

# Racial Inequities in 2018 Boulder Police Discretionary Stops: An Open Data Analysis Code ▾

*Peter Shaffery*

*June 9, 2019*

## Introduction

On May 1, 2019 the Boulder Police Department (BPD) released a (cleaned and processed) dataset (<https://www.denverpost.com/2019/05/01/people-of-color-arrests-boulder-police-stop/>) containing data on all *discretionary* stops made by Boulder police in 2018. Discretionary stops make up only a fraction of Boulder police activity, with the vast majority of police interactions being *non-discretionary*, eg. where an officer was dispatched based on a 911 call or following a warrant. Approximately 66,400 police interactions are non-discretionary (per an in-person workshop on the data organized by the City of Boulder on May 6, 2019), whereas 8,209 are discretionary (per the released data).

The data were compiled on the recommendation of the consulting firm Hillard Heintz, who in 2016 were hired to perform an internal study of arrest and citation activity within BPD. In an analysis of this 2018 data, Hillard Heintz concluded (<https://www.dailycamera.com/2019/04/30/data-people-of-color-more-likely-to-be-arrested-after-boulder-police-stops/>) that "Black people are twice as likely as White people to be stopped at an officer's discretion, and once stopped, they are twice as likely to be arrested". In addition to these results, the City of Boulder released an official report (<https://bouldercolorado.gov/police/stop-data-information>) which contained a few high-level takeaways: 1. The report represents only a single year of data and "cannot yet be put into context" (ie. trending up or down) 2. Small data sizes mean that small changes in stop or citation number may have meaningful effects in data trends 3. Laws and policies regulating discretionary action can have "significant positive effects", but will require coordinated action 4. Racial differences in stop rates differ by base population (eg. residents vs nonresidents)

In my analysis of the data I found these summaries to be true, but conservative descriptors of what the data shows. Agreeing with the official analysis of the dataset I find strong evidence that racial bias is occurring across all types of discretionary stop by BPD. Regardless of trend or context, in 2018 it was a real, urgent, and sizeable problem. If BPD were to stop Black individuals at the same rate as White individuals (relative to the base population estimates given in the report) it would have constituted 220 fewer stops in 2018.

Breaking down the stop data by "stop reason" highlights some of the reasons why this bias occurs in the top-level. While different stop types affect residents and non-residents differently, for both groups the stop type exhibiting the largest bias was "municipal violations". Speeding and traffic violations also made up a large component of the bias, as did "suspicious" stops. Indeed, Black individuals are stopped more frequently than White individuals in every type of stop listed except for "welfare checks" of non-residents, where BPD were slightly more likely to stop White individuals.

In addition to biases in the stop rates, I find that Black individuals who were searched for contraband were significantly less likely to have it than their White counterparts. This indicates that, on aggregate, BPD stop Black individuals with a lower threshold of (productive) evidence. It furthermore suggests that different perpetration rates cannot explain the apparent differences in other types of stop rates. Similar results also typically hold for White Hispanic individuals, although base demographic information was not provided for this group and so the results are more limited. Black Hispanic individuals were also largely absent from the dataset. It should be noted that race and ethnicity were assessed by the stopping officer.

## Front Matter

Biased or unreasonable police stops are an injustice that needs to be redressed immediately. While I'm glad that BPD have released this data and thereby taken a real step towards justice and transparency, it must be matched by appropriate changes in their day-to-day operations and procedures.

Before going any further it's relevant to consider a quote from Candice Lanier's excellent essay "Your Demand for Statistical Proof is Racist": (<https://thesocietypages.org/cyborgology/2015/01/12/fact-check-your-demand-for-statistical-proof-is-racist/>)

Perhaps statistics should be considered a technology of mistrust—statistics are used when personal experience is in doubt because the analyst has no intimate knowledge of it. Statistics are consistently used as a technology of the educated elite to discuss the lower classes and subaltern populations, those individuals that are considered unknowable and untrustworthy of delivering their own accounts of their daily life. A demand for statistical proof is blatant distrust of someone's lived experience. The very demand for statistical proof is otherizing because it defines the subject as an outsider, not worthy of the benefit of the doubt.

In short: nothing here is new, and if this analysis is the thing that changes your opinions on racial police activity in Boulder or in America then you're probably disregarding the lived experiences of your fellow humans. Black lives matter.

## About the Data

The original data is broken into two files: `police_stop_data_main_2018.csv` and `police_stop_data_results_2018.csv`. The first dataset contains a row for every individual discretionary stop made by Boulder PD. Let's see what columns the dataset includes:

```
## [1] "stopdate"      "stoptime"      "streetnbr"     "streetdir"
## [5] "street"        "Min"           "sex"           "race"
## [9] "ethnic"        "Year.of.birth" "enfaction"     "rpmainid"
```

We see columns for stop date, time, and duration (in minutes), as well as for race, ethnicity, and sex of the stopped individual **as perceived by the reporting officer**. We also have the column `enfaction`, which indicates if the stopped individual was a resident of Boulder or not, and the column `rpmainid`, which links each row to data in the second dataset.

