Joey Dubill, Kevin Chen, Shamroy Pellew
CSE 410/540 Machine Learning and Society
3121 words

Bias In Traffic Stops:
Veil of Darkness

Introduction

Between 2011 and 2015, there were over 60 million traffic stops in the US. On a typical day in the US, officers make more than 50,000 stops. Due to federal and state regulations, there is a requirement for some police departments to make the data on these traffic stops openly available.

In the past, people have used this data to show that these traffic stops can be racially biased, which flies in the face of our group's belief that all people are equal. So, we decided to look into traffic stops to reaffirm past results on racial discrimination, and in doing that, we came up with findings of our own.

In order to understand how traffic stops can be racially biased and identify bias in the data, we settled on the veil of darkness test after much deliberation. We originally planned to do our study on hot spot policing, but we ran into issues analyzing the data. Veil of darkness, on the other hand, was much more approachable. Crafted by researchers Grogger and Ridgeway in 2006, the veil of darkness test uses the assumption that traffic stops after dark should have a smaller ratio of black drivers stopped because police cannot racially identify drivers when there is less daylight. Using the veil of darkness test and the data analyzed using the Stanford Opening Policing Project, we were able to determine that there is bias in traffic stops.

Related Work

One of the works we used in our project was the Stanford Open Policing Project. This project included the data sets of police stops in all states and a tutorial program that helped us start our project. The tutorial code provided us with an easy way to visualize the data, and it gave us a digestible introduction to the R language. However, we had to adjust some of the code to work with different cities. So to better understand the algorithm and support those cities, we decided to replicate the veil of darkness test in Python.

Another work we used was the "Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness" (Jeffrey Grogger & Greg Ridgeway). In summary, the veil of darkness test suggests there are fewer traffic stops and searches conducted on people of color during times where it is harder to identify a person's race. This test was used to enforce our conclusion of bias in traffic stops after seeing there were higher stops and searches conducted on people of color compared to white people.

The next work we used was the hotspot policing article. While this did not have a direct link to the other works above, it did provide us with a possible reasoning why we might have seen increased stops or even bias in traffic stops. In general hotspot policing is when you allocate more police resources to an area with high criminal activity and most crime would occur in that area. Our theory was that police officers could have formed a bias to that hotspot and then in turn formed a bias to the people living in that area. Which would show an increased amount of searches for certain races related to the bias. However, we did not have an ample amount of time in order to find a way to prove our theory using the data set provided to us by the Stanford Open policing project.

Methodology

We first wanted to see if there was a group of people who were stopped more than another group. We were able to see this data by using the Stanford Open Policing project. We first replicated the tutorial test used by the Stanford Open Policing team which provided us with visual graphs and tables. After that we wanted to see if different regions of the United States had similar issues and if it was the same group or another group being targeted. We were able to adapt the Stanford Open Policing tutorial code to be used on two different cities which we picked was San Antonio Texas and Los Angeles California. We were able to output graphs for both cities and saw the trends for each of them. After seeing the visual data we wanted to validate our hypothesis that traffic stop bias exists.

The way we chose to validate our hypothesis was the Veil Of Darkness test which tests if a black drivers and hispanic drivers were targeted on purpose when their race could be identified. If we saw an decreased number of stops when race could not be identified as easily during the dark it would show a clear sign of police officers having a bias towards a certain race.

There could be a multitude of reasons why black drivers may be stopped more after dusk. There could be more black drivers heading home than white, black drivers could drive differently after a certain time. What's important to note though is that all of these reasons connect to "clock time."

In order to account for differences in driving behavior due to clock time, we can use changing daylight patterns to our advantage. For example, there are some parts of the year when 7PM will be in complete daylight and other times when it will be in complete darkness. Driving behavior however shouldn't change much between these two times. Geography could also play a role in traffic stops, some police districts may have lower conditions for search in predominantly black neighborhoods.

Taking all of these observations into account, and using OPP's model[3] as an example, we came up with a formula to find the effect that darkness, time, and location play on minority drivers being stopped:

$$is\_minority \sim is\_dark + cr(rounded\_minutes, df = 6) + location$$

First, we manipulate the raw data from the Open Policing Project to get it into the desired format. We load the data from CSV files using the pandas data analysis library, categorize the locations of traffic stops, and only consider black and non-black drivers. We filter out traffic stops that don't fall between our earliest and latest sunset, and the traffic stops with ambiguous lighting as the sun begins to set. For greater efficiency, we round the time of a traffic stop to five-minute intervals. However, this introduces discontinuity into our model, so we use the cr function to make up for the non-continuity and apply some smoothness using a spline. We used a spline with six degrees of freedom as the Open Policing Project, and Grogger & Ridgeway both found six to be the ideal value for degrees of freedom.

