7/19

DSSG + EDA in Political Campaign Media

Discussing using data science for non-tech problems and showcasing pandas for political campaign data analysis

Ian Castro, UC Berkeley School of Information
Guest Lecture for Astronomy 9, 7/19

About me

- Filipino, 1st generation graduate student from East Bay; pronouns: he/they
- Now:
    - Master of Information Management and Systems at the UC Berkeley I-School
    - National Equity Atlas @ PolicyLink
    - RA @ Urban Displacement Project
- Past:
    - BA Media Studies, BS Microbial Biology
    - Instructor @ UCB Division of Computing, Data Science, and Society

How I ended up in data science (with bio & media degrees)
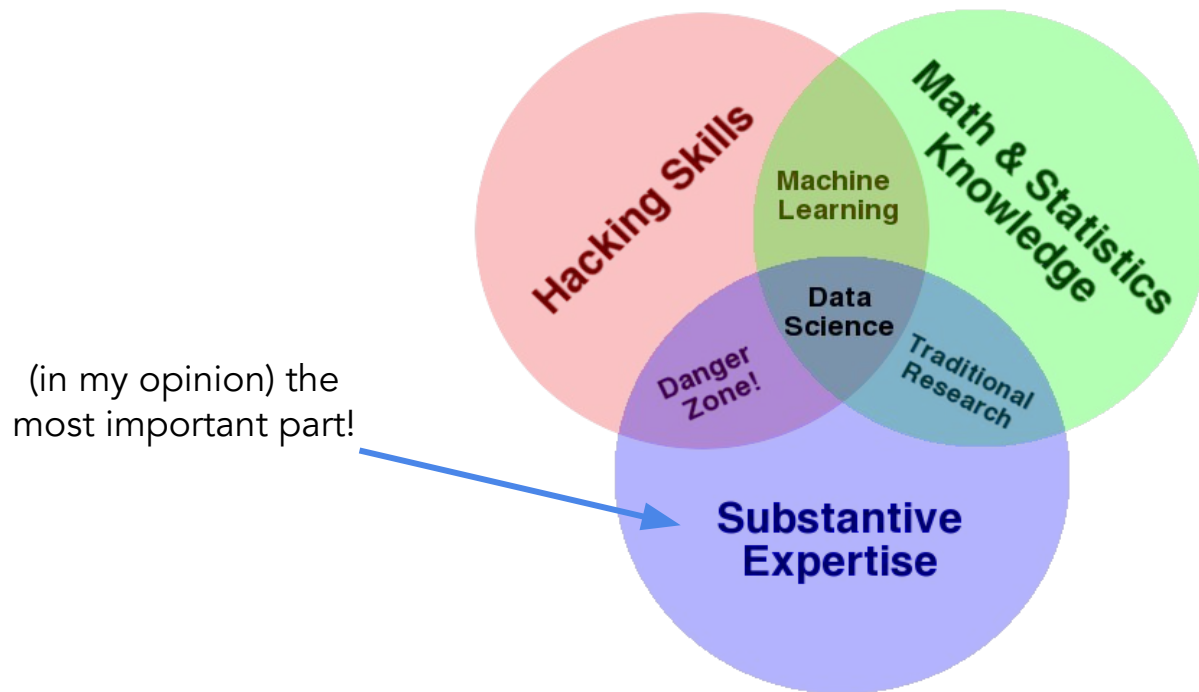

Took an intro class without any coding experience

Taught data science to pay the bills + worked on equity & accessibility in DS education

Had very supportive mentors & professors!

Berkeley
SCHOOL OF
INFORMATION

Joined a graduate program to learn more technical skills (it's never too late!)

What is data science?

(in my opinion) the most important part!

Data science for social good is when we use these skills to serve the public in an equitable way, solve important social issues, and help people & communities

also see: research justice, community based participatory research

by Drew Conway in 2010 (link)

What you can do with data science expertise for non-tech questions: some of my projects

Understanding Free Speech using Computation

Q: Has the US Supreme Court become more or less restrictive on the issue of free speech over its history?

