Gender bias in representation and publication rates across sub-fields

LSA 2019 Annual Meeting

Authors

- Hanna Muller
- Phoebe Gaston
- Bethany Dickerson
- Adam Liter
- Karthik Durvasula
- Mina Hirzel
- Kasia Hitczenko
- Maggie Kandel
- Paulina Lyskawa
- Jackie Nelligan
- Max Papillon
- Laurel Perkins
- Alicia Parrish

Bias in Linguistics

- Graduate students & faculty at Michigan State University, University of Maryland, UMass Amherst, NYU, Harvard
- Goals:
    - Collect data identifying where and why bias exists in the field.
    - Make that data publicly available.
    - Raise awareness and discuss solutions.

Outline

- Part 1: Evidence for a leaky pipeline in linguistics
- Part 2: Gender bias in publication rates
- Part 3: Potential causal factors

Leaky Pipelines

- Under-representation of women in STEM fields is known to be a problem, despite equal or over-representation at the undergraduate level.

- This pattern is the hallmark of a *leaky pipeline*:

    - Women disproportionately leave a field at each successive level.

Leaky Pipelines

- To what extent is this true in linguistics, specifically?
  - BIL collected representation data from 49 linguistics departments.
  - Available (anonymized) at biasinlinguistics.org

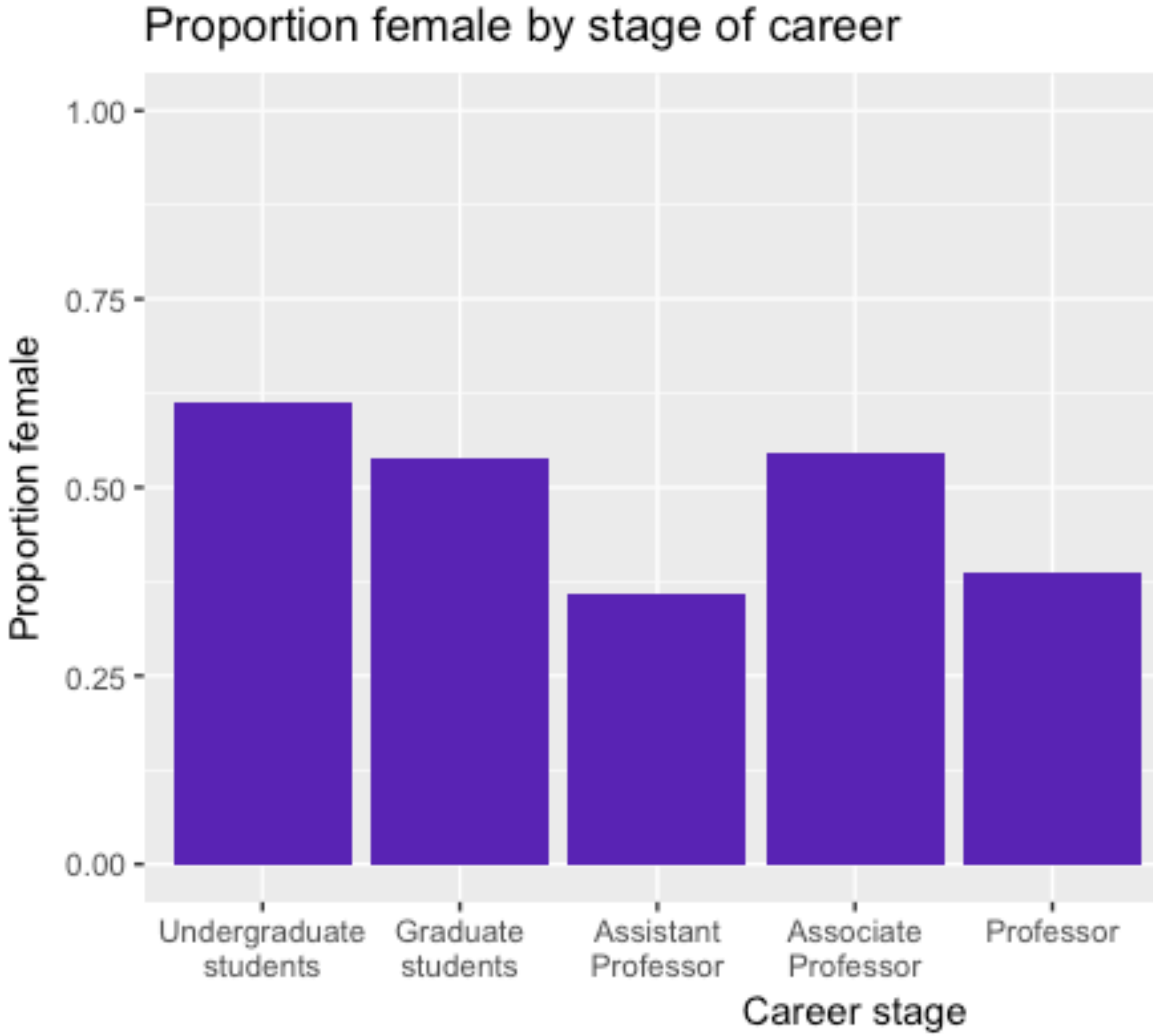
Methods

- Student demographics:
  - 29/49 department chairs provided a count of graduate students by gender and subfield.
  - 15/29 provided undergraduate data.
  - 995 students in our dataset.

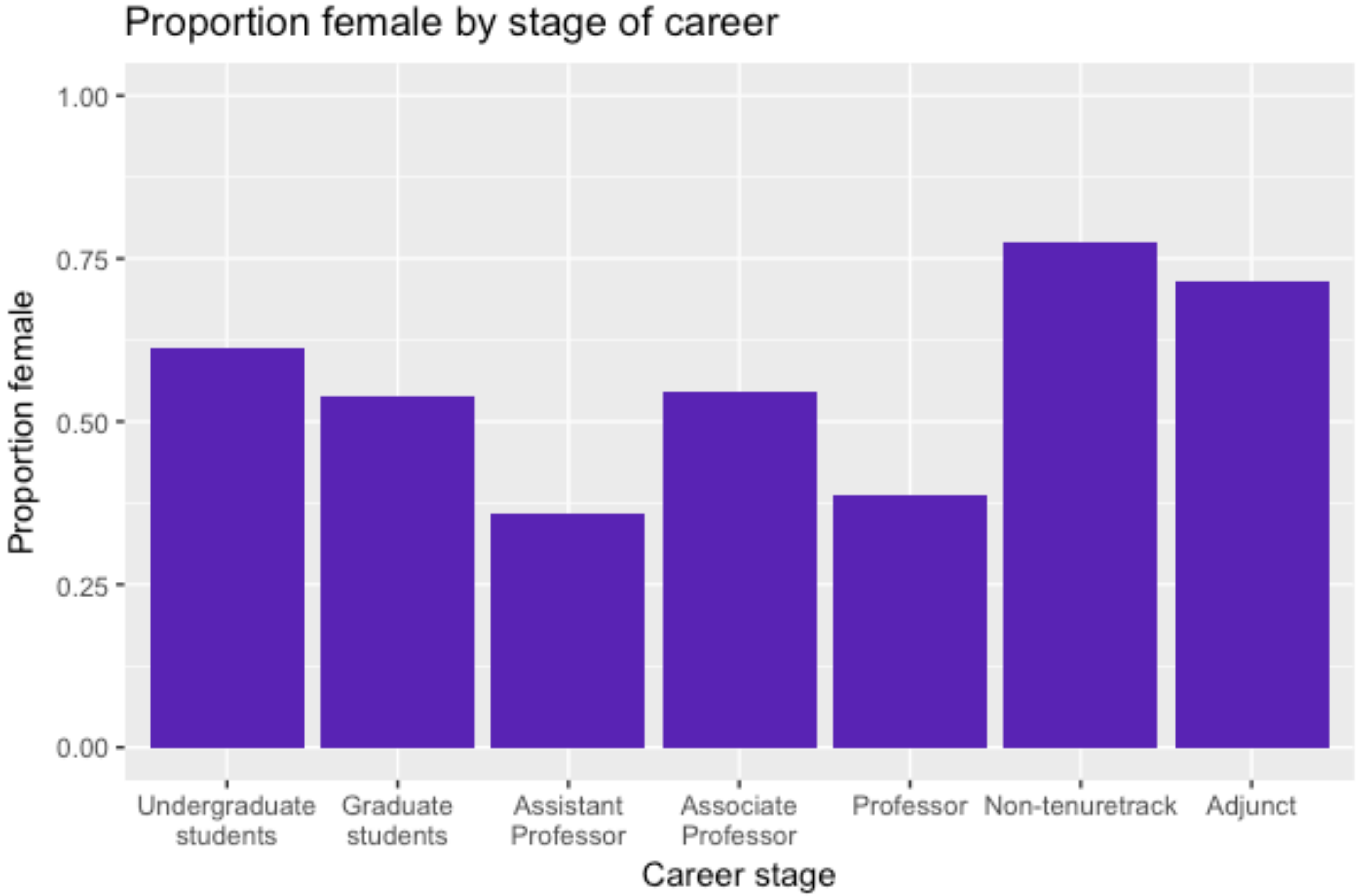
Methods

- Faculty demographics:
  - Sub-fields and positions taken from department websites for all 49 departments
  - 810 faculty members in our dataset
  - Hand-tagged for gender

Part 1


Proportion female by stage of career
Proportion female
1.00
0.75
0.50
0.25
0.00
Undergraduate students
Graduate students
Assistant Professor
Associate Professor
Professor
Career stage

Part 1


Proportion female by stage of career
Proportion female
1.00
0.75
0.50
0.25
0.00
Undergraduate students
Graduate students
Assistant Professor
Associate Professor
Professor
Non-tenuretrack
Adjunct
Career stage

Is this a leaky pipeline?

- We think yes: women are leaving at higher rates.
- Could this just be a hold-over from previous imbalances that have persisted due to the tenure system?
    - Unlikely, since there are also severe drop-offs in the earlier, inherently temporary stages.

Why would the pipeline leak?

- Systemic factors that lead women to “choose” to leave:
    - e.g., insufficient parental leave or childcare options
    - e.g., harassment, toxic work environments

Monroe et al (2008); Lober Newsome (2008); Mason et al (2013); Williams (2005)

Why would the pipeline leak?

- Discrimination in hiring decisions (overt or implicit)
- Hiring based on metrics that are themselves biased:
  - e.g., publication rates, citation rates, teaching evaluations, letters of recommendation, etc

Rivera (2017); Ceci & Williams (2015); Moss-Racusin et al (2012); Grunspan et al (2016); Trix & Psenka (2003); Madera et al (2009); Madera et al (2018); Schmader et al (2007); Knobloch-Westerwick & Carroll (2011); Maliniak et al (2013); King et al (2015); Schroeder et al (2013); Nittrouer et al (2018); MacNell et al (2014); Miles & House (2015); Boring et al (2016); Wagner et al (2016); Mengel et al (2017); Milkman et al (2015); van der Lee & Ellemers (2015); Witteman et al (2018)

Publication Rates

- Advancing in academia is heavily dependent on publication rate.
- If women are publishing less, this could be one factor limiting advancement.

Schucan Bird (2011); West et al (2013); Lariviere et al (2013); Filardo et al (2016); Theule Lubienski et al (2017)

Importance of small effects

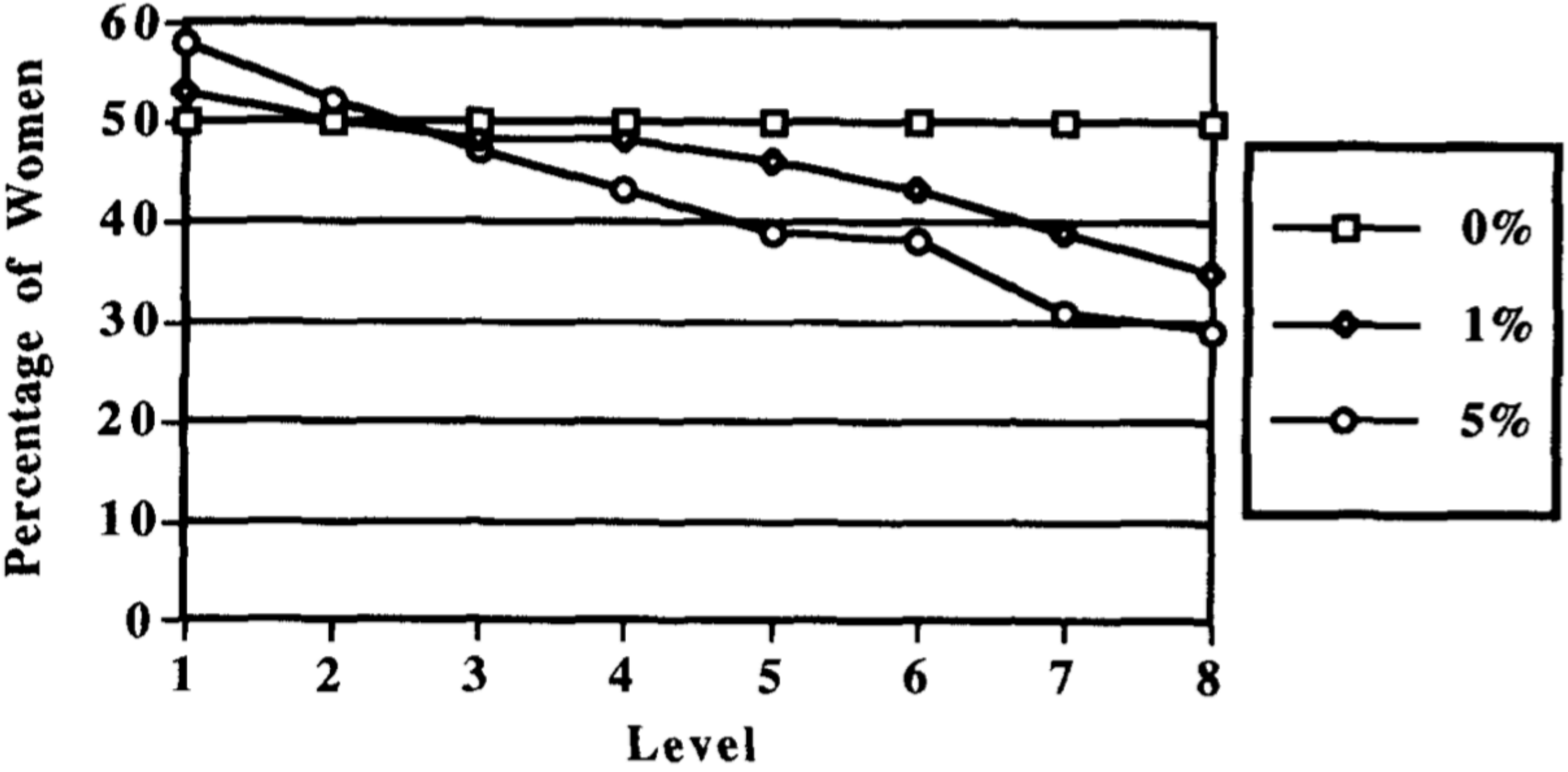
- How small of an effect should we care about?
- Simulations show that:
  - Small gender differences in performance scores will quickly propagate upwards in a workplace hierarchy.
  - This leads to large differences in promotion rates and therefore in representation at higher levels.

Martell, Lane & Emrich (1996)

Importance of small effects

Figure 1

Percentage of Women at Each Position Level, With 0%, 1%, and 5% of the Effect Size Variance Attributed to Sex



Martell, Lane & Emrich (1996)

Goal

- There is no previous data on gender bias in publication rates for linguistics.
- We are trying to establish whether bias exists.
- If so, does it vary by subfield?

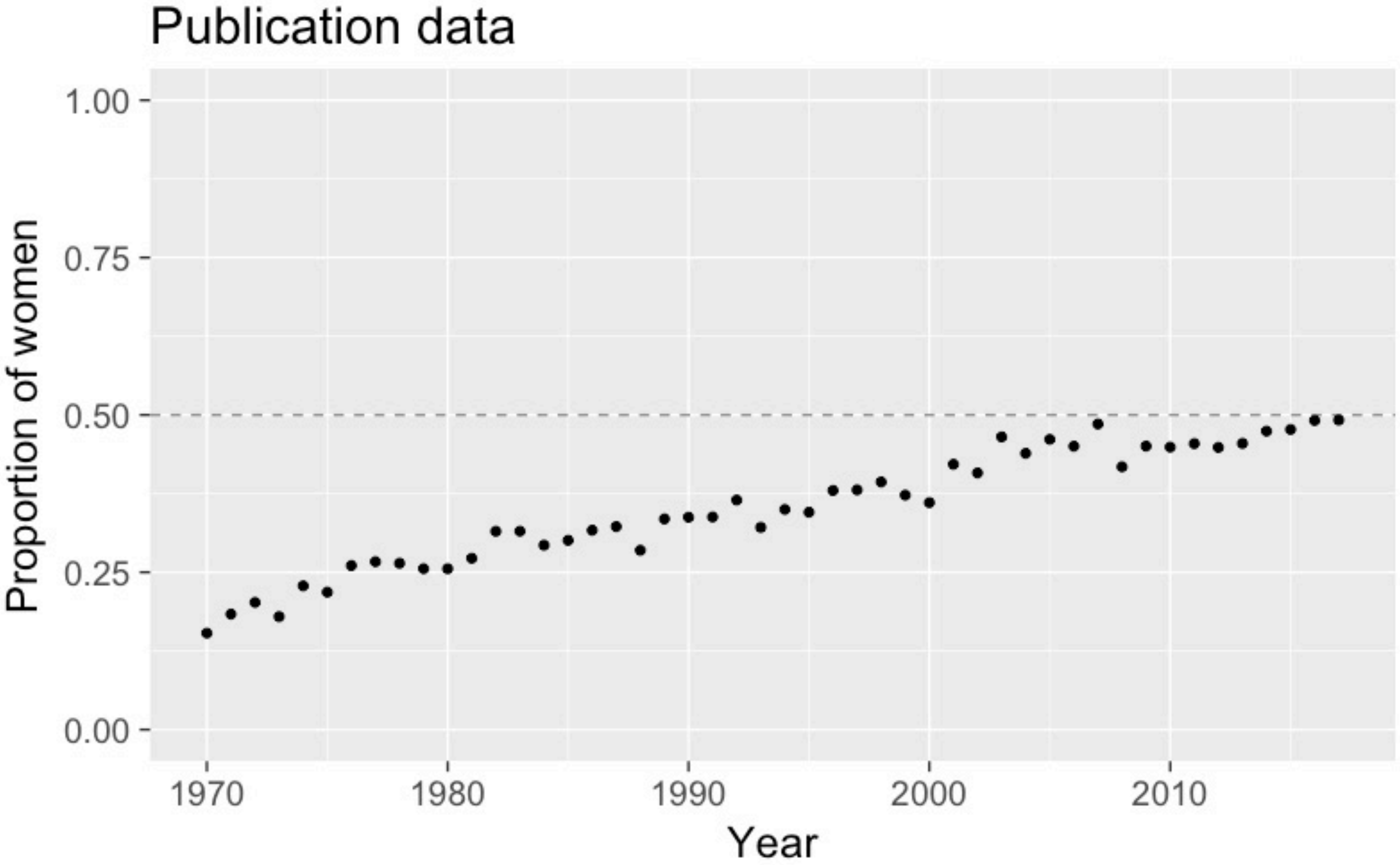
Methods/Data

- We looked at publishing rates for male and female linguists from 1970 to the present (using Crossref via the R package rcrossref).
- Extracted all available citation data (title, year, authors) from 31 journals across the following sub-fields:
    - Syntax, Semantics, Phonology/Phonetics, Language Acquisition, Psycholinguistics
    - plus domain-general linguistics journals that cover multiple sub-fields
    - Sociolinguistics & computational linguistics are excluded for lack/abundance of data, respectively.

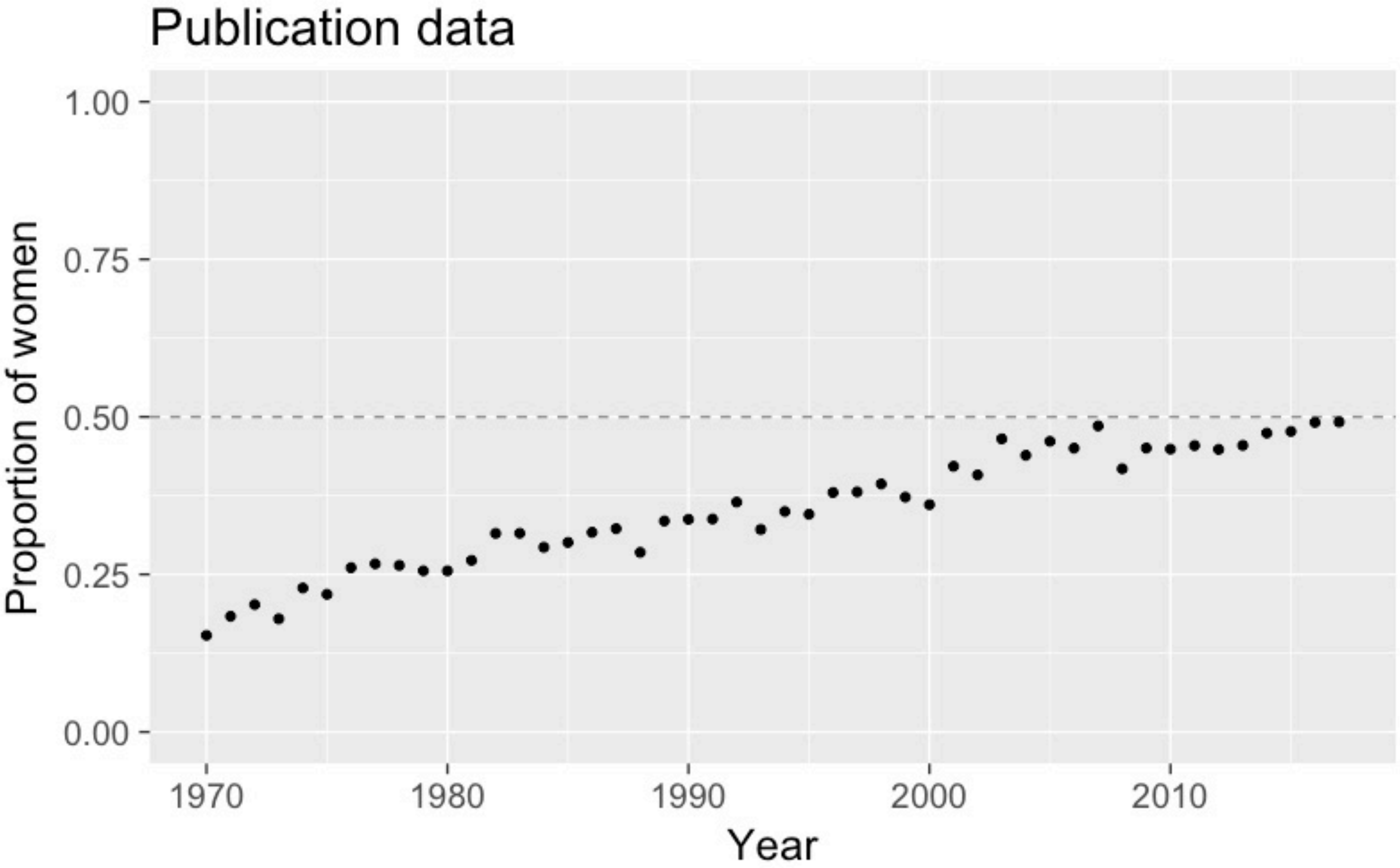
Methods/Data

- For each instance of authorship, we automatically tagged gender using the genderizeR package in R.
- Validated this by testing automatic tags for the 810 faculty linguists from the initial data set:
  - 97% accurate for the 90% of that group it tagged
- Result: 87,000 instances of gender-tagged authorship in the dataset

Publication proportion



Publication proportion



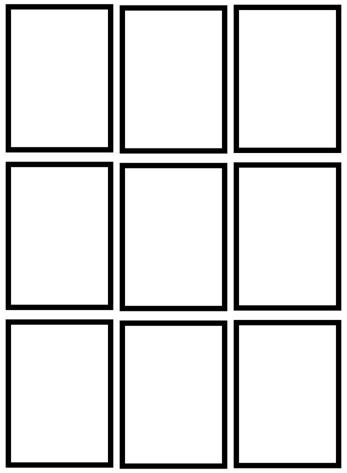
- From this, we can’t tell if there are fewer female linguists or female linguists publish less than male linguists.

