Gender and Politicians' Exposure to Profanity-Based Online Incivility

DACSS 603 Final Project

Madison Hess

Table of contents

Introduction

This research paper aims to investigate the association between gender and politicians' exposure to online incivility, specifically profanity-based incivility on Twitter. This analysis follows the study Politicians in the Line of Fire: Incivility and the Treatment of Women on Social Media (Rheault, Rayment, and Musulan 2019) using data obtained through Harvard Dataverse. The dataset consists of variables constructed from tweets addressed directly to Canadian politicians and U.S. Senators. These tweets were collected over a one-month period in 2017 by filtering for politicians' official Twitter handles, ensuring that the messages analyzed were sent directly to the public officials. The original study measures incivility using an approach that captures multiple forms of uncivil language, including profanity, insults, threats, and hate speech, but it does not examine these different types individually. This project replicates the original analysis in terms of structure, but considers a narrower definition of incivility by focusing specifically on profanity-based language. Profanity-based language in this context was measured using a LIWC (Linguistic Inquiry and Word Count) count of swear words in tweets. By isolating profanity from broader forms of incivility, this analysis provides a more focused examination of whether gender differences in online harassment persist when the definition of incivility is more restricted.

Literature Review

Online harassment has become a regular part of political life, especially as more political debate takes place on social media platforms. Recent survey evidence shows that political views are now the most common reason Americans believe they were targeted by online harassment, highlighting how hostile online political spaces have become (Vogels, 2021). These environments are especially challenging for women in politics. Krook and Sanín (2019) note that nearly all female members of parliament globally have experienced psychological violence at some point during their careers. Psychological violence includes sexist remarks, threats, intimidation, and harassment in this context. The authors argue that this type of abuse works to discourage women from political participation and to police who is seen as legitimate in positions of power.

Public opinion research reflects similar concerns. A recent Pew Research Center survey finds that a majority of Americans believe women face significant obstacles when seeking high political office, and most commonly that women have to do more than men to prove themselves (Horowitz & Goddard, 2023). Large majorities of the public also point to gender discrimination, lack of support from party leaders, and sexual harassment as major barriers for women in politics. These perceptions suggest that gendered challenges in politics are widely recognized, even beyond academic research.

While misogyny and political violence are not new, online spaces have changed how these behaviors are carried out. Social media makes harassment easier to direct at politicians, more visible to the public, and often more persistent. For many people, including myself, the online harassment faced by Hillary Clinton during the 2016 U.S. presidential election made these dynamics especially visible. Weaving et al. (2023) analyzed millions of tweets referring to Clinton and find that misogynistic language increased sharply after she announced her candidacy and continued throughout the campaign. Citing backlash theory, they argue that this rise in misogyny reflects a response to women seeking power and challenging traditional gender roles, particularly when those women are highly visible.

Importantly, the consequences of online incivility extend beyond politicians themselves. Vrielink and van der Pas (2024) show that exposure to sexist comments can reduce women's political ambition, meaning their interest in pursuing political roles or careers, even among individuals who are only observing the abuse. This literature highlights that online harassment is widespread, gendered, and has significant consequences. It also suggests that different forms of incivility may operate in different ways, making it important to examine specific types of harassment, such as profanity-based language, instead of treating incivility as a single, combined category.

Research Question

To state it clearly, my research question is: Are female politicians more likely than male politicians to be exposed to *profanity-based* online incivility on Twitter?

Hypothesis

I hypothesize that female politicians experience higher rates of profanity-based online incivility on Twitter compared to their male counterparts.

My hypothesis is informed by the nature of this project as a partial replication of prior research, as well as by existing literature on gendered political harassment. Krook and Sanín (2019) argue that violence against women in politics stems from misogyny and functions to police and enforce patriarchal norms by punishing women who are perceived to violate traditional gender roles. From this perspective, online harassment reflects broader patterns of symbolic and psychological violence directed at women in public life. In this context, profanity-based harassment can be understood as one form of this violence directed at women in public office.

Descriptive Statistics

The dataset consists of two samples of high-profile public officials from Canada and the United States. The Canadian sample includes 195 politicians (37% women), while the U.S. sample includes all 100 U.S. Senators (21% women). Tweets addressed directly to each politician were collected from Twitter using a streaming API during a one-month collection period in 2017. Incivility was identified using machine learning models trained on identifying uncivil content including profanity, insults, threats, personal attacks, or hate speech.