Approach: exploratory analysis of legal text (Supreme Court opinions) using <u>natural language processing (NLP)</u> tools

A Topic Analysis of the Supreme Court of the United States' Views on Free Speech and the First Amendment

Ian Castro
ANLP Fall 2021

Abstract

Throughout its history, the United States Supreme Court has addressed myriad issues regarding free speech and the First Amendment, taking both liberal (pro-individual) and conservative (pro-regulation) positions depending on the area. Using topic modeling and dictionary methods, I find that the Supreme Court has generally become more liberal on the issue of free speech over time, with three distinct eras shaped by the changing social contexts and political fears of the 20th century. I also demonstrate the effectiveness of natural language processing techniques to analyze Supreme Court case opinions for retrospective exploration, rather than prediction.

1 Introduction

In the United States, the First Amendment prevents Congress – or any other government actor – from infringing on the freedom of speech, press, religion, assembly, or petition. However, the interpretation of this amendment by the Supreme Court—who interprets the law of the land—has changed significantly over time. In one of the first major free speech cases, *Schenck v. United States* (1919), the Supreme Court held that this individual freedom of speech can be restricted if it poses "a clear and present danger;" Schenck had published a pamphlet that advocated against the draft in World War I. Over the next five decades, the Court used several other standards to decide when potentially dangerous speech can be limited, including the bad tendency test of *Gitlow v. New York* (1925). In *Brandenburg v. Ohio* (1969), the Court

created the current precedent for limited speech, which holds that speech can be limited if it "is directed to inciting or producing imminent lawless action and is likely to incite or produce such action" (Vile).

Many of these cases were decided in specific social and political contexts: World War I and II, the Red Scare, and the Civil Rights Movement, to name a few. These contexts, along with the politics of each time and the specific justices on the Court, have shaped the reasoning of these cases. Knowing this history is particularly interesting as we move into a more digital age, where the Court must apply the reasoning of past justices in past cases to new questions of misinformation, online speech, and censorship. Furthermore, knowing the ideology of the Court on this issue and how it has changed over time is important, as the Court has recently utilized the First Amendment to uphold individual rights, sometimes at the expense of marginalized groups and organizations. For example, the Court has protected rights of bakers to refuse service to same-sex couples (*Masterpiece Cakeshop v. CO Civil Rights Commission*, 2018) and restricted the ability of unions to collect dues from non-members (*Janus v. AFSCME*, 2018).

Therefore – in this study, I explore the evolution of the Supreme Court's ideology on this specific issue by using log-odds ratios, dictionary methods, and topic analysis. Given that past precedent plays a significant role in how the Court rules in future cases, how has the Court's ideology on free speech changed over time? What computational tools do we have to do this, and are they effective? Using these tools, I will show that the Court has become more liberal (i.e., pro-individual rights) rather than more

Topic modeling by Court discovers changes in issue areas

Dates	Court	Issue Areas
1910-1921	White	Seditious speech
1921-1930	Taft	Seditious speech
1930-1941	Hughes	Press freedom Incorporation
1941-1946	Stone	Public forums Religious speech
1946-1953	Vinson	Association Seditious speech
1953-1969	Warren	Assembly Dangerous speech Student speech Obscenity
1969-1986	Burger	Commercial speech School speech Press freedom
1986-2005	Rehnquist	Symbolic speech Political speech Religious speech Employee speech

Table 5. Issue areas for each court in the 20th century, discovered by topic modeling.
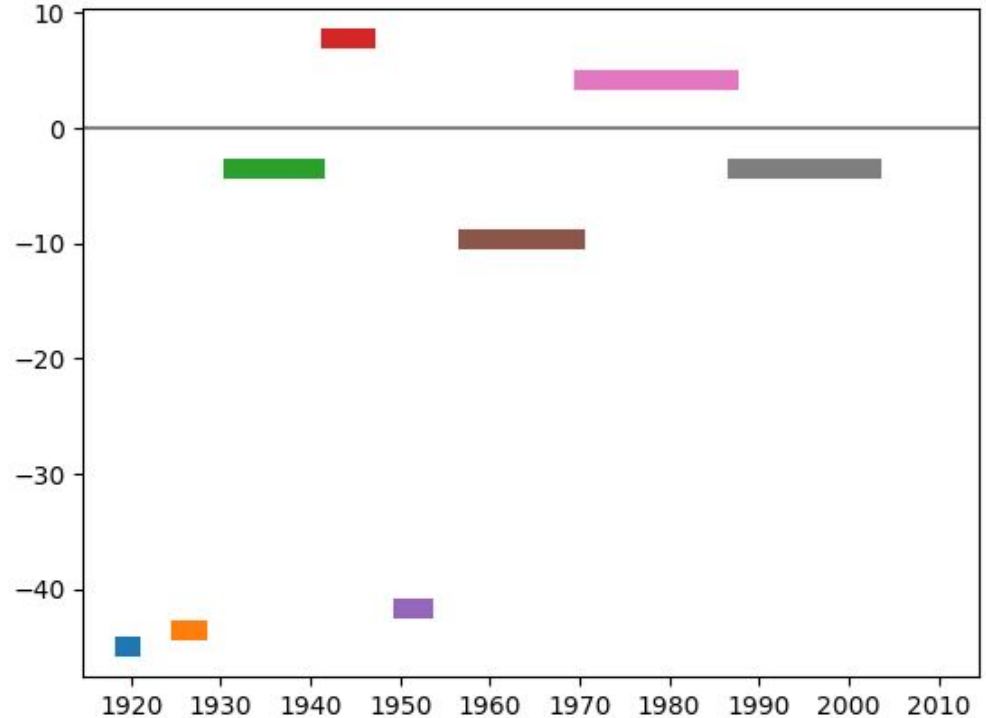
# Modeling & Mapping Inequities in Evictions

