Data Manual for "The Diversity of Conference Presenters in Virtual and In-Person Conferences"

Mitali Mathur, Jenna Stearns, and Keer Yang*

March 2025

1 Introduction

With the onset of the COVID pandemic in early 2020, academic conferences quickly pivoted to virtual formats for much of the next two years. One potential benefit of the move from in-person to virtual conferences may have been to improve the diversity of conference attendees.

In this project, we aim to investigate whether virtual conference formats improved the diversity of presenters at prominent national conferences in finance. We collect data on the gender and affiliation of conference presenters from four major annual finance conferences: the American Finance Association Annual Meeting, the Society for Financial Studies Cavalcade, the Western Finance Association Conference, and the Financial Management Association Annual Meeting. We scrape data from published conference programs from 2017 to 2023. These years cover in-person, fully virtual, and hybrid conferences. In particular, data from in-person conferences in 2017-2019 allows us to account for any pre-trends in conference participation over time.

This document is prepared as a data manual for the public distribution of our collected dataset[1]. Section 2 describes our data collection and cleaning process. Section 3 provides variable definitions for the dataset. Section 4 presents key summary statistics.

2 Data

2.1 Data Description

We web-scraped data from four main finance conferences across the years 2017-2023:

- AFA (American Finance Association) Annual Meeting
- SFS (Society for Financial Studies) Annual Meeting
- WFA (Western Finance Association) Annual Meeting
- FMA (Financial Management Association) Annual Meeting

2.2 Data Availability

Based on data availability and conference detail, we were limited by a few restrictions

- FMA 2017-2018 → We only had a list of presenters, no titles or authors
- WFA → Presenters were not indicated, so we only have presenter gender for single-authored papers

*Mathur is from the University of California, Davis, Davis, CA, E-mail: mrmathur@ucdavis.edu. Stearns is from the University of California, Davis, Davis, CA, E-mail: jestearns@ucdavis.edu. Yang is from the University of California, Davis, Davis, CA, E-mail: kkeyang@ucdavis.edu. Data collection was supported by an AFFECT/JFE grant from AFFECT and the American Finance Association awarded to Stearns and Yang. We are grateful for the support and responsible for all errors.

[1] *final_conference_data_public.xlsx*

- SFS 2017-2021 → Presenters were not indicated
- SFS 2022 → We do not have this data

2.3 Gender Merges

2.3.1 process Overview

The genders of presenters in this analysis were assigned based on the following process (done separately for each presenter and each author):

1. We took the first name and last name of each person in the AFA (self-reported females) data
2. We merged this data with the conference or author data on first and last name
3. We merged the AFA data with the full names (including middle names) of the presenters/authors in our data that weren't merged based on first and last name
4. We subsetted the dataset to the presenters/authors who weren't merged with AFA data
5. We merged this data with the Genderize io data via an API pull
6. We appended this data to the "female" subset of our merged AFA data

Therefore, the final dataset is based on a first pass of merging names from a self-reported list of people affiliated with AFA who identify as female, then a second pass of merging names with a universal database of names.

