taught by women [2]

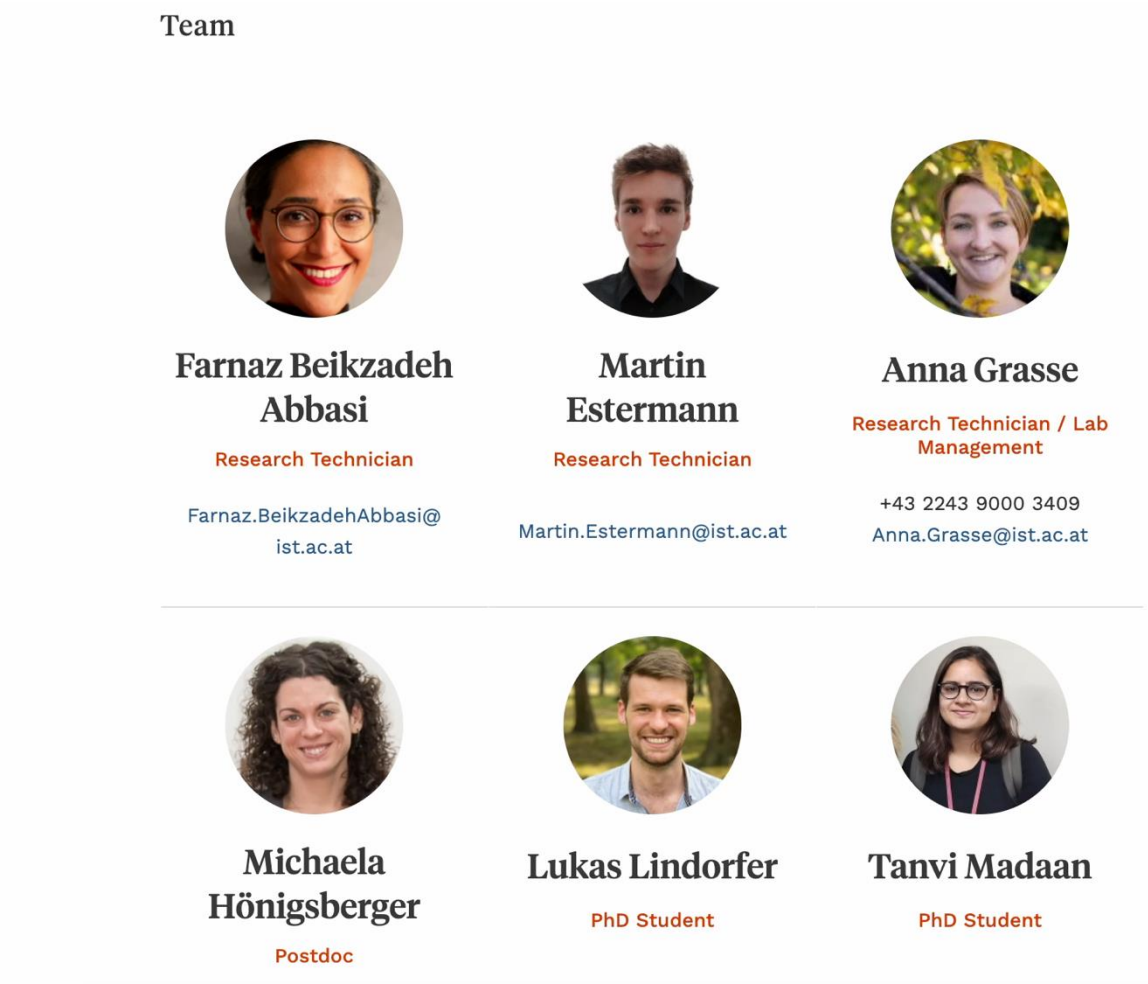
Contradictory study

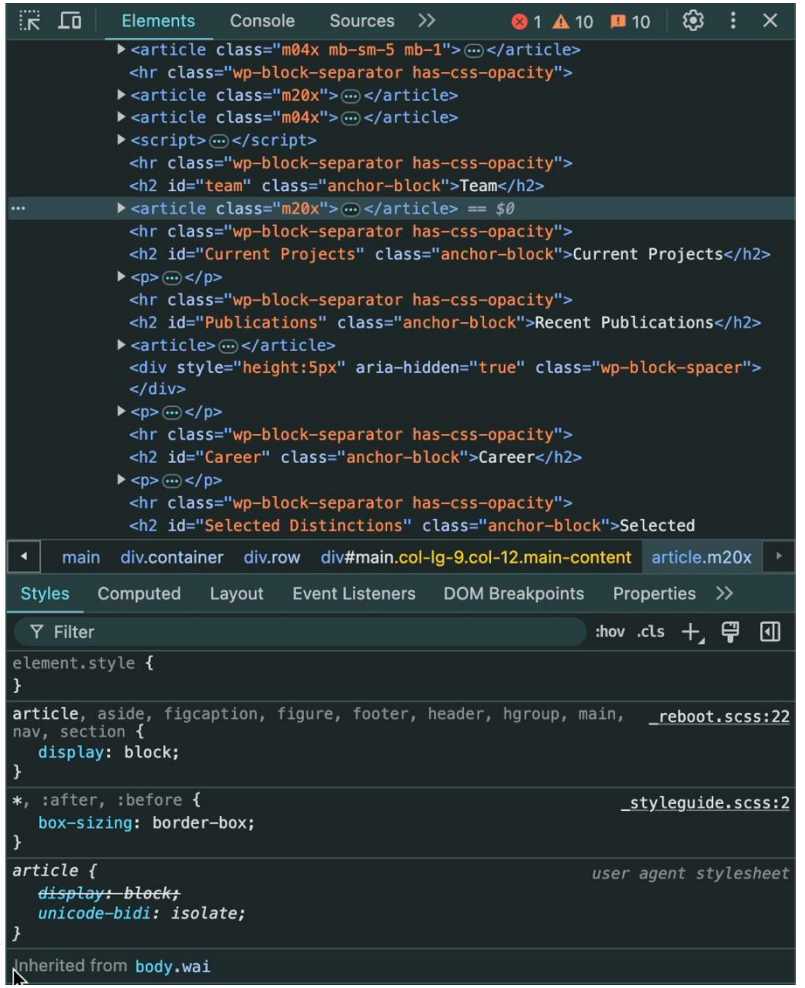
- No effect of number of female professors in the department on number of women choosing to major in that department [3]

