STAT 938 Final Project Presentation

XDS Consulting

Xamy Lopez Cortes   David Awosoga   Sharon Xue

2025-07-30

Table of contents

Introduction

D

Project Goals

D

- Analyze equity in U.S. traffic stops
- Identify potential demographic and geographic disparities in stop patterns
- Apply statistical methods to uncover trends over time and space
- Inform discussions on fairness and accountability in policing

Data Exposition

- Description:
  - 230 million traffic stops from 40+ states between 2006 - 2020 originally collected and made available by the Stanford Open Policing Project (Pierson et al., 2020)
- Augmentation:
  - State-level licence demographic information from the U.S. National Highway Traffic Safety Administration
- Management
  - Compression using Apache Arrow and the parquet format to reduce dataset from ~52 GB to ~4 GB

Question 1

S

Is there evidence that traffic stops disproportionately target men?

Are certain cities/locations “balanced” in terms of how traffic stops reflect the demography of that city/location?

Setup

- **Techniques:**
  - **Proportion calculation**: For each state, we computed the proportion of traffic stops attributed to a particular group (e.g., male drivers or stops per state), and compared it to the corresponding group's share of the licensed driver population.
  - **Spatial visualization (Choropleth map)**: We used spatial visualizations to map the differences between stop and license proportions across states. This allowed for quick identification of geographic patterns in over- or under-representation.

- **Techniques**:
    - **Cohen's h (Effect Size)**: statistically measure the difference between two proportions, independent of sample size.

$$h = 2 \cdot \arcsin(\sqrt{p_1}) - 2 \cdot \arcsin(\sqrt{p_2})$$

- Interpretation Thresholds

Cohen's *h* Range	**Interpretation**
h < 0.20	Negligible difference
0.20 < h < 0.50	Small effect
0.50 < h < 0.80	Medium effect
h > 0.80	Large effect

- **Workflow**
  - **Datascope**: Focused on overlapping time period (2014–2015) across major U.S. states. For each year, calculate the proportions(male stops, male drivers; licensed drivers proportion in the whole US for each state, stop proportion of the whole US for each state)
  - **Visualization**: Generate choropleth maps to visualize geographic differences in stop vs. license proportions.
  - **Cohen's h**: to statistically measure the difference

Results

Is there evidence that traffic stops disproportionately target men?

- Whole US
  - Male_stop_proportion = 66.534%
  - Male_licensed_driver_proportion = 49.415%
  - Cohen_h = 0.34852
  - Conclusion: Mildly overstop Male drivers

- State level
  - All states show positive gaps → more men are stopped than expected.
  - Similar pattern for both years.

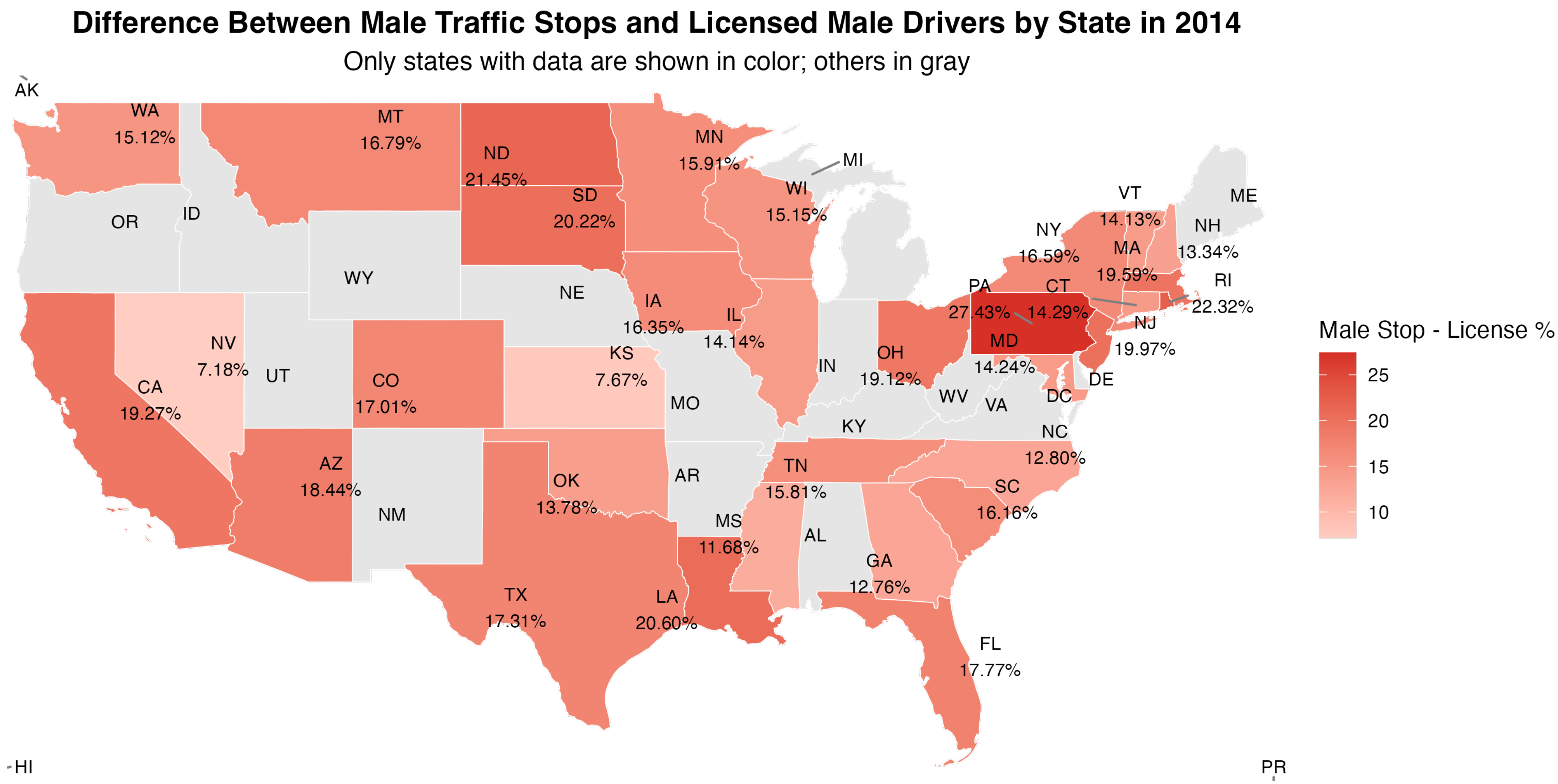

Figure 1: Difference Between Male Traffic Stops & Licensed Male Drivers by state in 2014