For this research project, the politician-level aggregated datasets for Canada and the United States were merged into a single dataset. Each observation represents an individual politician, with variables summarizing demographic characteristics and tweet volume/content.

```r
library(tidyverse)
```

```
Warning: package 'ggplot2' was built under R version 4.4.3
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   4.0.1     v tibble    3.3.0
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.1.0
```

```
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
# read in
canada <- read_csv("/Users/maddiehess/Desktop/FINAL_603/canada_aggregate.csv")
```

```
Rows: 195 Columns: 12
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): name, province, position
dbl (9): uncivil, follower_count, gender, party, visible_minority, federal, ...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
usa <- read_csv("/Users/maddiehess/Desktop/FINAL_603/usa_aggregate.csv")
```

```
Rows: 100 Columns: 11
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (2): name, state
dbl (9): uncivil, follower_count, gender, party, visible_minority, seniority...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
dim(canada)
```

```
[1] 195  12
```

```r
dim(usa)
```

```
[1] 100  11
```

```
# merge data
canada <- canada |>
  mutate(country = "Canada")

usa <- usa |>
  mutate(country = "United States")
```

Above, you can see the Canadian sample contains 195 observations and 12 variables, while the US sample contains 100 observations and 11 variables. After merging the two, only necessary variables were kept:

```
# keeping only necessary variables for analysis (some are only included in US set only or Can
politicians <- bind_rows(canada, usa) |>
  select(-uncivil, -state, -seniority, -position, -province, -federal, -swearjar)
```

Outcome Variable

- `profanity_rate` is the continuous outcome variable. It represents the proportion of tweets addressed to a politician that contain profanity, calculated as the number of tweets containing swear words (`liwc_swear`) divided by the total number of tweets received (`count`).

```
# creating profanity rate and factoring gender
politicians <- politicians |>
  mutate(
    profanity_rate = liwc_swear / count,
    gender = factor(gender, levels = c(0, 1),
                    labels = c("Male", "Female")
  ))
```

Let's take a look at the new dataset and other variables:

```
dim(politicians)
```

```
[1] 295   9
```

```
glimpse(politicians)
```

```
Rows: 295
Columns: 9
$ name             <chr> "Ahmed Hussen", "Al Hawkins", "Amarjeet Sohi", "Amrik~
$ follower_count   <dbl> 17634, 1122, 14103, 2895, 4151, 3067, 22789, 3371, 22~
$ gender           <fct> Male, Male, Male, Male, Male, Male, Female, Male, Mal~
$ party            <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 0,~
$ visible_minority <dbl> 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ count            <dbl> 1202, 21, 536, 50, 235, 81, 620, 58, 28, 241, 737, 18~
$ liwc_swear       <dbl> 42, 0, 22, 0, 6, 0, 20, 3, 0, 2, 52, 0, 2, 2, 165, 0,~
$ country          <chr> "Canada", "Canada", "Canada", "Canada", "Canada", "Ca~
$ profanity_rate   <dbl> 0.034941764, 0.000000000, 0.041044776, 0.000000000, 0~
```

Explanatory Variable

- `gender` is the primary explanatory variable of interest. It is coded as a binary factor indicating whether a politician is male or female.

Control Variables

- `follower_count` measures the number of Twitter followers a politician has and serves as a proxy for online visibility.
- `party` is a binary indicator of party affiliation (1 = Liberal in Canada, Democrat in the US)
- `visible_minority` indicates whether a politician belongs to a visible minority group (1 = visible minority) and is included to account for racial minorities.
- `country` indicates whether the politician is from Canada or the US and is included to account for differences in political context and social media environments between the two countries.

Other Variables

- `name` identifies each politician.
- `count` is the total number of tweets addressed to a politician and is used to construct the outcome variable `profanity_rate`.

All together, we have 295 observations and 9 variables in the new dataset.

DAG (Directed Acyclic Graph)

While gender is treated as the primary explanatory variable, this analysis does not claim a causal effect. Since the data we are dealing with is cross-sectional and not all relevant confounding factors are observed, the results should be interpreted as associations rather than causal effects.