1. Rask and Bailey 2010, 2. Bettinger and Long 2005, 3. Canes and Rosen 1995; *mostly in US undergraduate universities.

Do female PIs at ISTA have more female scientists in their groups?

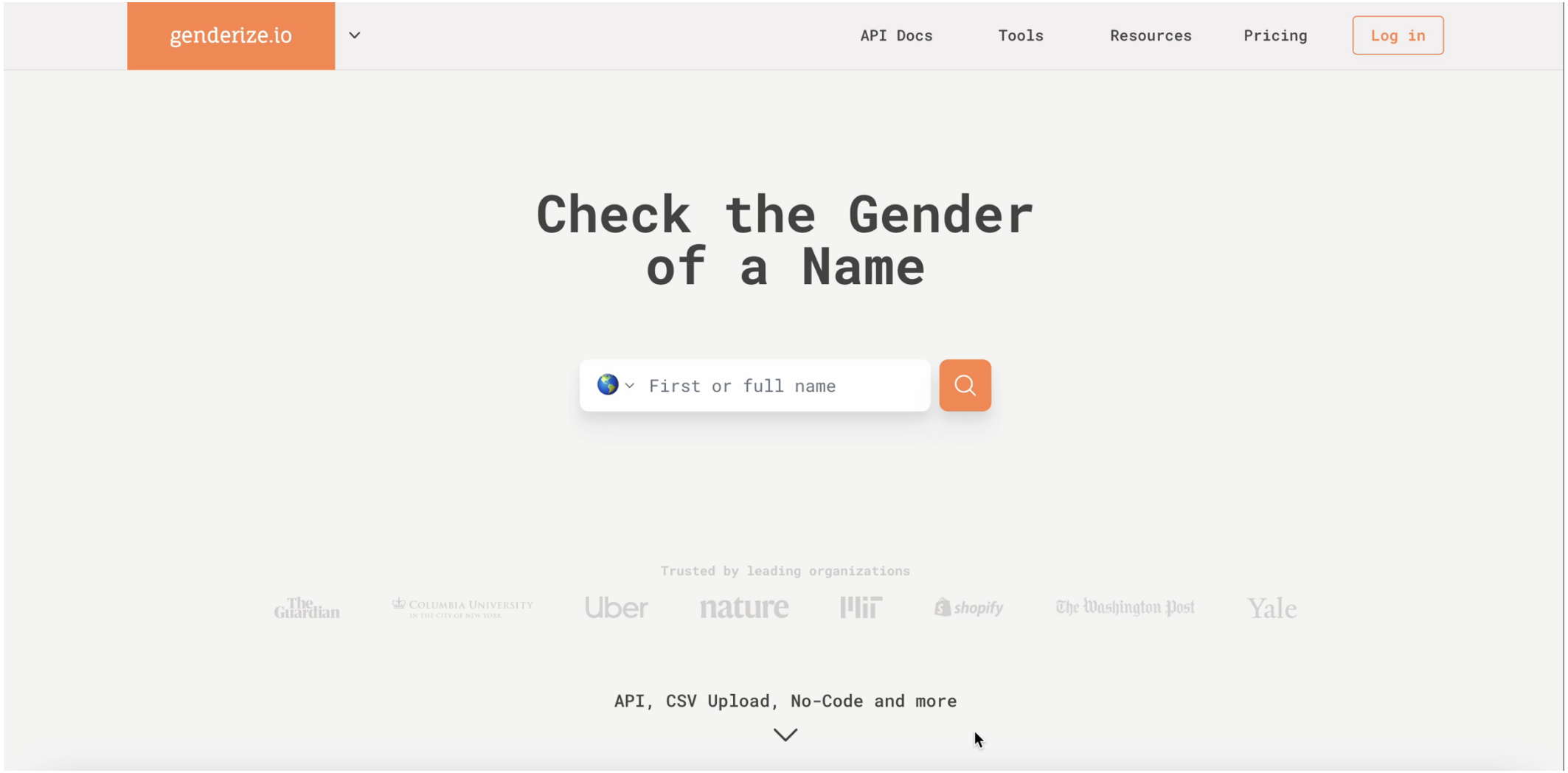
Methods – Gathering Names and Roles





BeautifulSoup: https://pypi.org/project/beautifulsoup4/

Methods – Genderise.io Assigning Sex



Genderise.io - https://genderize.io/

Methods – Field

PI–Field information received from the Communications team (thanks Lukas!)