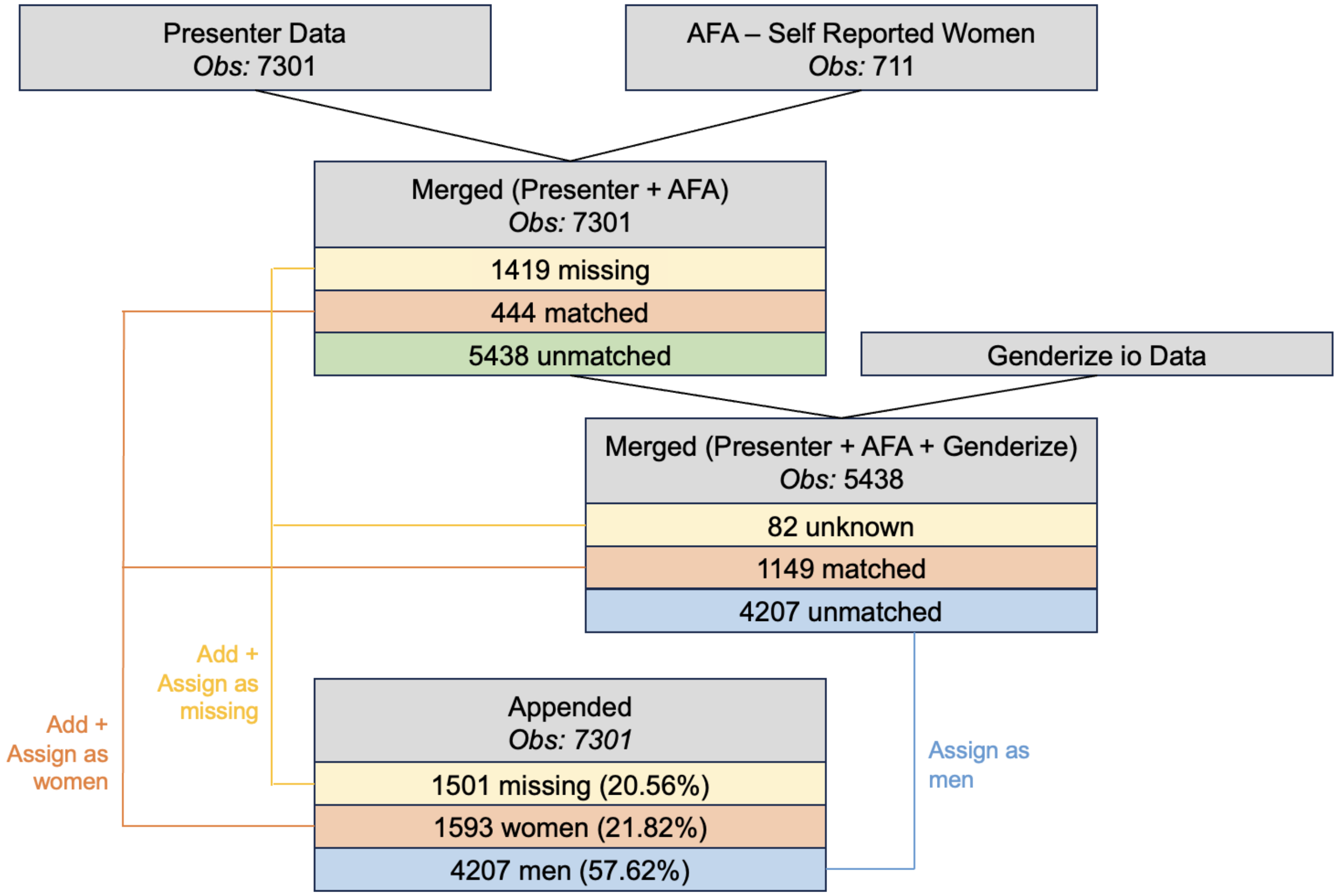
2.3.2 Different Thresholds

For each analysis, we used three different thresholds for assigning genders on the basis of names:

- *Merge 1* -
  - All names that were in the self-reported AFA dataset were assigned to female
  - All names that were not in the self-reported AFA dataset were then merged to Genderize io
    - These names were assigned to female if Genderize io assigned the name to "female" (Note that this occurs when the _genderize variable was "female" and the _prob variable had a probability of at least 50%)
    - These were assigned to male if Genderize io assigned the name to "male" (Note that this occurs when the _genderize variable was "male" and the _prob variable had a probability of at least 50%)
  - All names that were not matched to the Genderize io database and names that the Genderize io database could not assign a gender to have a missing gender characterization
- *Merge 2* -
  - All names that were in the self-reported AFA dataset were assigned to female
  - All names that were not in the self-reported AFA dataset were then merged to Genderize io
    - These names were assigned to female if Genderize io assigned the name to "female" *and* the probability the name was female in Genderize io was at least 70%
    - These were assigned to male if Genderize io assigned the name to "male" *and* the probability the name was male in Genderize io was at least 70%
  - All names that were not matched to the Genderize io database, names that the Genderize io database could not assign a gender to, and names in which the probability was less than 70% have a missing gender characterization