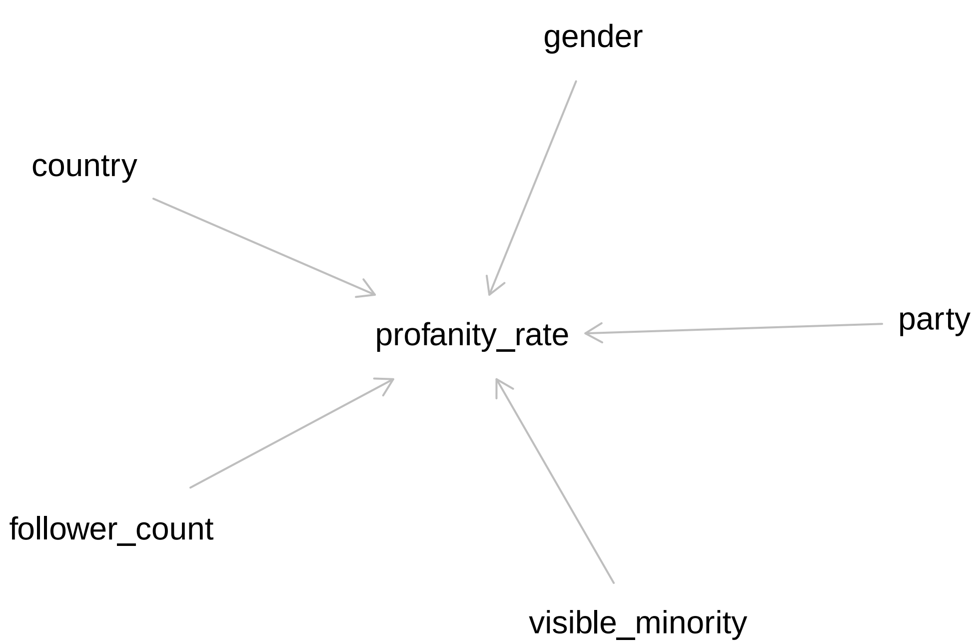
I formalize my assumptions below:

```
library(dagitty)

dag <- dagitty("dag {
  gender -> profanity_rate
  follower_count -> profanity_rate
  party -> profanity_rate
  visible_minority -> profanity_rate
  country -> profanity_rate
}")

plot(dag)
```

```
Plot coordinates for graph not supplied! Generating coordinates, see ?coordinates for how to
```



The DAG illustrates that gender is associated with profanity-based incivility, while also accounting for other factors such as follower count, party affiliation, visible minority status, and country, all of which may shape the tone of tweets directed at politicians.

Importantly, the DAG highlights the presence of potential unobserved factors that are not captured in the dataset but may be associated with gender and independently influence exposure to online incivility. These include differences in political prominence, media visibility beyond follower counts, issue positions, public controversy, or patterns of engagement on social media. These factors may affect how frequently a politician is targeted online and the type of language used in responses, but cannot be directly measured with the available data.

Model Fitting

```r
options(scipen = 999)

mod <- lm(
  profanity_rate ~ gender + follower_count + party + visible_minority + country,
  data = politicians
)

summary(mod)
```

```

Call:
lm(formula = profanity_rate ~ gender + follower_count + party +
    visible_minority + country, data = politicians)

Residuals:
     Min       1Q   Median       3Q      Max
-0.04984 -0.02213 -0.01030  0.00984  0.63448

Coefficients:
                         Estimate      Std. Error t value       Pr(>|t|)
(Intercept)           0.025911007655  0.006342640636   4.085 0.000057086700 ***
genderFemale         -0.014277335251  0.006834461796  -2.089         0.0376 *
follower_count        0.000000012230  0.000000006833   1.790         0.0745 .
party                 0.006260208891  0.006675495696   0.938         0.3491
visible_minority     -0.003815764280  0.010084959168  -0.378         0.7054
countryUnited States  0.046162201110  0.006967065452   6.626 0.000000000169 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.05337 on 289 degrees of freedom
Multiple R-squared:  0.1848,    Adjusted R-squared:  0.1707
F-statistic:  13.1 on 5 and 289 DF,  p-value: 0.00000000001663
```

```r
confint(mod, parm="genderFemale")
```

