# Data essay project

Intro

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Data description

The variable *citation_5yr_diff_circ_2022* is the dependent variable of the research and measures the number of times a judicial decision has been cited by other judges in appellate court decisions over the five-year period following its original publication. The number of citations serves as a key indicator of the influence and authority of the decision within the legal field. Citations are often used as a measure of the decision's legal significance, as they reflect how frequently other judges reference a particular decision when making their own rulings. A higher citation count typically suggests that the decision has had a greater impact on subsequent legal proceedings and is viewed as more authoritative or persuasive by other judges.

Number of citations

Total number of citations after five years
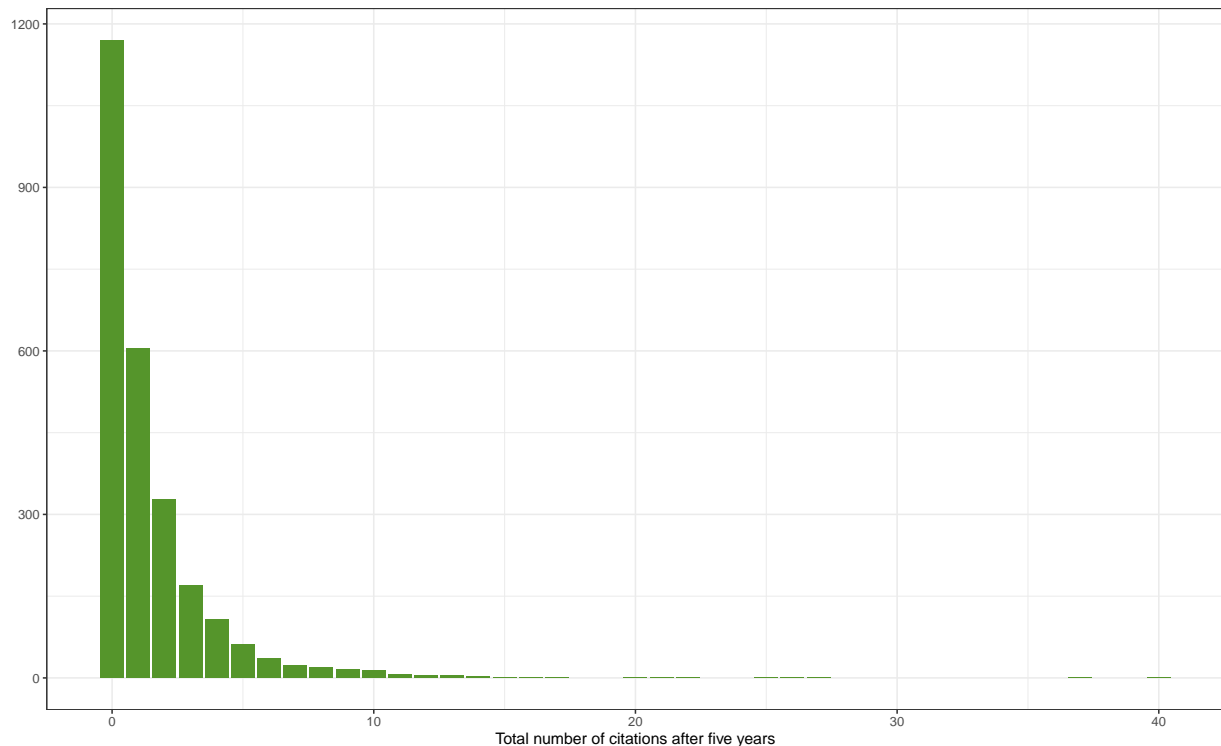
Main independent variables

The two main independent variables used in the analysis that will be presented in the following pages are *femaleOA* and *OAminority*, that is, the two dichotomous variables that indicate the gender and ethnic origin of the judge who issued the sentence.

The variable *femaleOA* is a dummy variable indicating the gender of the judge who authored each of the decision in the dataset. It is coded as 1 if the judge is female and 0 if the judge is a male. Looking at the values contained in the database, it is possible to observe that there are 2045 male judges and 637 female judges. Considering these values, it is possible to argue that, at least within our data set, the judicial system being analyzed is still mostly composed of men, given their clear predominance.