- *Merge 3* -
  - All names that were in the self-reported AFA dataset were assigned to female
  - All names that were not in the self-reported AFA dataset were then merged to Genderize io
    * These names were assigned to female if Genderize io assigned the name to "female" *and* the count the name was in the Genderize io database was at least 1000
    * These were assigned to male if Genderize io assigned the name to "male" *and* the count the name was in the Genderize io database was at least 1000
  - All names that were not matched to the Genderize io database, names that the Genderize io database could not assign a gender to, and names in which the count was less than 1000 have a missing gender characterization

An example flow of Merge 1 using the process described for presenters can be seen here:



2.3.3 Genderize io Performance

To test the success of Genderize io's performance, we ran the self-reported list of female names through Genderize io and assessed the performance of each of the three types of merges listed above. All names should be female, but the table below highlights how Genderize io classified all female names. The Genderize io approach performs fairly well as it identifies 71% - 88% of females in our dataset as female.

Merge type	Female	Male	Missing
Merge 1	631 (88.75%)	78 (10.97%)	2 (0.28%)
Merge 2	590 (82.98%)	56 (7.88%)	65 (9.14%)
Merge 3	507 (71.31%)	38 (5.34%)	166 (23.35%)

Table 1: Merge Type by Gender and Missing Data

3 Variable Definition

Name	Description	Format	Notes
conf	Conference	String	Options: AFA, FMA, SFS, WFA
conf_type	Conference Type	String	Options: In Person, Hybrid, Virtual
conf_year	Conference Year	String	Options: 2017-2023
conf_dates	Conference Dates	String	
conf_location	Conference Location	String	Missing if conference was fully virtual
session_name	Session Name	String	
session_chair	Session Chair's Name	String	
session_chair_aff	Session Chair's Affiliation	String	
session_chair2	Session Co-Chair's Name	String	
session_chair2_aff	Session Co-Chair's Affiliation	String	
title	Title of the Paper Presented	String	
presenter	Presenter's Name	String	
presenter_aff	Presenter's Affiliation	String	
presenter_fem_v1	Presenter's Gender - version 1 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
presenter_fem_v2	Presenter's Gender - version 2 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
presenter_fem_v3	Presenter's Gender - version 3 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
presenter_afa	Dummy if the presenter was in the AFA self-reported dataset	Numeric	1 = in the AFA dataset, . = not in the AFA dataset
presenter_genderize	Gender assigned by the genderize io algorithm	String	Male or Female
presenter_count	Number of times the name appeared in the genderize io database	Numeric	
presenter_prob	Probability the name is the assigned gender in the genderize io database	Numeric	
presenter2	Co-Presenter's Name	String	
presenter2_aff	Co-Presenter's Affiliation	String	
author1	Author 1's Name	String	
author1_aff	Author 1's Affiliation	String	
author1_fem_v1	Author 1's Gender - version 1 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author1_fem_v2	Author 1's Gender - version 2 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author1_fem_v3	Author 1's Gender - version 3 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author1_afa	Dummy if the presenter was in the AFA self-reported dataset	Numeric	1 = in the AFA dataset, . = not in the AFA dataset
author1_genderize	Gender assigned by the genderize io algorithm	String	Male or Female
author1_count	Number of times the name appeared in the genderize io database	Numeric	
author1_prob	Probability the name is the assigned gender in the genderize io database	Numeric	
author2	Author 2's Name	String	
author2_aff	Author 2's Affiliation	String	