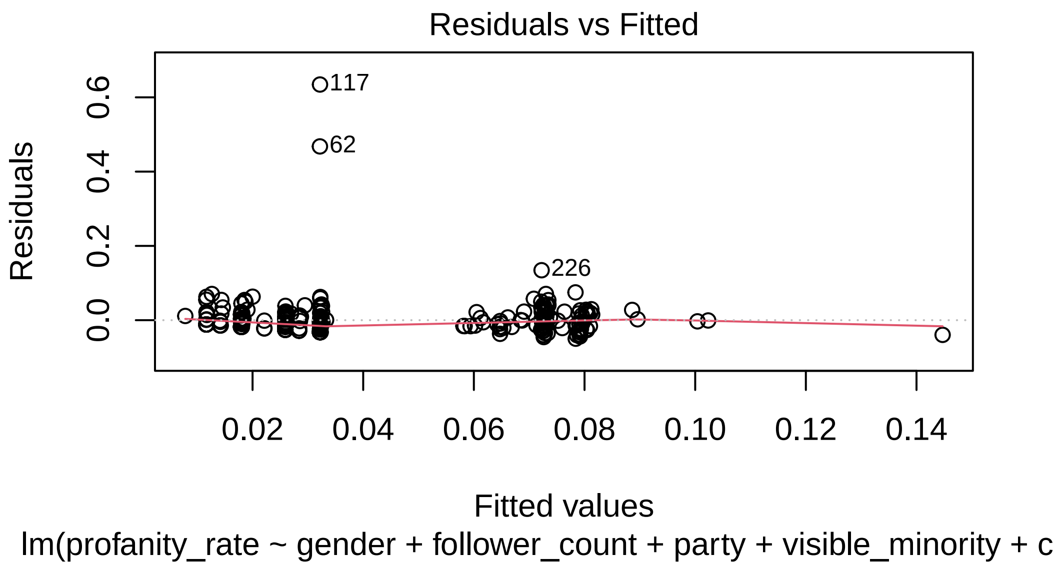
```
                   2.5 %        97.5 %
genderFemale -0.02772897 -0.0008257036
```

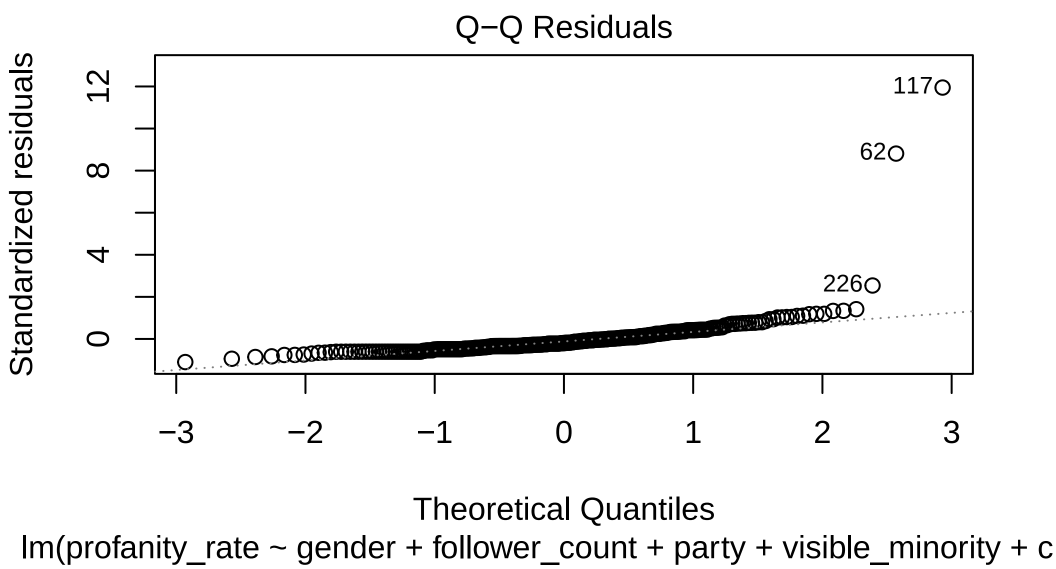
The coefficient of the explanatory variable gender (Female) is -0.0143, and the 95% confidence interval is $[-0.0277, -0.00083]$.