The variable *OAminority* is a dummy variable indicating whether the judge belongs to an ethnic minority. It is coded as 1 if the judge is non-white and 0 if the judge is white. Looking at the values contained in the database, it is possible to observe that there are 2426 white judges and 478 non-white judges. Taking these values into consideration, it is possible to observe how the judicial system appears to be composed of a clear prevalence of white judges rather than judges belonging to ethnic minorities.

Consequently, by observing the values reported in the graph below it is possible to verify the distribution of the number of citations for both the main independent variables, namely gender and ethnic group. Looking at the graph it is possible to notice the high presence of white-male judges, who are more than all the other three categories added together. On the contrary, the group that appears to be at the center of the analysis, namely women belonging to ethnic minorities, appears to be the one with the least representation within the data set.

Number of citations by race and gender
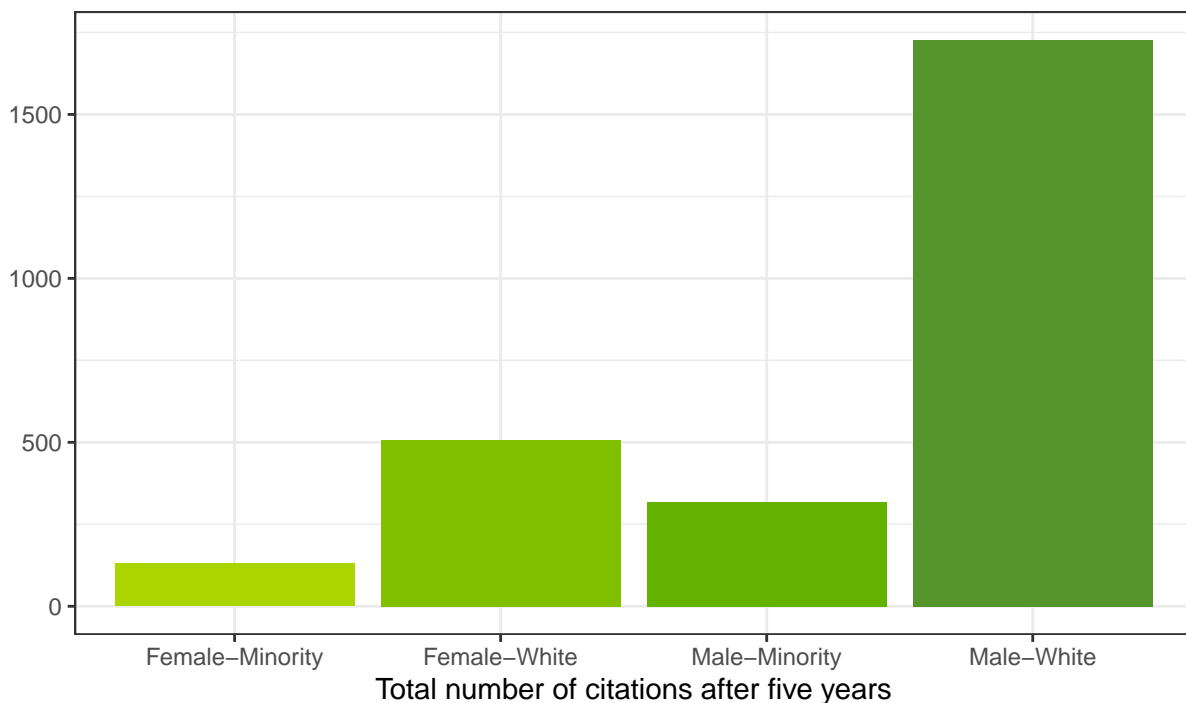
Control variables

- *authorities_deep* Reliance on existing precedents is widely recognized as a key indicator of legal quality, with decisions that heavily draw upon precedent often perceived as more rigorous and thorough in their legal reasoning. The variable *authorities_deep* serves as a crucial control variable in analyses aimed at understanding the factors driving citations of judicial decisions. By accounting for the extent to which a decision is rooted in precedent, researchers can more accurately differentiate between citations motivated by the inherent quality of the legal reasoning and those potentially influenced by demographic attributes of the judge. This variable helps to capture the depth and quality of the decision-making process, as decisions that reflect a strong foundation in established legal principles are typically more persuasive and therefore more likely to be cited in future cases. Controlling for reliance on precedent ensures that the evaluation of citations is not skewed by variations in legal quality, allowing for a more nuanced understanding of how demographic factors might influence citation patterns.