author2_fem_v1	Author 2's Gender - version 1 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author2_fem_v2	Author 2's Gender - version 2 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author2_fem_v3	Author 2's Gender - version 3 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author2_afa	Dummy if the presenter was in the AFA self-reported dataset	Numeric	1 = in the AFA dataset, . = not in the AFA dataset
author2_genderize	Gender assigned by the genderize io algorithm	String	Male or Female
author2_count	Number of times the name appeared in the genderize io database	Numeric	
author2_prob	Probability the name is the assigned gender in the genderize io database	Numeric	
author3	Author 3's Name	String	
author3_aff	Author 3's Affiliation	String	
author3_fem_v1	Author 3's Gender - version 1 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author3_fem_v2	Author 3's Gender - version 2 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author3_fem_v3	Author 3's Gender - version 3 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author3_afa	Dummy if the presenter was in the AFA self-reported dataset	Numeric	1 = in the AFA dataset, . = not in the AFA dataset
author3_genderize	Gender assigned by the genderize io algorithm	String	Male or Female
author3_count	Number of times the name appeared in the genderize io database	Numeric	
author3_prob	Probability the name is the assigned gender in the genderize io database	Numeric	
author4	Author 4's Name	String	
author4_aff	Author 4's Affiliation	String	
author4_fem_v1	Author 4's Gender - version 1 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author4_fem_v2	Author 4's Gender - version 2 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author4_fem_v3	Author 4's Gender - version 3 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author4_afa	Dummy if the presenter was in the AFA self-reported dataset	Numeric	1 = in the AFA dataset, . = not in the AFA dataset
author4_genderize	Gender assigned by the genderize io algorithm	String	Male or Female
author4_count	Number of times the name appeared in the genderize io database	Numeric	
author4_prob	Probability the name is the assigned gender in the genderize io database	Numeric	
author5	Author 5's Name	String	
author5_aff	Author 5's Affiliation	String	
author5_fem_v1	Author 5's Gender - version 1 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author5_fem_v2	Author 5's Gender - version 2 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched

author5_fem_v3	Author 5's Gender - version 3 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author5_afa	Dummy if the presenter was in the AFA self-reported dataset	Numeric	1 = in the AFA dataset, . = not in the AFA dataset
author5_genderize	Gender assigned by the genderize io algorithm	String	Male or Female
author5_count	Number of times the name appeared in the genderize io database	Numeric	
author5_prob	Probability the name is the assigned gender in the genderize io database	Numeric	
author6	Author 6's Name	String	
author6_aff	Author 6's Affiliation	String	
author6_fem_v1	Author 6's Gender - version 1 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author6_fem_v2	Author 6's Gender - version 2 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author6_fem_v3	Author 6's Gender - version 3 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author6_afa	Dummy if the presenter was in the AFA self-reported dataset	Numeric	1 = in the AFA dataset, . = not in the AFA dataset
author6_genderize	Gender assigned by the genderize io algorithm	String	Male or Female
author6_count	Number of times the name appeared in the genderize io database	Numeric	
author6_prob	Probability the name is the assigned gender in the genderize io database	Numeric	
author7	Author 7's Name	String	
author7_aff	Author 7's Affiliation	String	
author7_fem_v1	Author 7's Gender - version 1 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author7_fem_v2	Author 7's Gender - version 2 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author7_fem_v3	Author 7's Gender - version 3 gender merge	Numeric	1 = Female, 0 = Male, . = Unmatched
author7_afa	Dummy if the presenter was in the AFA self-reported dataset	Numeric	1 = in the AFA dataset, . = not in the AFA dataset
author7_genderize	Gender assigned by the genderize io algorithm	String	Male or Female
author7_count	Number of times the name appeared in the genderize io database	Numeric	
author7_prob	Probability the name is the assigned gender in the genderize io database	Numeric	
author_count	Total Number of Authors	Numeric	
discussant	Discussant's Name	String	
discussant_aff	Discussant's Affiliation	String	
discussant2	Co-Discussant's Name	String	
discussant2_aff	Co-Discussant's Affiliation	String	
discussant3	Co-Discussant's Name	String	
discussant3_aff	Co-Discussant's Affiliation	String	

Table 2: Variable Definition

4 Summary Statistics

4.1 Overview of Presenter Genders

Conference	Year	Type	Female Presenters	Male Presenters
AFA	2017	In Person	60	173
AFA	2018	In Person	51	194
AFA	2019	In Person	50	199
AFA	2020	In Person	61	177
AFA	2021	Virtual	57	182
AFA	2022	Virtual	51	190
AFA	2023	In Person	63	178
FMA	2017	In Person	172	363
FMA	2018	In Person	174	368
FMA	2019	In Person	113	262
FMA	2020	Virtual	109	319
FMA	2021	Hybrid	139	366
FMA	2022	Hybrid	206	526
FMA	2023	In Person	194	418
SFS	2017	In Person	2	14
SFS	2018	In Person	7	9
SFS	2019	In Person	0	15
SFS	2020	Virtual	3	19
SFS	2021	Virtual	4	10
SFS	2022	In Person	NA	NA
SFS	2023	In Person	37	95
WFA	2017	In Person	6	15
WFA	2018	In Person	7	18
WFA	2019	In Person	2	26
WFA	2020	Virtual	9	15
WFA	2021	Virtual	5	19
WFA	2022	In Person	6	17
WFA	2023	In Person	5	20