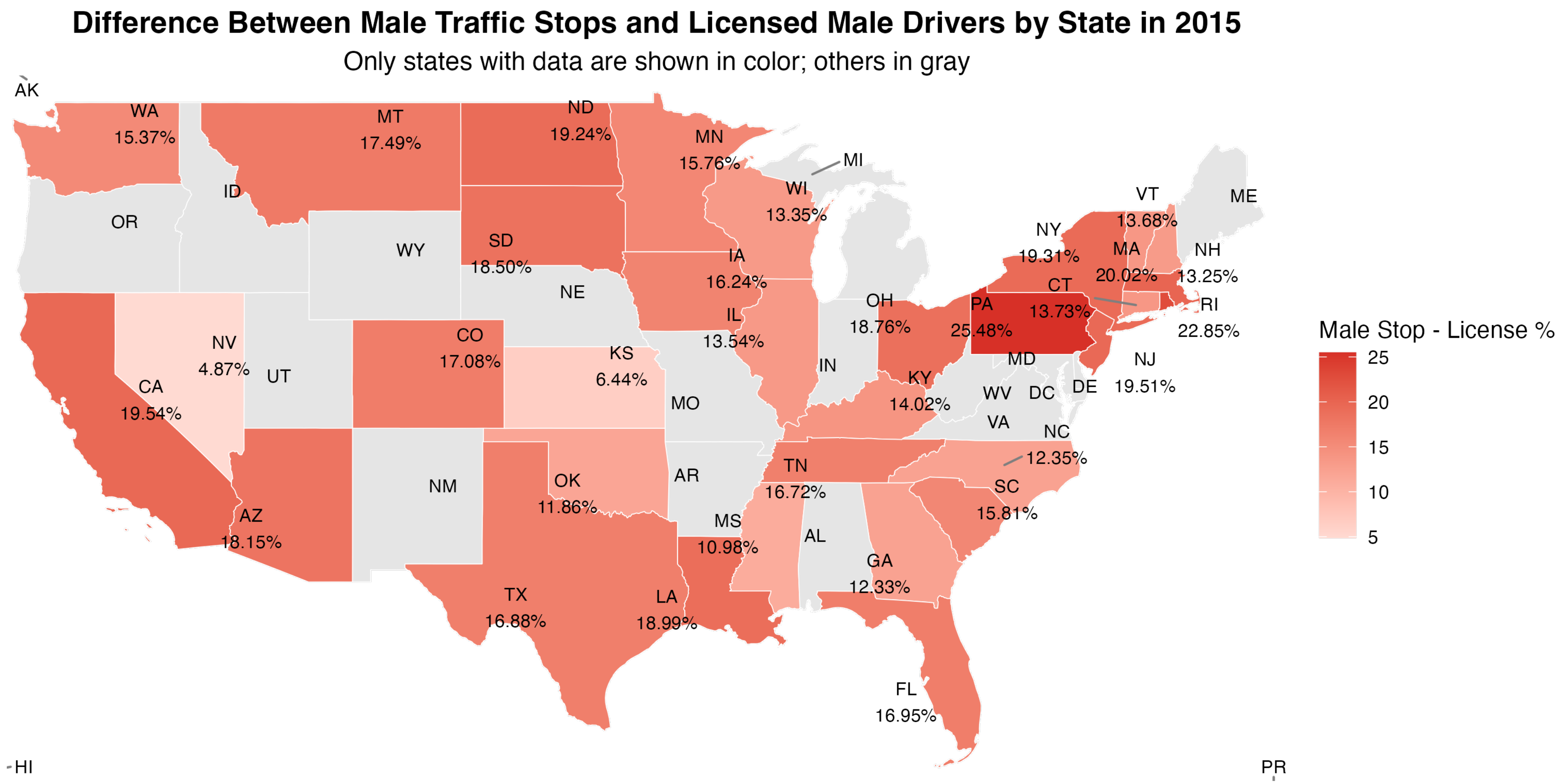

Figure 2: Difference Between Male Traffic Stops & Licensed Male Drivers by state in 2015

Key Takeaways

- **Nationwide Trend:** Most states, as well as the U.S. overall, show **mild overrepresentation of male drivers** in traffic stops compared to their share of licensed drivers.
- **Balanced States:** Only **Nevada (NV)** and **Kansas (KS)** show **gender-balanced enforcement**, with male stop rates aligning closely with male driver proportions.
- **High Overstop State: Pennsylvania (PA)** exhibits a **notably higher overstop rate** for male drivers.

Real-world Takeaways

- **Operational Patterns:** The mild overstop of male drivers may be linked to **driving behavior patterns**

Results Cont.

Are certain cities/locations “balanced” in terms of how traffic stops reflect the demography of that city/location?

- Datascope: 2014-2015
- 39 states
- 3 states did not have any overlap with other major states
- Choropleth maps show the difference between stop and license proportions by state in 2014 and 2015.
- Red indicates more stops than expected, blue indicates fewer stops than expected, white/gray denotes minimal difference.