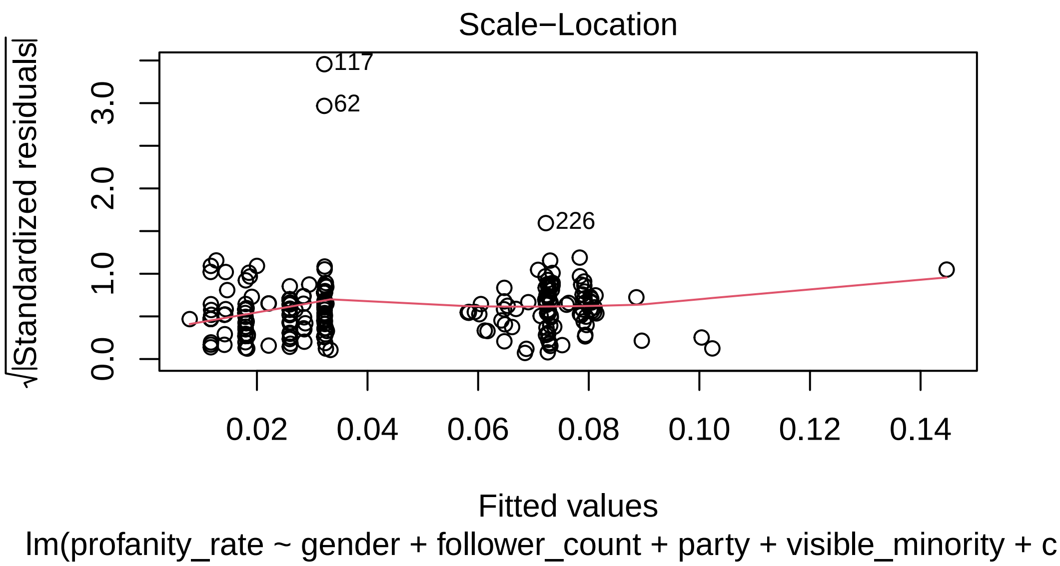
This result supports the idea that gender is associated with exposure to profanity-based online incivility, but in the opposite direction of my hypothesis. Specifically, female politicians receive lower rates of profanity-based tweets than male politicians, holding follower count, party affiliation, visible minority status, and country constant. The coefficient indicates that, on average, female politicians receive a profanity rate that is approximately 1.4 percentage points lower than that of male politicians. Because the confidence interval does not include zero, this estimate is statistically significant at the 0.05 level.

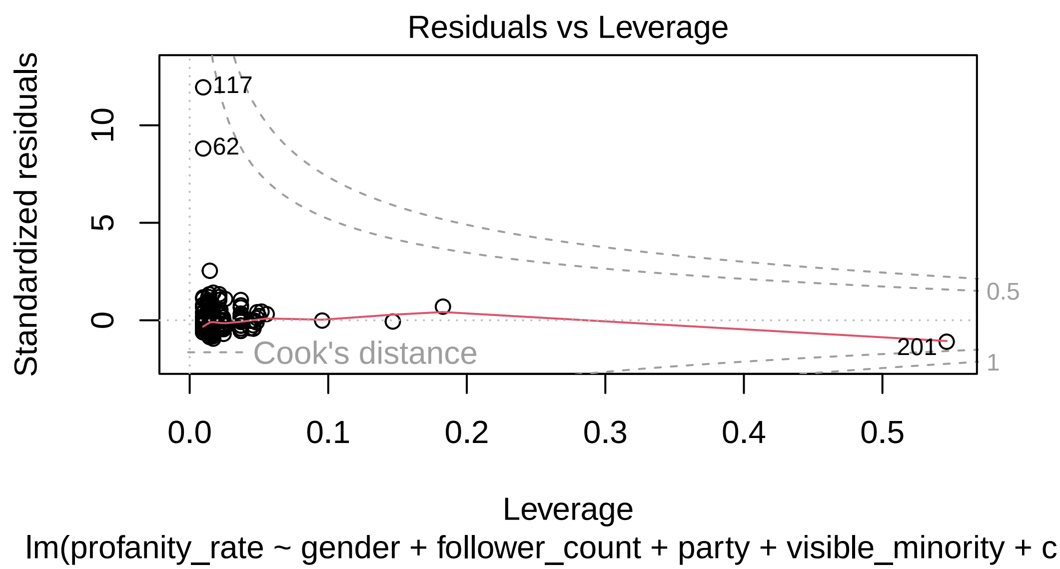
Diagnostics and Model Evaluation

```r
plot(mod)
```