Representation estimate

- We need some estimate of how many male vs female linguists are currently active in the field.

Representation estimate

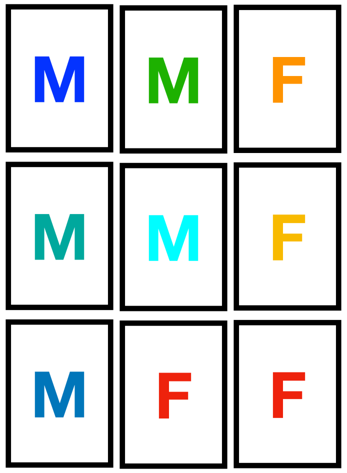
- We need some estimate of how many male vs female linguists are currently active in the field.



2015

Representation estimate

- We need some estimate of how many male vs female linguists are currently active in the field.

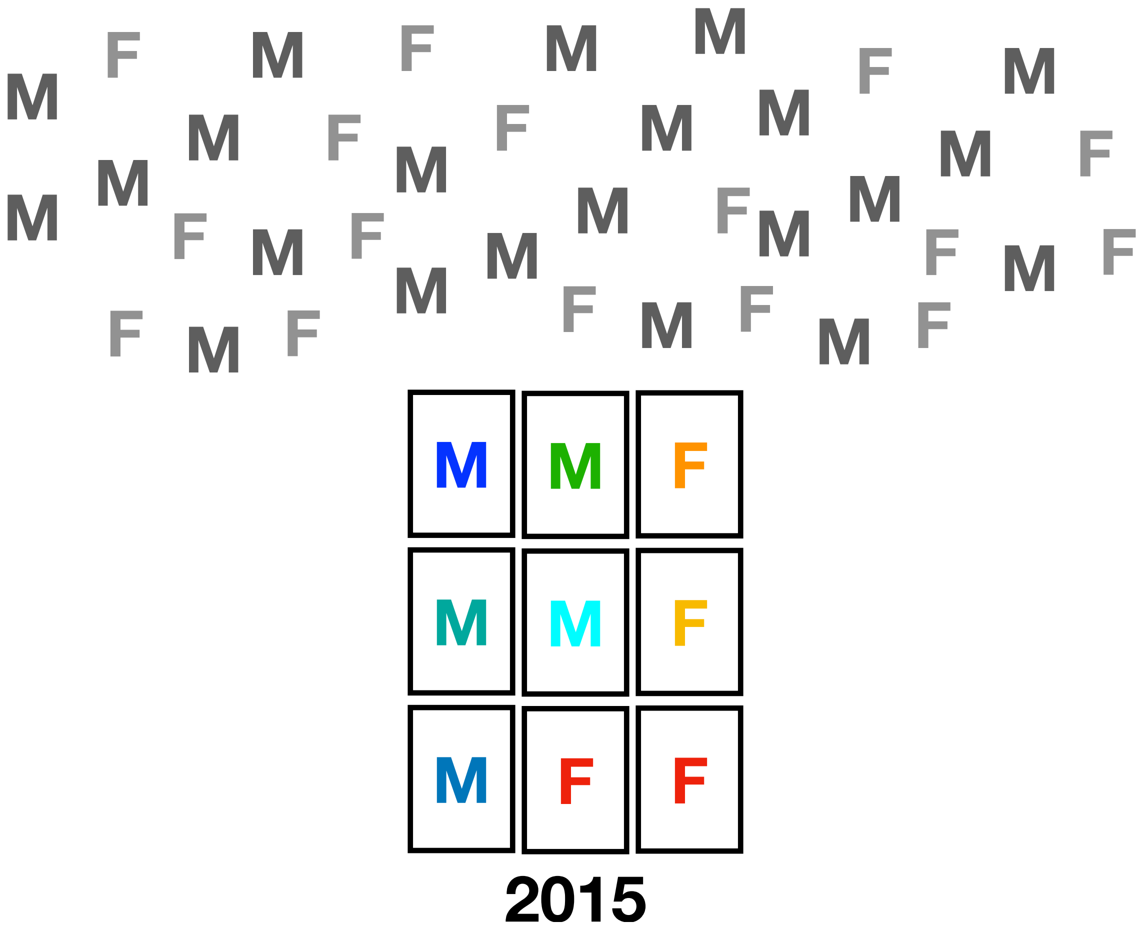


M M F
M M F
M F F

2015

Representation estimate

- We need some estimate of how many male vs female linguists are currently active in the field.

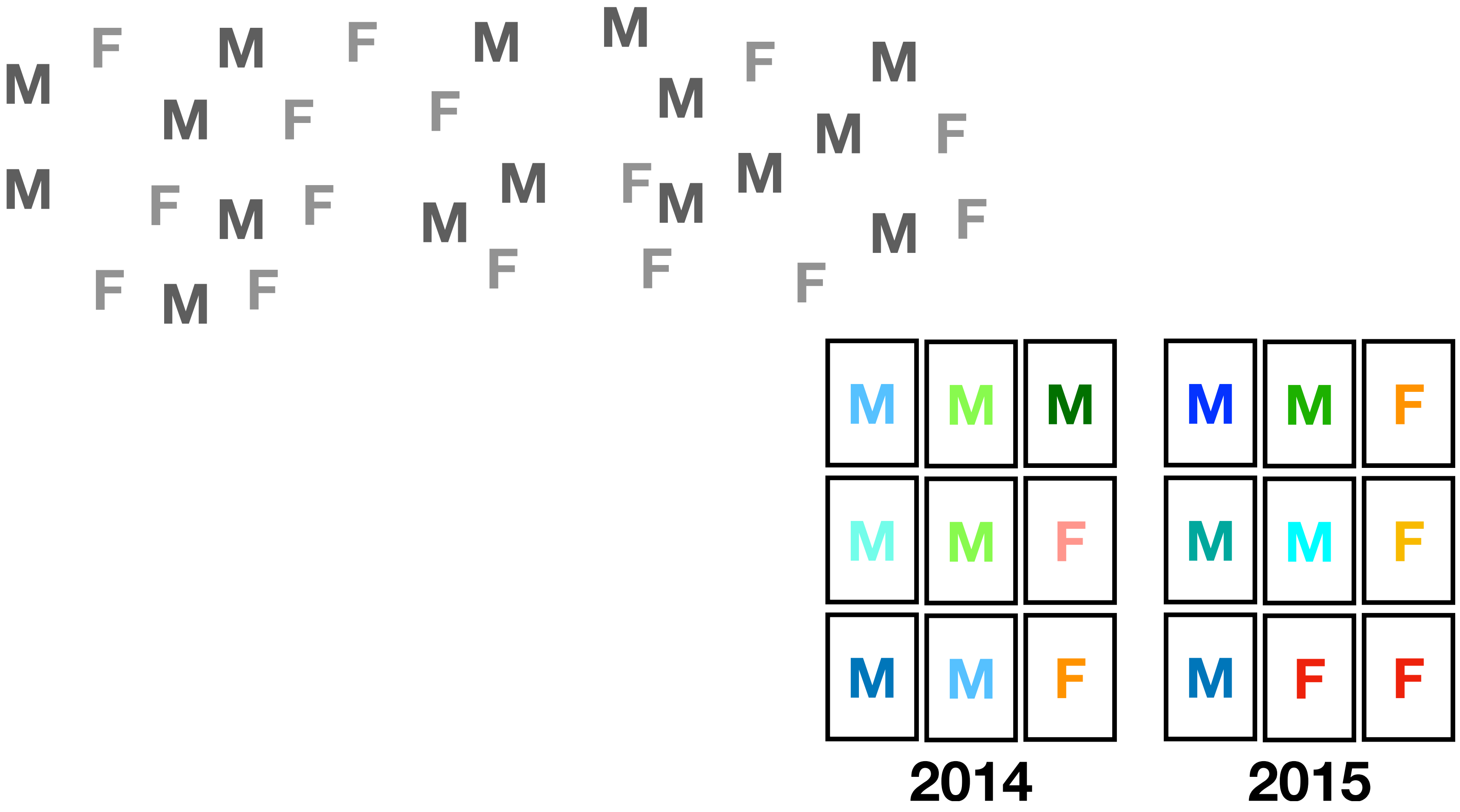


M	M	F
M	M	F
M	F	F

2015

Representation estimate

- We need some estimate of how many male vs female linguists are currently active in the field.



M	M	M
M	M	F
M	M	F

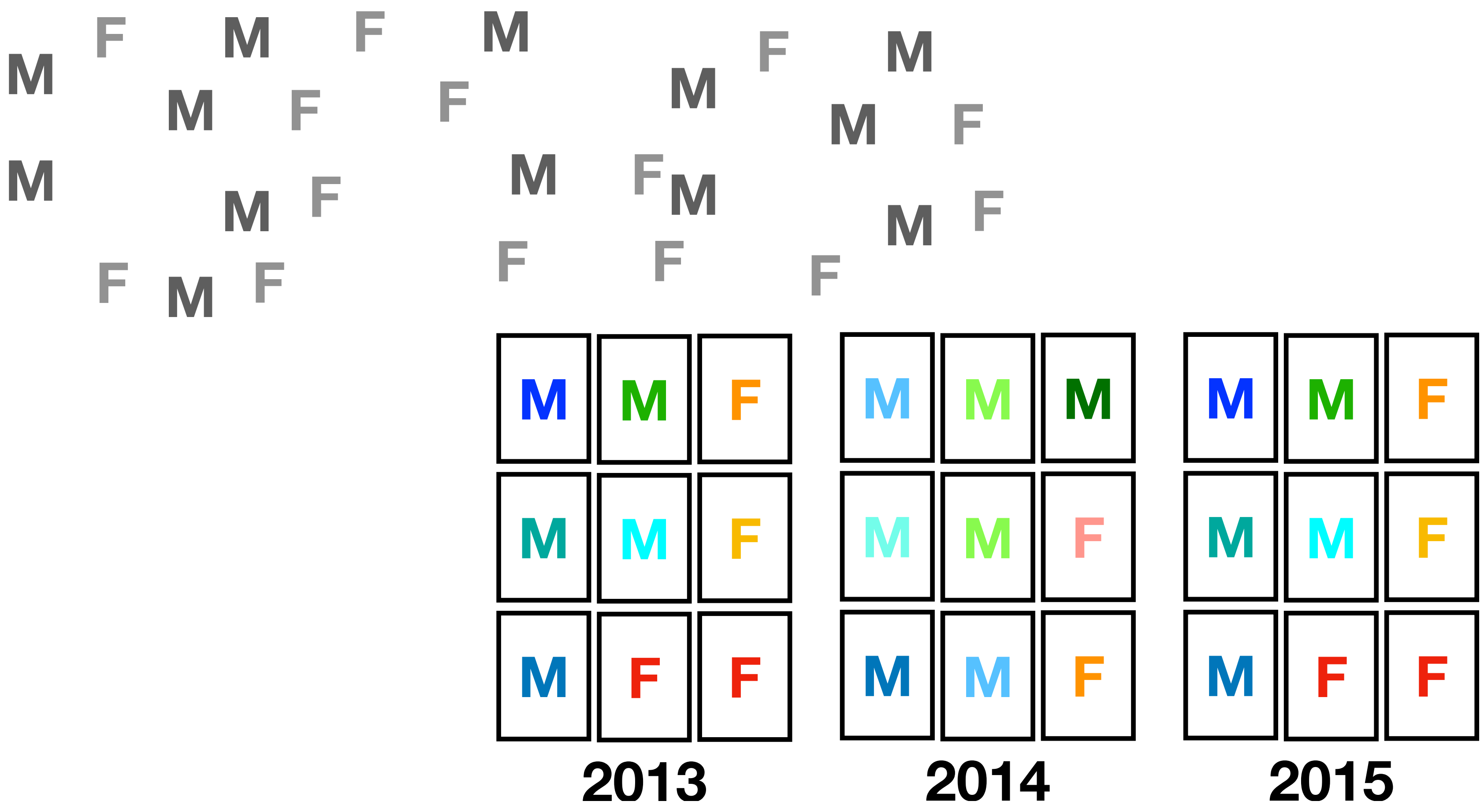
2014

M	M	F
M	M	F
M	F	F

2015

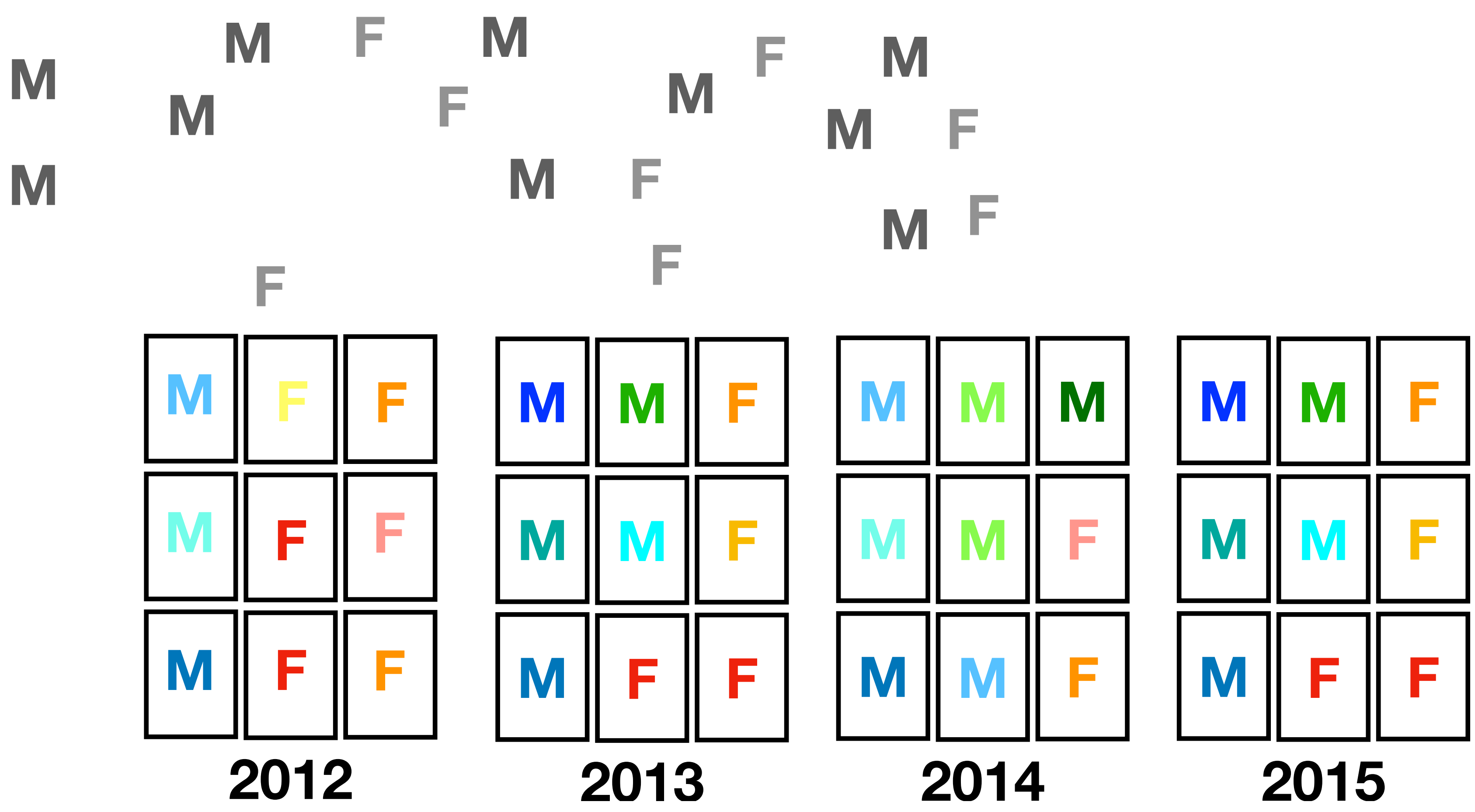
Representation estimate

- We need some estimate of how many male vs female linguists are currently active in the field.



Representation estimate

- We need some estimate of how many male vs female linguists are currently active in the field.



M	F	F
M	F	F
M	F	F

2012

M	M	F
M	M	F
M	F	F

2013

M	M	M
M	M	F
M	M	F

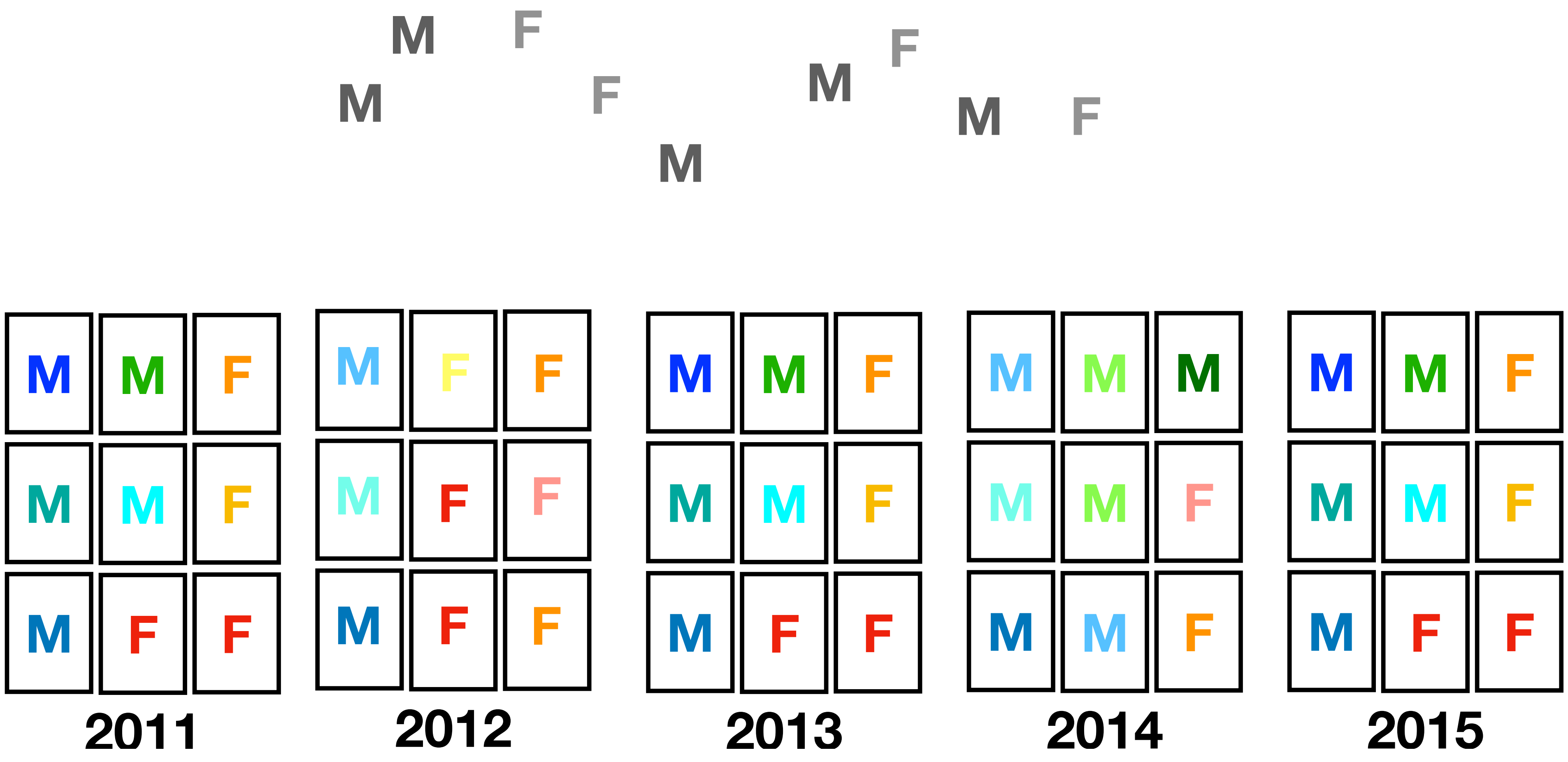
2014

M	M	F
M	M	F
M	F	F

2015

Representation estimate

- We need some estimate of how many male vs female linguists are currently active in the field.



M	M	F
M	M	F
M	F	F

2011

M	F	F
M	F	F
M	F	F

2012

M	M	F
M	M	F
M	F	F

2013

M	M	M
M	M	F
M	M	F

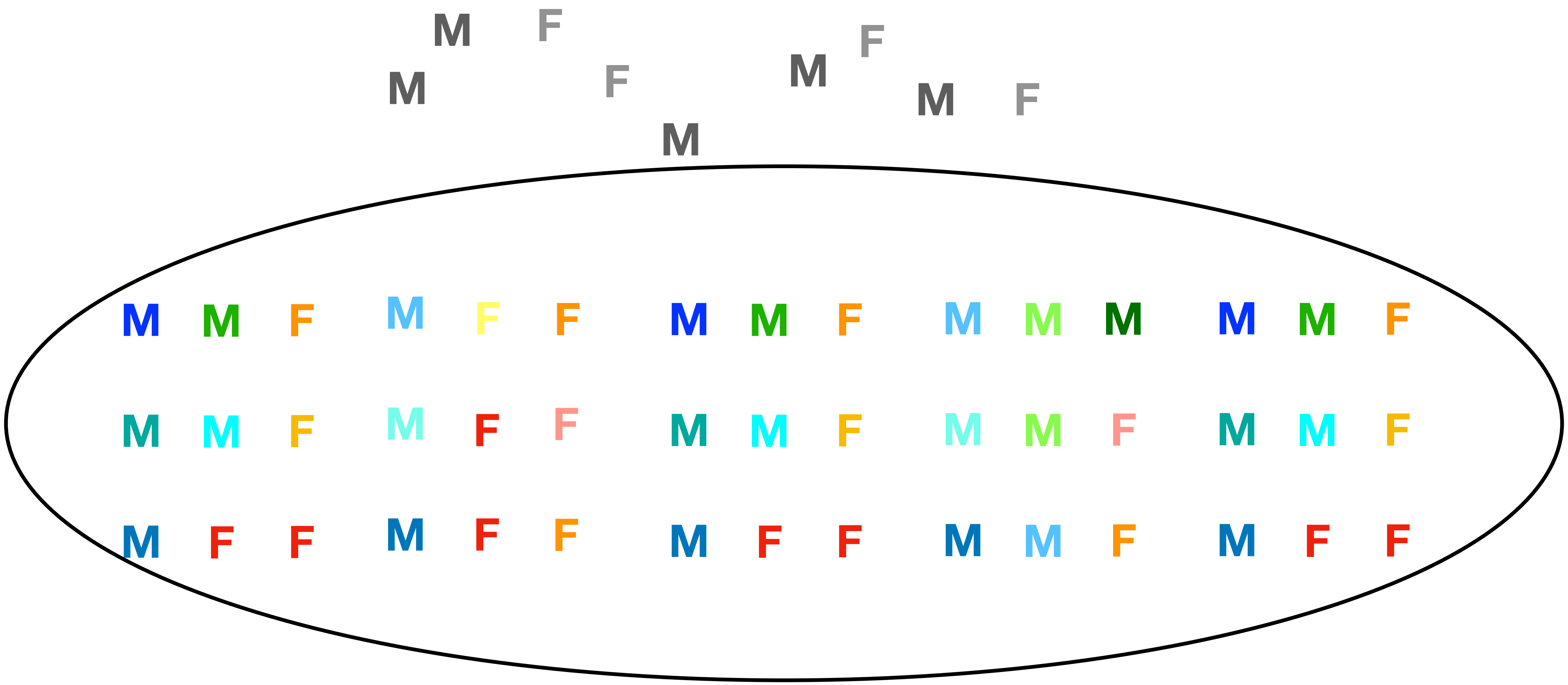
2014

M	M	F
M	M	F
M	F	F

2015

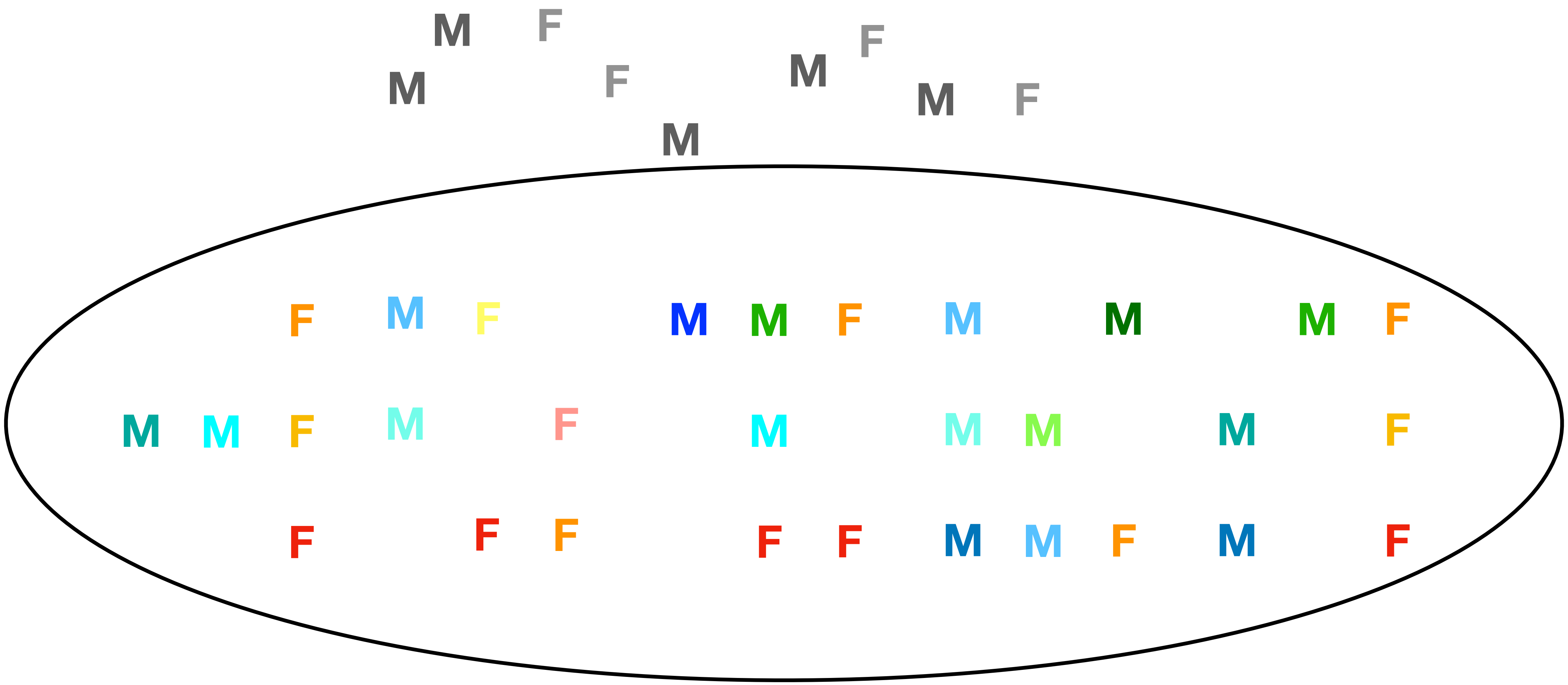
Representation estimate

- We need some estimate of how many male vs female linguists are currently active in the field.