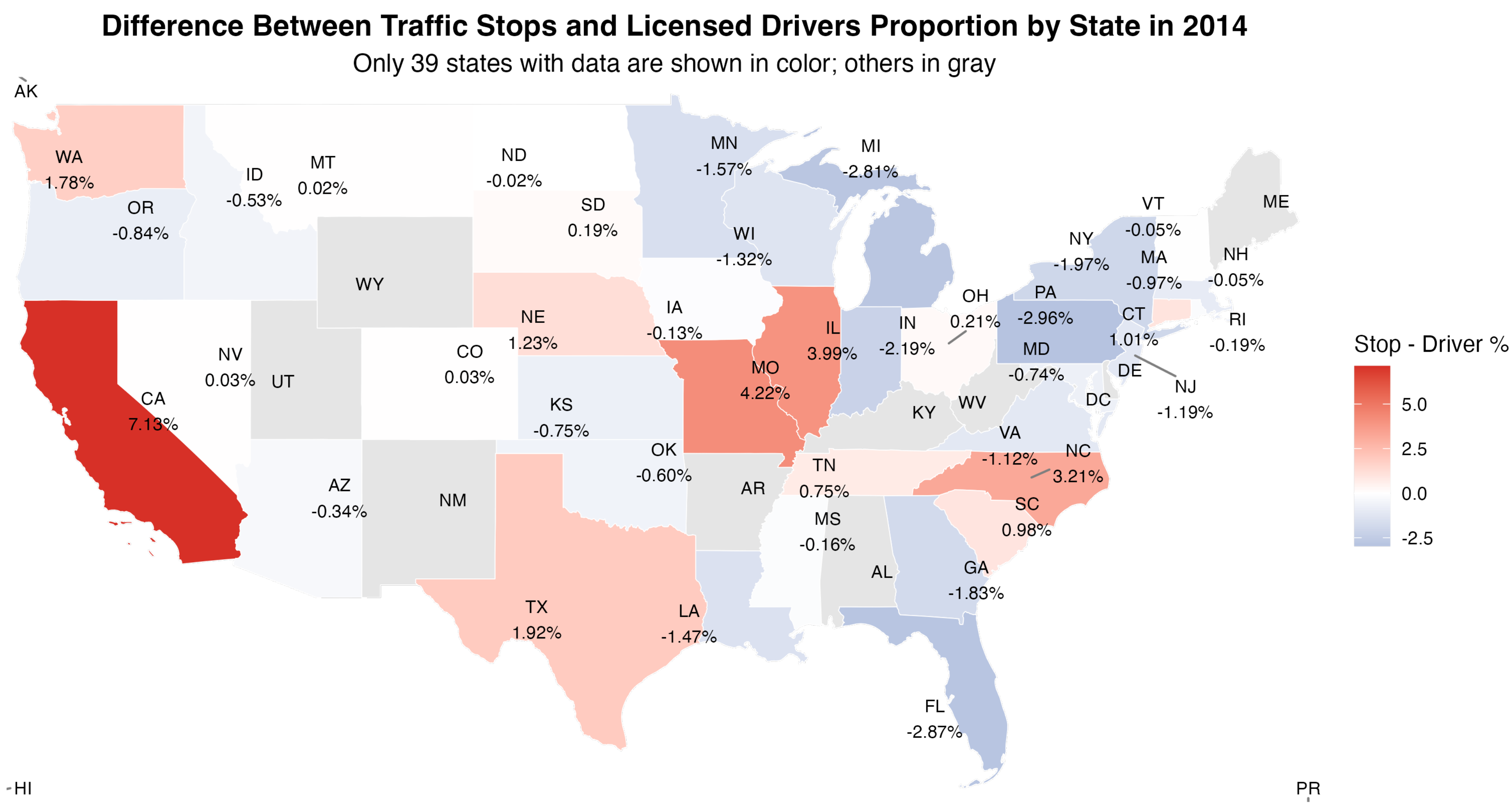

Figure 3: Differences between Traffic Stops & Licensed Drivers Proportion by State in 2014

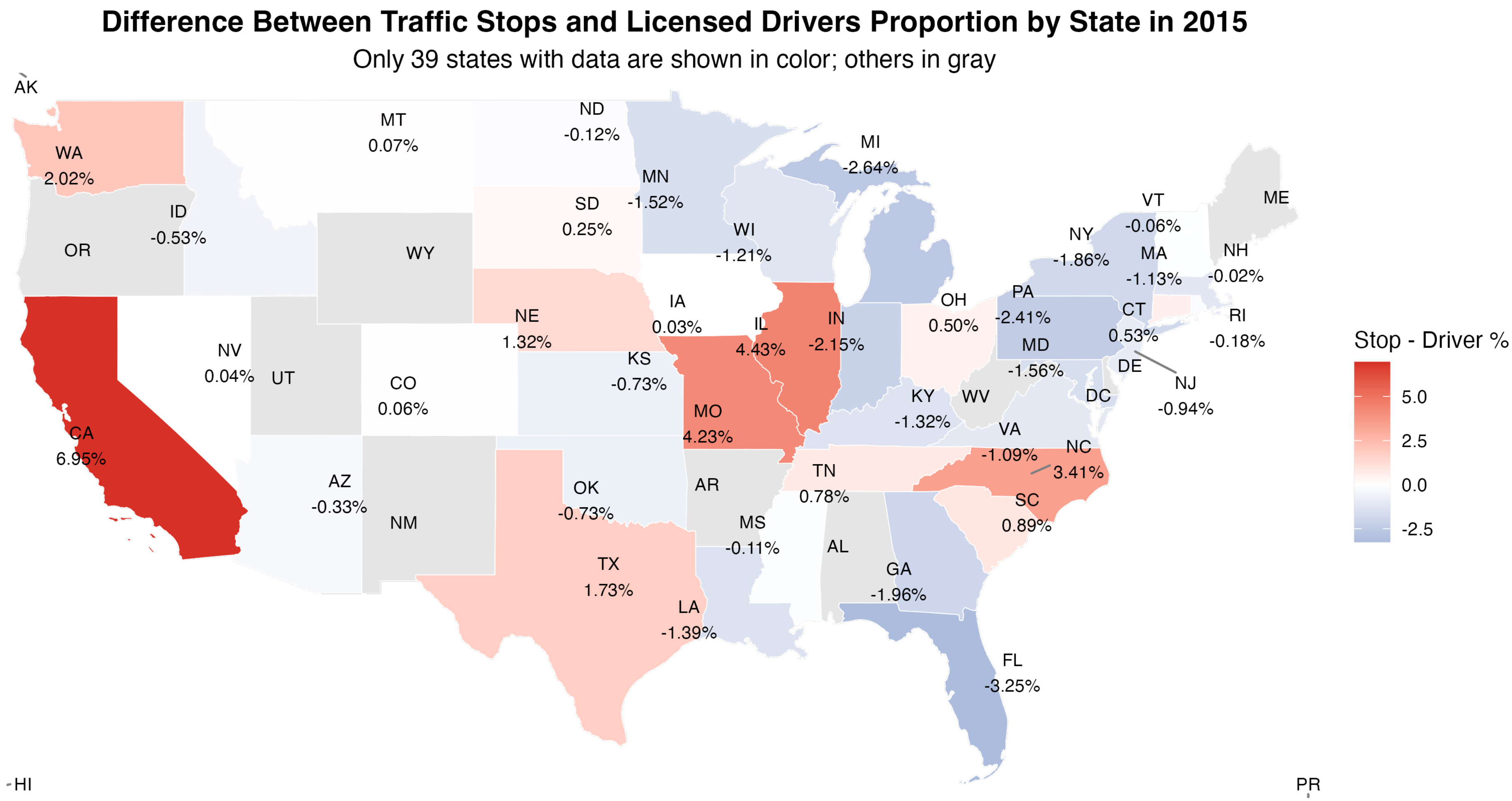

Figure 4: Differences between Traffic Stops & Licensed Drivers Proportion by State in 2015

- Consistently **Under-Policed States** (fewer stops than expected in both years):
  - Pennsylvania(PA)
  - Florida(FL)
  - Michigan(MI)
  - Indiana(IN)
- Consistently **Over-Policed States** (more stops than expected in both years):
  - California(CA)
  - Illinois(IL)
  - Missouri(MO)
  - North Carolina(NC)

State	2014	2015
Missouri(MO)	Mildly Imbalanced	Mildly Imbalanced
Michigan(MI)	Mildly Imbalanced	Balanced
Indiana(IN)	Mildly Imbalanced	Mildly Imbalanced

- **Most states are statistically balanced** in terms of traffic stop proportions vs. licensed driver proportions.
- **Only a few states** showed *mild imbalance* across both years, These imbalances are small in magnitude and may have limited real-world impact.
- **No strong evidence of systemic geographic bias** in traffic stops across states
- **Small imbalances may reflect local operational factors** (e.g., highway traffic, policing resource allocation)

Question 2

D

Are there any day-to-day variations in traffic stops (e.g., more/fewer stops on weekends; more/fewer stops at certain times of the day)? Why?