- *OAls_elite* Judges who graduated from elite law schools often enjoy a perception of enhanced authority and credibility, which can significantly influence the frequency with which their decisions are cited. This association stems from the prestige and perceived competence linked to attendance at these institutions. By incorporating the variable *OAls_elite* into the analysis, researchers can account for the citation advantage that may arise from a judge's educational background, independent of the substantive quality of their decisions or their demographic attributes. This control ensures that the analysis distinguishes between citations driven by the judge's perceived authority—rooted in educational prestige—and those influenced by demographic factors. Educational pedigree can contribute to external perceptions of competence and authority and controlling for this variable allows for a more accurate assessment of how demographic characteristics, such as the two main independent variables of this analysis, influence the recognition of judicial work, free from the overshadowing effects of educational prestige.

- *OAtenure_ln* Judicial tenure plays a crucial role in shaping a judge's experience and institutional knowledge, which can enhance the authority and influence of their decisions. Judges with longer service are often perceived as more experienced and knowledgeable about court processes, making their rulings more likely to be cited. Including *OAtenure_ln* as a variable in the analysis helps disentangle the effects of professional experience from those of demographic characteristics. This measure, represented as the logged number of years a judge has served on the Court of Appeals, provides an indicator of expertise and seniority. By accounting for this factor, the analysis controls for variations in citation patterns driven by professional experience, ensuring that observed differences are not mistakenly attributed to demographic attributes. In this way, the inclusion of tenure mitigates bias and allows for a clearer understanding of how demographic characteristics influence citation rates, independent of the judge's institutional knowledge or professional longevity.

Description	Quality.of.decision	Elite.school	Years.in.court
Min.	0	0	0.000000
Median	4	1	2.833213
Max.	54	1	3.828641

The inclusion of these variables ensures that the model accounts for key factors influencing citation patterns beyond demographic attributes. This approach minimizes omitted variable bias, strengthens causal inference, and allows for a more accurate assessment of the impact of gender and minority status on judicial influence. These variables collectively enable the isolation of the hypothesized effects, ensuring that any observed differences are robust and attributable to the judges' demographic characteristics rather than extraneous factors.

Analysis

Incluce the note at the end of *01 model.R* in the text here

The code fits two models—a Poisson regression (m1) and a Negative Binomial regression (m2)—to count data (citation) with predictors (female, minority, and their interaction). The Poisson model assumes the mean equals the variance, while the Negative Binomial model allows for overdispersion (variance > mean). Log-likelihoods of the two models (L1 and L2) are compared using a Likelihood Ratio Test (LRT), with the test statistic following a chi-squared distribution with 1 degree of freedom. The null hypothesis (H0) is that the Poisson model is sufficient, while the alternative (H1) suggests the Negative Binomial provides a better fit. If LRT > qchisq(0.95, df = 1) evaluates to TRUE, H0 is rejected, indicating significant overdispersion and the superiority of the Negative Binomial model.

$$citation = \beta_0 + \beta_1 * female + \beta_2 * minority + \beta_3 * female * minority + \beta_4 * authorities + \beta_5 * elite + \beta_6 * tenure$$

Table 2: Regression Results

	Dependent variable:	
	Number of citation	
Constant	0.409***	0.0004
	(0.038)	(0.105)
Female	−0.046	−0.153*
	(0.080)	(0.079)
Ethnic minority	0.112	0.060
	(0.094)	(0.092)
Years in Court		−0.001
		(0.034)
Quality of decision		0.065***
		(0.006)
Elite school		0.100
		(0.061)
Female*Ethnic minority	−0.433**	−0.285
	(0.187)	(0.183)
Observations	2,373	2,373
Log Likelihood	−3,939.913	−3,878.208
θ	0.642*** (0.032)	0.717*** (0.038)
Akaike Inf. Crit.	7,887.826	7,770.415
Note:	*p<0.1; **p<0.05; ***p<0.01	

