Data visualization project

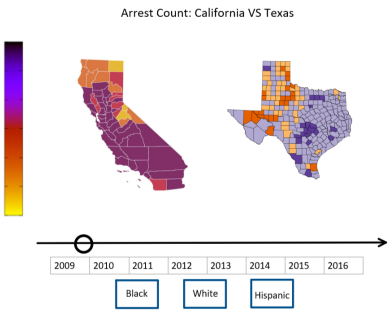
1 Introduction

In this project, we want to investigate a potential variation in racial bias in police stops from 2009 to 2016. We also want to explore, in case such variations exist, if those are constant across different states of America. To simplify our investigation, we will focus our research on California (mostly Democrat) and Texas (mostly Republican).

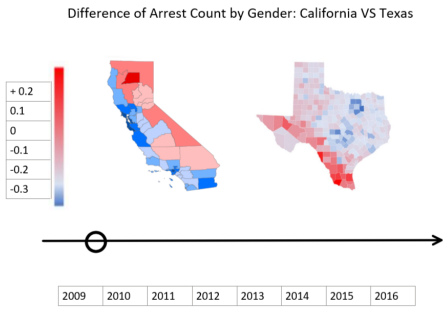
2 Visualization ideas and sketches

2.1 Map view ethnicity

It will be possible to visualize the respective number of police arrest of Black, White and Hispanic persons in the different counties of California and Texas. With the help of a cursor it will be possible to observe the evolution of the number of arrests between 2009 and 2016.



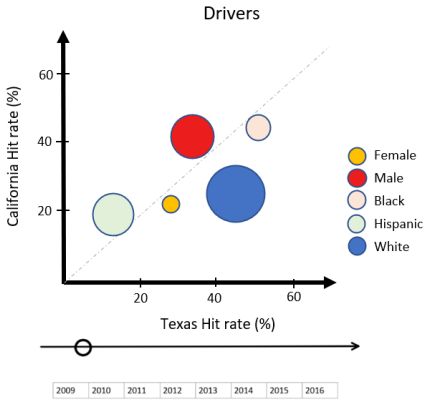
2.2 Map view gender



It will also be possible to similarly visualize the numbers of arrests considering the gender of the person arrested. This could showcase a potential evolution of the percentage of women apprehended across time. We might also observe a difference in the ratio of arrested women between the democratic state of California and the republican state of Texas that we will represent with a color gradient.

2.3 Hit rate to represent potential bias

In this second view, we want to focus on illustrating potential bias by measuring the hit rate (an arrested person is found guilty of an infraction). The size of the bubbles will represent the number of arrests and, the (x,y) position, will illustrate the hit rate of each population in Texas compared to the respective hit rate in California.



3 Tools we will need

The tools we will need are the d3.js library in order to design our different visualizations, the topoJSON library to visualize the maps of the California and Texas states along with their counties. We will also need to find the respective topoJSON of the California and Texas counties. As for the lectures we will need to complete our visualization, we will need the lecture on maps for the map plot, and the lecture on tabular data for the hit rate plots. We will also need the lectures on d3.js and javascript.

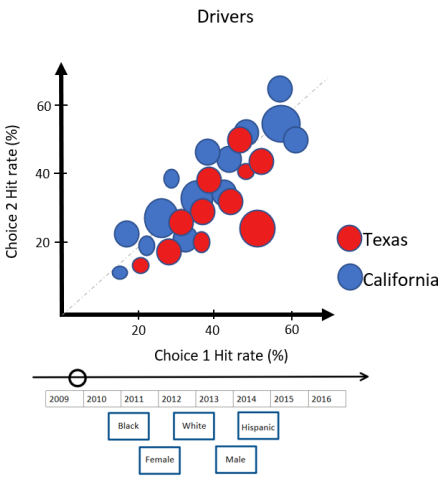
4 Project breakdown structure

4.1 Core visualization

Our project will be structured around two main visualizations. In the first one, we want to visualize a map of arrests for the different ethnicities and genders across time. With the second visualization, we will illustrate bias among the different population between Texas and California using the hit rate of police arrests. As the data for this part of the project is available for a small portion of cities, we will select a representative city for each state. This will furthermore simplify data pre-processing and aggregation.

4.2 Extra ideas

We would like to add multiple cities for both states if we can to obtain more accurate results. We would also like to add a new feature to this visualization where it would be possible to compare any 2 particular populations (ex : Hispanic living in Texas and White living in California) and visualize their respective hit rate.



Références

[1] E. Pierson, C. Simoiu, J. Overgoor, S. Corbett-Davies, D. Jenson, A. Shoemaker, V. Ramachandran, P. Barghouty, C. Phillips, R. Shroff, *et al.*, "A large-scale analysis of racial disparities in police stops across the united states," *Nature human behaviour*, vol. 4, no. 7, pp. 736–745, 2020.

[2] C. Simoiu, S. Corbett-Davies, and S. Goel, "The problem of infra-marginality in outcome tests for discrimination," *The Annals of Applied Statistics*, vol. 11, no. 3, pp. 1193–1216, 2017.