M F
M F
F
M F
M M F
M
M M F M F F M M F M M M M M F
M M F M F F M M F M M F M M F
M F F M F F M F F M M F M F F

Representation estimate

- We need some estimate of how many male vs female linguists are currently active in the field.



Representation estimate

- How to compare representation and publication rates?

Representation estimate

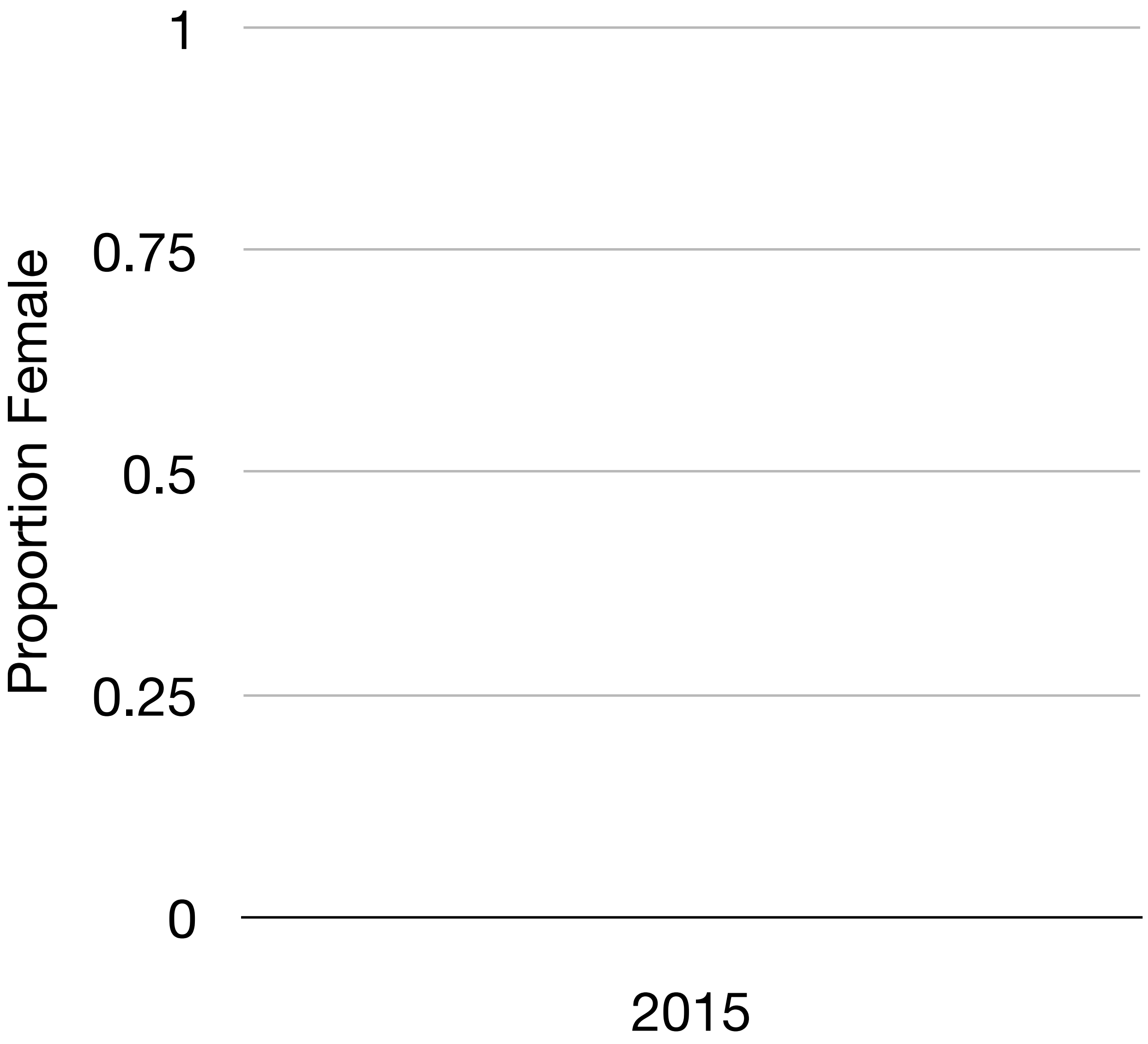
- Equal representation and publication rates

M	**F**
M	**F**
M	**F**
M	**F**
M	**F**
M	**F**

Representation estimate

- Equal representation and publication rates

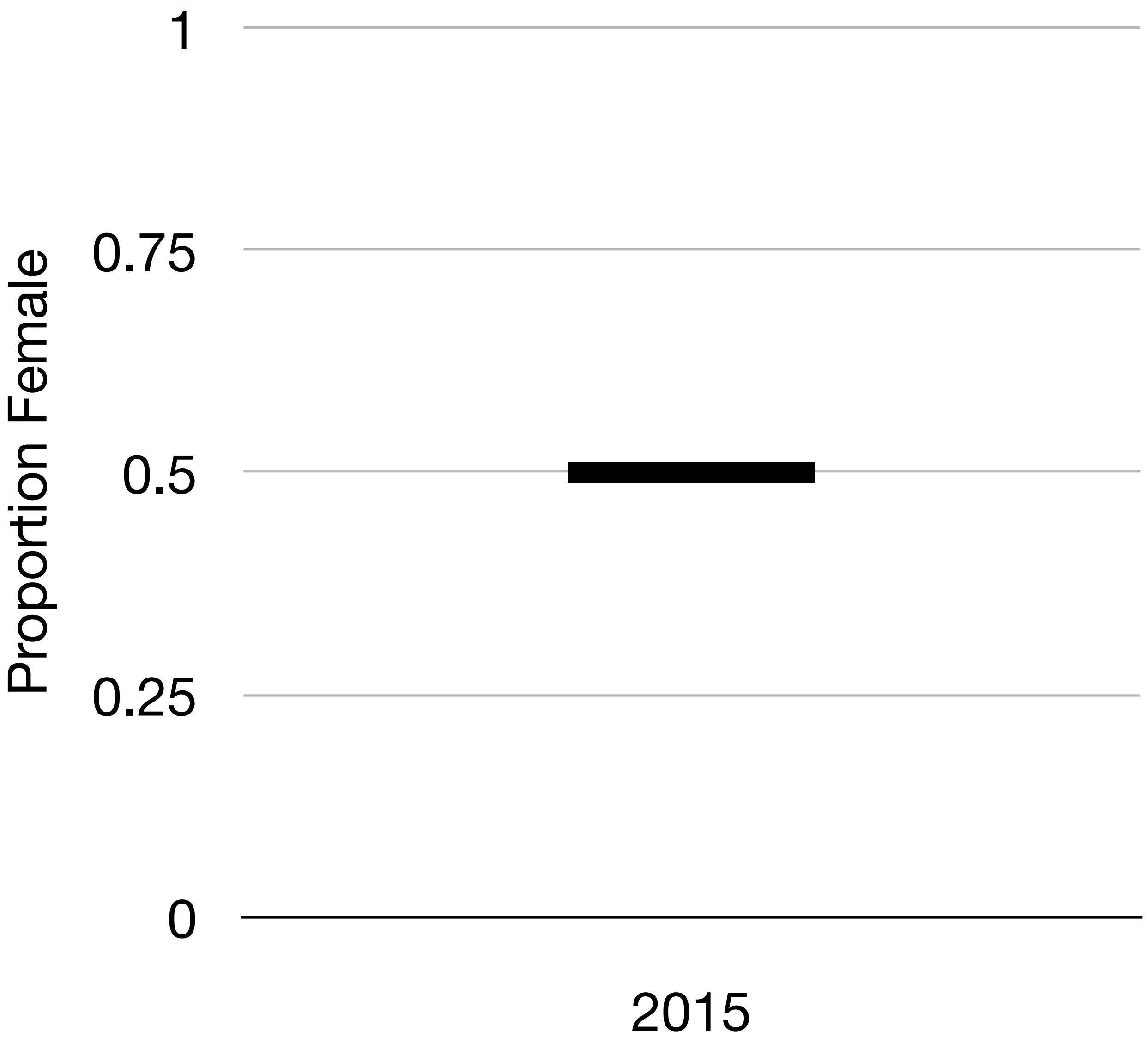
M	**F**
M	**F**
M	**F**
M	**F**
M	**F**
M	**F**



Representation estimate

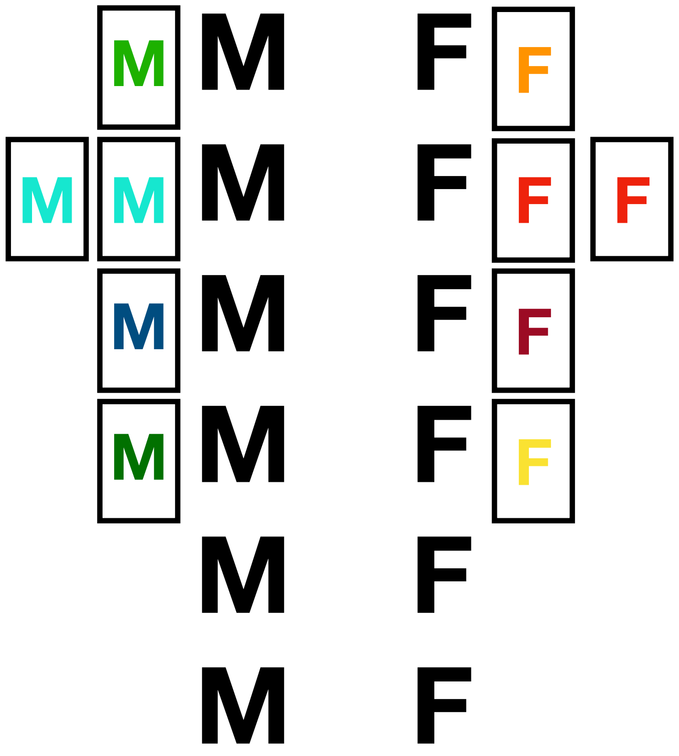
- Equal representation and publication rates

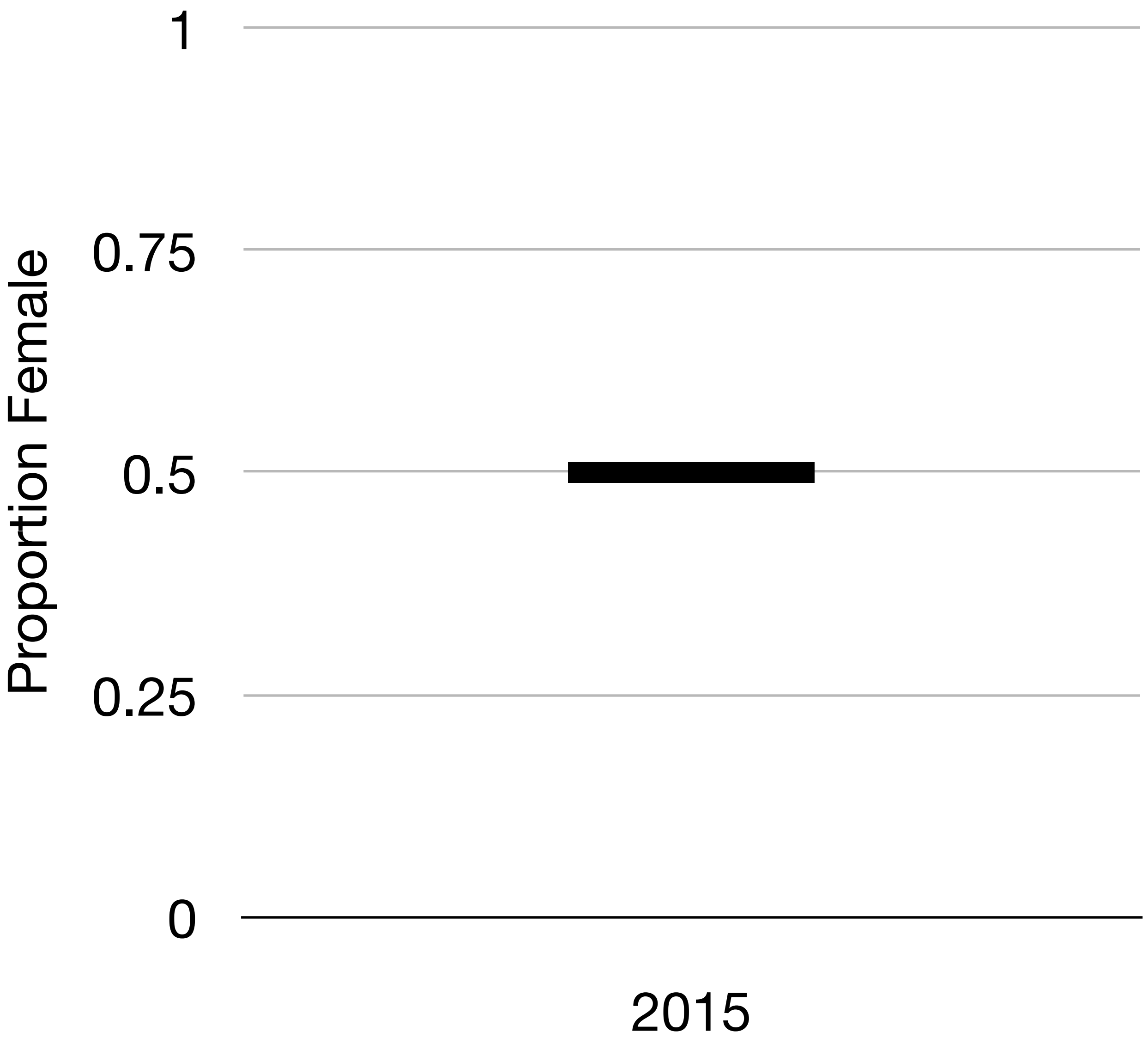
M F
M F
M F
M F
M F
M F



Representation estimate

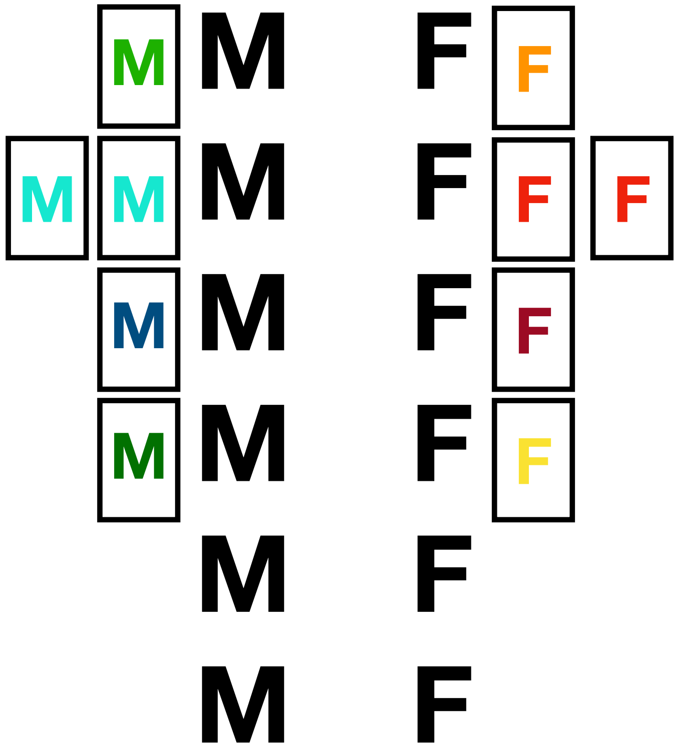
- Equal representation and publication rates

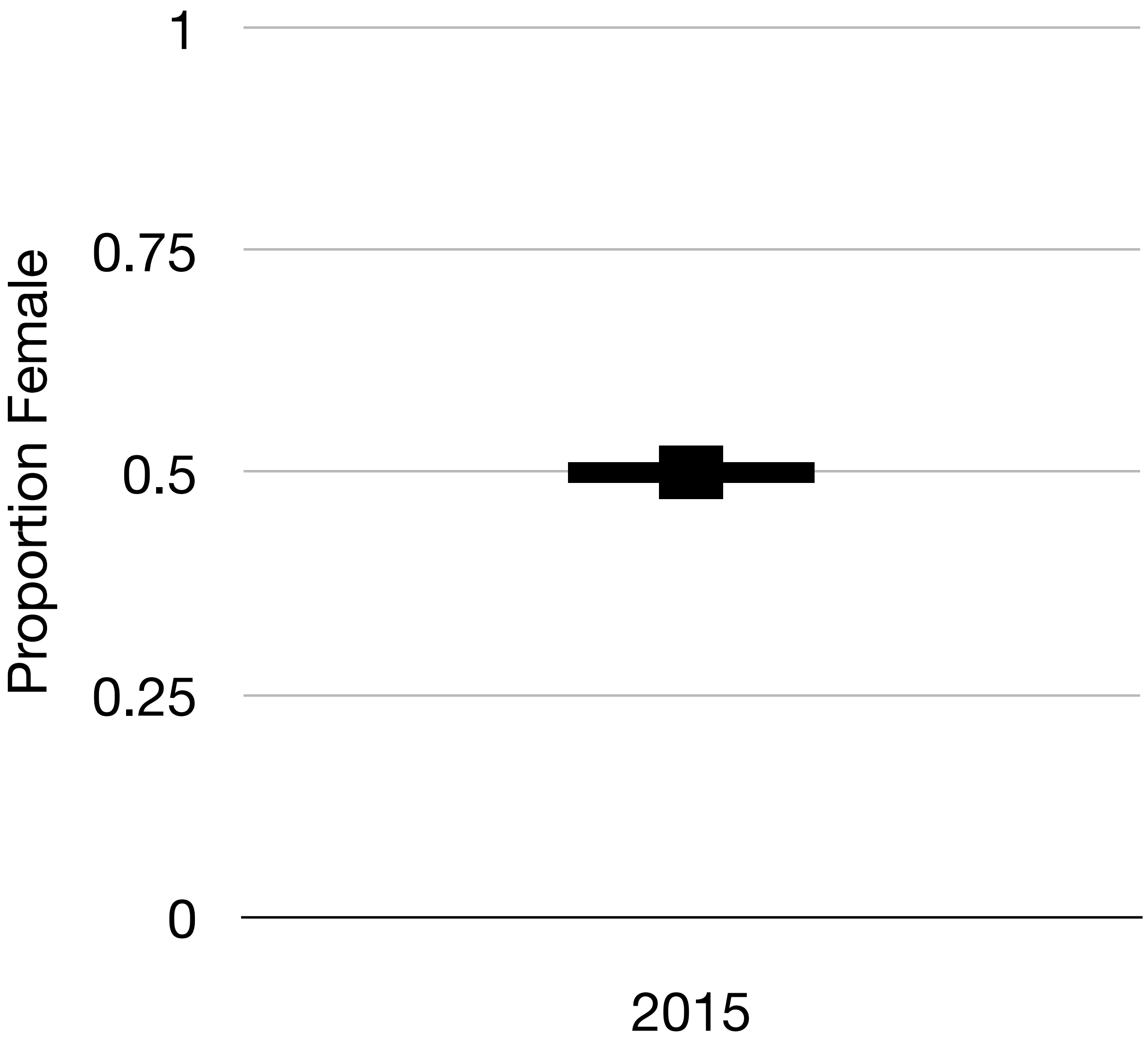




Representation estimate

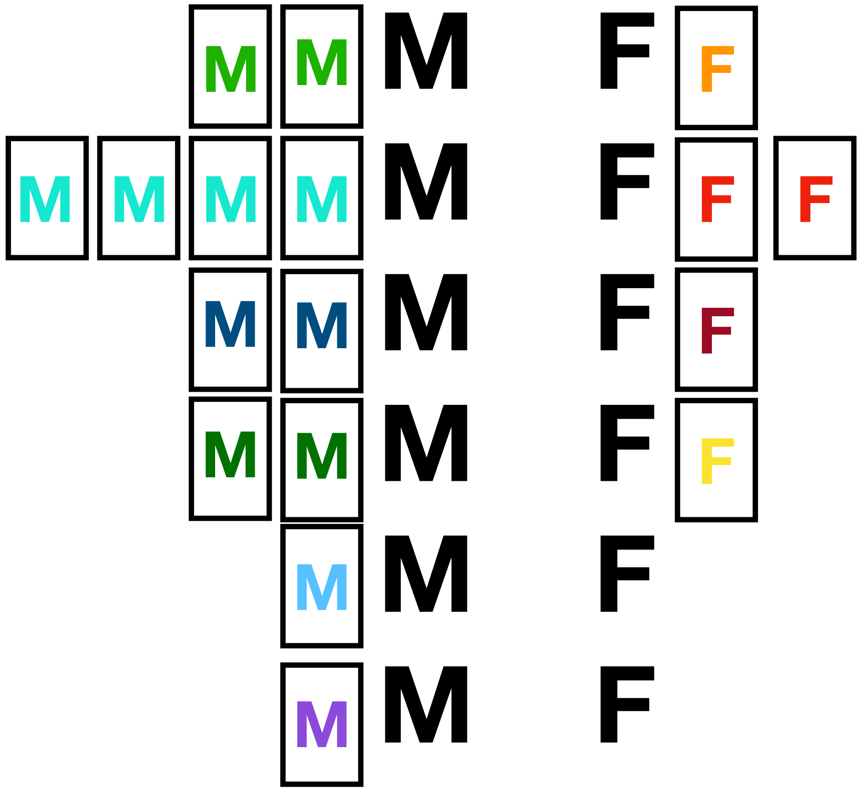
- Equal representation and publication rates