Setup

- Technique:
  - χ^2- goodness of fit test to assess whether traffic stops occur uniformly across specified time intervals
  - Cohen's W used to measure the effect size of observed differences to assess practical significance
- Assumptions:
  - Traffic stops occur independently of one another
  - Expected frequency of stops during a given interval is at least 5

Results

D

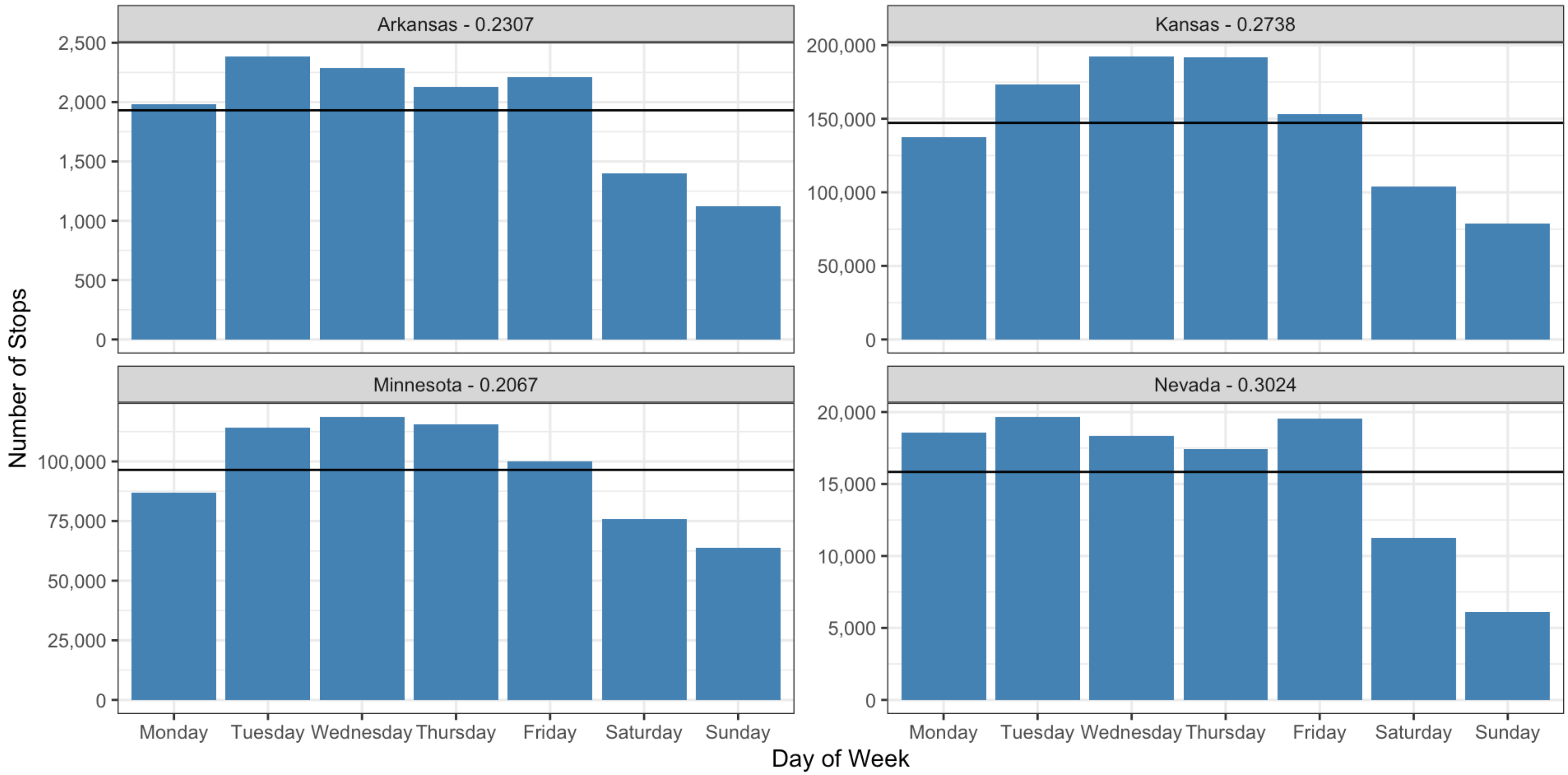

Figure 5: Day-to-day variation of stops for states with largest Cohen's W effect size