Table 3: Conference Presenter Data Using Merge 1

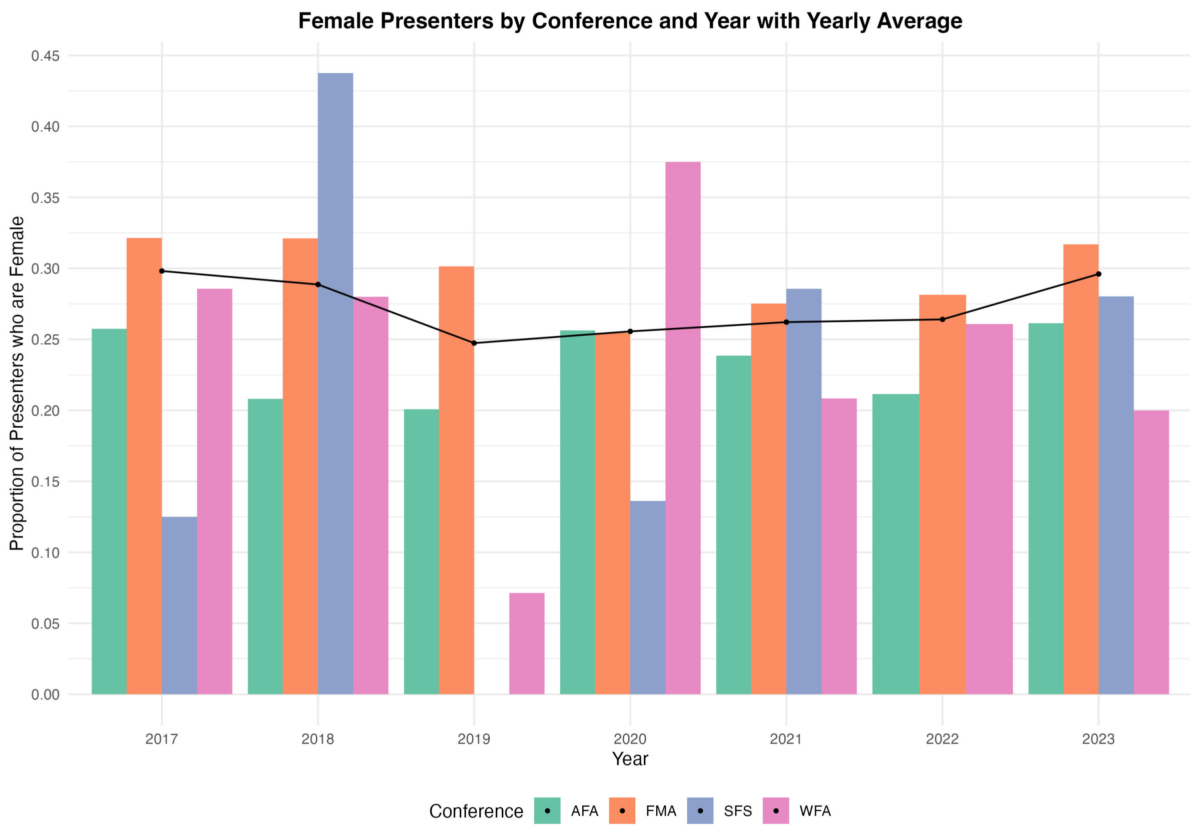


Figure 1: Conference Presenter Data Using Merge 1