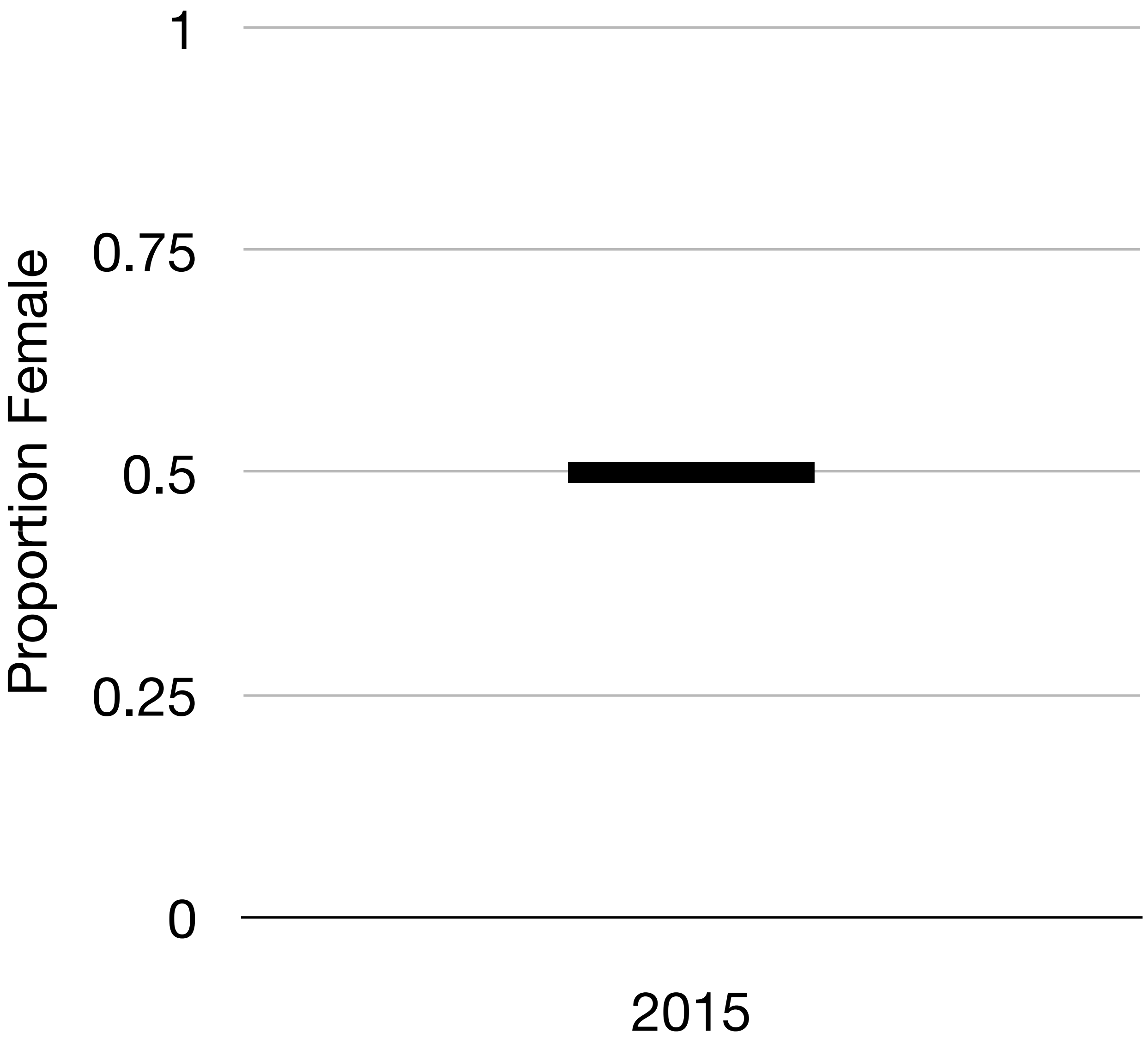




Representation estimate

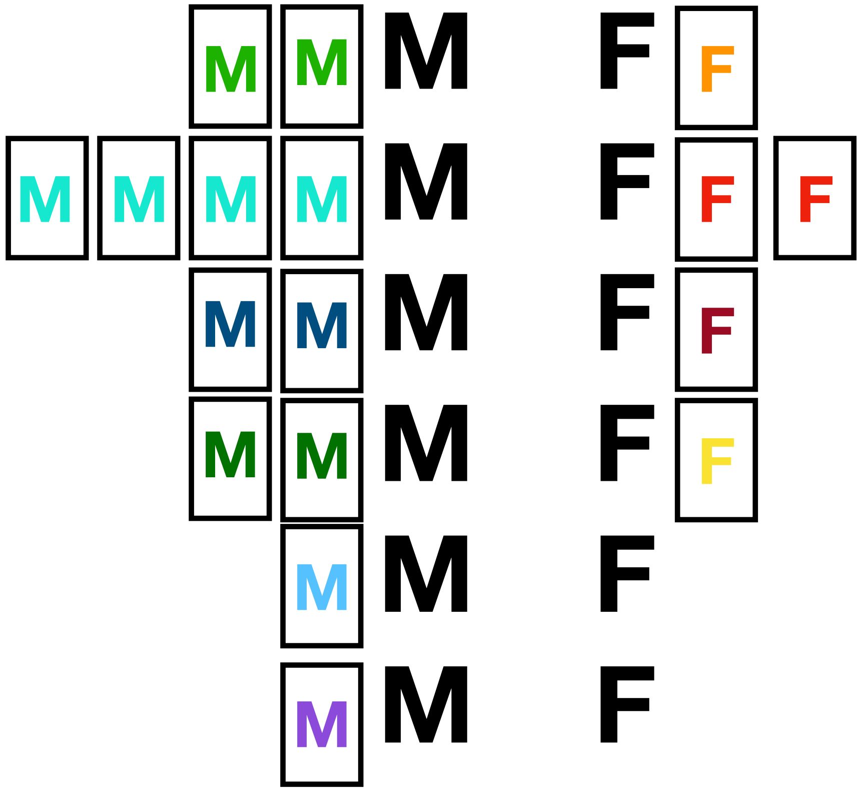
- Equal representation but unequal publication rates

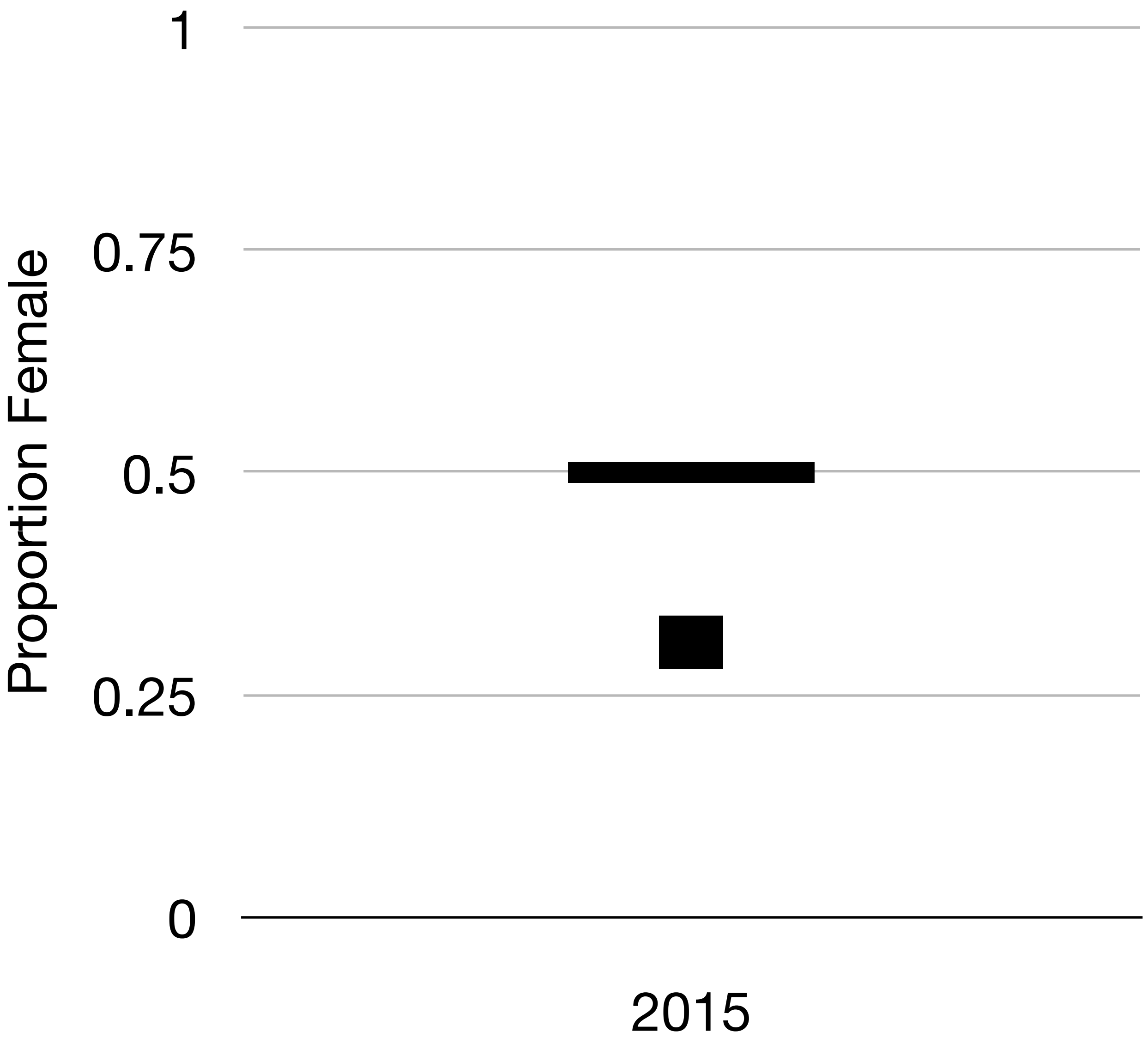




Representation estimate

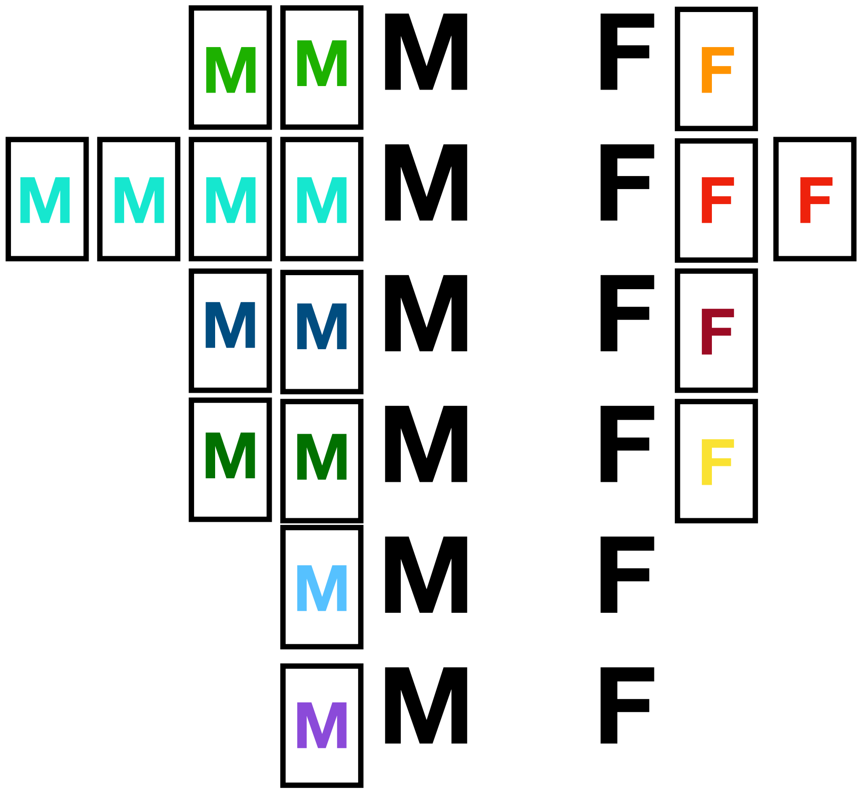
- Equal representation but unequal publication rates

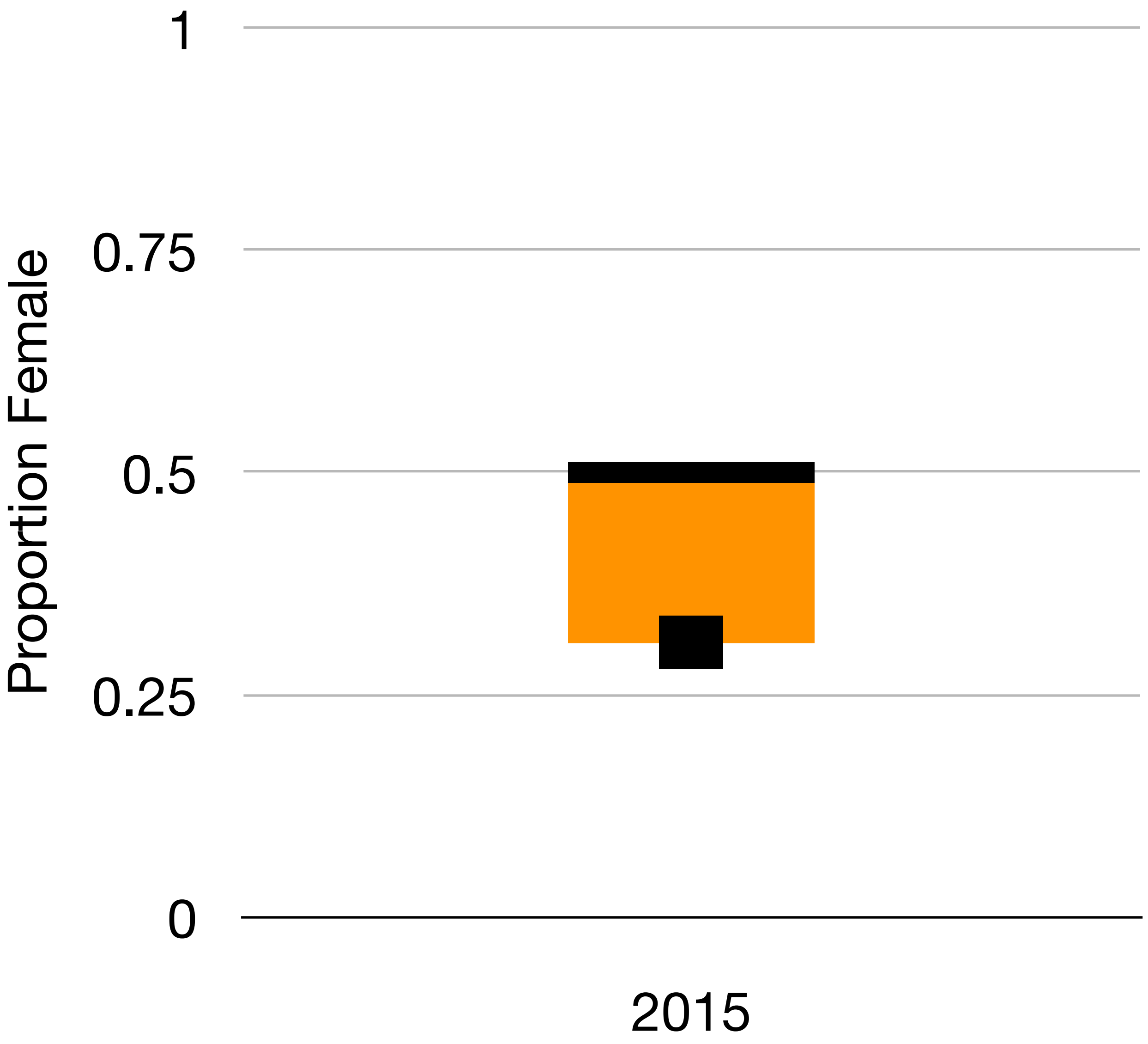




Representation estimate

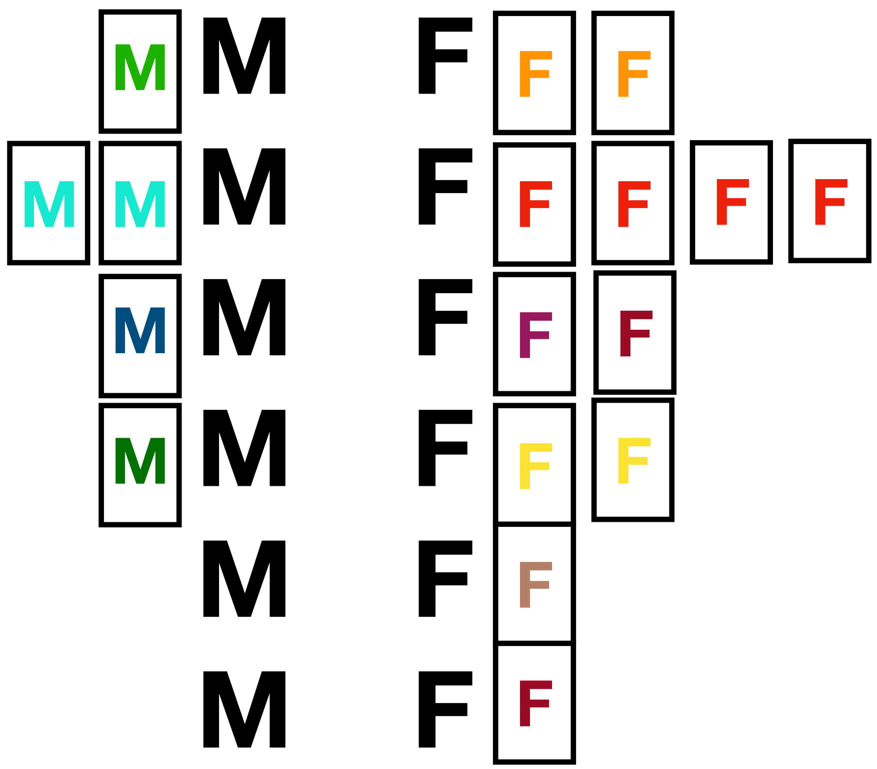
- Equal representation but unequal publication rates

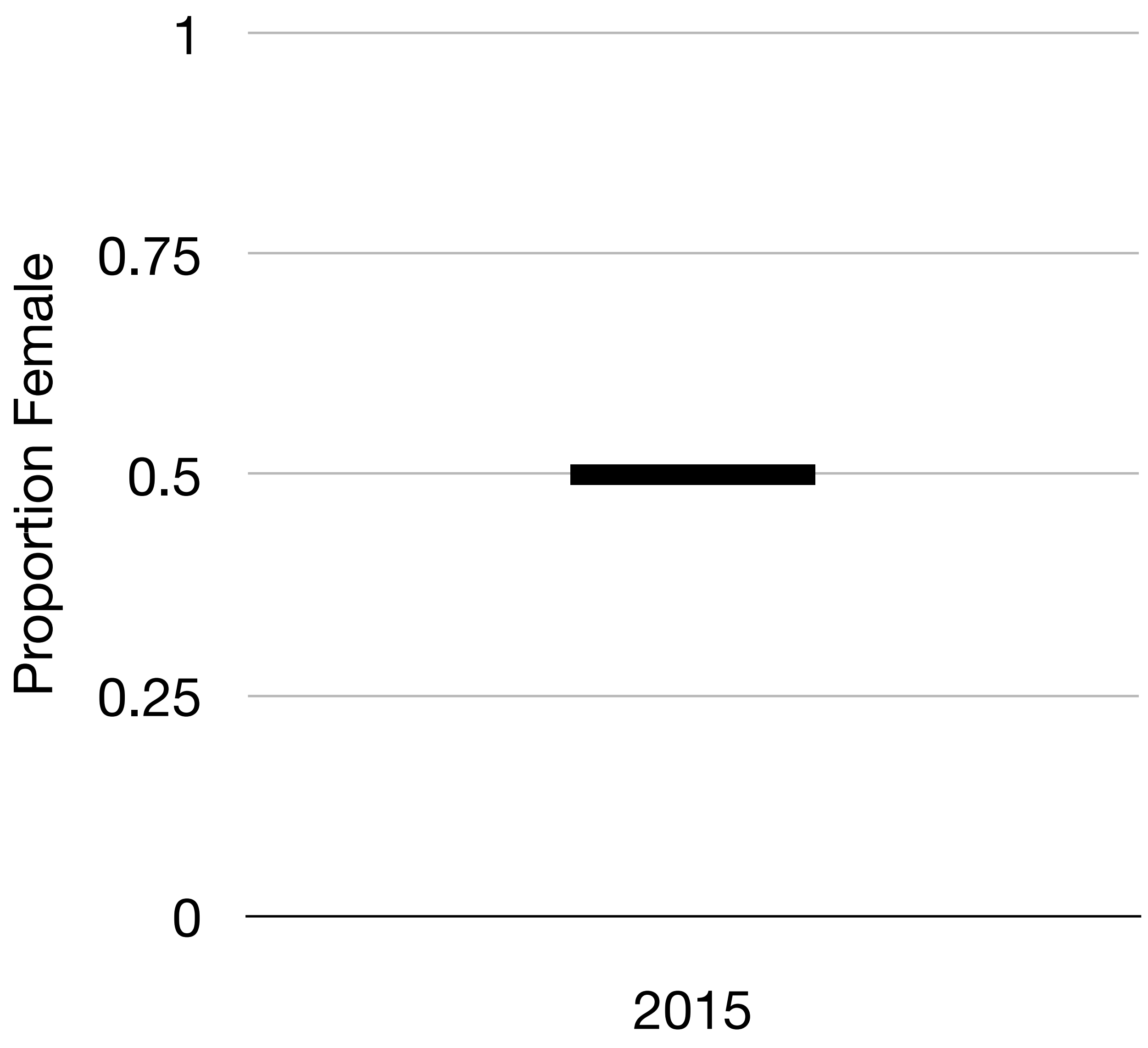




Representation estimate

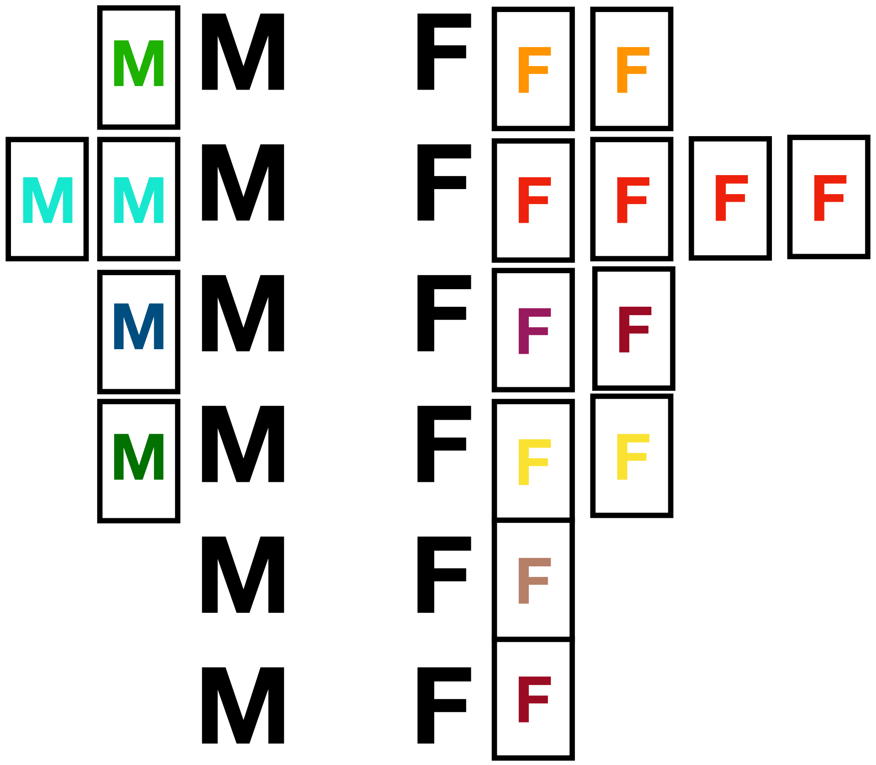
- Equal representation but unequal publication rates