Total: 85 PIs (6 without a group)

Astronomy – 4 –> Physics

Earth Science – 2 –> Chemistry

Merging Role, Sex and Field data –

	Name	Role	Group	Sex	PI	PI_sex	Field
0	Dan Alistarh	Professor	Alistarh	male	True	male	ComputerScience
1	Jiale Chen	PhD Student	Alistarh	male	False	male	ComputerScience
2	Alexander Fedorov	PhD Student	Alistarh	male	False	male	ComputerScience
3	Eugenia Iofinova	PhD Student	Alistarh	female	False	male	ComputerScience
4	Eldar Kurtic	Research Technician	Alistarh	male	False	male	ComputerScience
...	...	...	...	...	...	...	...
682	Nona Boustan	PhD Student	Zilberman	female	False	male	Biology
683	Elizabeth Hollwey	Research Technician	Zilberman	female	False	male	Biology
684	Minerva Trejo Arellano	Postdoc	Zilberman	female	False	male	Biology
685	Bingqing Cheng	Assistant Professor	Cheng	female	True	female	Physics
686	Zezhu Zeng	Postdoc	Cheng	male	False	female	Physics

687 rows × 8 columns

Hypotheses

- **A0** – There is *no correlation* between female sex ratio and ascending academic roles
- **A1** – There is a *correlation* between female sex ratio and ascending academic roles

- **B0** – There is *no difference* between number of males and females in a field
- **B1** – There is a *difference* between number of males and females in a field

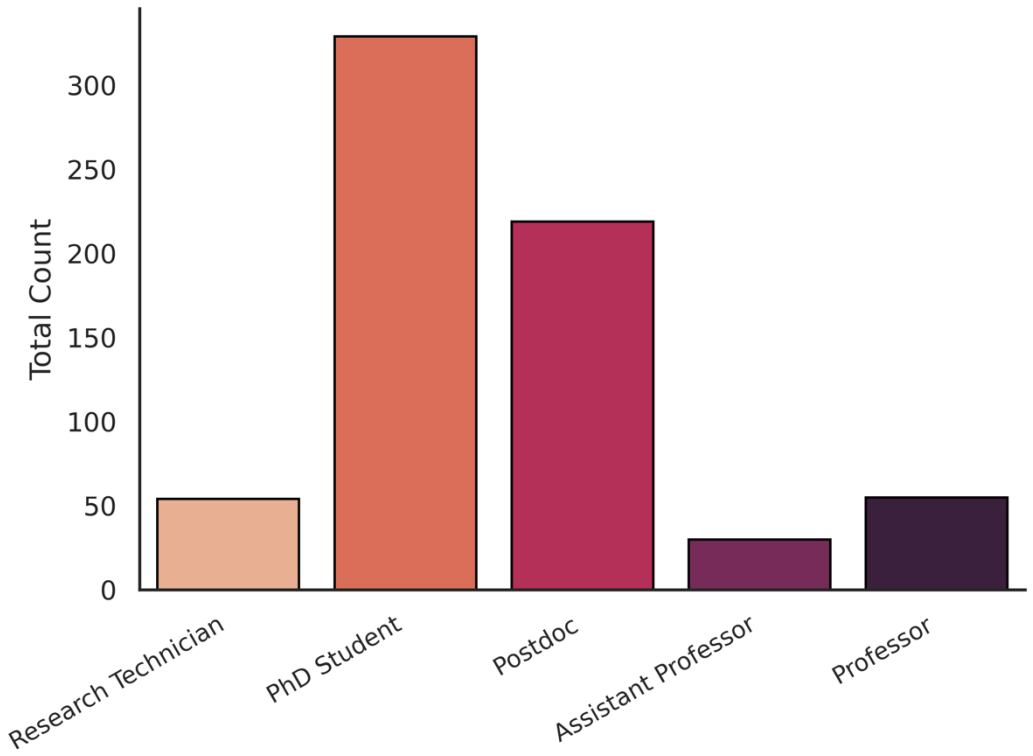
- **C0** – Female and male PIs have *similar* proportion of female students in their group
- **C1** – Female and male PIs have *different* proportions of female students in their group

*At ISTA in November 2024

Distribution plots

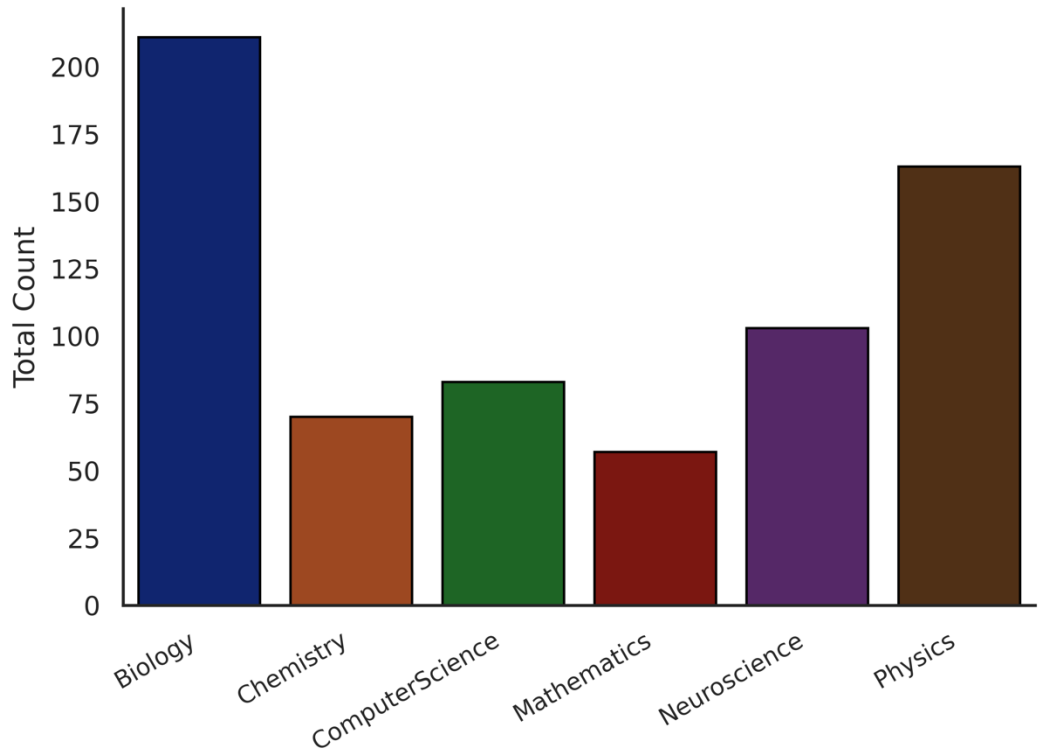
Total: **687** people (excluding visiting scientists, interns and A2Ps)