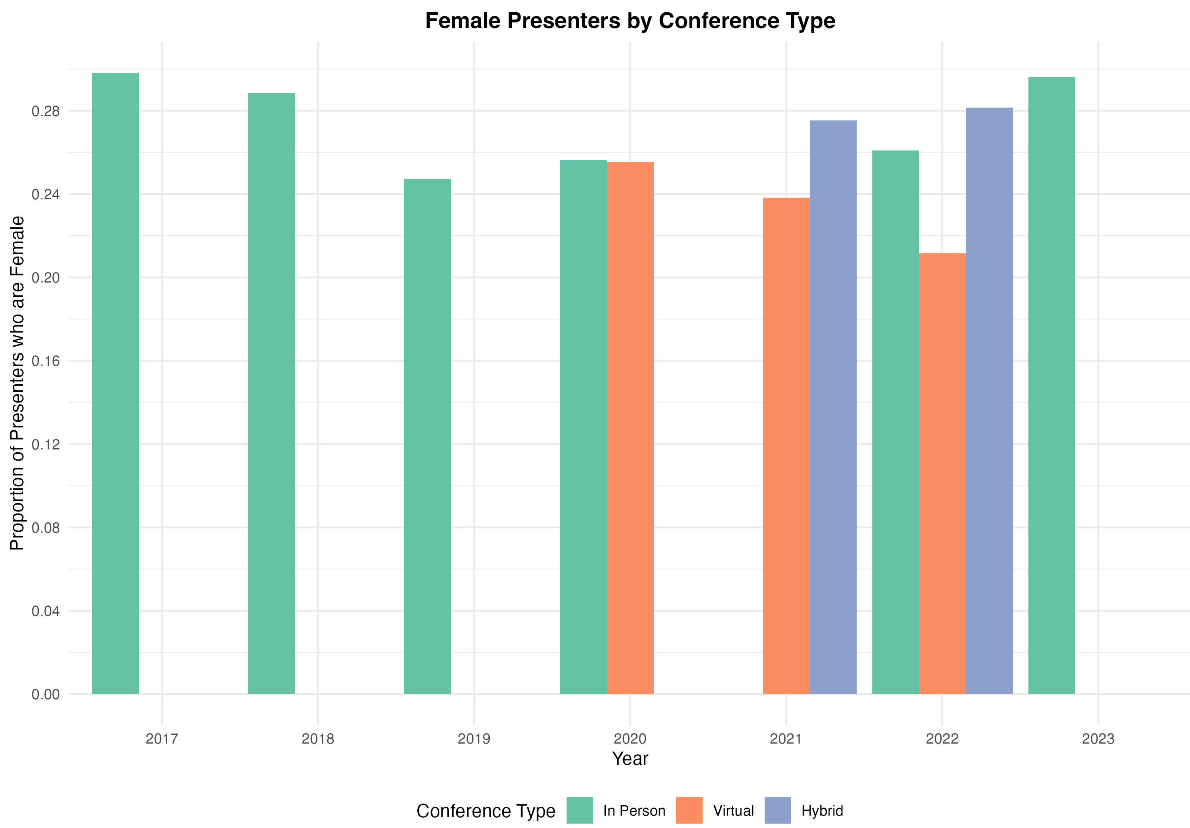


Figure 2: Conference Presenter Data Using Merge 1