Residuals vs Fitted
Residuals
0.0
0.2
0.4
0.6
117
62
226
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Fitted values
lm(profanity_rate ~ gender + follower_count + party + visible_minority + c


Q−Q Residuals
Standardized residuals
0
4
8
12
117
62
226
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(profanity_rate ~ gender + follower_count + party + visible_minority + c





In the Residuals vs. Fitted plot, the residuals are centered around zero and the line is nearly horizontal. This indicates that linearity assumption is reasonably satisfied. The clustering of points around fitted values of approximately 0.02 and 0.08 reflects groups of politicians with similar predicted profanity rates. Most residuals are small, suggesting the model predicts profanity rates fairly well for most observations, although a small number of outliers indicate cases where profanity exposure is higher than the model predicts.

The Q–Q plot shows that most of the residuals follow the reference line closely, which suggests that the residuals are fairly normally distributed. However, a few points in the upper tail fall far from the line, indicating the presence of some large positive outliers. These outliers suggest that the model underpredicts profanity rates for a small number of politicians (previously seen through the Residuals vs. Fitted plot as well). While normality is not perfect due to a few extreme cases, the assumption is reasonable for most observations.

The Scale–Location plot shows that residual spread is mostly consistent across fitted values, suggesting roughly constant variance. The red line dips slightly in the middle and rises toward the ends, which suggests that the model fits best for politicians with mid-range predicted profanity rates and slightly less well at very low or very high predicted values. However, this pattern is mild, and the overall spread of points remains fairly even.

In the last Residuals vs. Leverage plot, a small number of cases have higher residuals or leverage and are flagged by the Cook's distance contours. However, none of these observations cross the highest Cook's distance thresholds, suggesting that no single case has a severe impact on the model estimates.

```r
adj_r2 <- summary(mod)$adj.r.squared
adj_r2
```

```
[1] 0.1706709
```

The adjusted R-squared value for the model is 0.1706709. This value indicates that the model explains approximately 17% of the variation in profanity-based incivility. While the model captures some meaningful patterns, most of the variation remains unexplained, suggesting that other factors not included in the analysis also influence politicians' exposure to online profanity.

Conclusion

This study finds that gender is significantly associated ($p < 0.05$) with profanity-based online incivility, but not in the direction originally hypothesized. The coefficient for female politicians is -0.0143, with a 95% confidence interval of $[-0.0277, -0.00083]$, indicating that female politicians receive on average a 1.4 percentage point lower rate of profanity-based tweets than male politicians, holding other factors constant.

This result is notable because it suggests that isolating profanity as a specific form of incivility produces a different pattern than broader measures of online abuse used in prior research. Rather than being more frequently targeted by profanity, women may be more likely to experience other forms of incivility, such as sexist or non-profane insults, that are not captured by this measure. Finally, because this analysis is associative rather than causal, future research should further examine how different types of online incivility vary by gender and *why* profanity in particular may be more commonly directed toward male politicians.

References

Horowitz, J. M., & Goddard, I. (2023, September 27). Views of obstacles for women seeking high political office. Pew Research Center. https://www.pewresearch.org/social-trends/2023/09/27/views-of-obstacles-for-women-seeking-high-political-office/

Krook, M. L., & Sanín, J. (2016). Violence against women in politics: A defense of the concept. Política y Gobierno, 23(2).

Rheault, L., Rayment, P., & Musulan, A. (2019). Politicians in the line of fire: Incivility and the treatment of women on social media. Policy & Internet, 11(4). https://doi.org/10.1177/2053168018816228

Rheault, L., Rayment, P., & Musulan, A. (2019). Replication data for: Politicians in the line of fire: Incivility and the treatment of women on social media [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/B97TGX

Vogels, E. A. (2021). The state of online harassment. Pew Research Center. https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/

Vrielink, J., & van der Pas, D. J. (2024). Part of the job? The effect of exposure to the online intimidation of politicians on political ambition. Political Studies Review, 22(4). https://doi.org/10.1177/14789299241258099

Weaving, M., Alshaabi, T., Arnold, M. V., Blake, K., Danforth, C. M., Dodds, P. S., Haslam, N., & Fine, C. (2023). Twitter misogyny associated with Hillary Clinton increased throughout the 2016 U.S. election campaign. Scientific Reports, 13, Article 5266. https://doi.org/10.1038/s41598-023-31620-w