
DATA VISUALIZATION

LUCAS SMITH

SUMMER 2025



INTRODUCTION & MOTIVATION

- Background in healthcare
- Partner is a physician
- Interested to see overall trends
- Thought that this could be a challenging dataset to work with



INTRODUCTION TO DATA

- Received from TidyTuesday
- Contains research journal article high level data, from 2010-2023
- Only containing gynecological research journals
- 318 observations, of 65 features each
- Mainly around information such as
 - Journal name, title, year of publication
 - Races and ethnicities included in the research
- Required careful manipulation of race and ethnicity
 - Standardization
 - Aggregation



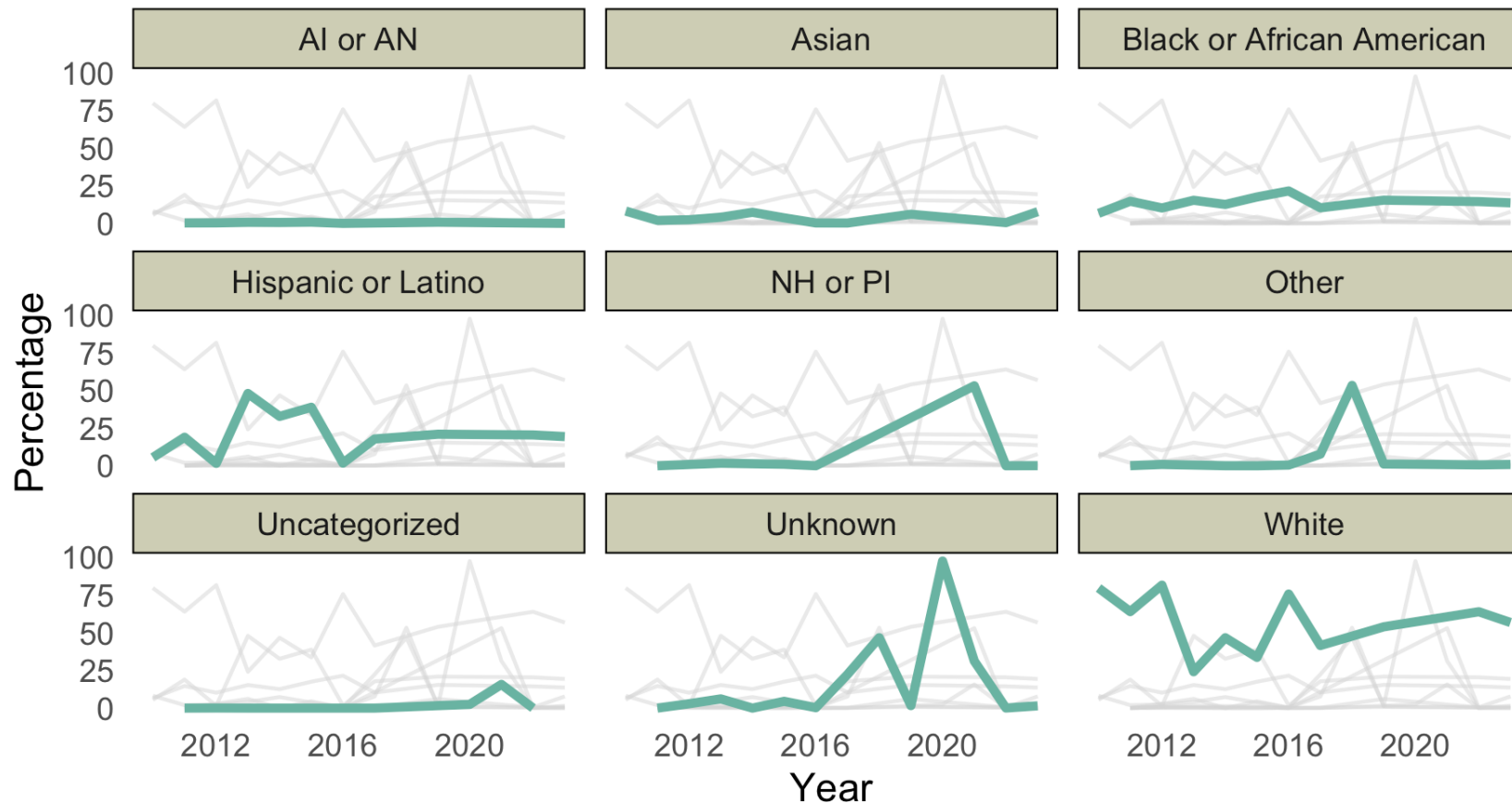
MAIN QUESTIONS

1. How have ethnic and racial inclusion into research changed across time?
 2. How is the length of a study noticeable between different journals?
-
- The overarching concept is to understand how trends could be seen over time.