Results

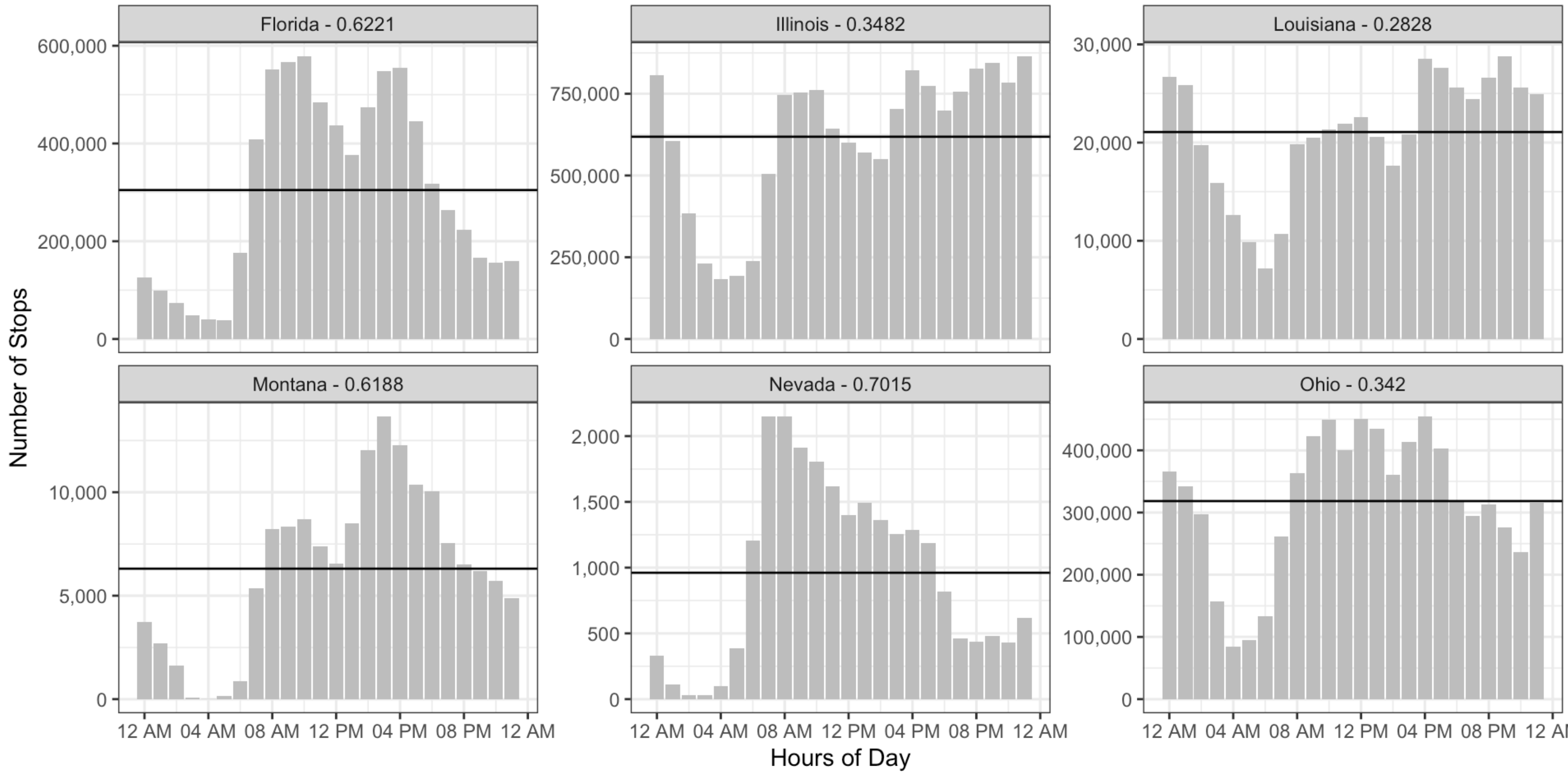

Figure 6: Hour-to-hour variation of traffic stops for states with largest and smallest Cohen's W effect size

Results

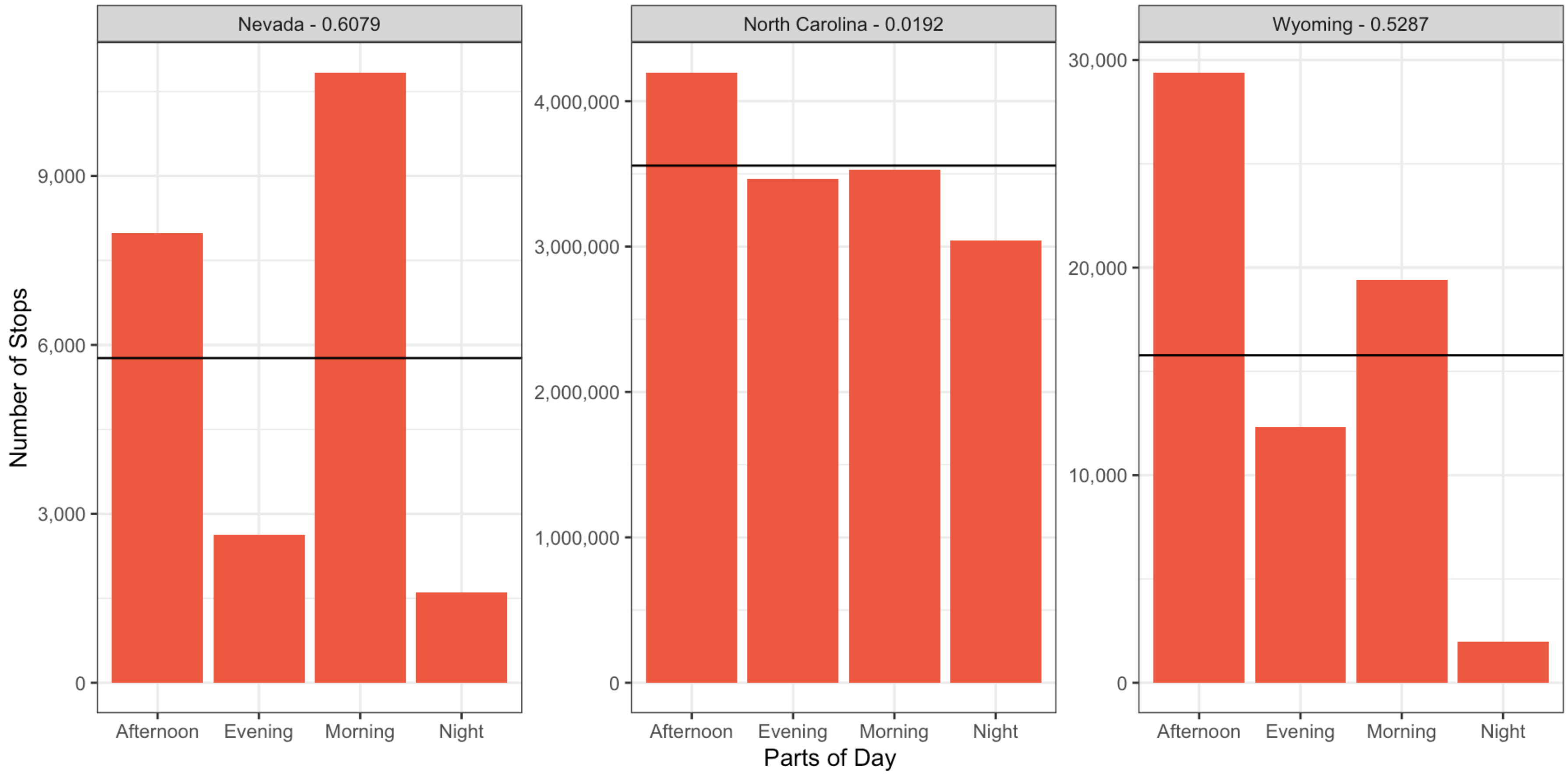

Figure 7: Part-of-Day variation of traffic stops for states with largest and smallest Cohen's W effect size

Discussion

- Minnesota, Kansas, Arkansas, and Nevada have the most traffic stop imbalance throughout the week.
- Friday is the most commonly observed date for traffic stops, and weekends generally display large differences in stops compared to weekdays.
- There is a noticable hour-to-hour variation in traffic stops, with the most stops typically occurring in the afternoon.
  - However, there is a smaller difference between the number of stops in the morning, afternoon, evening, and night.

Question 3

X

Can we identify trends in behaviours over the long term? Do some cities/locations see increases or decreases? What could that tell us?

Setup

- Time series analysis examines the components of data through time, including **trend**, **seasonality**, and **noise** (NIST, 2013).
- The *Seasonal and Trend decomposition using Loess* (STL) performs time series decomposition by separating a series into **seasonal**, **trend**, and **remainder** components using *locally weighted regression* (Loess) (Hartmann et al., 2023).