Although both OPP and Ridgeway's models included darkness, the papers never specified how the researchers determined sunset times, and they weren't included in the traffic stops data either. So, because the is_dark variable in our model relies on sunset times, we had to look to external sources. During our search, we found the popular Sunrise Sunset website. This website touts itself as a way to "make it easy [for] everybody to access Sun related information through simple tools that [offer accurate] information." It also included an API to make it more effortless to do so, and we used it to determine the times when civil twilight ends for a given day and longitude and latitude. Longitude and latitude were determined using Python Geocoder, an open source geocoder library that reverse geocoded the location names.

Civil twilight is important for our predictions because it removes ambiguity regarding darkness. Civil twilight is defined as the period after sunset when there is still sufficient sunlight for outdoor activities, so we only determine if it was dark using civil twilight and if a traffic stop happens after it means it was dark enough for vision to be impaired.

Next, using our formula and all of the data we've gathered, we instantiate a binomial family model and perform generalized linear regression on our model. The results of this model can then be inspected, including the coefficients of each independent variable. These coefficients represent the weight each independent variable holds on the dependent variable. The dependent variable here is whether or not a stopped driver is black. Ideally, darkness would have no effect on whether a stopped driver is black, but we can determine if it does by the coefficient of the is_dark independent variable. Due to a zero tolerance policy on discrimination, any significant difference signifies racial bias.

See below for the veil of darkness used to compute the results in the results section.

<u>Veil of Darkness Code</u>

```
from datetime import datetime
from dateutil import tz, parser
from sklearn import model_selection
from sklearn.metrics import classification_report
import geocoder
import json
import math
import pandas as pd
import requests
import statsmodels.api as sm
import statsmodels.formula.api as smf
import time


class OpenPolicing:
   def __init__(self, download_directory=""):
       self.latlng = None
       self.twilight_by_date = None
       self.factors_dict = None
       self.download_directory = download_directory
       pd.set_option('display.max_columns', None)

   def veil_of_darkness(self, data_file, year, minority_demographic="black",
degree=6, consider_district=False, twilight_cache={}, debug=False):
       path_to_data_file = data_file[0] + ".csv"
       to_geocode = data_file[1]

       df = pd.read_csv(self.download_directory + path_to_data_file)
       if year:
           self._run_vod(path_to_data_file, to_geocode, df, year,
minority_demographic, degree, consider_district, twilight_cache, debug)
       else:
           years = df["date"].map(lambda date: date.split("-")[0]).unique()
           print(f"Path To Data File: {path_to_data_file}")
           print(path_to_data_file)
           for data_year in years:
               print(f"Data Year: {data_year}")
               self._run_vod(path_to_data_file, to_geocode, df.copy(deep=True),
int(data_year), minority_demographic, degree, consider_district, twilight_cache,
debug)

   def _run_vod(self, path_to_data_file, to_geocode, df, year, minority_demographic,
degree, consider_district, twilight_cache, debug):
       self.latlng = geocoder.osm(to_geocode).latlng
       self.twilight_by_date = {}
       self.factors_dict = {}

       df, num_initial_rows, num_valid_date_time_rows, num_non_duplicate_rows,
num_year_rows = self._clean_data_frame(df, year)
       earliest_sunsets = df[((df["date_time"].dt.month == 11) &
(df["date_time"].dt.day >= 25)) | ((df["date_time"].dt.month == 12) &
(df["date_time"].dt.day <= 15))]
       earliest_sunsets["date_time"].map(lambda date_time:
self._date_time_to_darkness(date_time))

       latest_sunsets = df[((df["date_time"].dt.month == 6) &
(df["date_time"].dt.day >= 10)) | ((df["date_time"].dt.month == 7) &
(df["date_time"].dt.day <= 15))]
```

```
self._date_time_to_darkness(date_time))

       print("Converting date_time to rounded_minutes...")
       df["rounded_minutes"] = df["date_time"].map(lambda date_time:
self._date_time_to_rounded_minutes(date_time))

       feature_cols = ["is_dark", "rounded_minutes"]
       if consider_district and "district" in df.columns:
           df["location"] = df["district"].map(lambda district:
self._as_factor(district))
           feature_cols += ["location"]

       is_minority = df["subject_race"].map(lambda x: 1 if x == minority_demographic
else 0)
       return df.loc[:, feature_cols], is_minority


with open("cache.json", "a+") as f:
   op = OpenPolicing("/CSE540/traffic-stops-project/data/")
   twilight_cache = {}
   f.seek(0)

   if f.read(1):
       f.seek(0)
       twilight_cache = json.load(f)

   op.veil_of_darkness(
       ("ca_oakland", "Oakland California USA",),
       2015,
       consider_district=True,
       twilight_cache=twilight_cache,
       debug=True,
   )
```

Results

Using this model, we were able to reaffirm the findings of Pierson et al.[1], we found that there seems to be racial profiling going on in Texas statewide's stops. The negative coefficient of the is_dark variable means that darkness decreases the likelihood a stopped driver is black. So, because it is too dark for an officer to identify a driver's race, they are likely discriminating against black drivers during the daytime.