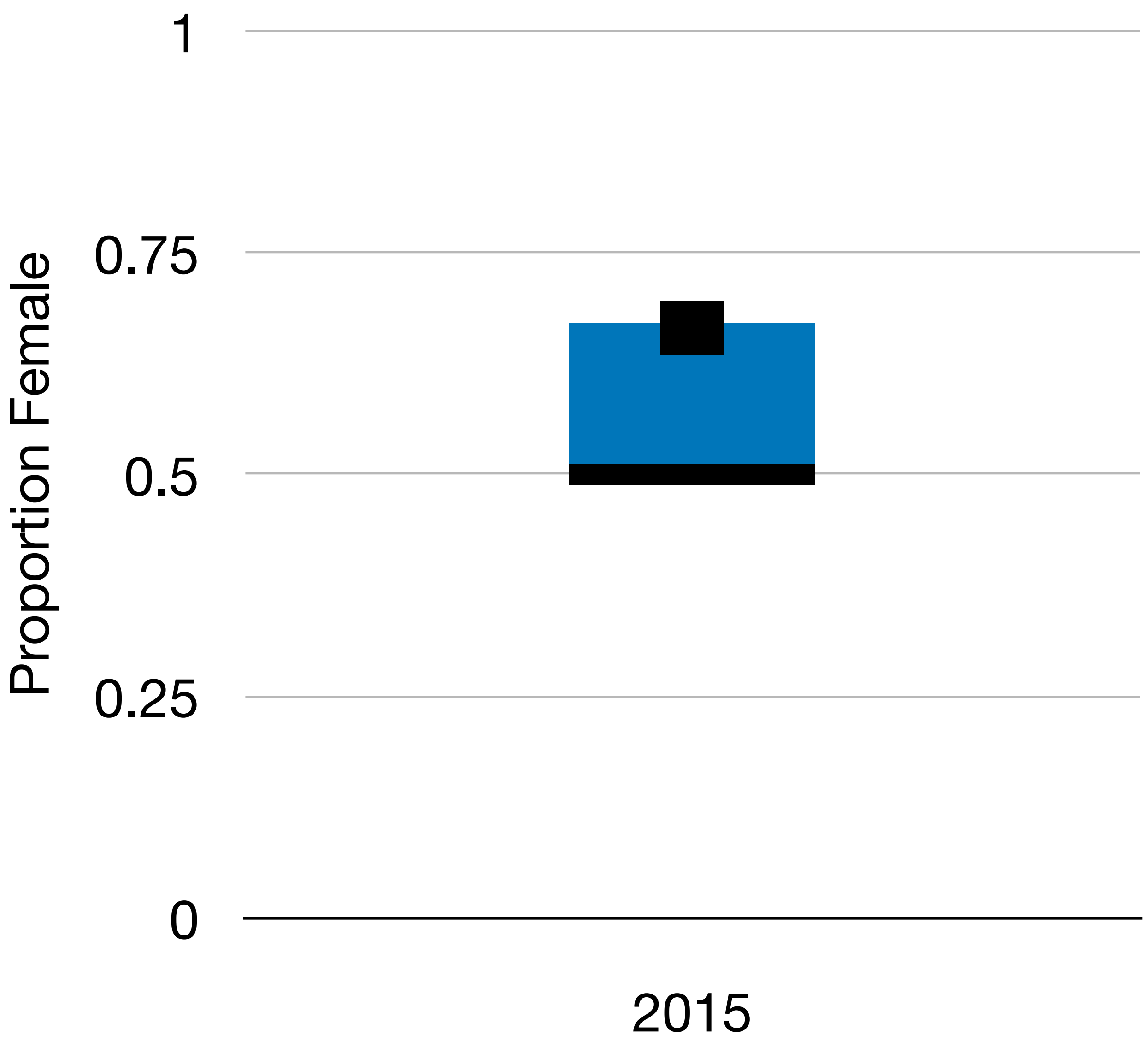




Representation estimate

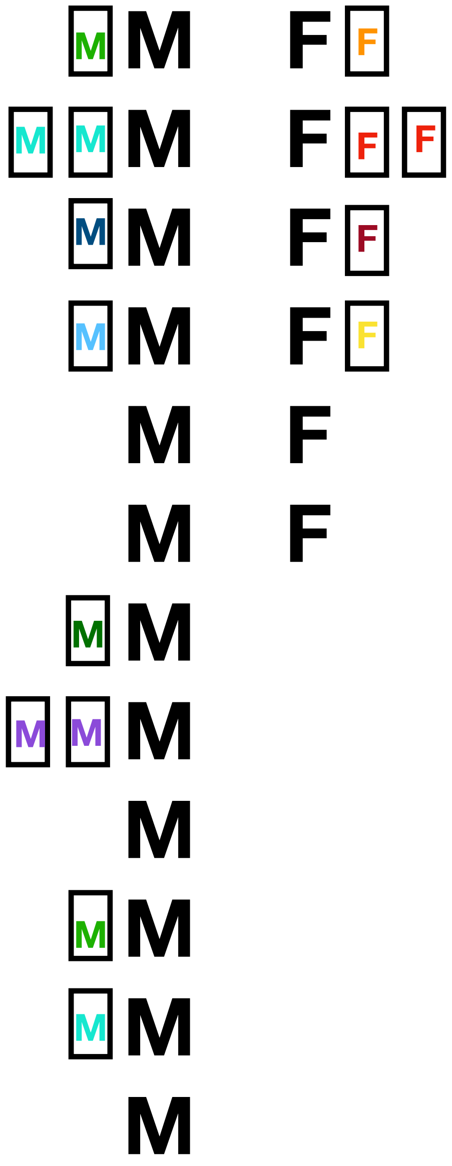
- Equal representation but unequal publication rates

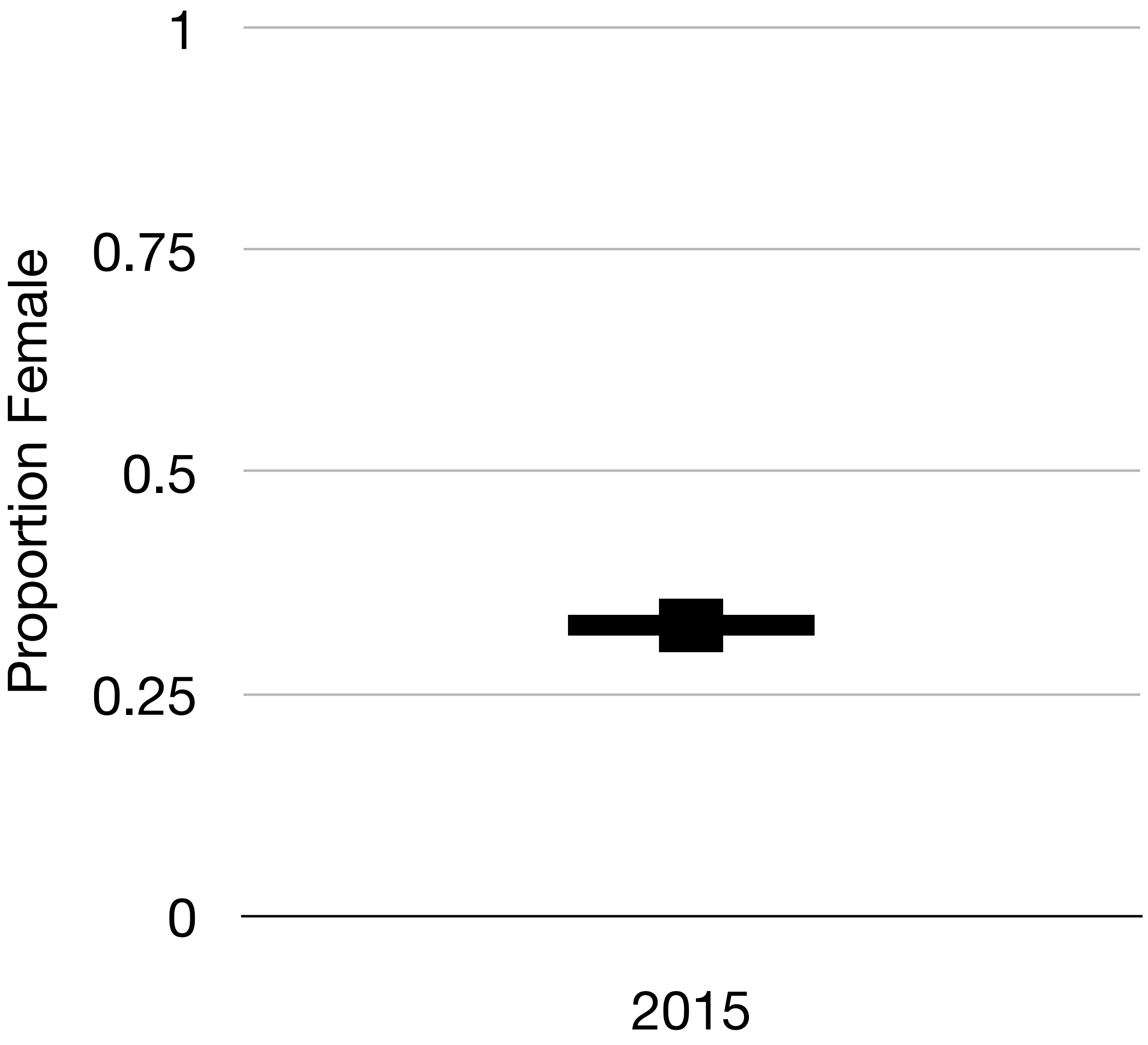




Representation estimate

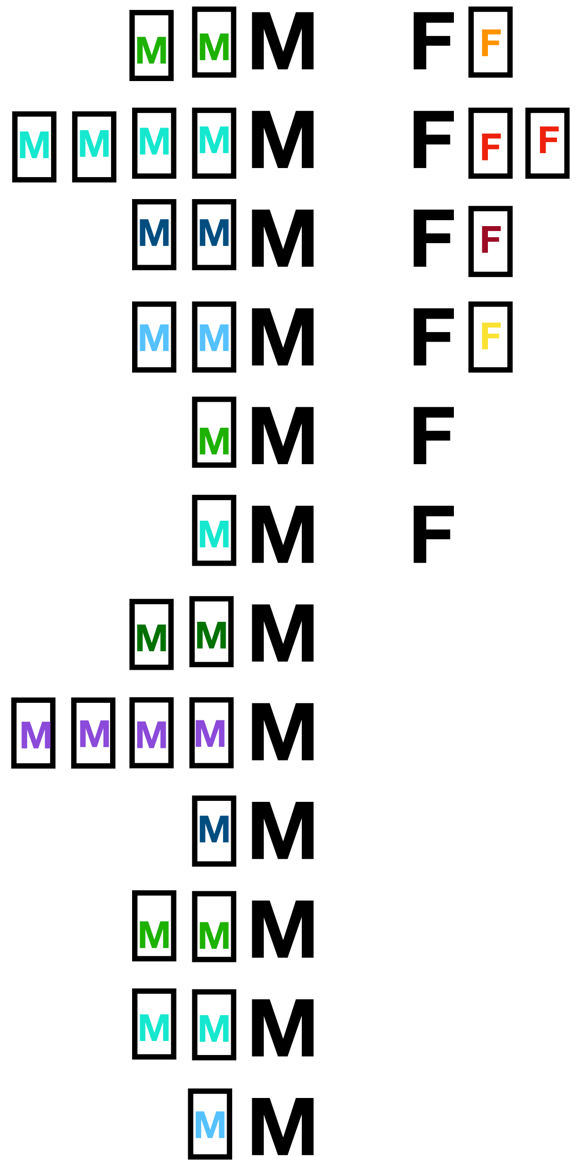
- Unequal representation but equal publication rates

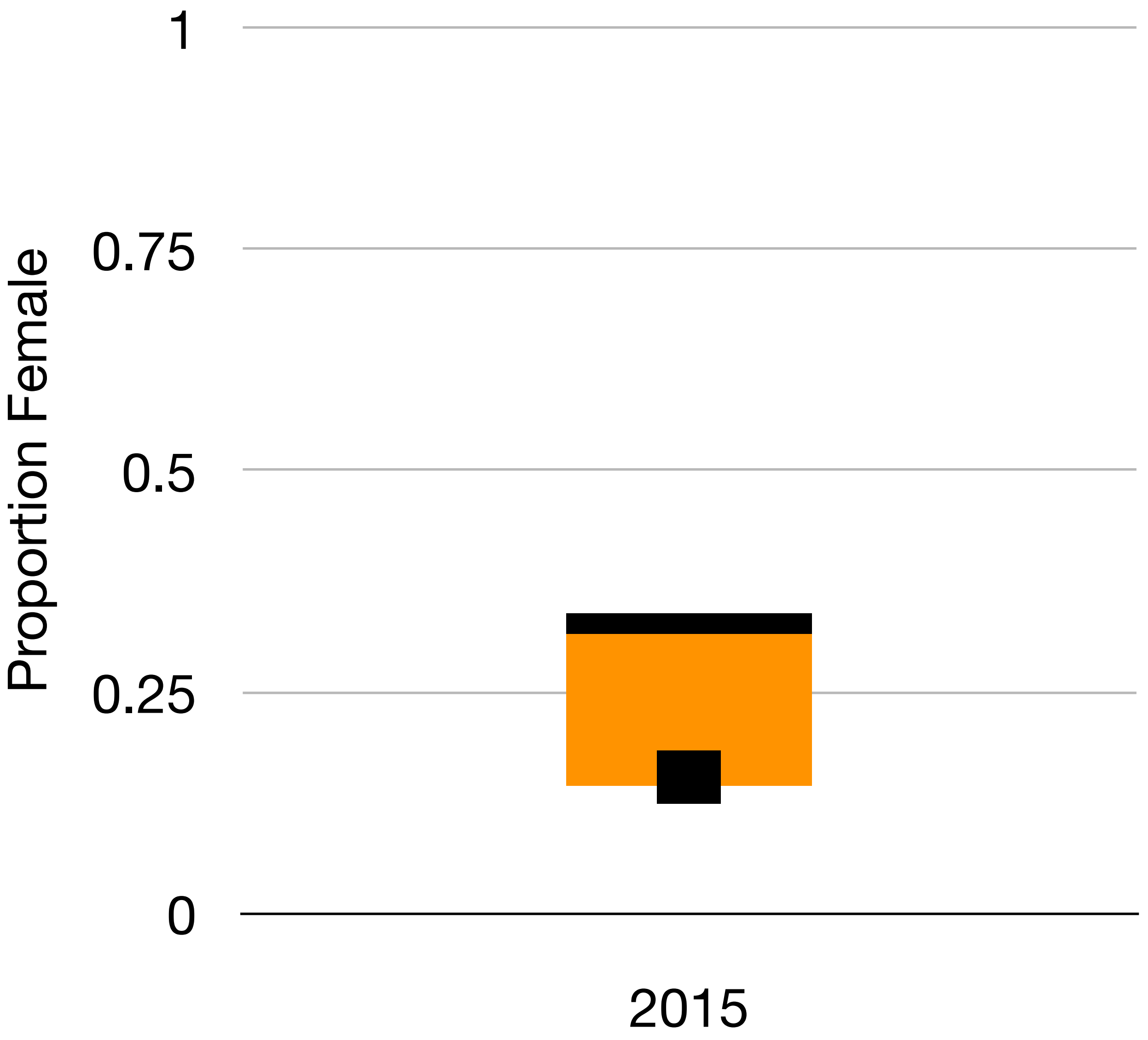




Representation estimate

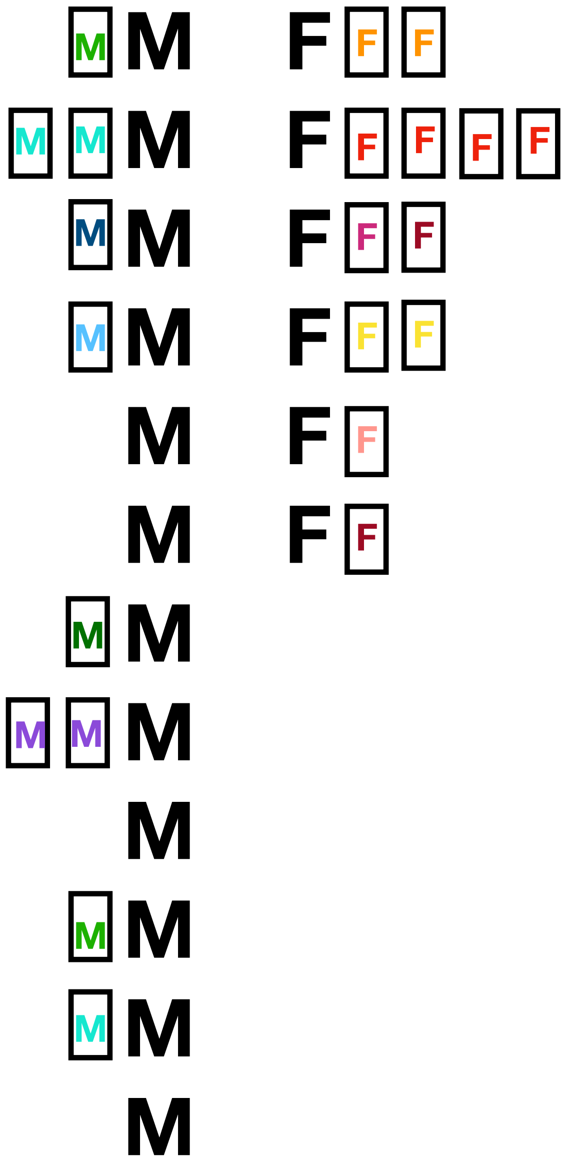
- Unequal representation but equal publication rates

