Publication Analysis Summary

1. Total number of publications on WOS: 11,236.

2. Total number of valid publications used for genderize analysis: 10576.

**Criteria for exclusion**: Publications without a proper first name/abbreviated first name for both their corresponding and first authors were excluded from the genderize analysis.

3. Out of the valid publications, overall male vs. female CA and FA:

* + Overall Male CA: 7250 = 66.66%
  + Overall Female CA: 3626= 33.34%
  + Overall Male FA: 5208=54.20%
  + Overall Female FA: 4400= 45.80%

Overall Publications for CA: 10461

* + Male Authors: 7250
  + Female Authors: 3626
  + Unknown Gender: 839

Overall Publications for FA: 10299

* + Male Authors: 5208
  + Female Authors: 4400
  + Unknown Gender: 691

4. Out of the high-confidence publications (Gender prediction probability >= 80%), male/female CA and FA:

For Corresponding Authors (CA), Total authors: 9555, Total Publications=8872

* + High-confidence Male CA: 6532 =68.36%
  + High-confidence Female CA: 3023=31.64%

For First Authors (FA), Total authors: 8301, Total publication: 8301

* + High-confidence Male FA: 4561 = 54.95%
  + High-confidence Female FA: 3740=45.05%

Why the total unique count of publications and the sum of unique CA and FA publications might not match up as expected:

* **Overlap in Roles**: Some publications might list the same individual as both the Corresponding Author (CA) and the First Author (FA). When we count unique publications within CA and FA categories separately, such publications are counted once in each category. However, in the overall unique publication count, each publication is counted only once, regardless of the authorship role overlap.
* **Distinct Counts for Each Role**: The unique count for CAs and FAs refer to the number of unique publications where an individual is identified specifically in that role. If a publication has different individuals serving as CA and FA, it will contribute to the count in both categories. This can lead to a situation where the sum of unique publications by author role (CA + FA) exceeds the total number of unique publications because some publications are being counted twice.

Tasks to be performed:

* ~~Using overall AND high prediction forecast find the number of~~ **~~male CA/year~~**~~, predict the number for 2020 and 2021 and find out if there is a significant difference between expected and real numbers. (show this data in excel and p values) : First with your model and then OLS (Refer to paper:~~ [~~https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8043146/#app1~~](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8043146/#app1) ~~)~~
* ~~Using overall AND high prediction forecast find the number of~~ **~~male FA/year~~**~~, predict the number for 2020 and 2021 and find out if there is a significant difference between expected and real numbers. (show this data in excel and p values) : First with your model and then OLS~~
* ~~Using overall AND high prediction forecast find the number of~~ **~~female CA/year~~**~~, predict the number for 2020 and 2021 and find out if there is a significant difference between expected and real numbers. (show this data in excel and p values) : First with your model and then OLS~~
* ~~Using overall AND high prediction forecast find the number of~~ **~~female FA/year~~**~~, predict the number for 2020 and 2021 and find out if there is a significant difference between expected and real numbers. (show this data in excel and p values) : First with your model and then OLS~~
* ~~Use Overall results to: (Refer to paper~~ [~~https://www.nature.com/articles/s43856-023-00417-3~~](https://www.nature.com/articles/s43856-023-00417-3) ~~)~~
* ~~1) find the correlation between female CA and female FA~~
* ~~2) Male CA and female FA~~
* ~~3) Male CA and male FA~~