Time Series Visualizations (USA)

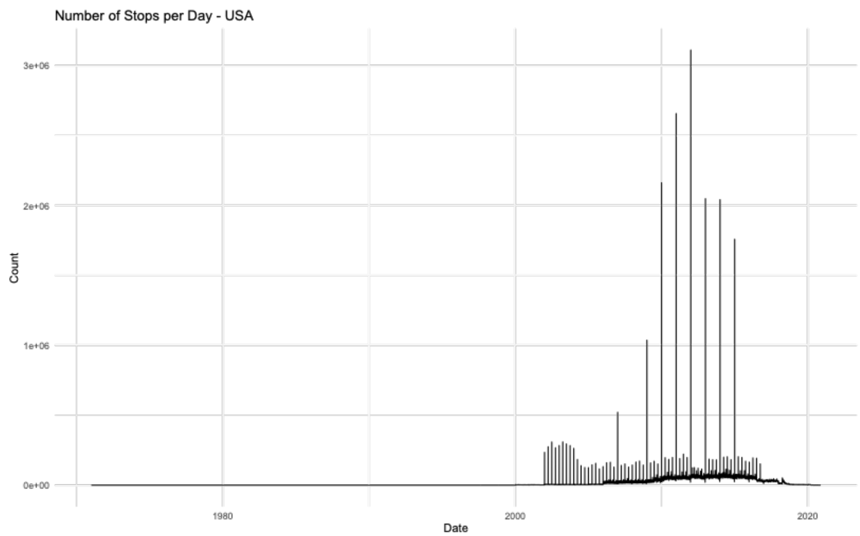

Figure 8: Time series of number of stops per day in USA. Few data before the year 2002.

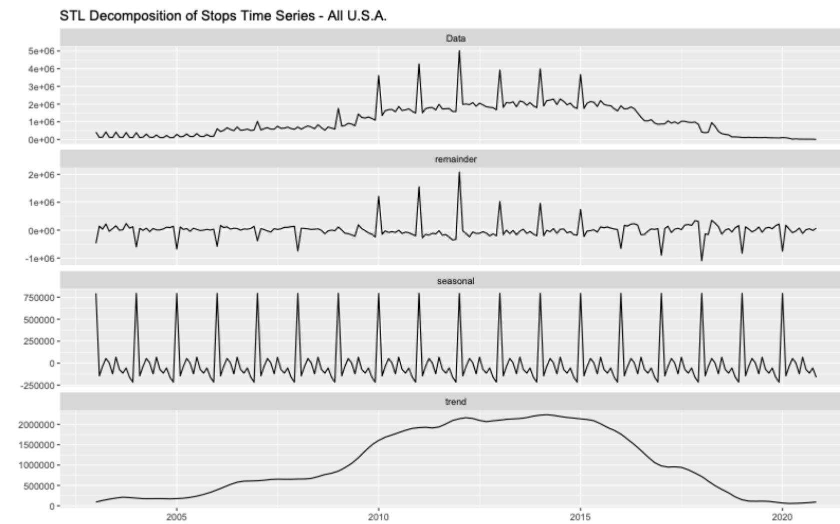

Figure 9: STL for number of stops per day in USA.

Trend Results

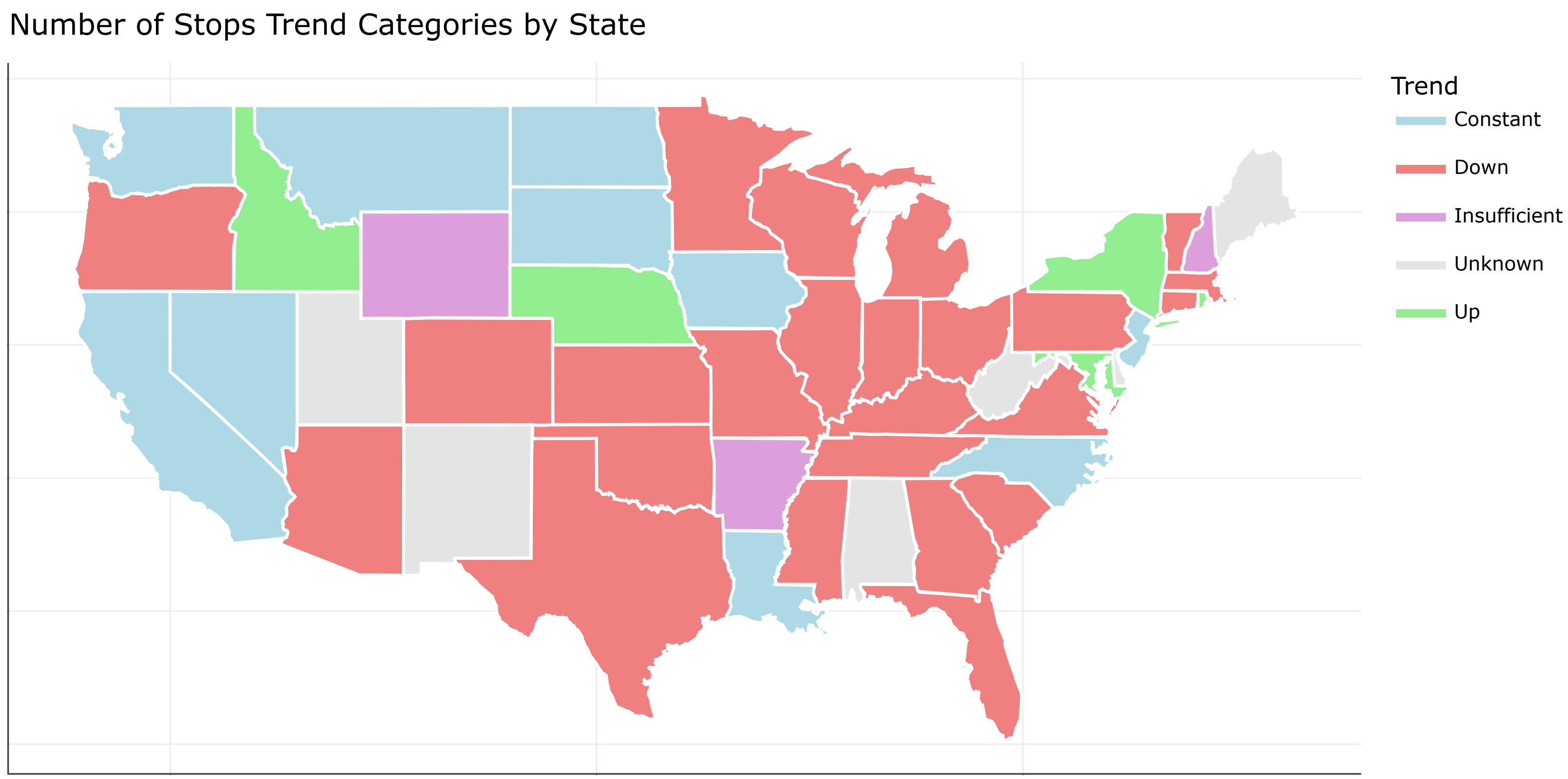

Figure 10: Most states have shown a downward trend or remained constant in the number of stops in recent years.