Academic Role



Ordinal categories

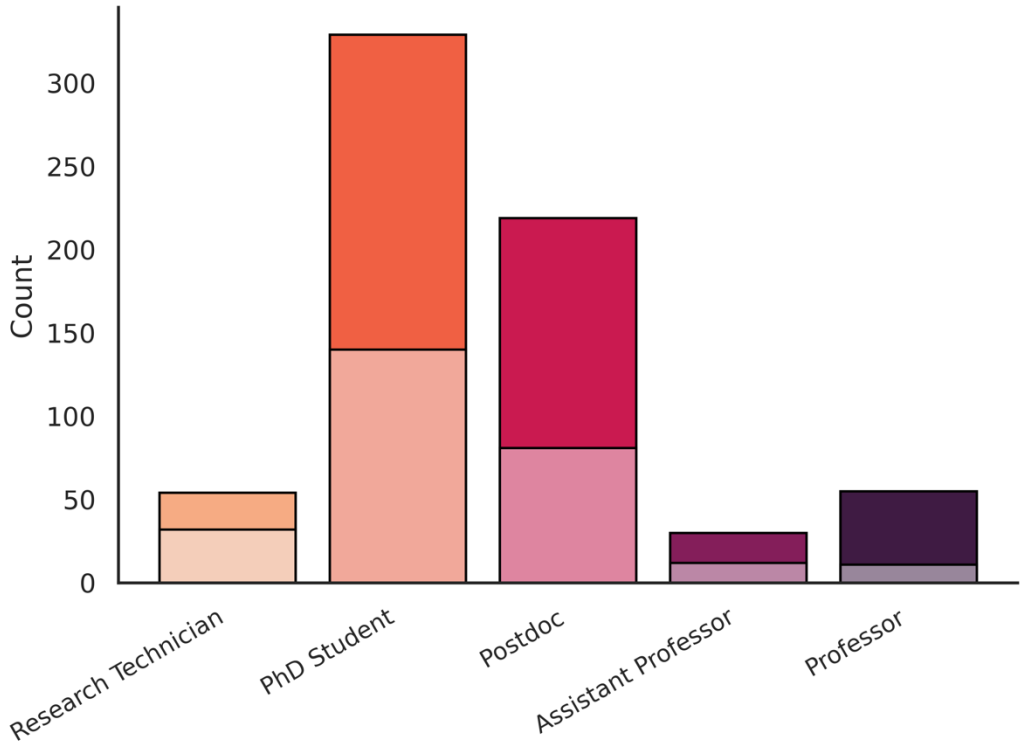
Field



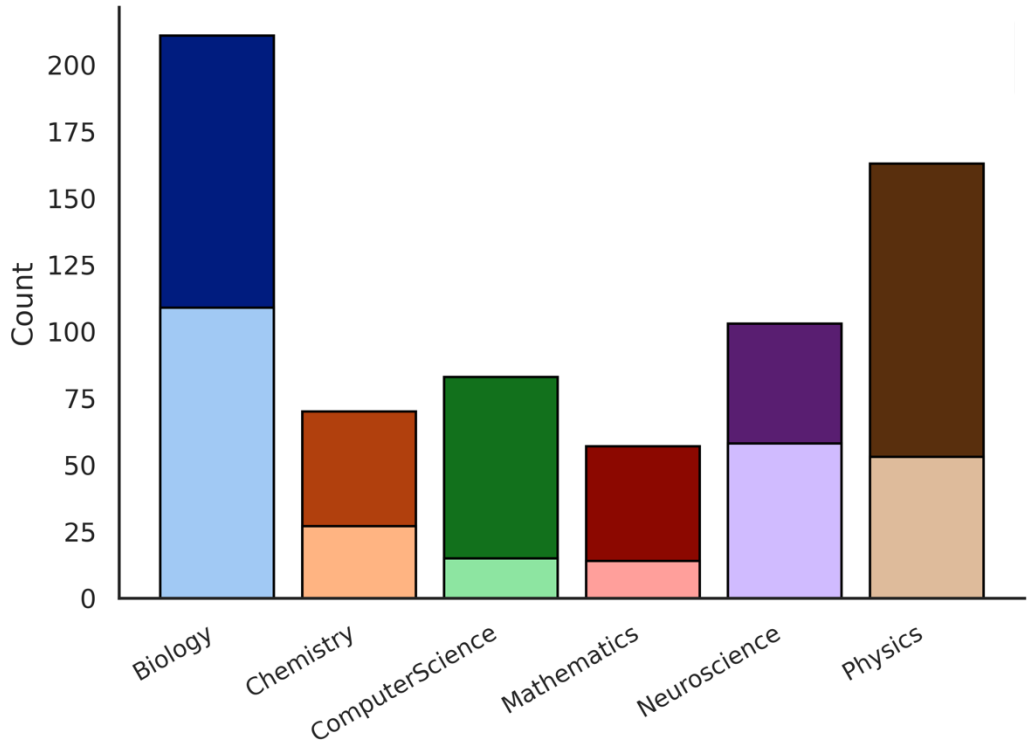
Nominal categories

Number of females by role and field

Academic Role

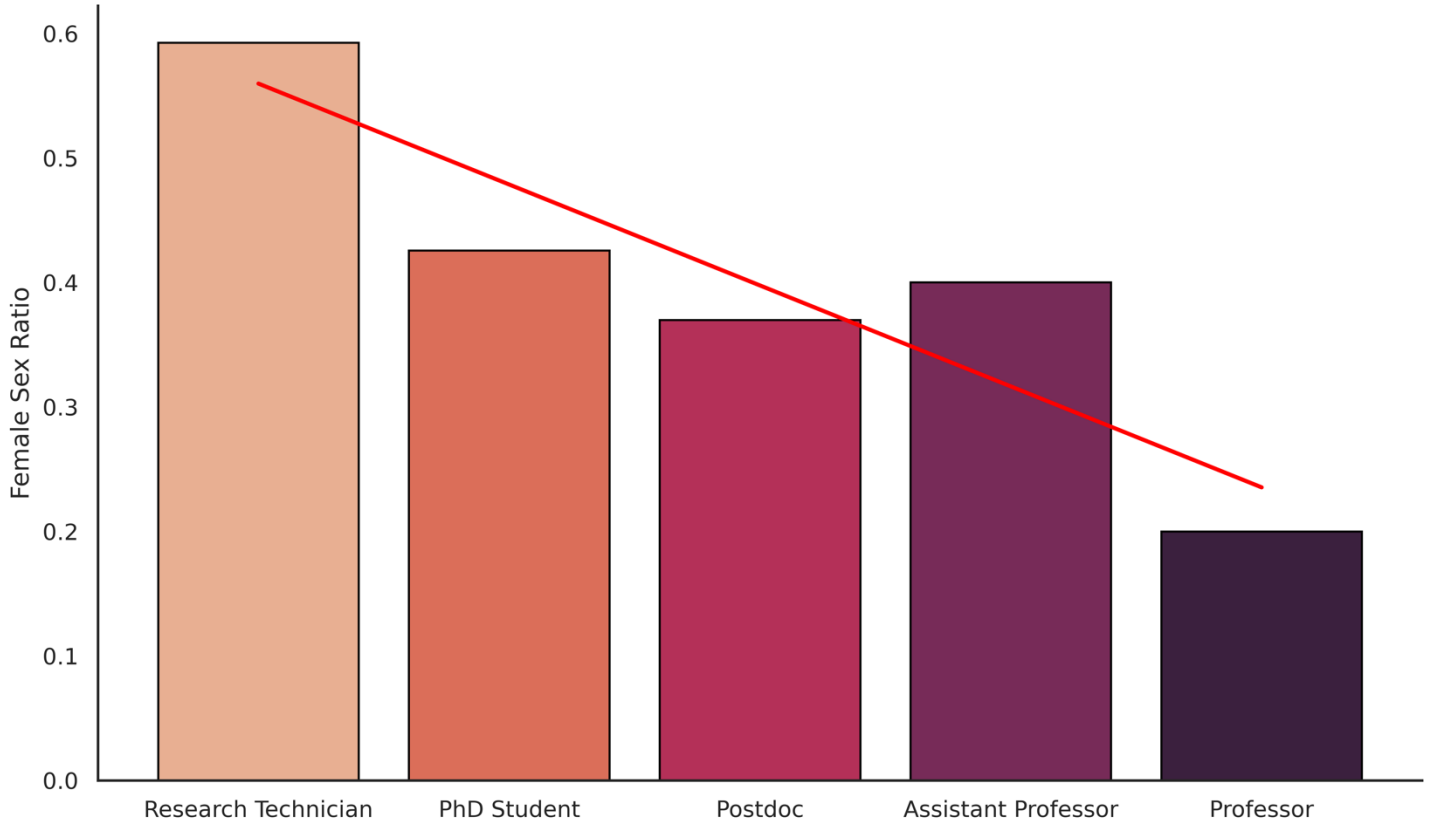


Field



*Lighter shade indicates number of female scientists

A: Leaky Pipeline?



Spearman correlation

Between female sex ratio and position of the role

$\rho_s = -0.899$

p-value = **0.0373**

A1 – There is a *correlation* between female sex ratio and ascending academic roles