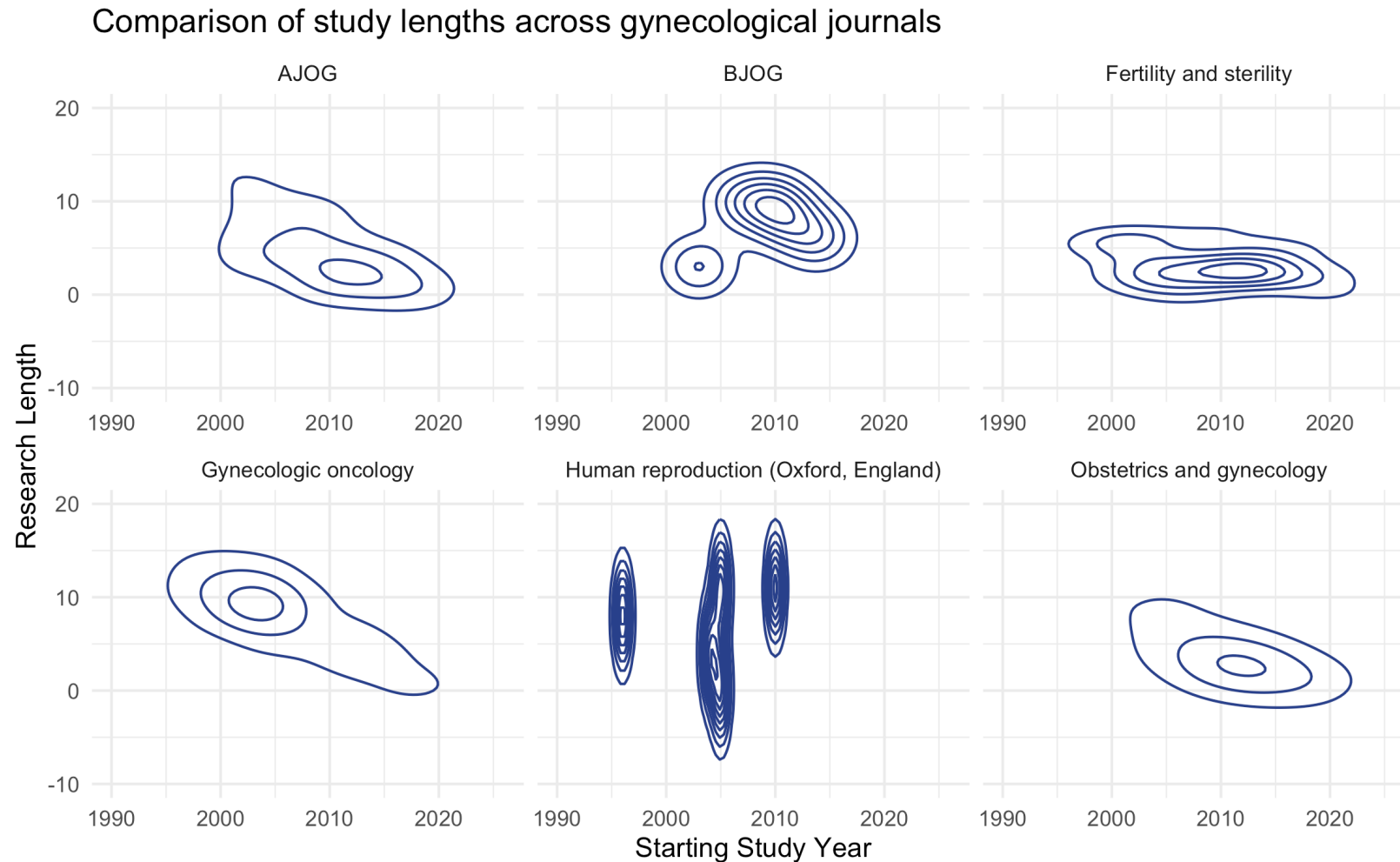


VISUALIZATION I

Comparison of proportional race disparities
in gynecological studies, 2010-2023



VISUALIZATION 2



CONCLUSION

- Comparison of race/ethnicity across time
 - Not much change
 - 2020: highlight of 98% of values were listed as Unknown
- Comparison of study length
 - Journals seemed to have trends of lengths
 - *Family and sterility* being less than 5
 - *BJOG* around 10
 - Human reproduction started in incremental years:
 - 2000, 2005, 2010
- Future
 - During write up, found longer study times – viz 2 based on start years; would want more data
 - Similarly, for viz 1 would want more data for a longer period than just 20 years