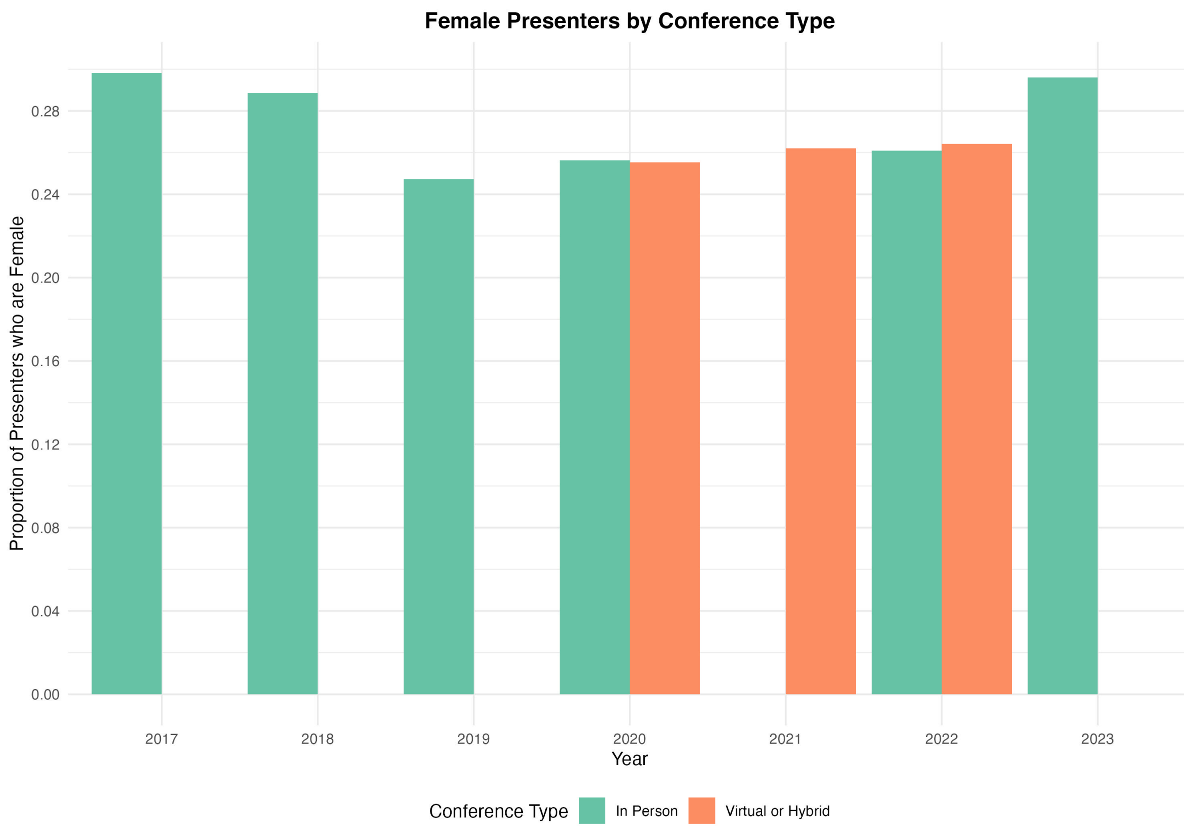


Figure 3: Conference Presenter Data Using Merge 1

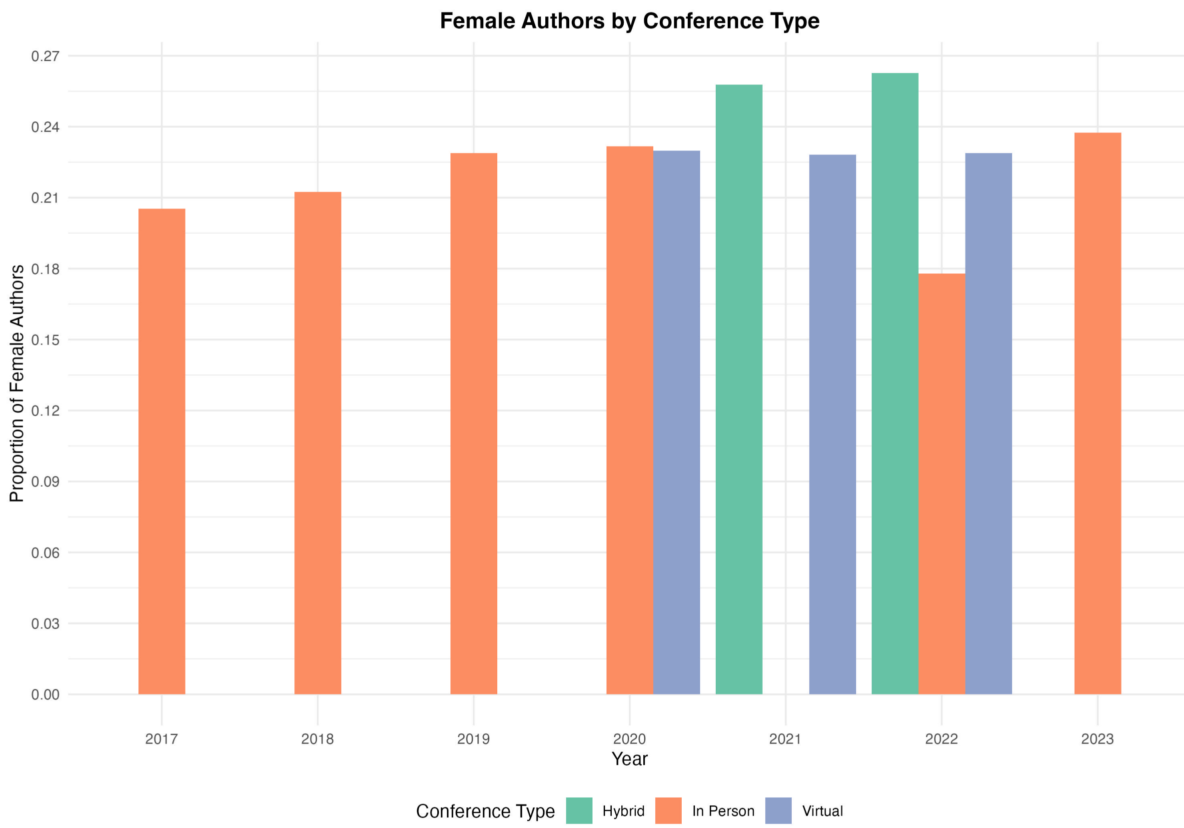