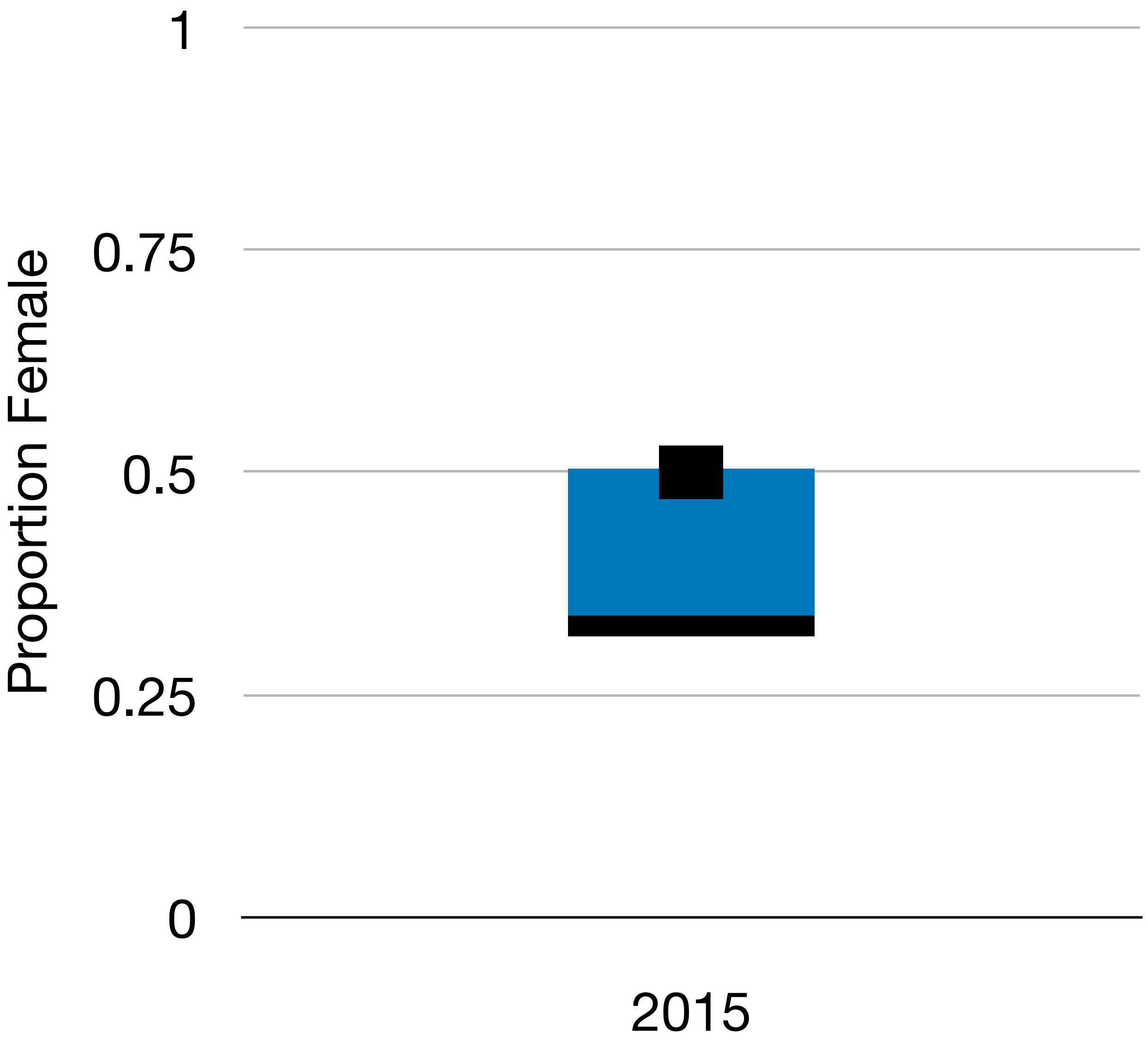




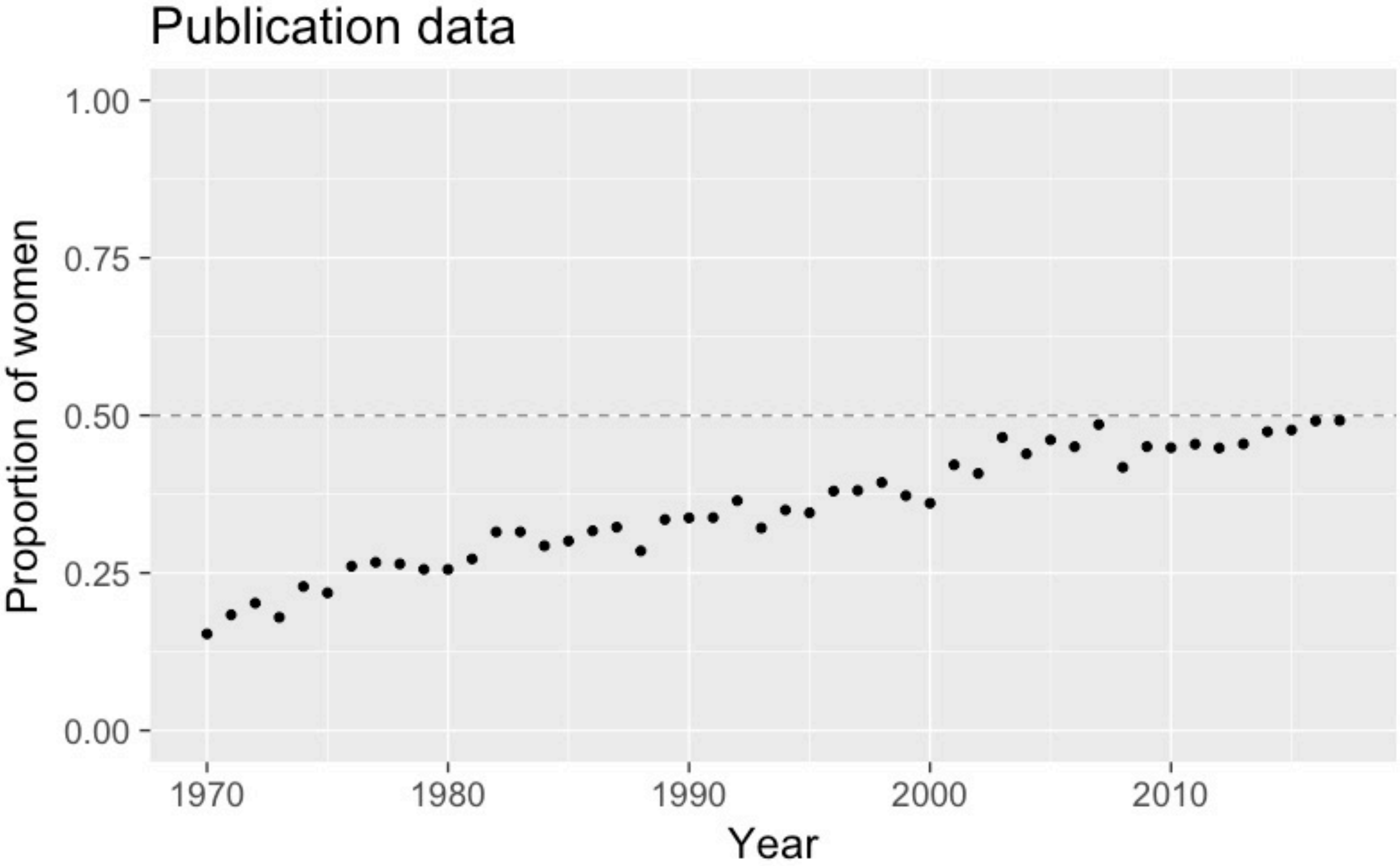
Representation estimate

- Unequal representation but equal publication rates



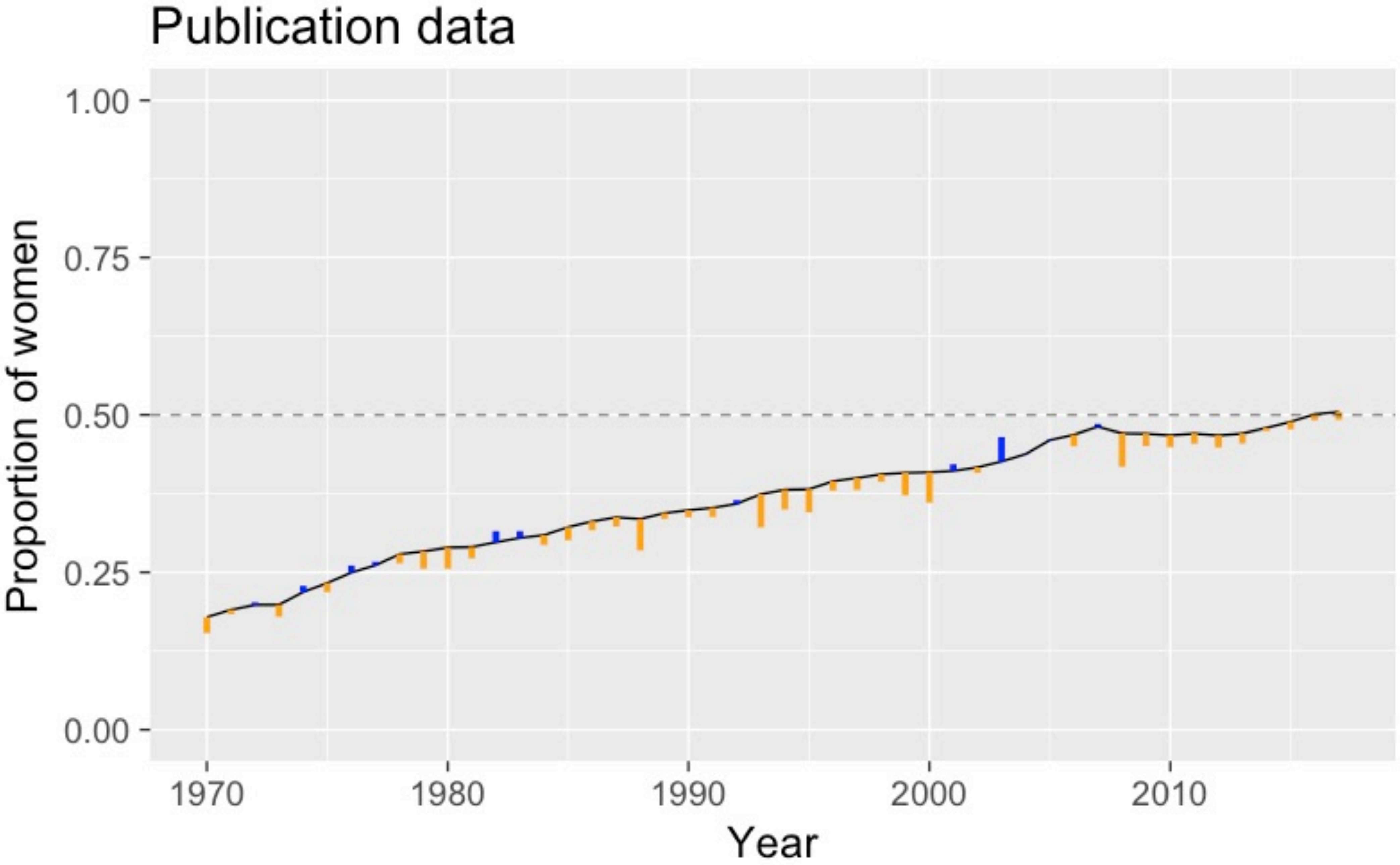


Publication rates

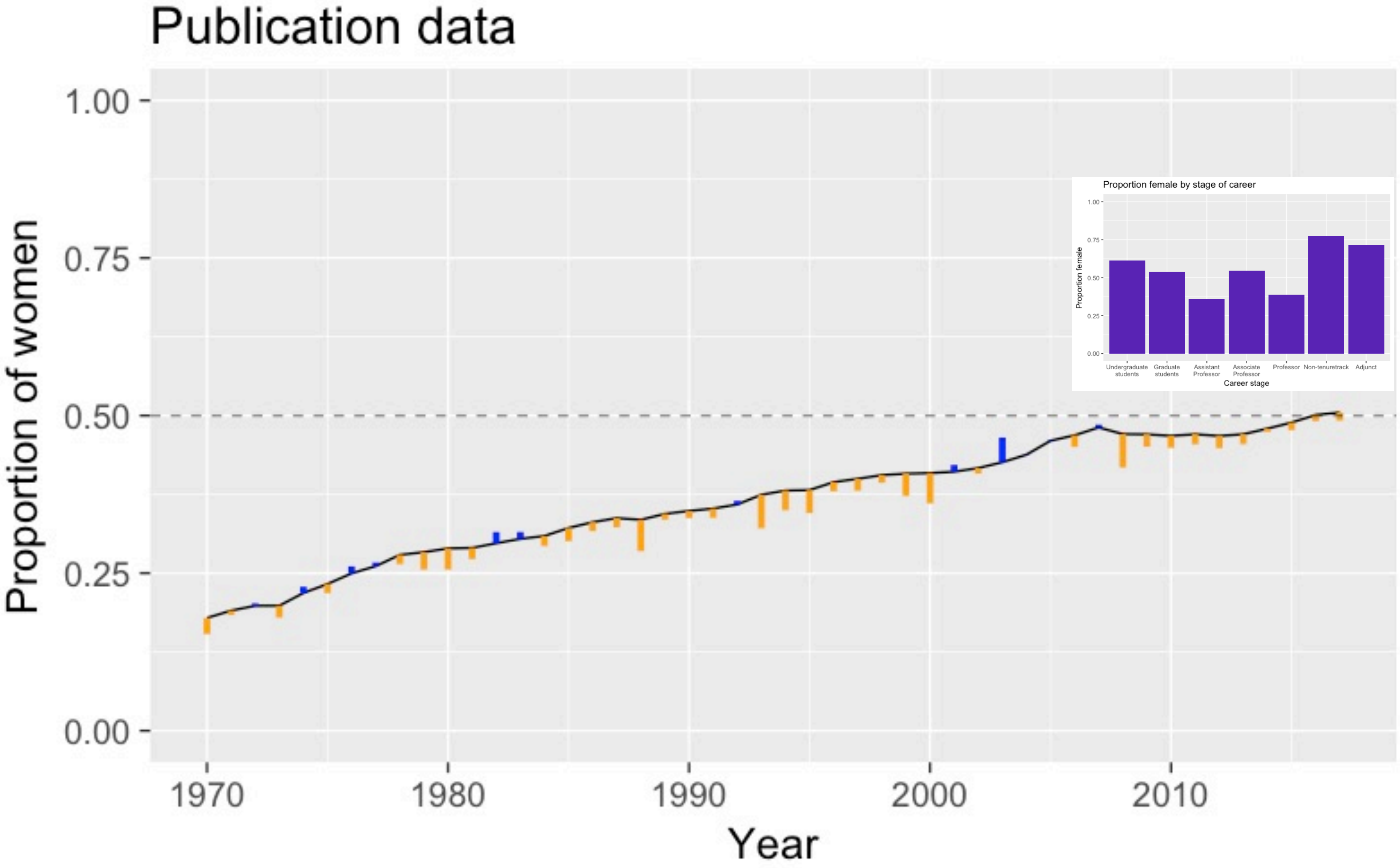


Part 2

Publication and representation rates

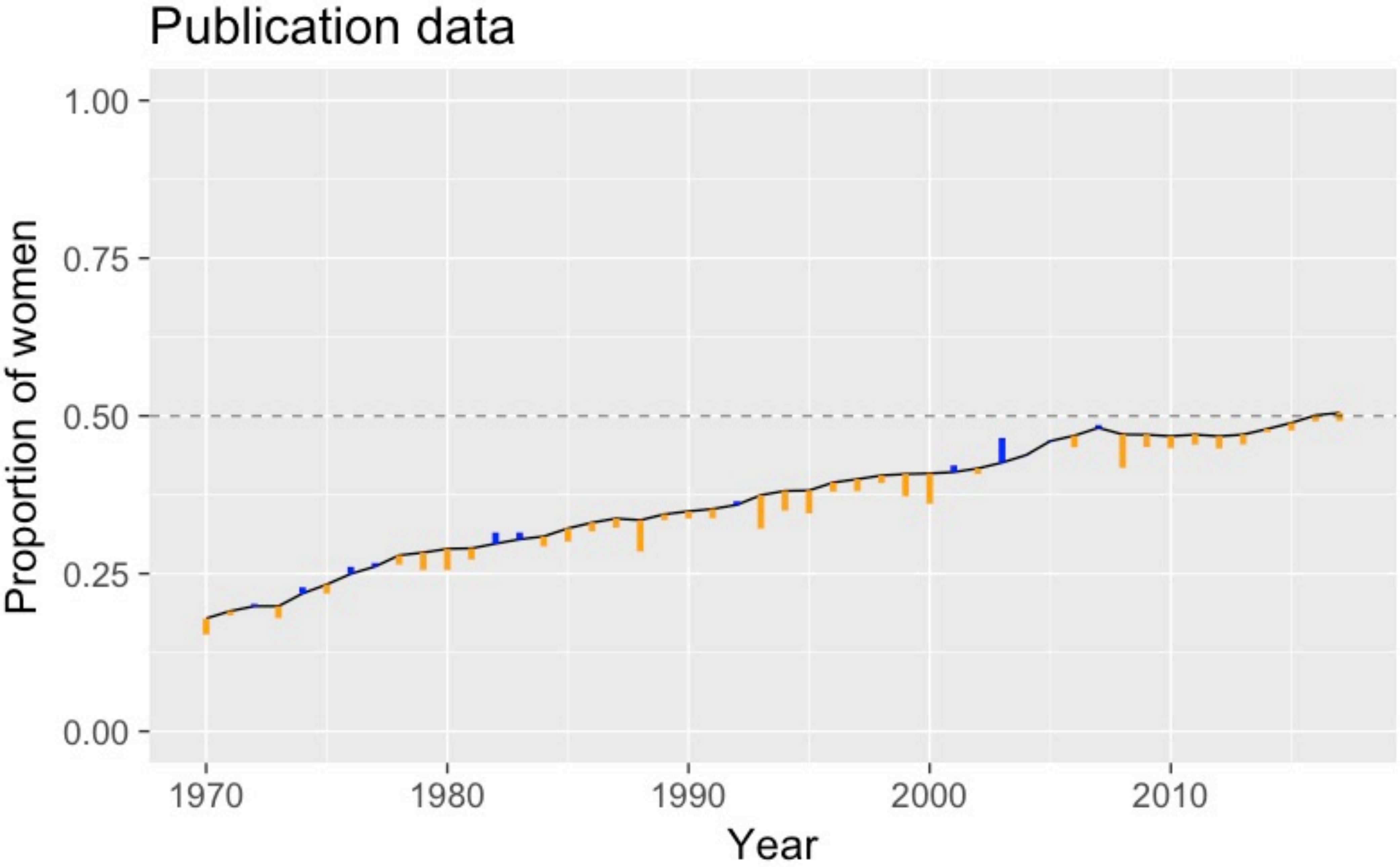


Publication and representation rates



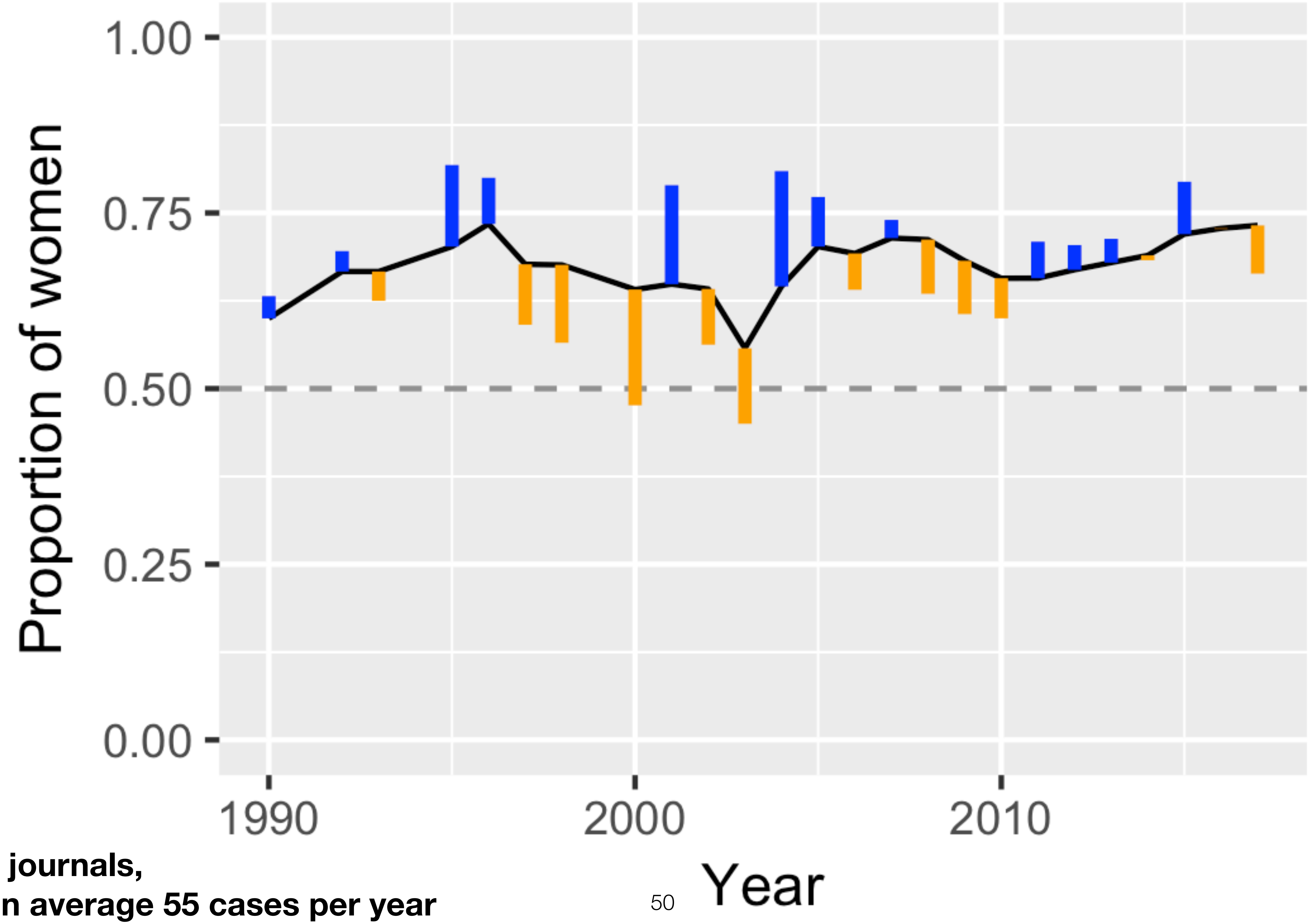
Part 2

Publication and representation rates



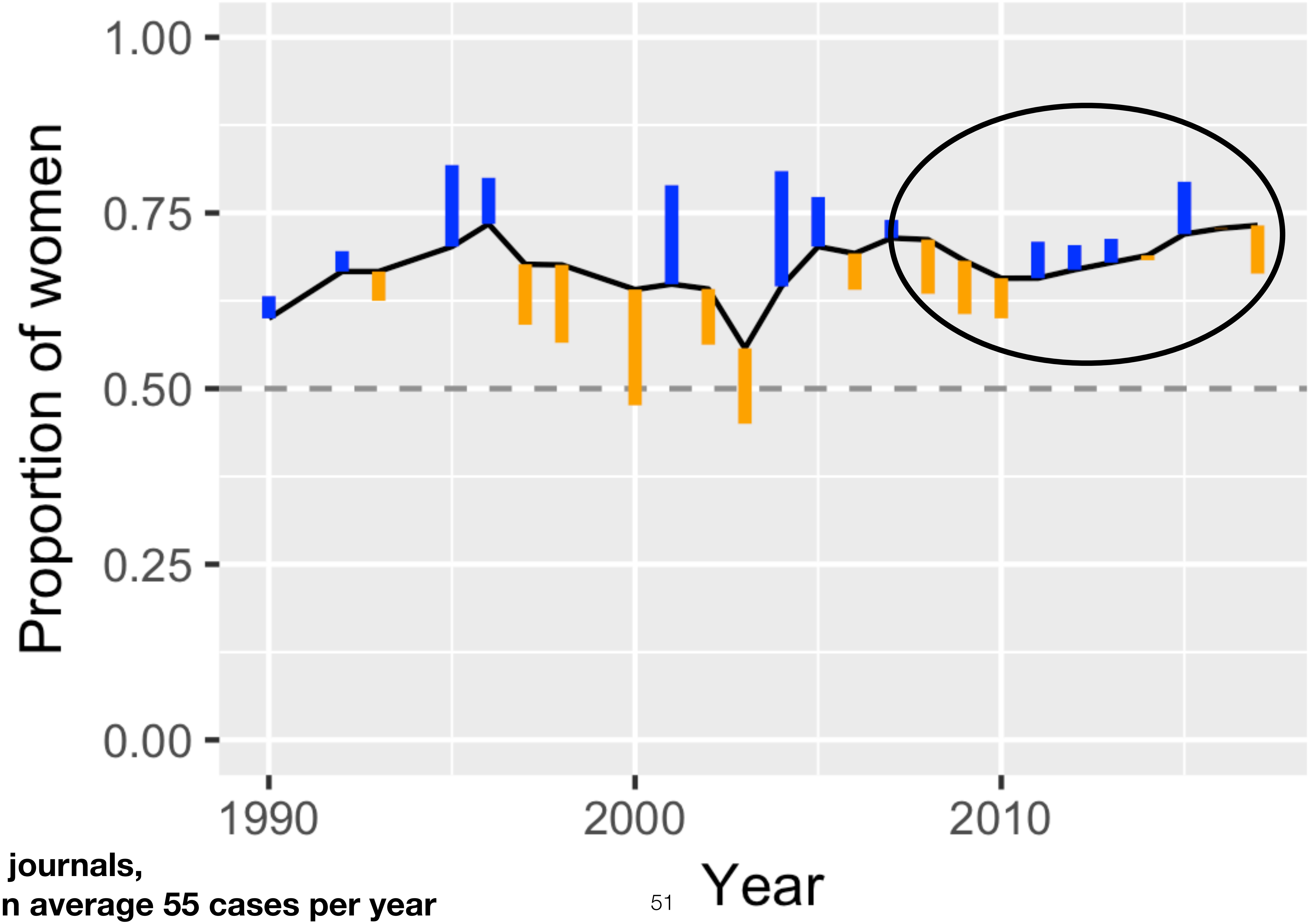
Part 2

Acquisition



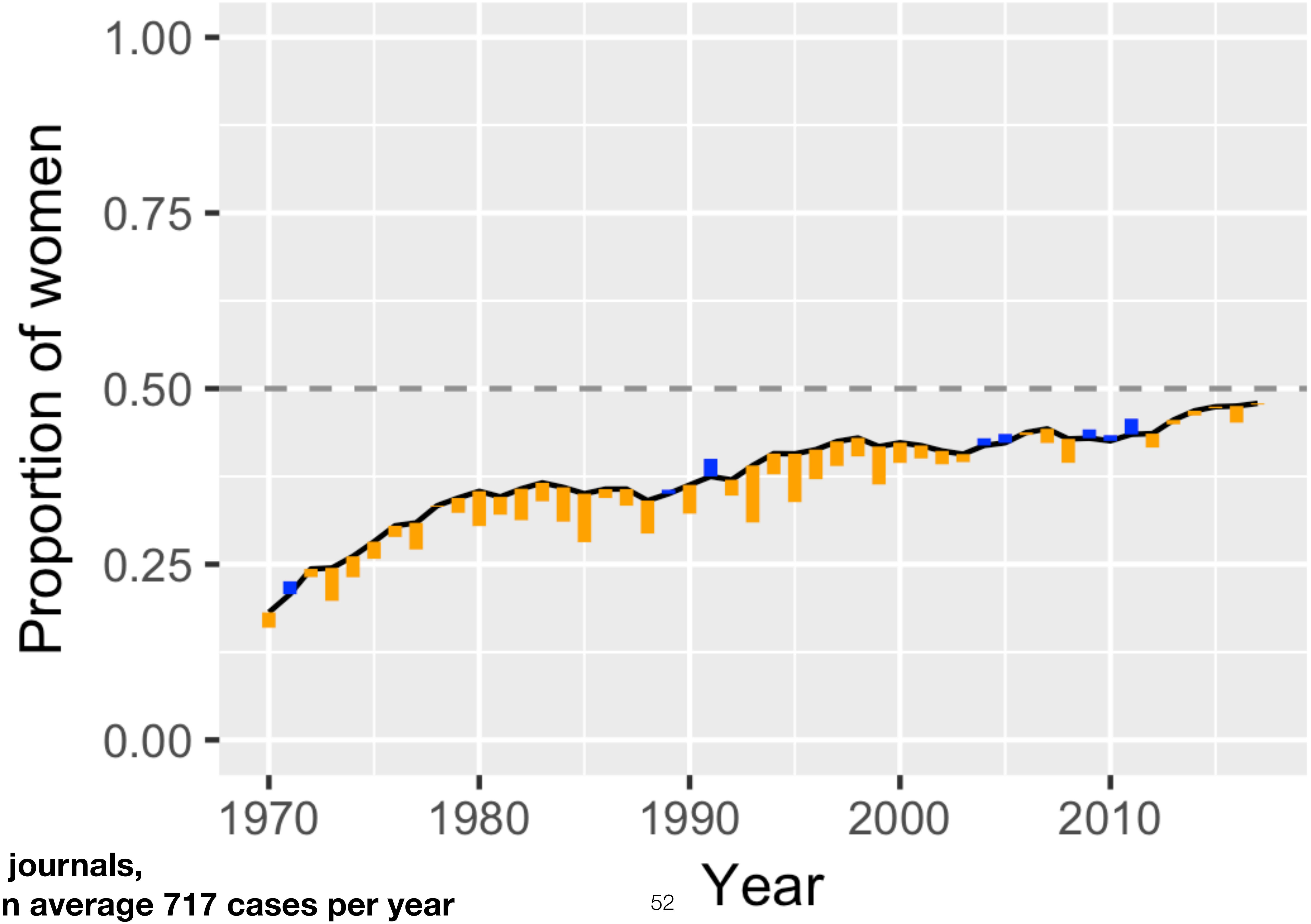
**3 journals,
on average 55 cases per year**