*The regression line is only for visualization

B: Female scientists across fields

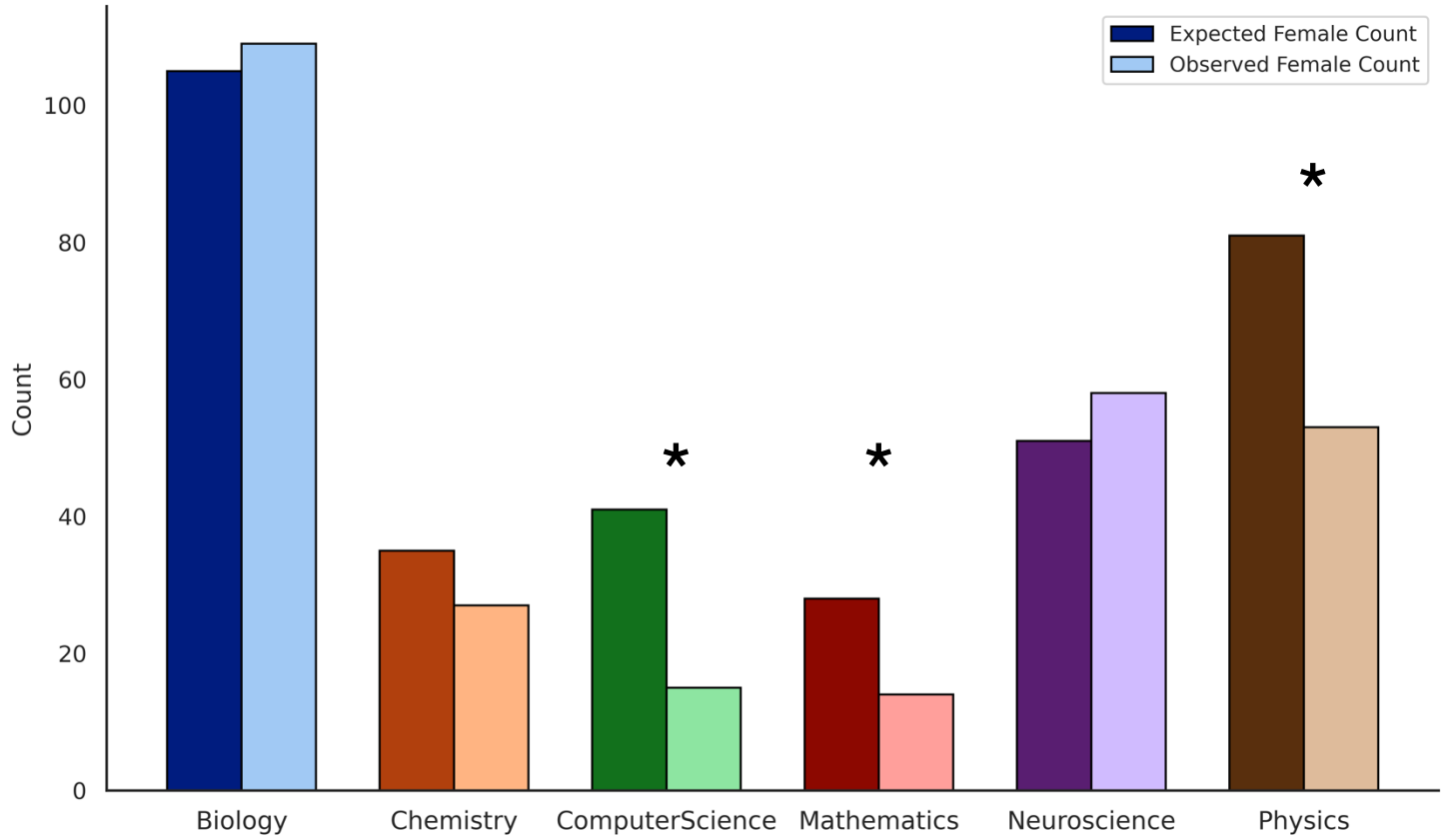
In each field, compare the number of male and female scientists using **One-dimensional Chi-square test**