Q: In urban areas (especially those affected by gentrification), who is being evicted, where are they being evicted, and what inequalities exist between different groups?

Approach: Look at court orders and use NLP + statistics to predict the demographics of the person being evicted, and compare across neighborhoods

In conjunction with Urban Displacement Project and PolicyLink (Read about our work here and here!)
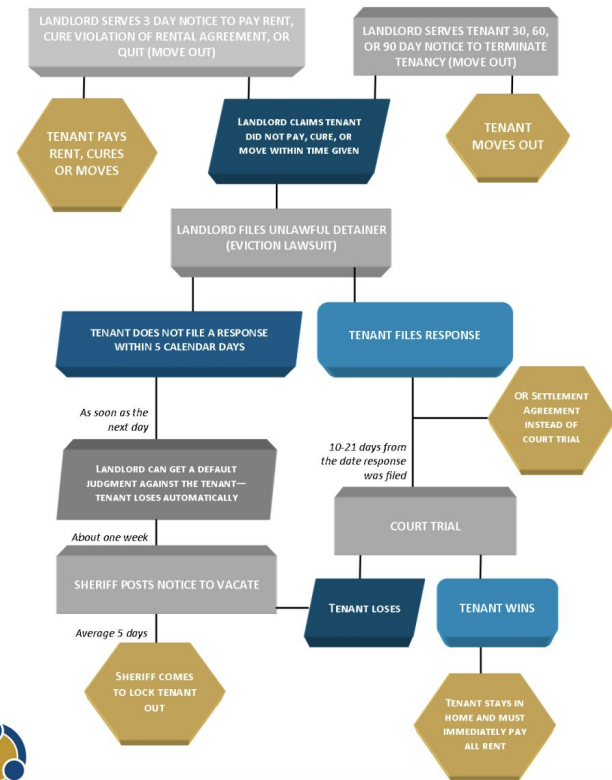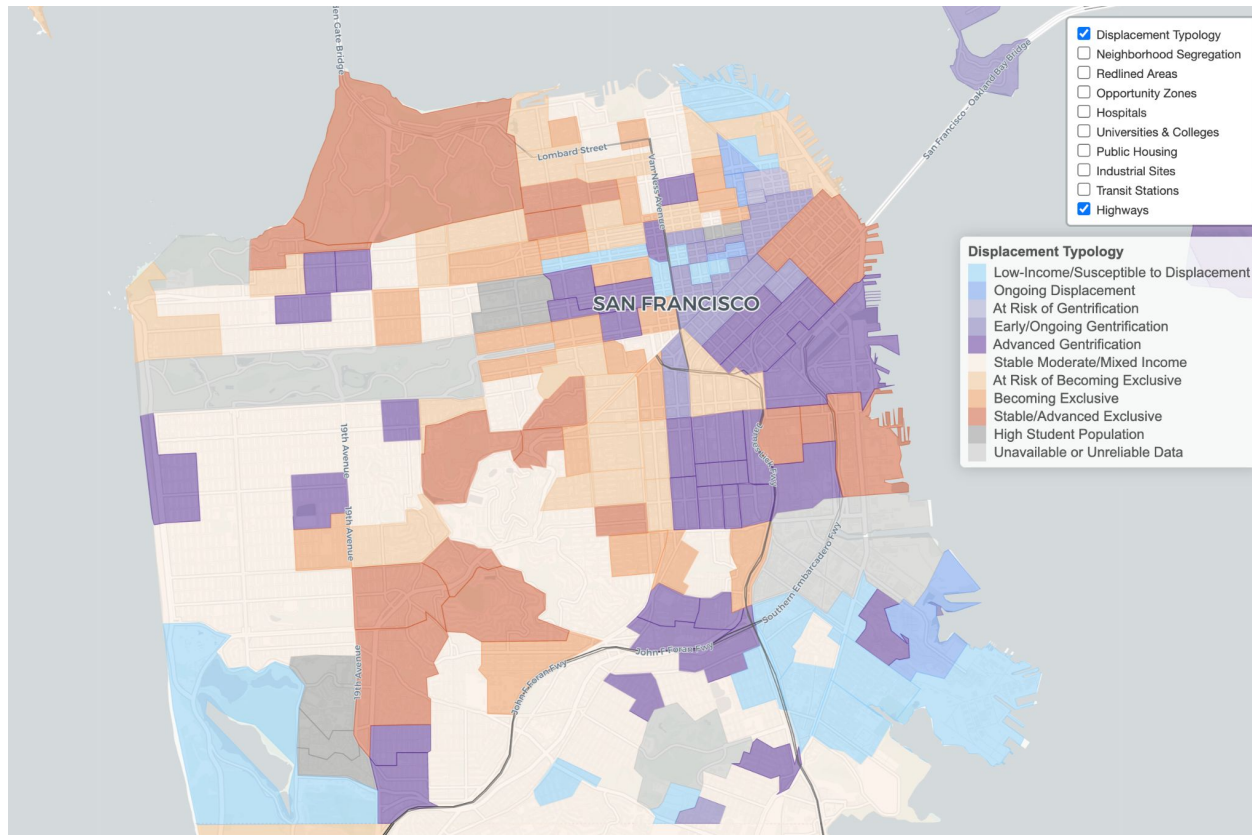
SF neighborhoods experiencing or at highest risk of gentrification/displacement tend to be predominantly Black & Hispanic (the Mission, Bayview-Hunters Point)