Seasonal Results

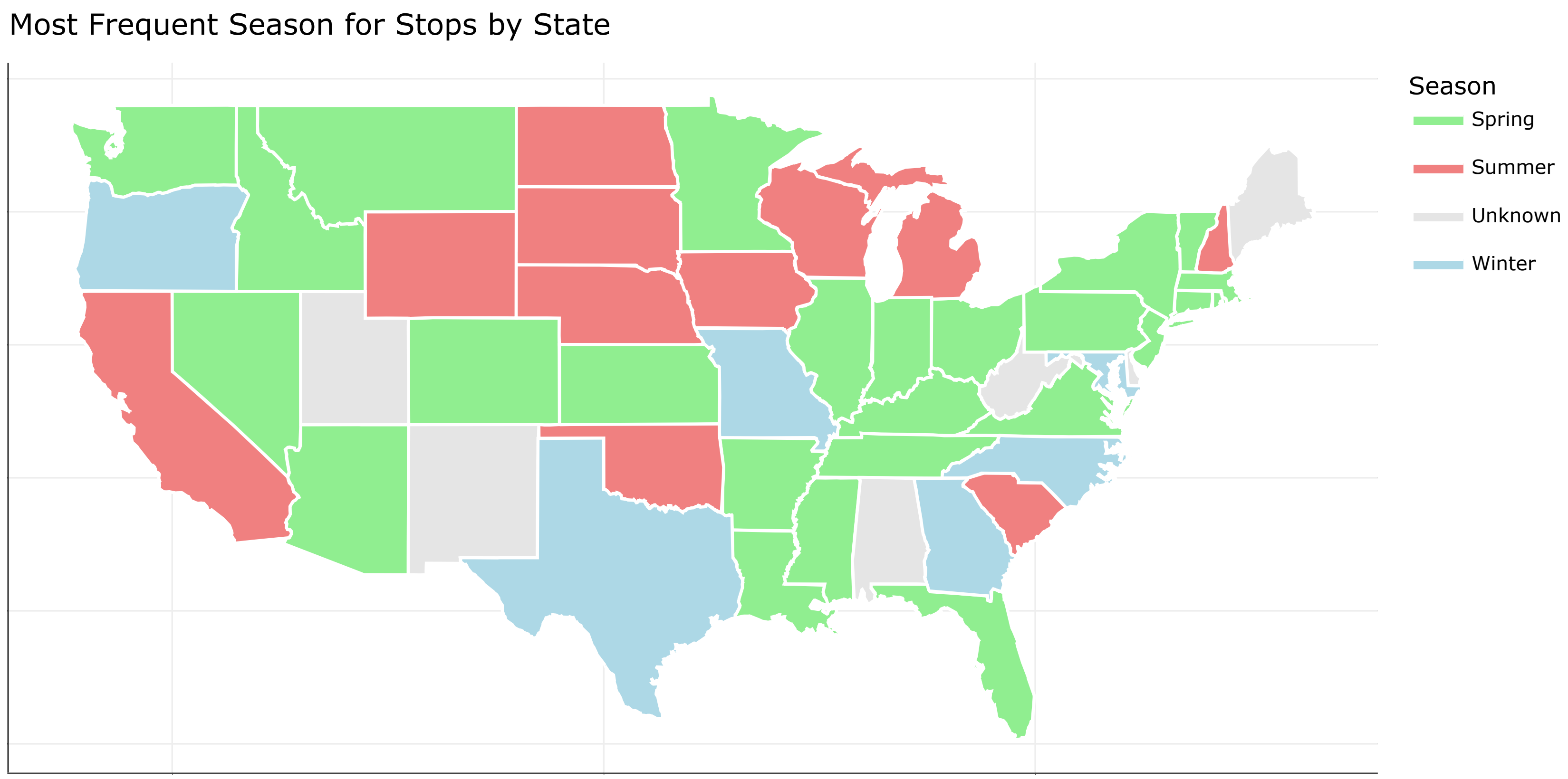

Figure 11: The highest number of daily stops occurs in spring and summer across all states.

Case 1: Texas - Sandra Bland Act (SB 1849)

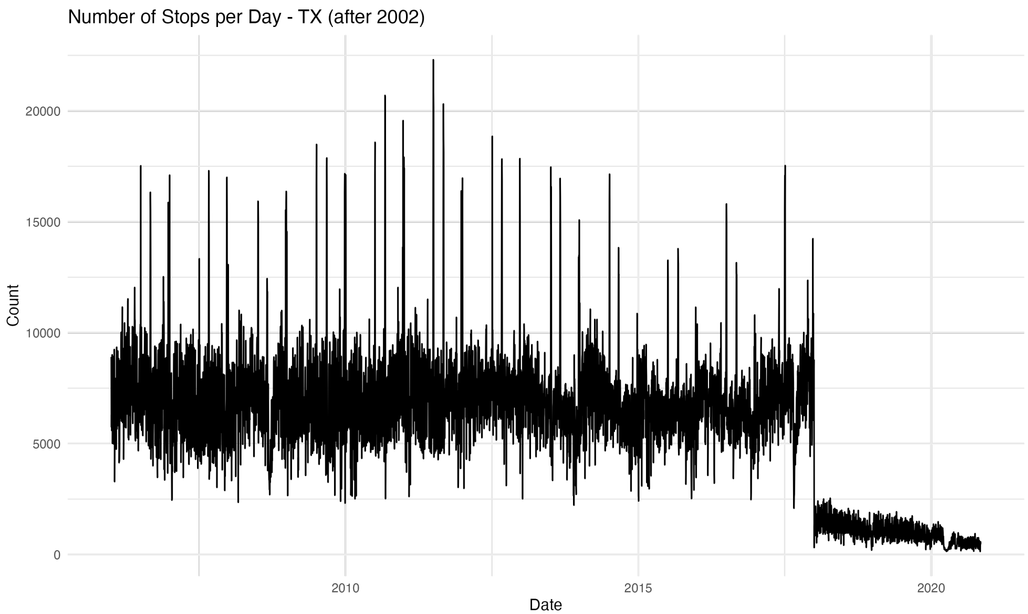

Figure 12: An abrupt decrease is observed after 2017.

- The **Sandra Bland Act (SB 1849)** was enacted in 2017, and it:
  - Stricter criteria now required for vehicle stops and searches in Texas (e.g, asking for consent and preventing pretextual stops) (Whitmire, n.d.).