Acquisition



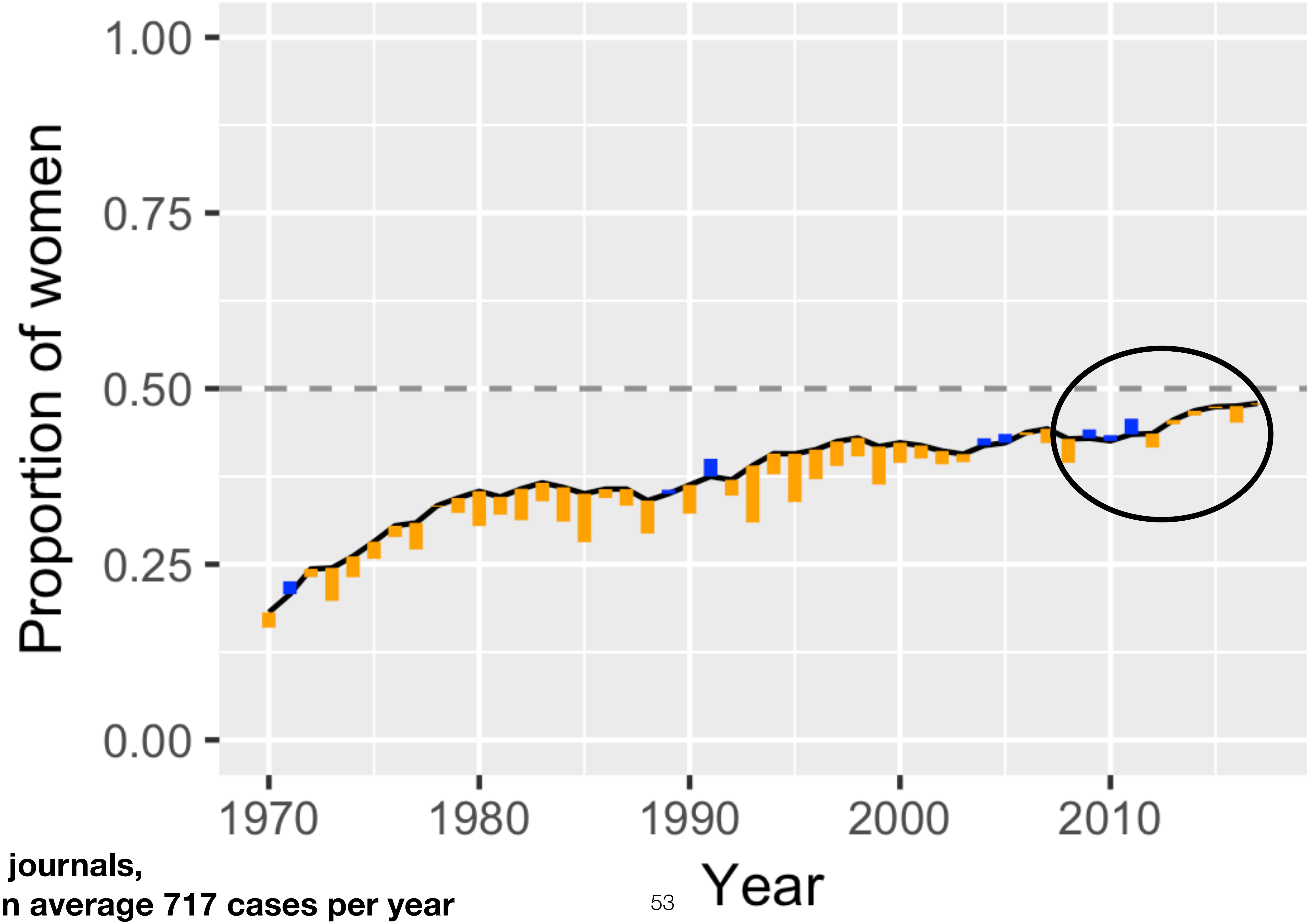
3 journals,
on average 55 cases per year

Phonology/Phonetics



**7 journals,
on average 717 cases per year**

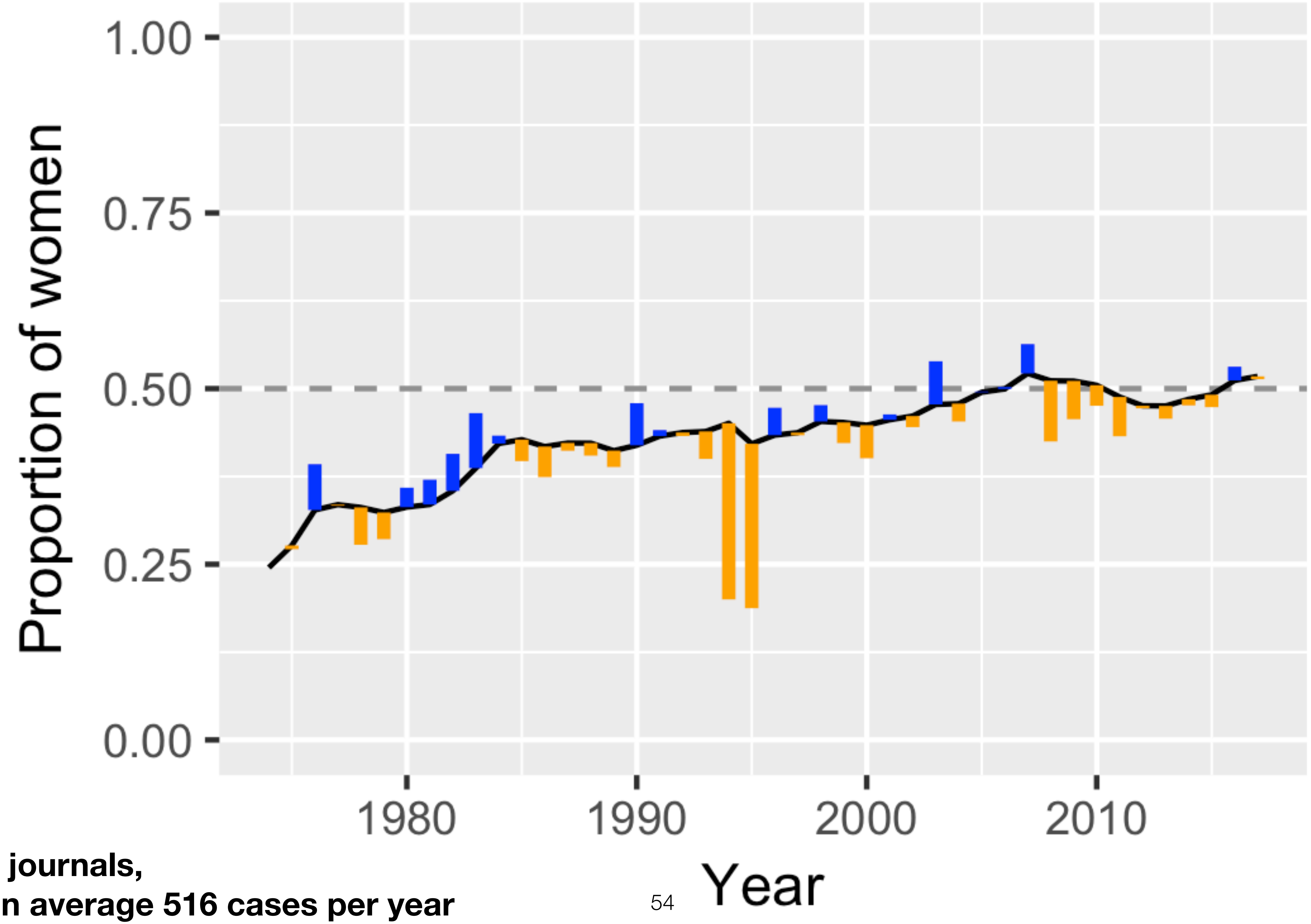
Phonology/Phonetics



7 journals,
on average 717 cases per year

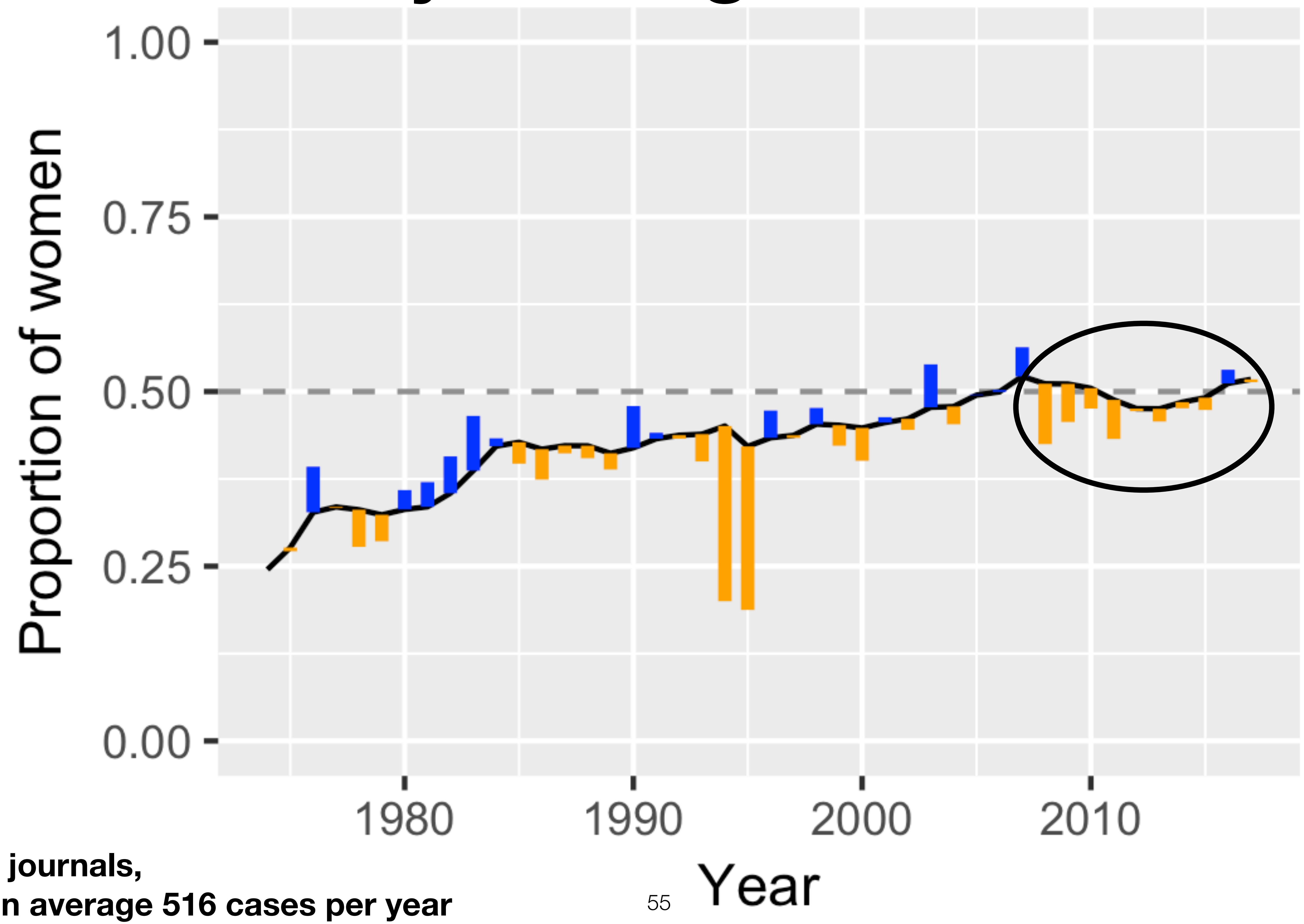
Part 2

Psycholinguistics



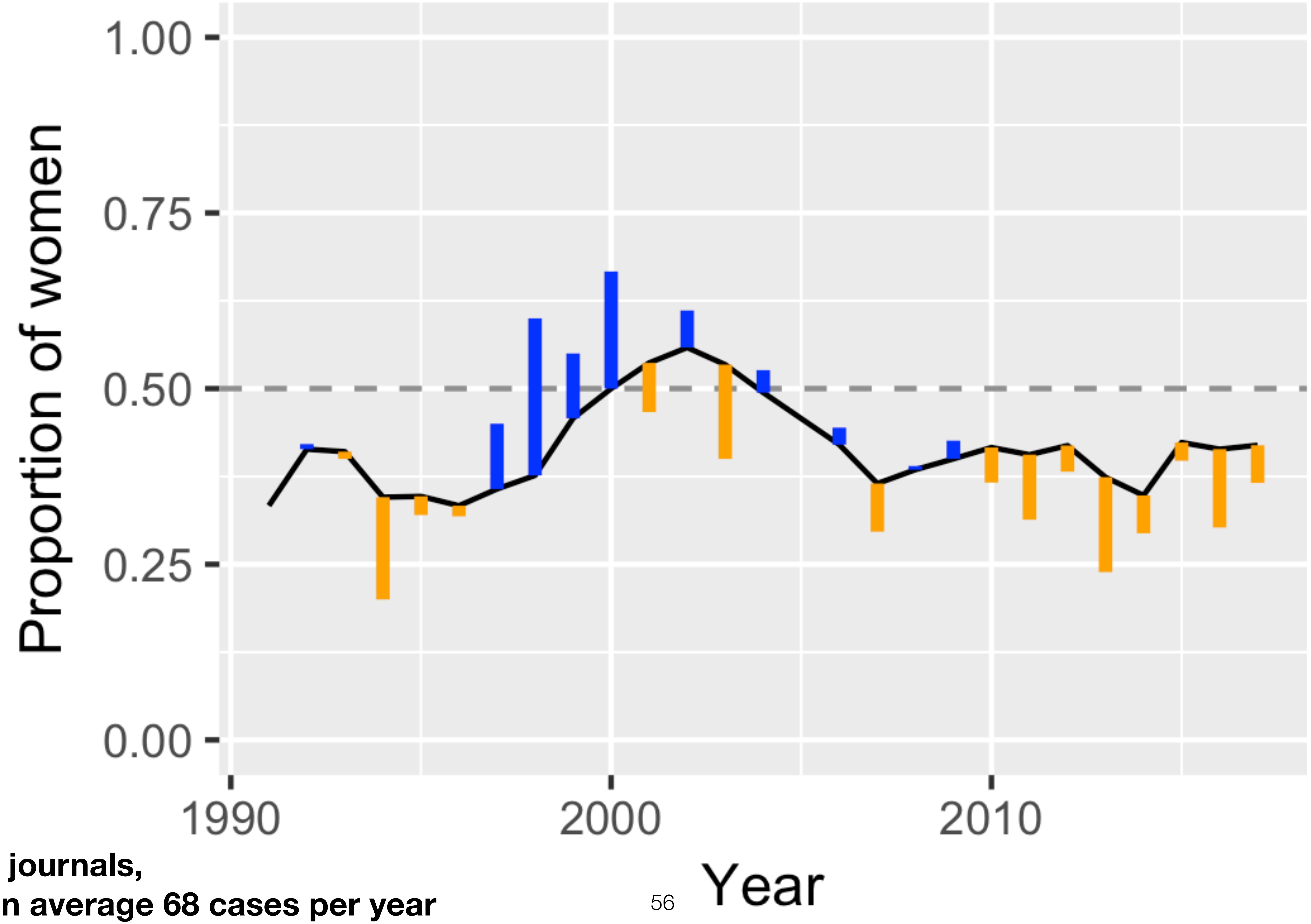
4 journals,
on average 516 cases per year