Goal: Explain how content moderation algorithms work in conjunction with human moderators, using TikTok for Younger Users as a case study

Why is science/technical communications important?
1) Society and technology are always linked
2) Audiences (and even coworkers) may not have the background or experience to understand these complex topics

Let's work through a case study:
imagine we're running a presidential campaign for a progressive candidate in Pennsylvania

(this exercise comes from Political Science 106A with Dan Schnur here at Berkeley)

A brief primer on the use of data in politics & campaigns

- The term "statistics" comes from "state" - collection of data was used by governments to manage populations and militaries
  - In politics: public opinion polls, voter demographics, voting histories
- 2012 was an important election for data science!
  - Obama's campaign = one of the first to invest in and use "big data" analytics ($1 billion)
  - Data allows campaigns to decide: where to raise funds, what issues to focus on in ads, what voters to target, and more!
    - Problems later on: Cambridge Analytica
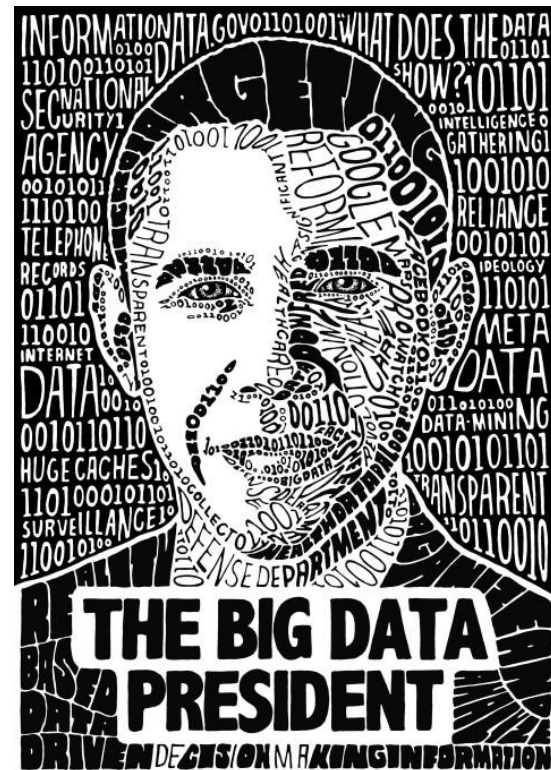
What is "campaign media"?