We found the opposite to be true for Oakland California, and were able to reaffirm the findings of Grogger and Ridgeway[2] using our model. We found no evidence of racial profiling in the city's traffic stops, and although the is_dark coefficient is negative here, the standard error is significantly larger.

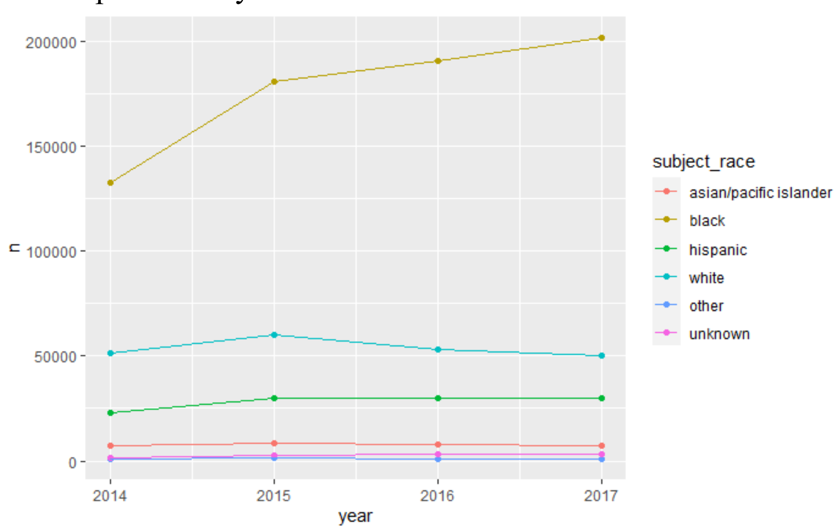
The results of running our model on the data also support the Stanford Open Policing Project's analysis of traffic stops in Philadelphia, Pennsylvania. OPP found that darkness lessens the chance a black driver is stopped by police with a coefficient of -0.20676109 and a standard error of 0.02171533. Our model predicted an is_dark coefficient of -0.0420 and 0.017. Both models imply racial bias, but we found that darkness has less of an impact on a black driver being stopped. This could be due to our model rounding time to five-minute intervals, which the OPP didn't do before running their model.

Similar to the Open Policing Project, we also ran our model while taking location into account as a possible reason black drivers are being stopped. OPP found a coefficient of -0.16660152 and a standard error of 0.02502906 while taking the police district into account, and we found a darkness coefficient of -0.0425 and a standard error of 0.017. Although our model again agrees with OPP's and suggests racial bias, it seems to find that location has less of an impact on traffic stops in Philadelphia. Another important thing to note is that our model's prediction has a larger standard error, however, our standard error is at least 2.47 standard errors away. This means our standard error isn't significant although it's larger than the one calculated by OPP.

Given all of this evidence, we believe our recreation of the veil of darkness test works. As a result, we were also curious if we could find discrimination in traffic stops in the cities in which the Open Policing Project found racial disparities in traffic searches.

The Stanford Open Policing project provided us graphs and tables visualizing the data. The cities we analyzed and looked at were Philadelphia Pennsylvania, San Antonio Texas and Los Angeles. In Philadelphia there proportionally more stops for blacks than other races. San Antino also showed a similar trend with Hispanics being stopped far more than any other race. Los Angeles also showed that Hispanics were stopped more than the other races.

Graphs and Tables Obtained by Using the Stanford Open Policing Projects:
Philadelphia Pennsylvania :