For instance, in Computer Science -

	Obs	Exp
Female	15	41
Male	68	42

Chi2_stat = 32.58

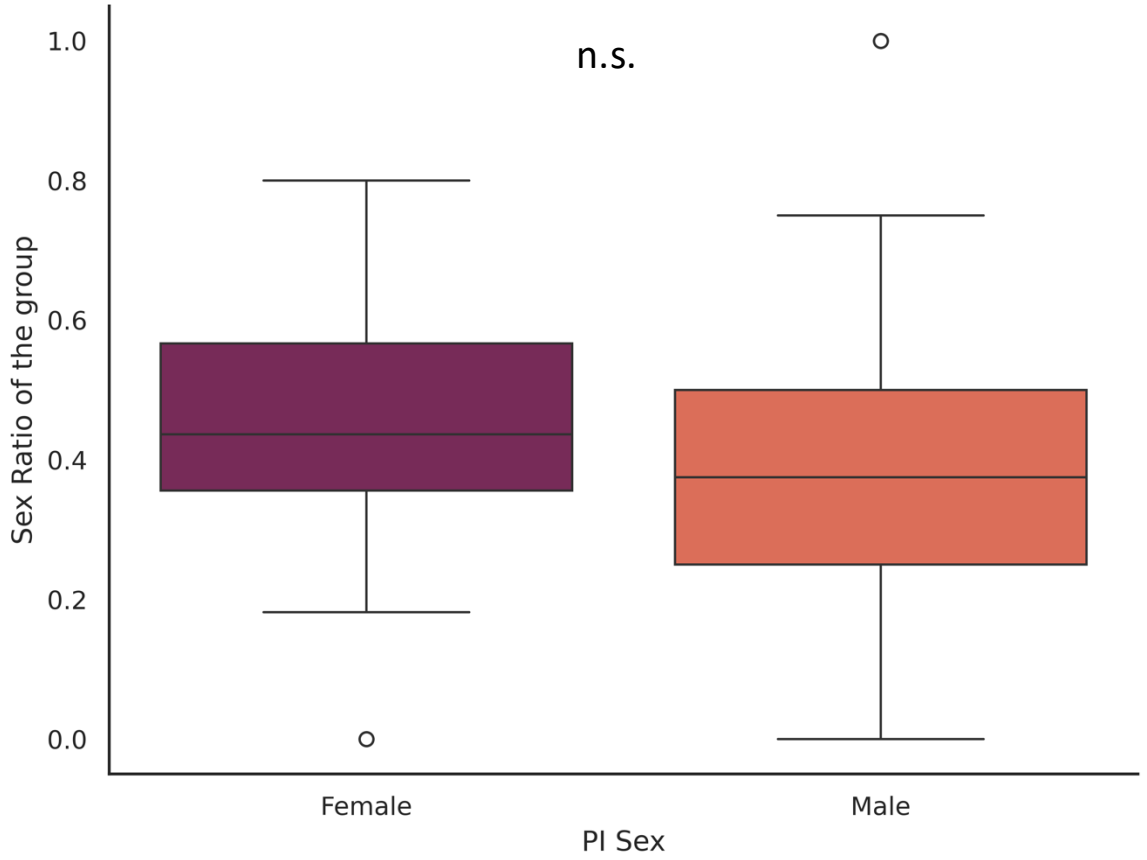
p-value = **1.142e-08**



B1 – There is a *difference* between number of males and females in some fields

*Significance tested with Bonferroni Correction

C: Do Female PIs have more female group members?



Assumptions for parametric test

1. Shapiro-Wilk test for normality

	stat	p-value
Female	0.948	0.337
Male	0.967	0.121

2. Levene test for equal variance: stat = 0.494, p-value = 0.483
3. Independence

Student's T-test for comparison of means

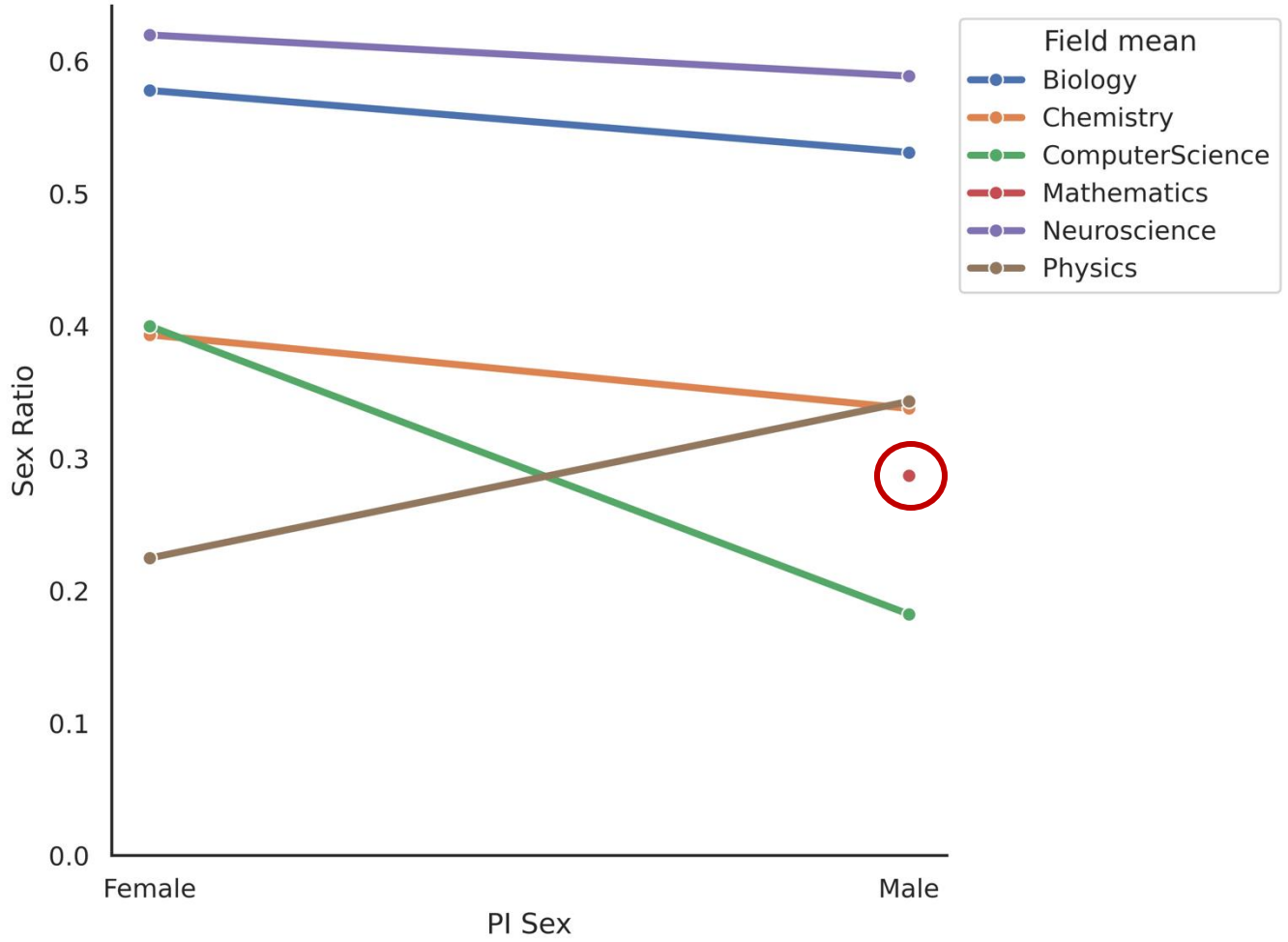
statistic = 1.039, p-value = 0.301, df = 77

C0 – Female and male PIs have *similar* proportion of female students in their groups

C: Breaking it down by field

There are 0/9 female PIs in mathematics and 1/11 in Computer Science.

I couldn't do pair-wise MannWhitneyU test for each field due to lack of data points.