"New Media"

- Social media, online messaging, emails, texts
- Involves "microtargeting" and personalized ads
- Lots of data available online allow us to generate individualized voter profiles

"Old Media"

- Television/radio, mail flyers
- Mostly legacy media, lots of regulation created over time (e.g., FCC for television)
- Targeted via geography or "media market" due to technical + industry limitations
    - example: KTVU vs KTLA
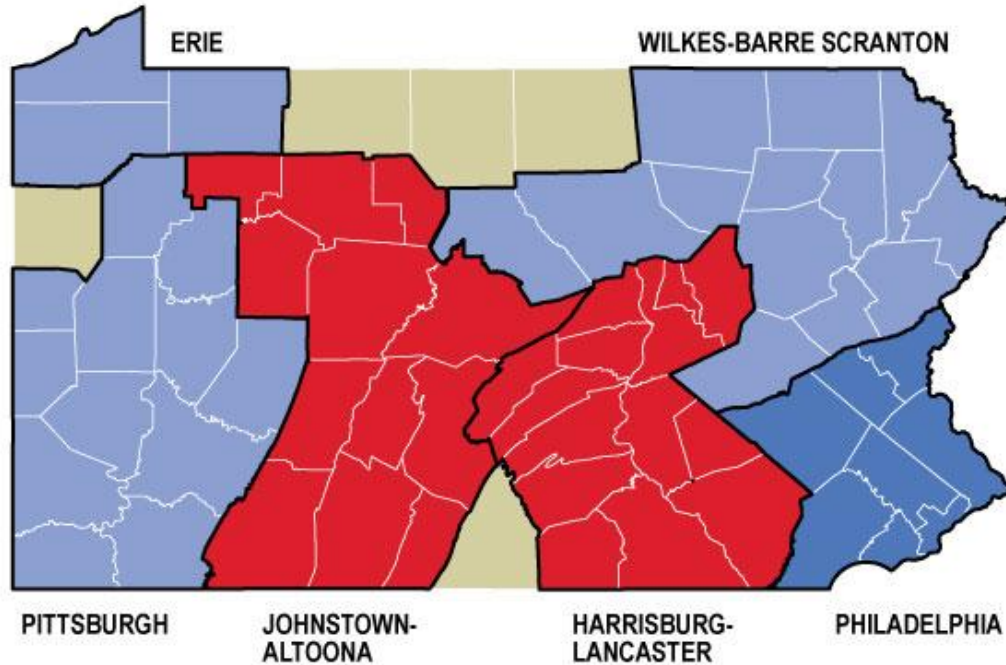
Pennsylvania's media landscape

Why care about Pennsylvania (for winning the presidency)?

- 20 electoral votes
- Used to be part of the Democrats "Blue Wall"
  - Voted for Democratic presidents from 1988-2012
  - Changed in 2016 with Trump, back to Biden in 2020 (by only .01%!)
- Interesting demographics and economy: **white, working class,** with big **union** presence and heavy reliance on **manufacturing** and healthcare
  - Left-leaning urban population centers (Philadelphia, Pittsburgh), right-leaning rural regions (most of state)
- Tradition of **divided state government** (ex. Democratic Governor, Republican state legislature)

What we're going to do today:

All Departments of State collect detailed county-level voting data.

So, using only that data for PA, we should be able to answer

1. Where should we focus our campaign messaging (for "old" media)?
2. What type of campaign strategy should we employ?
3. How did Trump beat Clinton in 2016?
4. How did Biden beat Trump in 2020?

???

Saints: consistent single-party voters for your party

Sinners: consistent single-party voters for the opposition party

Salvageables: voters who have voted for either party (ex. third party, undecided, split-ticket)

Some vocabulary

General campaign media strategies

Base-Motivation Strategy

(more common with polarized bases)

Goals:

- Convince your saints to turnout to vote in large numbers
- Convince sinners NOT to vote for your opponent

Attack ads, negative campaigning

Salvageable Strategy

(aka: 80-20 strategy)

Goals:

- Convince salvageables to vote for your candidate (flip them)
- Convince your saints to turnout to vote in large numbers

Policy & issue-focused ads, positive messaging

Now, let's finally do some analysis: follow along on my notebook
https://tinyurl.com/astro9politics

Interested in my work, going to graduate school, or anything else?

Email: castro.ian@berkeley.edu
Website: castroian.github.io