Psycholinguistics



**4 journals,
on average 516 cases per year**

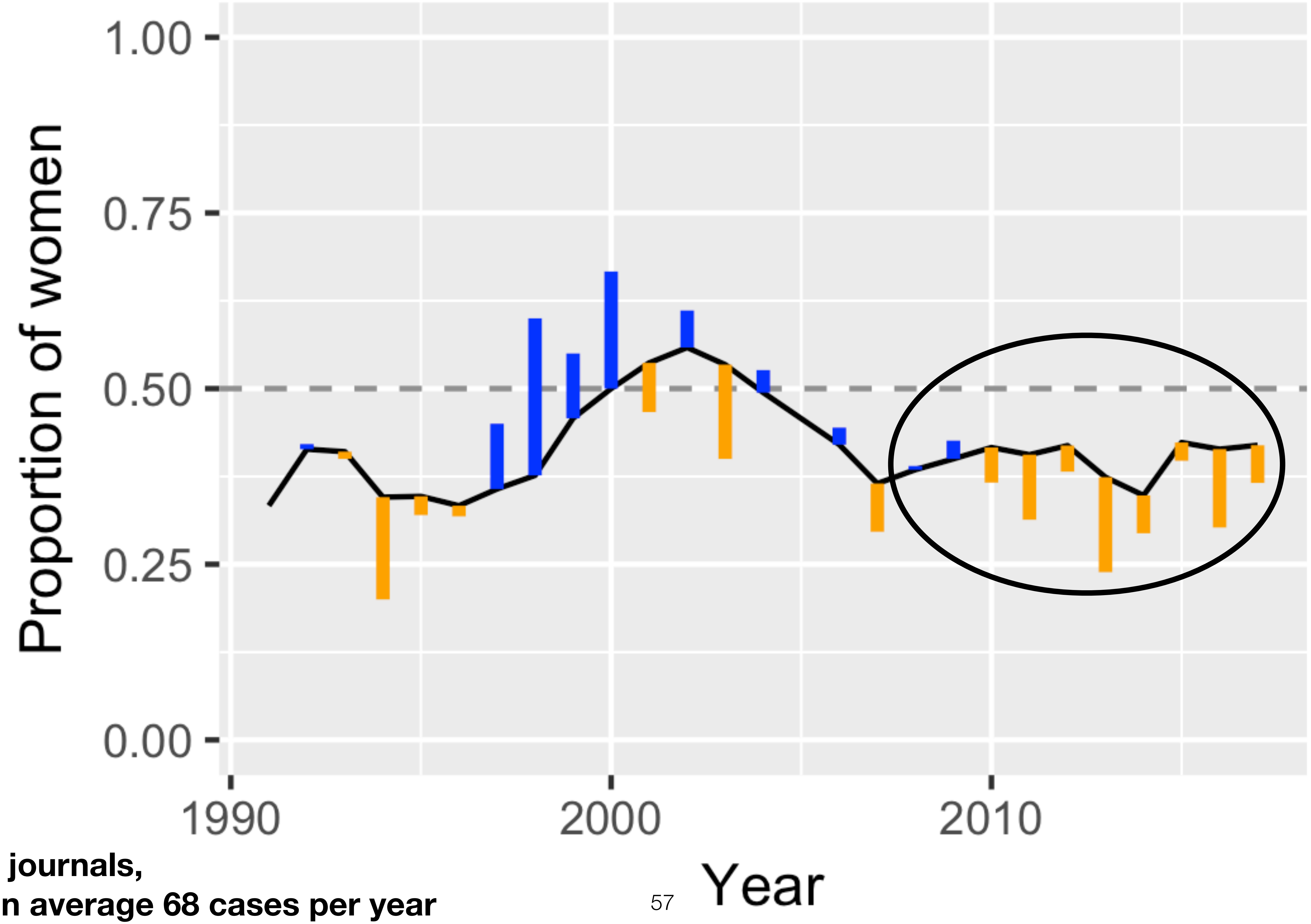
Semantics



2 journals,
on average 68 cases per year

Part 2

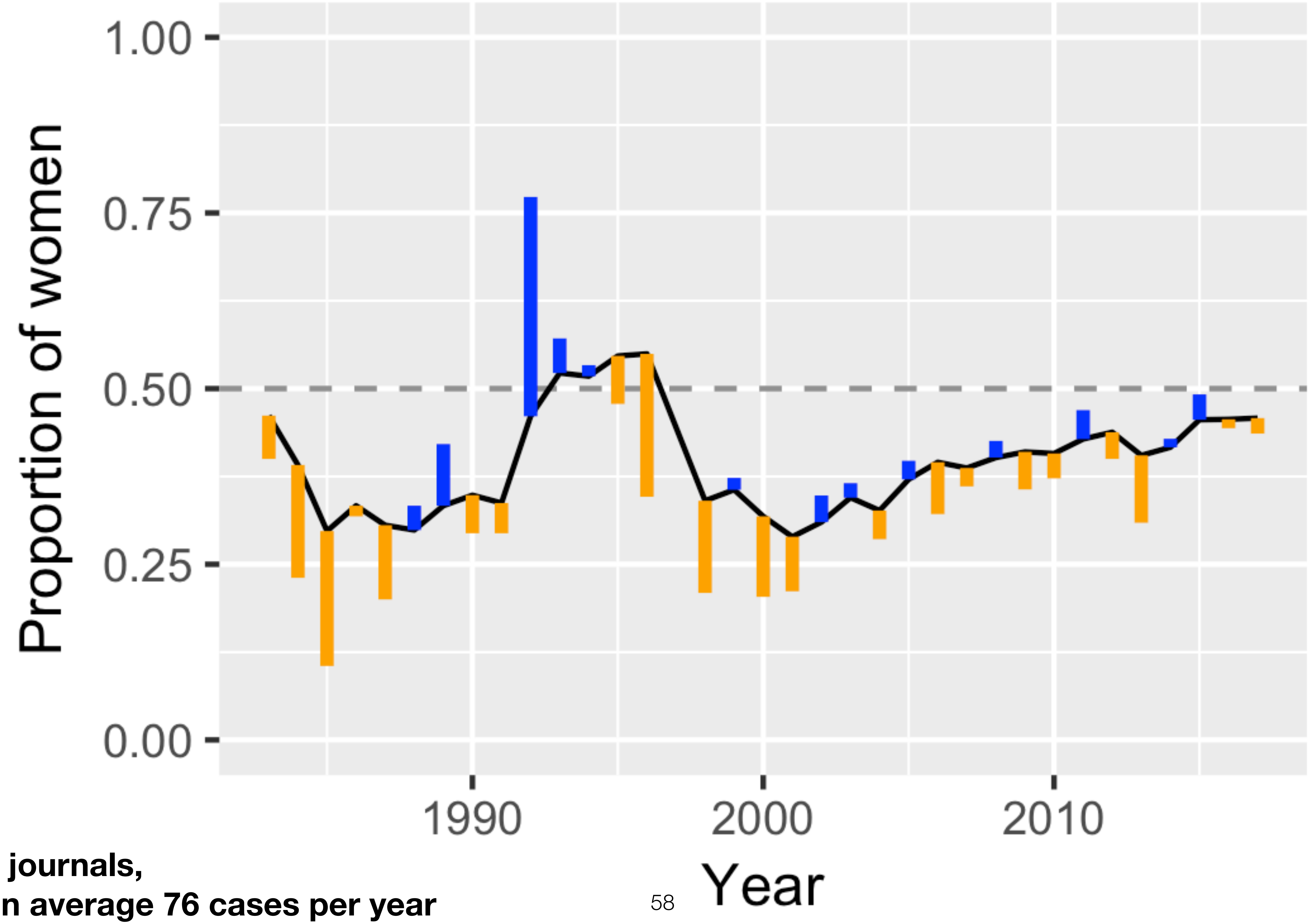
Semantics



**2 journals,
on average 68 cases per year**

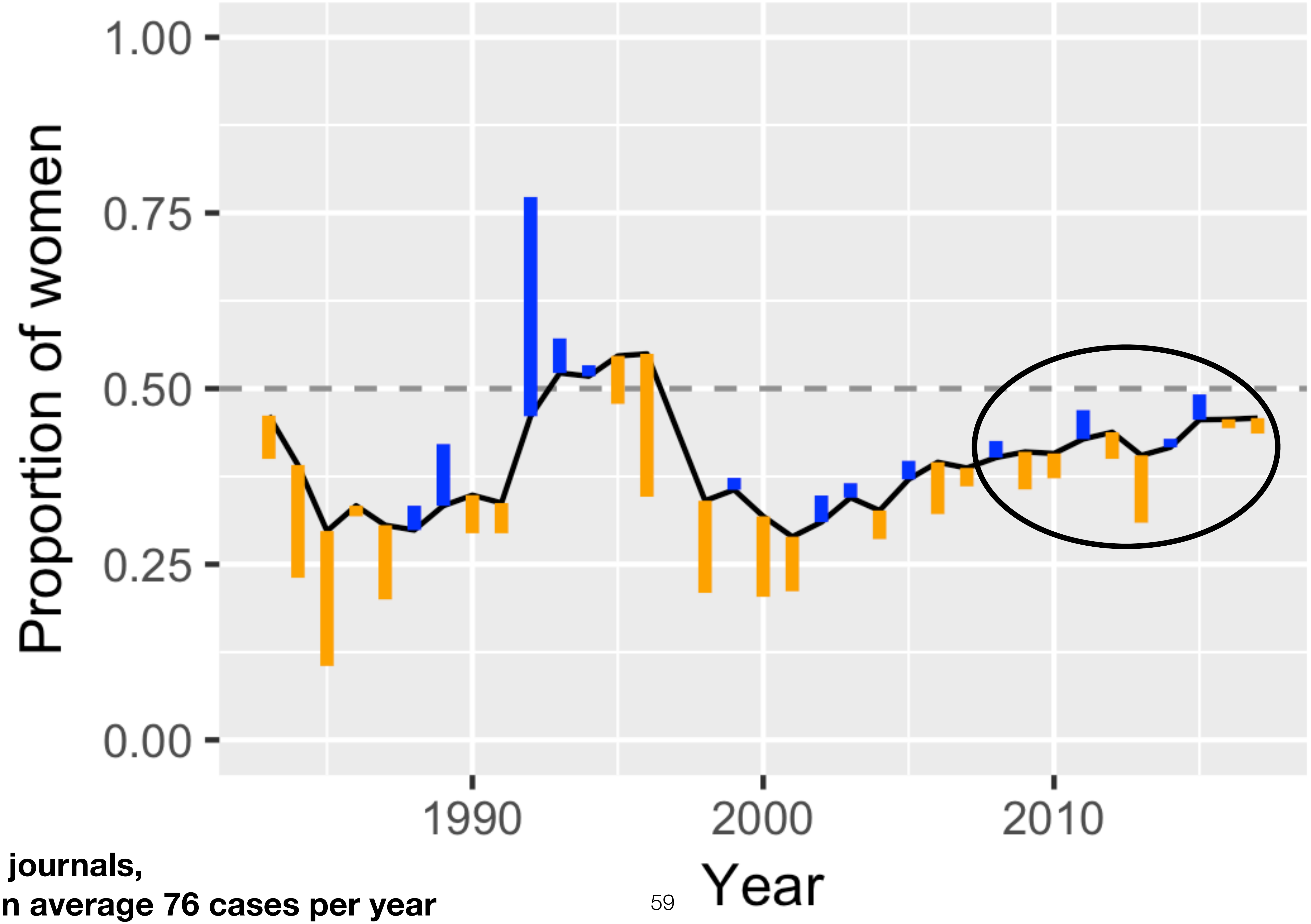
Part 2

Syntax



**6 journals,
on average 76 cases per year**

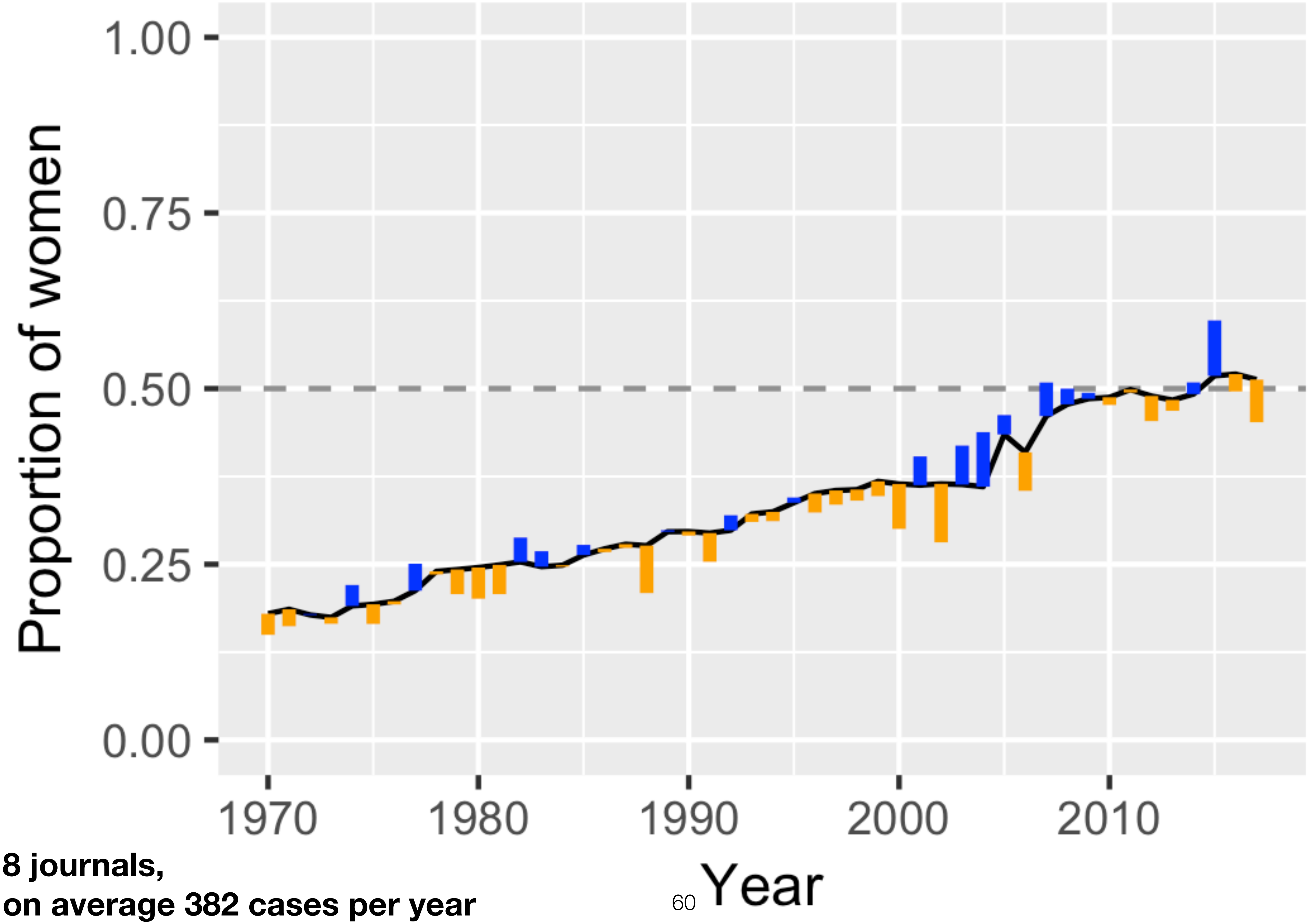
Syntax



6 journals,
on average 76 cases per year

Part 2

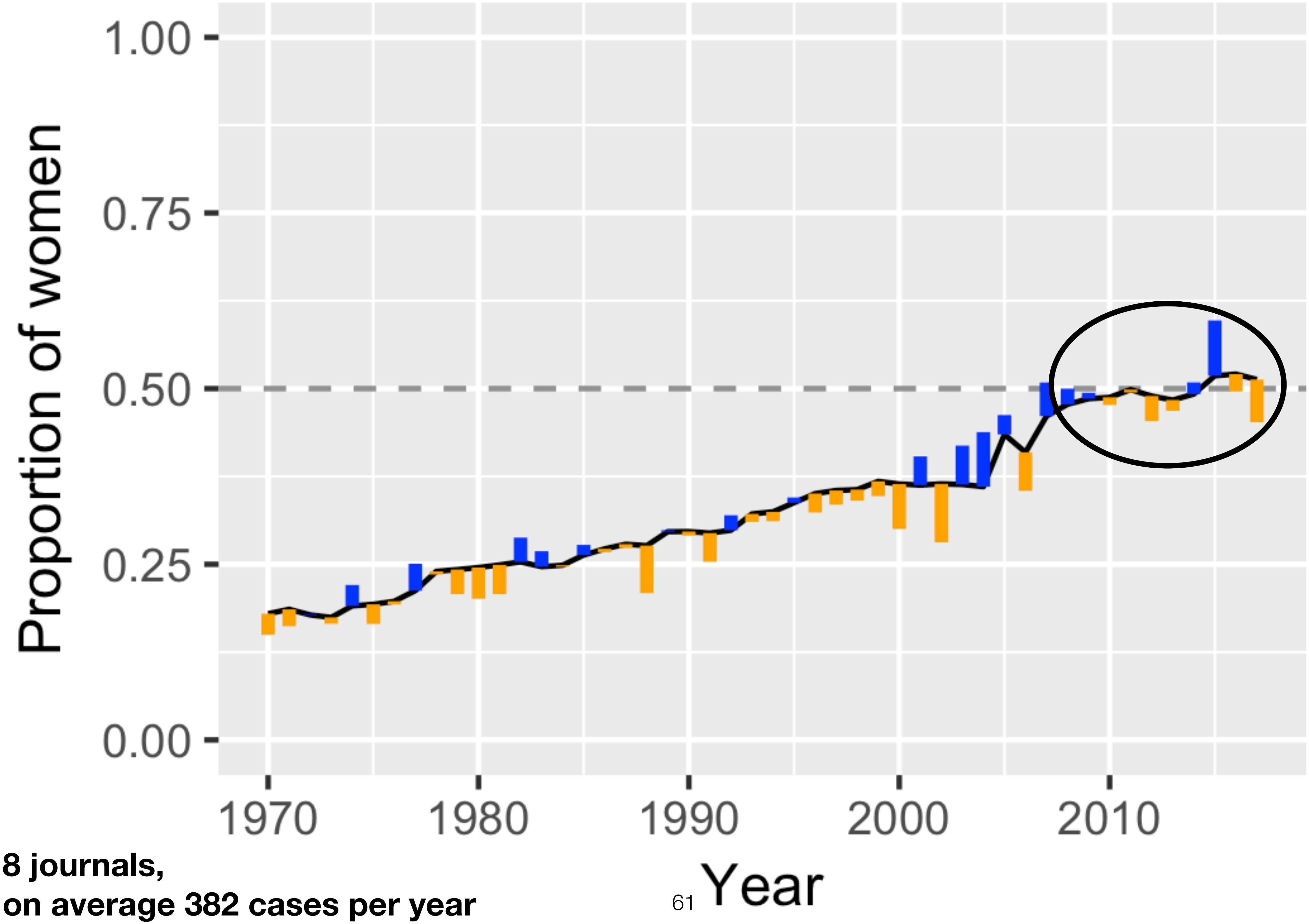
Domain-general



8 journals,
on average 382 cases per year

Part 2

Domain-general



8 journals,
on average 382 cases per year

Do women publish less?

Yes.

Why do women publish less?

- One possibility is differences in submission rates, because of:
  - Trade-off with other obligations (service, teaching)
  - Prioritizing quality over quantity (perhaps because they’re forced to)

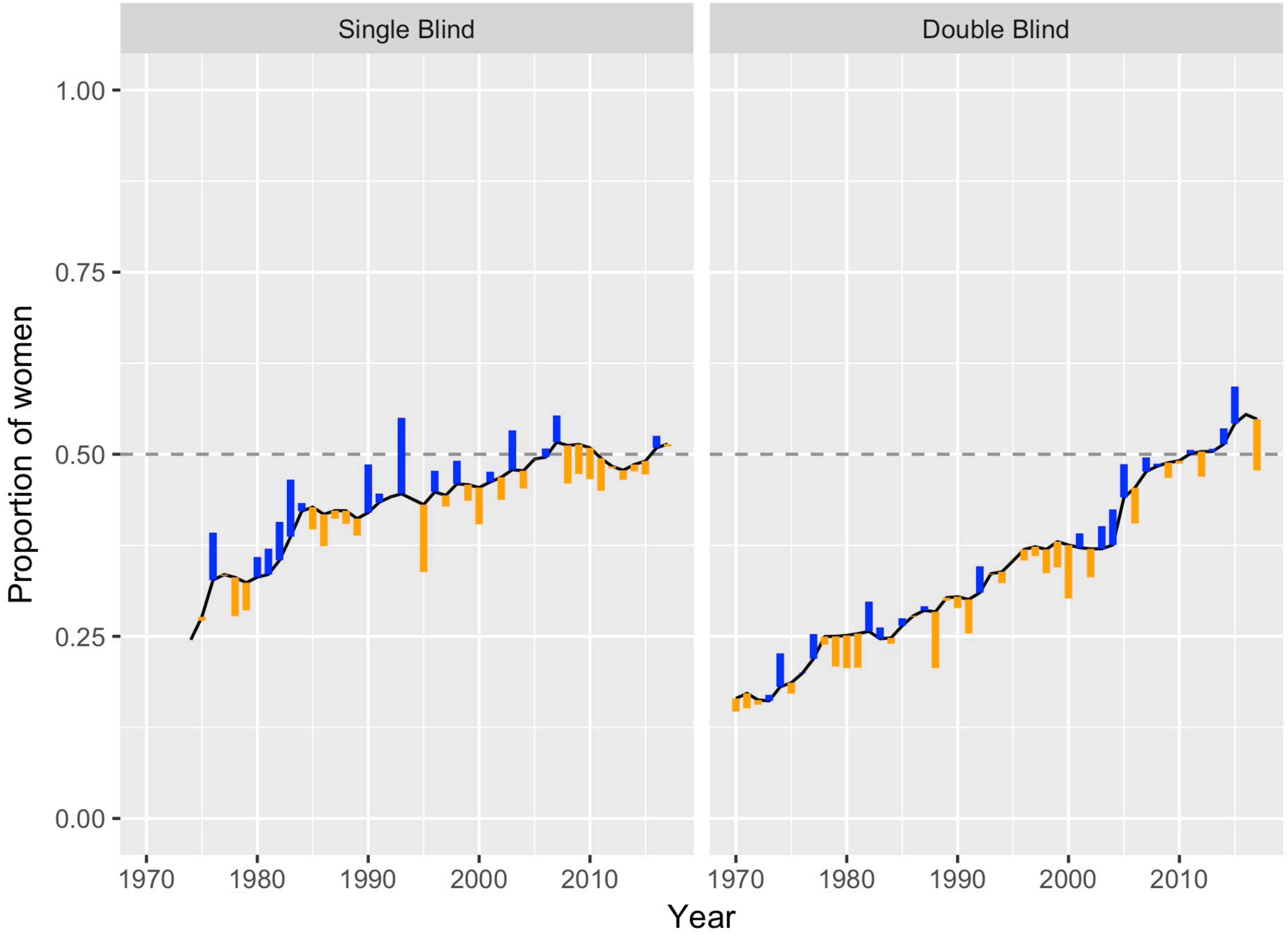
O’Meara et al (2017); Guarino & Borden (2017); Hengel (2018)

Why do women publish less?

- Alternative:
  - submission at equal rates for male and female linguists
  - higher rejection rate for female linguists
- One potential indicator:
  - differences in publication rates between single-blind and double-blind journals

Knoblauch-Westerwick et al (2013)

Single-blind vs double-blind



Role models/leaky pipeline

- Underrepresentation in faculty positions is itself likely a factor in perpetuating the leaky pipeline.
    - In chemistry, female PhD students working with female advisors are more productive and more likely to become faculty themselves.
    - Recent longitudinal study on female undergraduate majors in the geosciences shows a massive effect of female mentorship on retention.

Hernandez et al (2018); Gaule & Piacentini (2018); Sheltzer & Smith (2014)

Role models/leaky pipeline

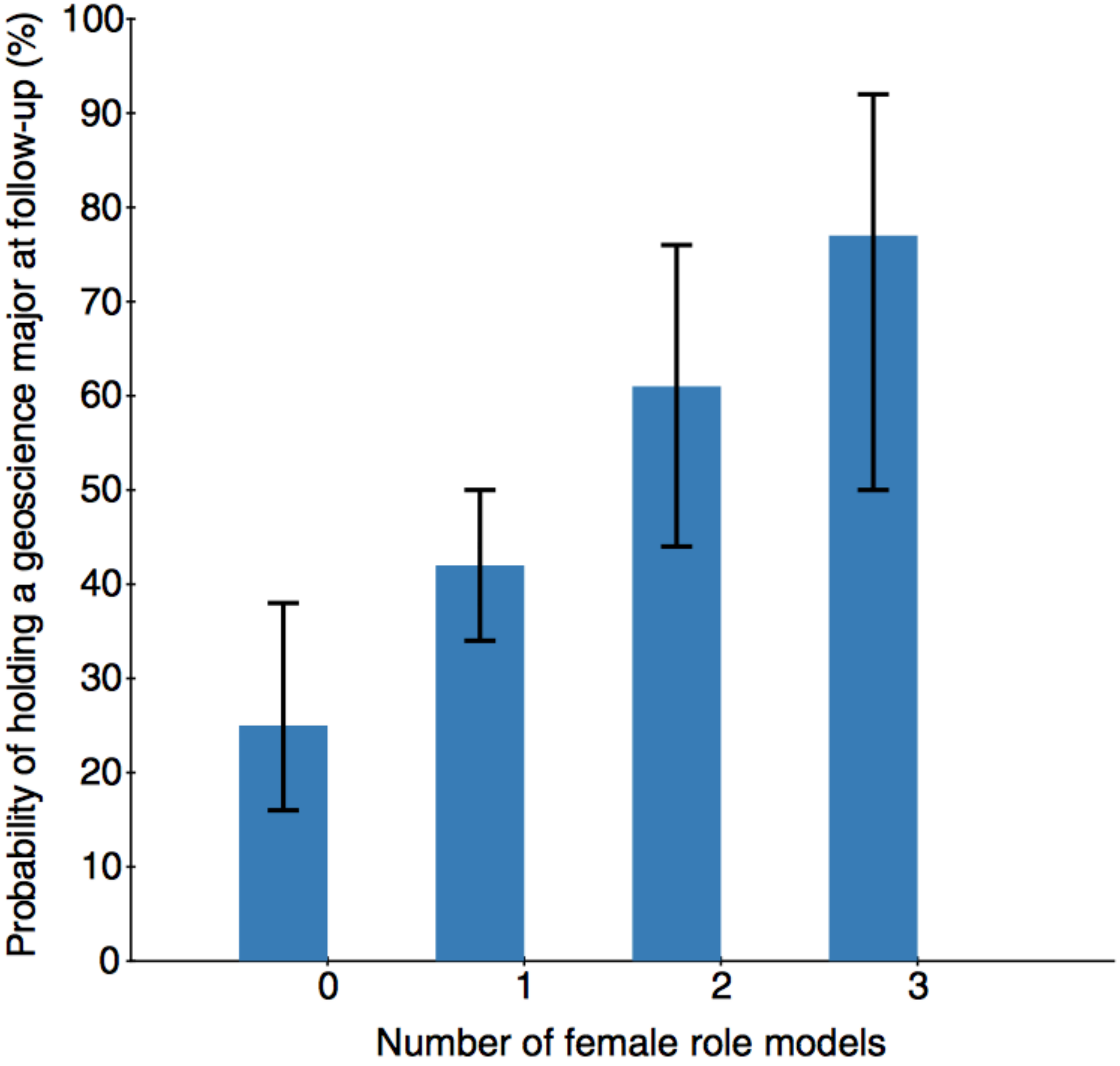
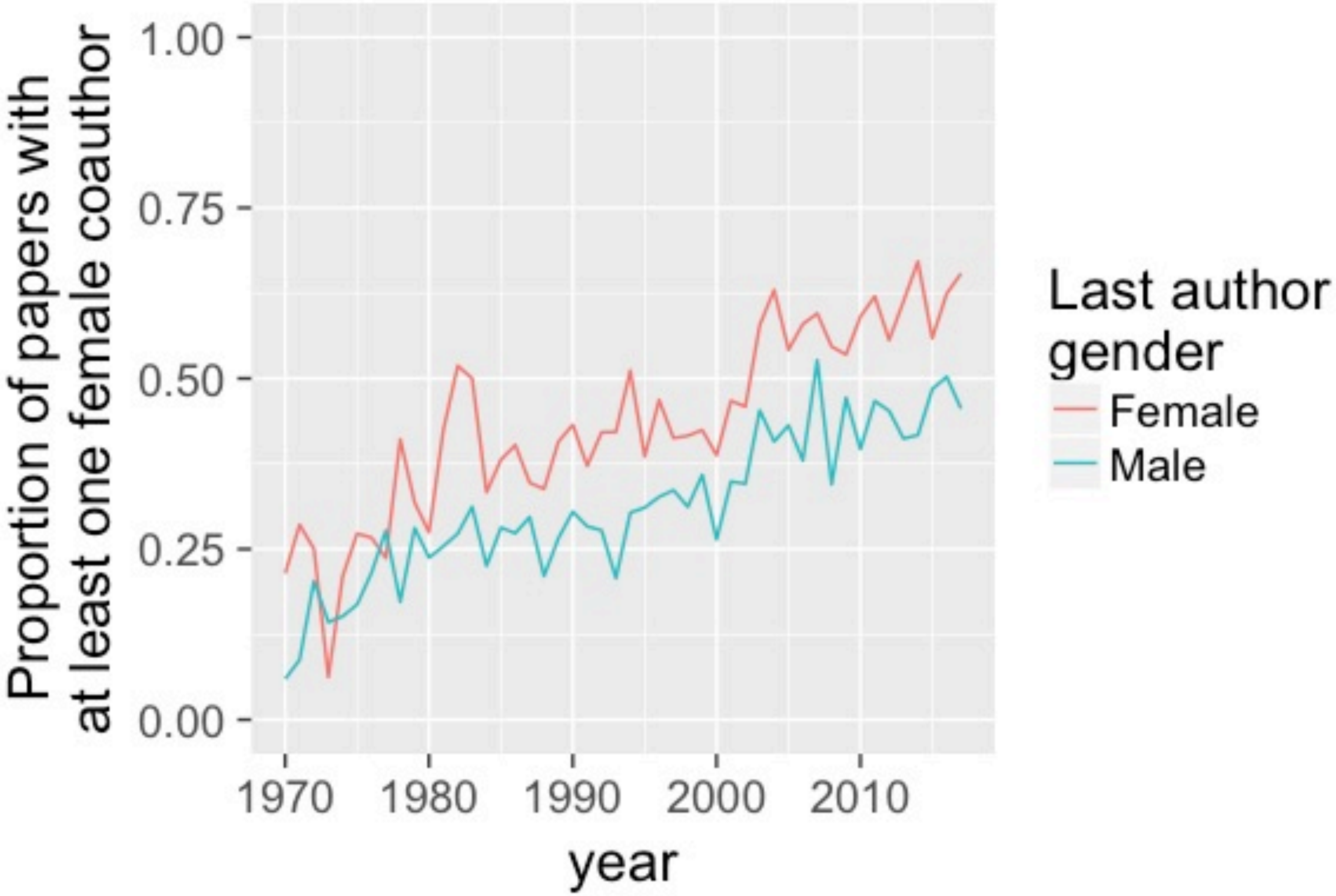


Figure 2. Probability of holding a geoscience-related major at follow-up as a function of the number of female STEM career role models. Predicted values and confidence-interval error bars computed from a weighted multilevel model for the number of role models. Error bars represent 95% confidence intervals.

Hernandez et al (2018)

Female co-authorship



Glass ceiling in NLP

- Growing disparity in proportion of male/female mentors
- Gender gap in time required to achieve mentor status
- Female mentorship increases likelihood of female researchers becoming mentors themselves

Schluter (2018)

Summary

- Women are increasingly under-represented at each successive career stage.
- In many sub-fields women are under-publishing given their representation estimate.
- Male mentors are less likely than female mentors to publish with female co-authors.

Limitations

- If we want to understand why there are fewer female faculty, publications are just one small piece of the problem.

- Information in publication process that we're lacking: submission rates, time under review, etc.

- Technical issues: noise in the data, name matching, gender tagging (possible bias), etc.

Reproducibility

- Much of the data is available at biasinlinguistics.org and we will continue to add what we’ve done.
- Making our analysis pipeline available so that others can do this e.g. for other sub-fields, more journals, etc.

Next steps

- Citation rates, submission rates, related fields, etc.
- Survey on grad student experiences
- What should we as a field do with this information?
  - Hiring/tenure committees taking publication asymmetry into account.
  - Advisor awareness of asymmetry for female grad students in particular.

Thank you!

- Virginia Valian
- Bill Idsardi
- Alyson Reed
- David Robinson
- Brian Joseph
- Kai von Fintel
- Andries Coetzee
- Donca Steriade
- Joe Pater
- Michelle Erskine
- Lara Ehrenhofer
- Savithry Namboodiripad
- Corrine Occhino
- Lynn Hou
- Anne Charity Hudley
- Kristen Syrett
- Kerry Ann O’Meara
- Cognitive Neuroscience of Language Lab
- Language Science Lunch