Figure 4: Conference Author Data Using Merge 1

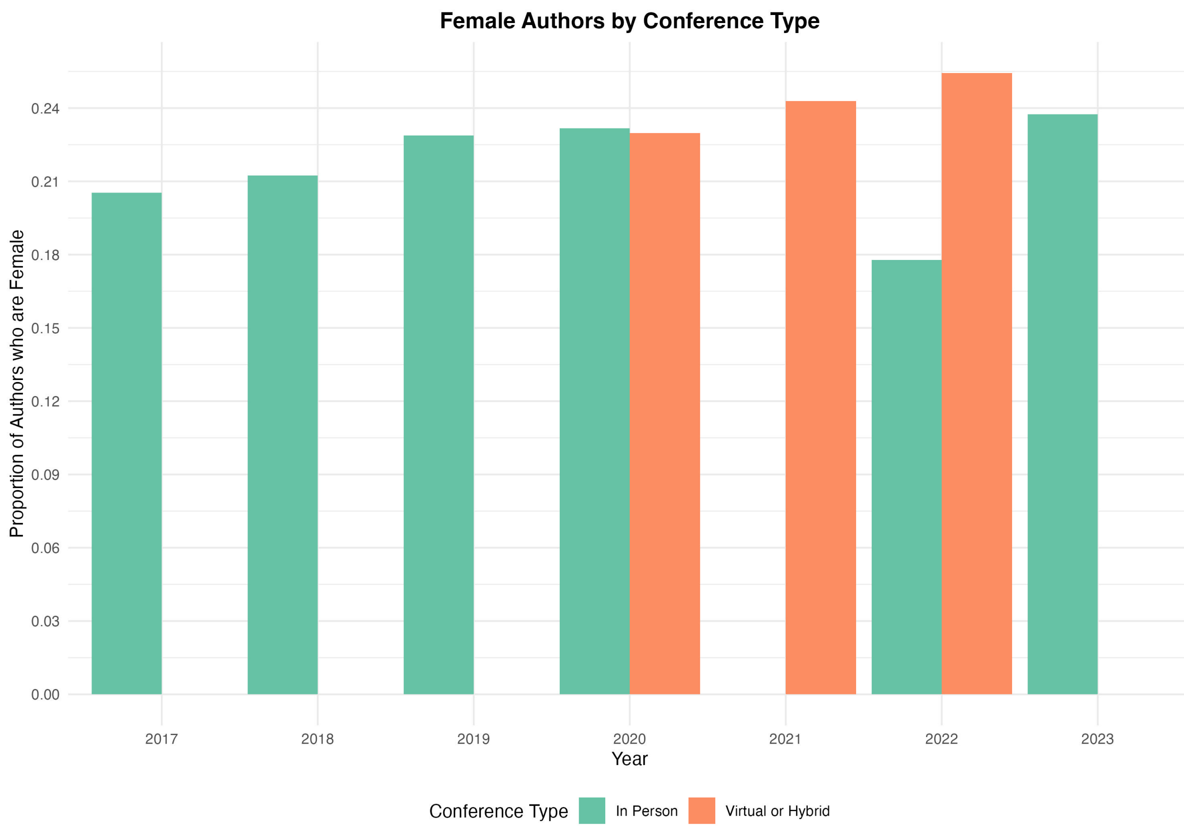


Figure 5: Conference Author Data Using Merge 1