

Final Project

2025-10-05

Part 1: Brainstorm

Idea 1

How does the distribution of affordable Housing Connect units by income eligibility level vary across NYC boroughs, and how does this compare to borough-level rent burden and median rents?

Affordable housing is one of NYC's most urgent problems. Housing Connect is the city's lottery-based affordable housing program and a central strategy to address this issue. While Housing Connect produces units targeted at different income brackets (extremely low-income to middle-income), these units may not correspond to the rent affordability needs of borough residents. Comparing the distribution of affordability levels across boroughs alongside rent burden rates will help assess whether Housing Connect developments are addressing local housing needs. Sarvi also worked in Affordable Housing (underwriting credit loans for affordable housing projects from New York to Florida) and has seen first-hand how central of an issue affordable housing is within cosmopolitan cities especially. The waitlists for affordable housing units within New York City especially span thousands of people.

Potential Datasets:

- Affordable Housing Production by Building (HPD)
Number of units in each affordability category (Extremely Low Income, Very Low Income, Low Income, Moderate Income, Middle Income, Other). Borough location of each building.
- American Community Survey (via tidycensus)
Borough-level median rent, household income, and rent burden rates.
- NYC Department of Homeless Services (DHS)
Homelessness data by borough to connect housing affordability to broader housing needs.
- Local Law 44 — Rent Affordability (HPD)
Additional rent affordability categories by building.

Analysis Plan:

- Aggregate data by borough
 - Count total Housing Connect units in each affordability category for each borough.
- Descriptive statistics
 - Calculate proportions of units by affordability level in each borough.
 - Compare those proportions to borough-level median rents and rent burden rates.

This project will produce a borough-level analysis of how Housing Connect affordable housing units are distributed by income eligibility level, and whether they align with borough-specific rent burdens and housing needs. The results can help policymakers target affordable housing production more effectively.

Idea 2

Question: For the final project, my proposed research question would be: Do individuals experiencing greater economic insecurity show more support for tariffs and trade protection? I am currently taking a SIPA class on International Political Economy, and in class, we are covering protectionism and reciprocal tariffs. I would be interested in seeing how economic anxiety contributes to support for protectionism, if at all. I think this is a particularly relevant topic in today's current political climate, as Trump pushes a protectionist trade agenda. Analysis here would involve using regression models (logistic if dependent variable is binary or an OLS regression if it's a continuous scale). The study will estimate the relationship between perceived job risk and support for trade protectionism. We could also use descriptive statistics such as a histogram to show trade attitude differences across income categories. Additionally, we could utilize ggplot2 to visualize predicted probabilities of tariff support by economic security at party ID.

Potential Datasets: - CES (Cooperative Election Study) 2020: its a survey that is a representative sample of U.S. adults. This survey covers demographics, partisanship, issue positions, and policy preferences—discussing several economic and trade-related questions.

- ACS 2020: looks at county-level unemployment and median income. Here, we could test whether respondents in more trade-exposed areas support protectionist policies more.
- MIT Election Lab: looks at county-level election results. Here, we study how protectionism correlates with local Trump/Biden support.

ANES (American National Election Studies) 2016 / 2020: a survey series that tracks political attitudes and behaviors. This data would provide insight onto how Americans perceive globalization and trade, and also their policy preferences Here, we can see whether economic insecurity and trade attitudes move together over time or vary by partisanship.

Idea 3

Question: How does gender identity influence perceptions of political leadership competence?

For the final project, another idea for a research question would be how gender identity influences how voters see the competency of potential candidates. Gender has been at the forefront of the latest Presidential election—between Kamala Harris and Donald Trump—as well as in 2008 and 2016, when Hillary Clinton was running for President. It would be interesting to investigate how the competence of female candidates is perceived within the voter base. This question is also important as more openly LGBTQ+ candidates run for office, there is an interesting intersection between gender identity and political ideology and voting behavior that warrants further research.

Potential analysis could include a regression table—which would show perceived competence by gender identity. Also a word cloud could be created to see media descriptors about gender identity. We could collect media articles that mention political leaders of various gender identities and use sentiment analysis to figure out how certain adjectives that allude to competence differ by gender identity.

Potential Datasets:

- AP-NORC VoteCast: To analyze voter perceptions of political candidates' competence, and how those perceptions vary by gender identity, ideology, and demographics.
 - Respondent demographics: gender identity, age, education, income, race, ideology, party identification.
 - Perception questions: evaluations of candidates' competence, honesty, empathy, leadership ability, etc.

- Candidate-level data: which candidate each respondent voted for or favored (e.g., Kamala Harris, Donald Trump).
- Behavioral data: turnout, vote choice, issue priorities.
- Media Cloud Explorer: To study how media coverage frames the competence of political leaders with different gender identities. We can extract and visualize adjectives/descriptors linked to candidates and create word clouds showing how competence-related terms differ by gender identity.
 - News articles from major outlets mentioning key political figures
 - Article metadata: publication date, source, country, topic.
 - Full text of articles (for text analysis).
- Census Data: To provide demographic context for understanding regional variation in perceptions of leadership competence.
 - Population-level statistics by state, county, metro area:
 - * Gender composition
 - * Education levels
 - * Income
 - * Political participation proxies
 - * Age distribution
 - These can be merged with the geographic identifiers in VoteCast

Idea 4: Using the V-Dem Dataset: democracy dynamics

I am a political science student, and have used Rstudio in the past to study democracy dynamics in the world. I have come up with multiple questions, using variables in the V-dem dataset, that are interesting in the current political landscape.

As the V-dem dataset encompasses all countries of the world, the following questions could be adapted to a global context/or a specific case analysis (democracies of the world, USA, autocracies of the world, the EU... etc).

Question 4.1: Are education levels linked to greater media freedom?

I would expect higher education levels to be linked to greater media freedom because educated populations are more likely to value information, seek transparency, and hold governments accountable. In societies where people are literate and informed, there tends to be more pressure on authorities to allow diverse viewpoints and independent reporting. This question is relevant because understanding this link can highlight how investments in education may indirectly support press freedom, which is essential for democracy, accountability, and civic engagement. Can data explain this?

Question 4.2: Does media freedom lead to better social outcomes (e.g. health, education)?

I would expect media freedom to improve social outcomes because a free press can expose mismanagement, inform citizens about health and education policies, and encourage governments to respond to public needs. When media can operate independently, it helps ensure that social programs are more effective and accessible. This question is relevant because it shows that protecting press freedom is not only a democratic goal but can also have direct, tangible benefits for people's wellbeing, making it a key factor in development and policy planning.

and here come my favorite ones..:

Question 4.3: Does higher women's political representation contribute to stronger democracy and lower levels of political conflict?"

I have seen some time ago, that women in power are correlated with less conflicts overall. Women in power tend to focus more on social protection, and less on war. Is this true? Can data prove this?

Question 4.4: Does higher women's political representation contribute to lower levels of corruption?

Studies of close elections in Brazil indicate that women elected as Brazilian mayors are 29-35% less likely to engage in corruption - They also use clientelism less, resulting in approximately 64 fewer temporary employees - women mayors contribute to a 1 percentage point reduction in premature births. - In Africa, female leaders have been shown to empower women to use social accountability

What about in other parts of the world? Is this a real recurrent dynamic? or simply a “coincidence”?

Potential Datasets <https://v-dem.net/data/the-v-dem-dataset/> A very comprehensive dataset, includes data on every single country of the world on democracy dynamics