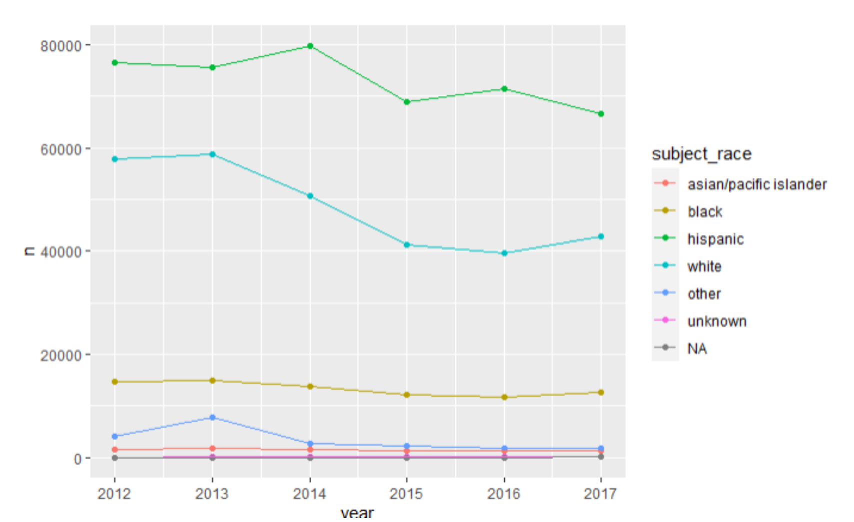


A tibble: 6 x 3

subject_race <fctr>	search_rate <dbl>	frisk_rate <dbl>
asian/pacific islander	0.02613240	0.01849370
black	0.05920740	0.06223491
hispanic	0.05053183	0.04660605
white	0.04063881	0.03021657
other	0.03195489	0.02349624
unknown	0.03830528	0.03656413

6 rows

San Antonio Texas:

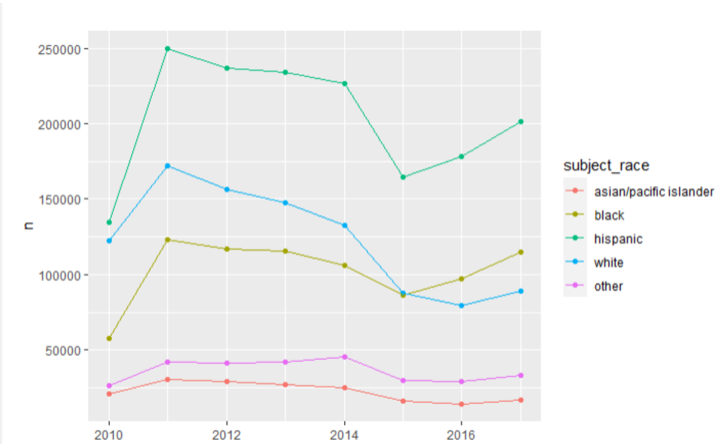


A tibble: 7 x 4

subject_race <fctr>	n <int>	num_people <dbl>	stop_rate <dbl>
asian/pacific islander	1427	110864	0.01287163
black	12597	648846	0.01941447
hispanic	66577	221777	0.30019795
white	42907	548312	0.07825289
other	1769	NA	NA
unknown	173	NA	NA
NA	89	NA	NA

7 rows

Los Angeles California:



A tibble: 5 x 4

subject_race <fctr>	n <int>	num_people <dbl>	stop_rate <dbl>
asian/pacific islander	16937	110864	0.1527728
black	114806	648846	0.1769387
hispanic	201172	221777	0.9070914
white	89335	548312	0.1629273
other	32883	NA	NA

5 rows

Conclusion

The visual data obtained from the Stanford Open Policing project suggests that there is a bias since african americans and hispanic drivers were shown to be stopped more in all three cities we tested. The search rate was also higher for african americans and hispanics compared to white drivers which suggests that not only there was a stop bias but also a search bias.

Based on the results, it looks like our reproduction of the veil of darkness test was a success. Caveats exist though, for example, darkness may be temporarily limited by street lights or headlights, allowing police to identify a driver's race. The type of vehicle driven may correlate to race, and a police officer may recognize a driver's race using this correlation. Seasonal changes may also play a part in driving behavior, and a possible solution may be examining traffic stops around daylight savings time. DST allows for a period to be dark one day and then light the next, and because it happens from one day to the next there aren't seasonal changes. VOD does provide a useful measure of bias though, but some factors are tough to consider.

We believe future work should involve optimizing our model to run faster. Testing out different interval times and different degrees of freedom could majorly reduce runtime at the cost of a minor decrease of accuracy.

Hopefully find a solution/preventive measures in traffic stop bias. We had difficulties with hot spot policing, primarily due to the data. We initially looked at the veil of darkness, but continued on with search rates, hit rates, etc. This test provided meaningful insight into how traffic stops are biased. Running the data on every state was our goal, but we ran into issues with the features

Citations

1. Jeffrey Grogger & Greg Ridgeway (2006) Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness, Journal of the American Statistical Association, 101:475, 878-887, DOI: 10.1198/016214506000000168
2. Pierson, E. et. al (2020, July). *A large-scale analysis of racial disparities in police stops across the United States*. https://5harad.com/papers/100M-stops.pdf.
3. The Stanford Open Policing Project. (2017). Openpolicing.Stanford.Edu. https://openpolicing.stanford.edu/tutorials/
4. David Weisburd and Cody W. Telep(May 2014). Hot Spots Policing: What We Know and What We Need to Know, Journal of Contemporary Criminal Justice DOI:10.1177/1043986214525083 https://www.researchgate.net/publication/269598433_Hot_Spots_Policing