C: Using GLMs

$$model = glm(sex_{ratio} \sim PI_{sex} + Field, data, family = 'binomial')$$

- Used Binomial distribution for proportion data
- Tried different predictor variables and interactions

```
glm(formula = Sex_ratio ~ PI_sex + Field, family = "binomial",
    data = data)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.20263    0.57853   0.350   0.7262
PI_sexmale         -0.02231    0.56590  -0.039   0.9685
FieldChemistry     -0.75814    0.90223  -0.840   0.4007
FieldComputerScience -1.55562    0.88007  -1.768   0.0771 .
FieldMathematics   -1.09013    0.87783  -1.242   0.2143
FieldNeuroscience   0.21095    0.78012   0.270   0.7868
FieldPhysics       -0.98990    0.64476  -1.535   0.1247
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 20.504  on 78  degrees of freedom
Residual deviance: 13.822  on 72  degrees of freedom
AIC: 93.36

Number of Fisher Scoring iterations: 4
```

No significant results from any of the models.

Back to the hypotheses

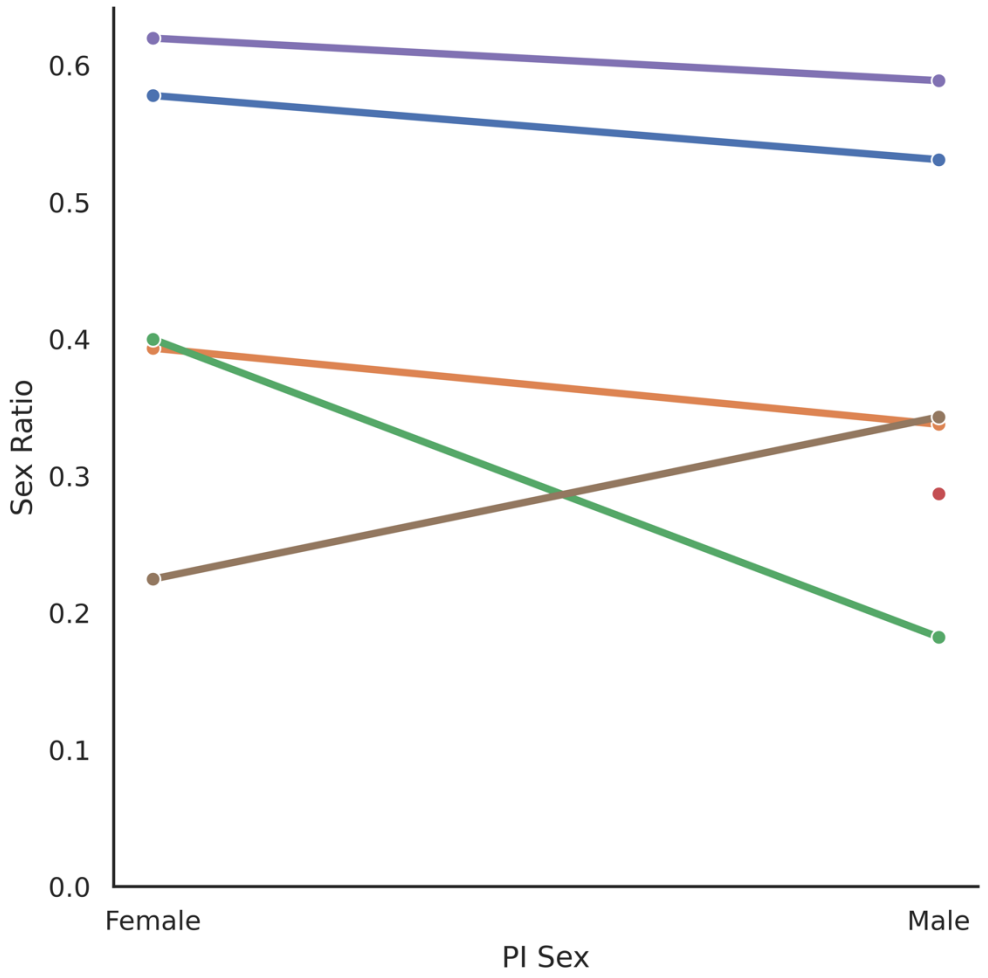
- A0 – There is *no correlation* between female sex ratio and ascending academic roles
- **A1 – There is a *correlation* between female sex ratio and ascending academic roles**

- B0 – There is *no difference* between number of males and females in a field
- **B1 – There is a *difference* between number of males and females in a field**

- **C0 – Female and male PIs have *similar* proportion of female students in their group**
- C1 – Female and male PIs have *different* proportions of female students in their group

Discussion

- The leaky pipeline is still leaky.
- We need to talk about perceptions about ‘soft’ and ‘hard’ sciences and how it affects the diversity within.
- There seems to be no effect of sex of the PI on the sex ratio in the group, which can be good.
- But given this graph (and some conversations), I think more data in math, CS and physics could show an effect.



Caveats

1. Field boundaries can be changed for PIs
2. Excluded interns
3. It's a snapshot – no analysis of past students or the current 2024 batch
4. Could check the effect of group size and age
5. Excluded PIs without a team in analysis of question C, but not A or B
6. ISTA is still growing

Reflections

- My categories are binary, but **biological sex is a spectrum**.
- One of the biggest factors that is important (but outside the scope of this analysis) is capturing the **intersectionality** in terms of nationality, race, language, sexuality and other dimensions that might play a more significant role.
- I had a lot of fun working on it, because it's an **answerable question** – whatever the answer.

Thank you for listening!

Questions?