Hypothesis 1

Hypothesis 2

Hypothesis 3

Simulation (aka scenario)

To assess the presence of significant differences between gender and ethnic groups as set out in the previous hypotheses, it is necessary to control for the difference between the expected values generated from the models developed in the previous paragraph and from the same personal characteristics highlighted in the hypotheses. Therefore, four different simulations will have to be developed in order to correctly evaluate the differences between the three hypotheses presented. The first simulation that will be developed will consider women (female = 1) belonging to ethnic minorities (non-white = 1). The result of this simulation will be used to calculate all three first differences with the expected values of the simulations using white men, white women and non-white men.

Hypothesis 1

- Intercept: 1
- Female: 1
- Minority: 1
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 1*1

Compare

- Intercept: 1
- Female: 0
- Minority: 0
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 0*0

Hypothesis 2

- Intercept: 1
- Female: 1
- Minority: 1
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 1*1

Compare

- Intercept: 1
- Female: 1

- Minority: 0
- Authorities: median
- Elite: median
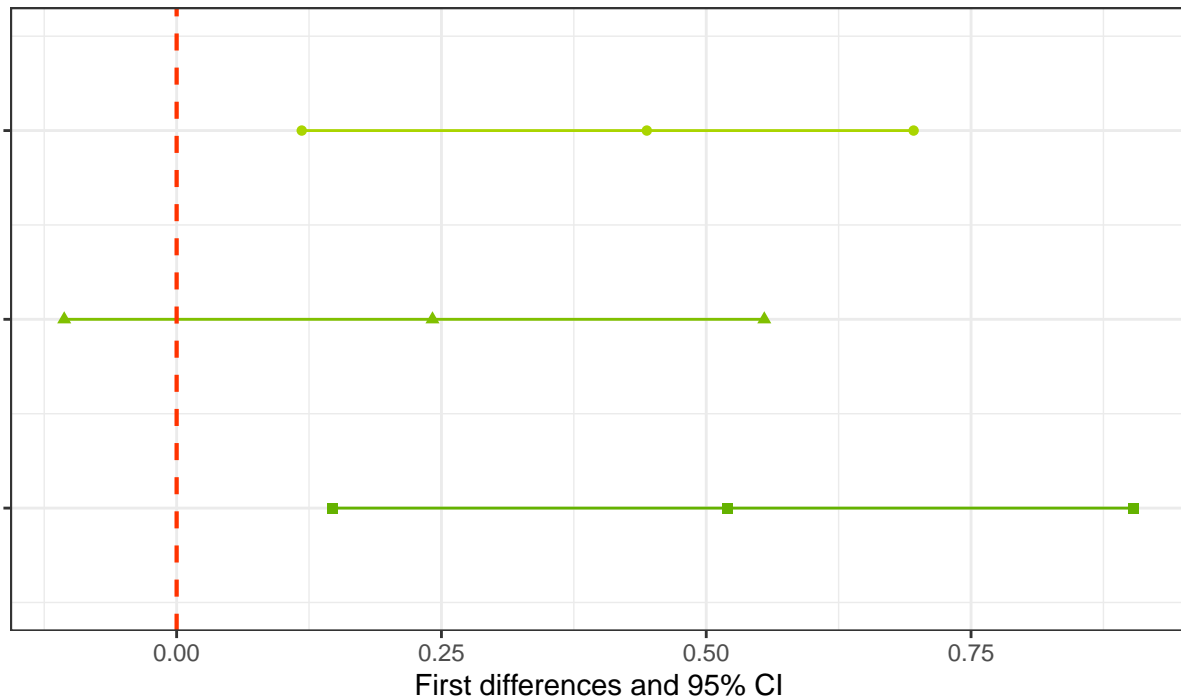- Tenure: median
- Interaction: 1*0

Hypothesis 3

- Intercept: 1
- Female: 1
- Minority: 1
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 1*1

Compare

- Intercept: 1
- Female: 0
- Minority: 1
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 0*1

First differences for the three models

White male, white female and non–white male minus non–white female

First differences and 95% CI

Robustness check

Conclusion