Case 2: Arizona – SB 1070

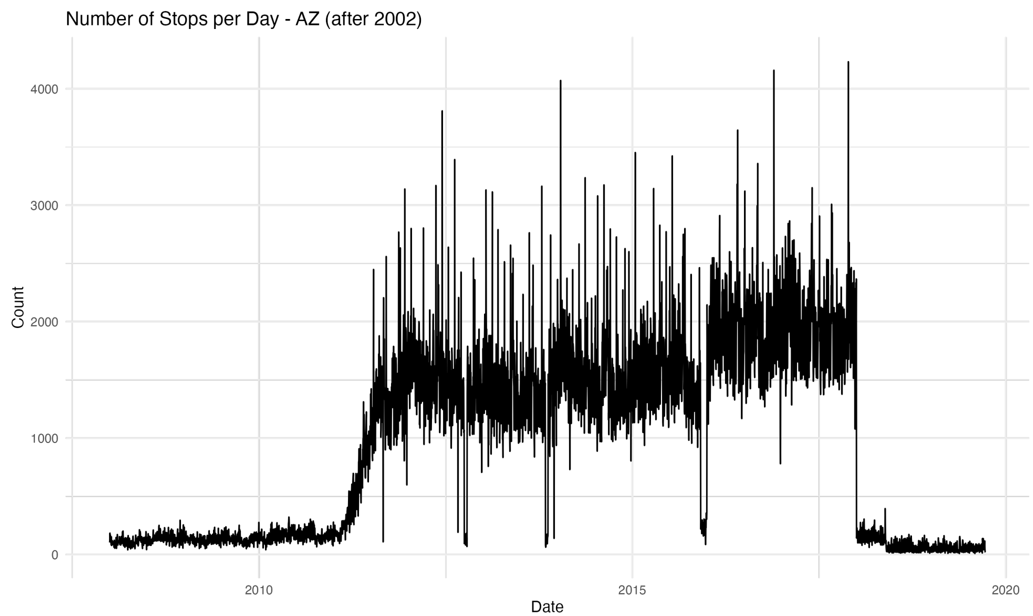

Figure 13: Stops increased from 2012 to 2016, then declined.

- **SB 1070** (enacted in 2010) stopped in 2016 (Schwartz, 2016). It required police to determine immigrant status of someone 'reasonable suspicious' of not being legally in U.S.A., even in traffic stops.
- In 2012 the Supreme Court of U.S.A. nullified 3 of the 4 law provisions, but one section (2B - 'Show your papers') was still intact (Newman, 2017).
  - It required police to detain anyone suspected of being undocumented until immigration status was verified.

Case 2: Arizona – Enforcement & Court Supervision

- **Sheriff Joe Arpaio** (2008-2011) launched large-scale traffic sweeps (*crime saturation patrols*) (Arizona, 2011).
- Maricopa County had been under court supervision since a judge concluded in 2013 that sheriff's deputies racially profiled Latinos (Collins, 2023).

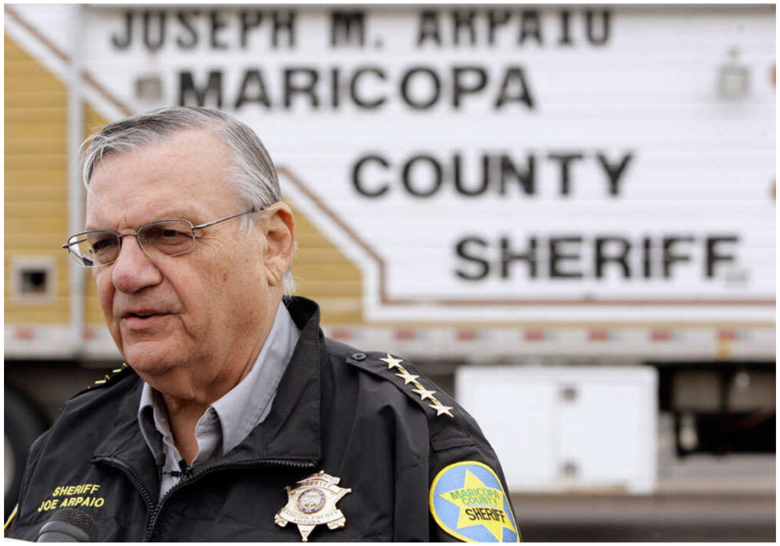

Figure 14: Maricopa County Sheriff Joe Arpaio

Case 3: Connecticut - Alvin W. Penn Law and UConn Audit

- The **Alvin W. Penn Racial Profiling Traffic Stop Law** (enacted in 1999) required state and local police to collect and report traffic stop data to the state (UConn, 2021).
  - It prohibited stops based solely on race, ethnicity, age, gender, or sexual orientation.
- Using this data, the University of Connecticut audited stops from 2014 to 2021 (Collins, 2023).

Figure 15: Connecticut General Assembly

Case 3: Connecticut - Record Falsification

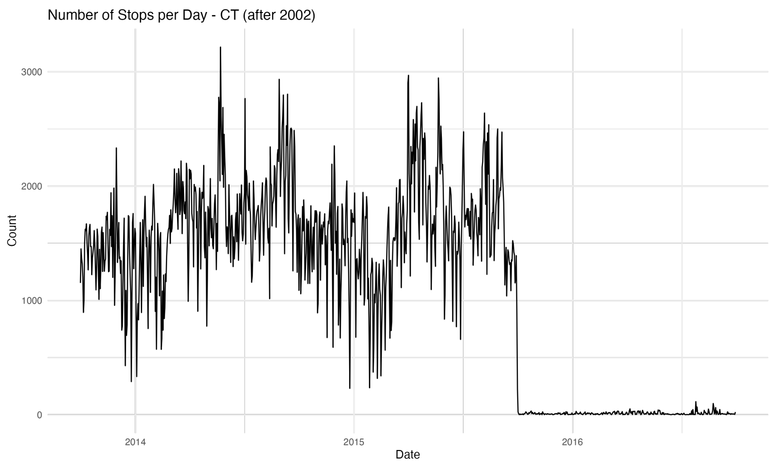

Figure 16: Number of stops per day drastically decreased after 2015.

- The study found at least 25,000 falsified traffic stop records (possibly up to 58,000) (Collins, 2023).
- White drivers were overrepresented; Black and Hispanic drivers were underreported.
- The study could not confirm whether falsifications were intentional or due to human error.

D

Conclusion

Limitations: Incomplete panel data, lack of racial demographic information for licenses

Future work: Multilevel spatial modelling, linear mixed effects models

| *Georgetown Immigration Law Journal | Georgetown Law*. https://www.law.georgetown.edu/immigration-law-journal/in-print/volume-31-issue-3-spring-2017/arizonas-anti-immigration-law-and-the-pervasiveness-of-racial-profiling/?

NIST. (2013). *6.4. Introduction to time series analysis*. https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm. https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, *4*(7), 736–745. https://doi.org/10.1038/s41562-020-0858-1

Schwartz, D. (2016). Arizona police will no longer stop people to check immigration status | reuters. In *Reuters*. https://www.reuters.com/article/us-arizona-immigration/arizona-police-will-no-longer-stop-people-to-check-immigration-status-idUSKCN11M05D/

UConn. (2021). Connecticut racial profiling prohibition project (CTRP3). In *Institute for Municipal and Regional Policy*. https://imrp.dpp.uconn.edu/ctrp/

Whitmire, J. (n.d.). In *85(R) SB 1849 - A BILL TO BE ENTITLED AN ACT*. https://capitol.texas.gov/tlodocs/85R/billtext/pdf/SB01849I.pdf