The structure of `police_stop_data_results_2018.csv` is slightly more complex. This file contains information on the stops listed in the `main` dataset, as well as any potential outcomes of each. Let's see what columns it has:

```
## [1] "appkey"      "appid"      "itemcode" "itemdesc" "addtime"
```

The `appid` column is used to link rows in `results` to rows in `main` (through the `rpmainid` column in `main`). A big difference here is that each value of `rpmainid` appears only once in `main`, whereas a single `appid` might be listed in multiple rows in `results`, corresponding to multiple outcomes of the same stop. The `appkey` column contains one of seven different values, 'RPT1' through 'RPT7', which indicate what kind of information is contained in the row:

Appkey	Data
RPT1	Stop type
RPT2	Stop reason
RPT3	Search conducted
RPT4	Search authority
RPT5	Contraband found
RPT6	Result of stop
RPT7	Charge issued

The corresponding info is then stored in `itemdesc`. Each pair of `appid` and `appkey` might be listed in multiple rows, eg. if there's more than one reason for the stop. This is not a very convenient data structure. It would be far better if all of the information was present in a single dataset, and if each row corresponded to exactly one stop, ie. if the

data was tidy (<https://vita.had.co.nz/papers/tidy-data.pdf>). Fortunately my friend Sam (<https://github.com/samzhang111>) has gone ahead and done just that (<https://github.com/community-data-project/boulder-police-2019-analysis/tree/master/data>). This basically swings the each `appkey` / `infodesc` pair out into its own column, with a 1 in that column indicating that the `appkey` for that stop had the `infodesc` value. For example, if in the un-tidied data one row has an `appkey` of "rpt1" and an `infodesc` of "pedestrian" that gets converted into a "1" under the column `rpt1.pedestrian`. Let's see the first few rows just of the "rpt2" information (which corresponds to the stop reason) to get a feel for what the tidying process produced.

```
## rpt2.disturbance rpt2.equipment.violation rpt2.municipal.violation
## 1 0 0 0
## 2 0 0 0
## 3 0 0 1
## 4 0 0 1
## 5 0 0 1
## 6 0 0 1
## rpt2.noise.violation rpt2.state.violation rpt2.suspicious
## 1 0 0 0
## 2 0 0 0
## 3 0 0 0
## 4 0 0 0
## 5 0 0 0
## 6 0 0 0
## rpt2.traffic.parking.violation rpt2.traffic.reckless.careless
## 1 0 0
## 2 0 1
## 3 0 0
## 4 0 0
## 5 0 0
## 6 0 0
## rpt2.traffic.reddi.observed.pc rpt2.traffic.right.of.way.violation
## 1 0 FALSE
## 2 0 FALSE
## 3 0 FALSE
## 4 0 FALSE
## 5 0 FALSE
## 6 0 FALSE
## rpt2.traffic.speeding rpt2.welfare.check
## 1 TRUE 0
## 2 FALSE 0
## 3 FALSE 0
## 4 FALSE 0
## 5 FALSE 0
## 6 FALSE 0
```

A few of the columns are "TRUE/FALSE" coded instead of "1/0" coded, but this makes no difference, R treats them as the same thing.

## Data Analysis

A useful resource that I relied on while performing this analysis was Methods for Assessing Racially Biased Policing ([https://www.rand.org/content/dam/rand/pubs/reprints/2011/RAND\\_RP1427.pdf](https://www.rand.org/content/dam/rand/pubs/reprints/2011/RAND_RP1427.pdf)) by Ridgeway (RAND Corporation) and MacDonald (University of Pennsylvania). They define two approaches for determining police bias, "benchmark analysis" and "outcome tests". I use both here, beginning with a benchmark analysis of all discretionary stops.

## Assessing Top-Level Bias

The basic logic of a benchmark analysis compares the percentage of stopped individuals of race  $R$  to the underlying demographics of the policed population. If race  $R$  is  $Y\%$  of the population, but comprises  $X\%$  of the police stops, then this is taken to indicate that race  $R$  is over-policed (or vice versa). This approach faces a number of challenges however, primary among being that for a city like Boulder, the total policed population includes not only the census population (ie. "residents"), but also commuters, students, and unhoused folks. The City of Boulder therefore put in a lot of work estimating demographic information about this larger community, and their findings can be found in their 2018 Annual Report (<https://boulder.colorado.gov/police/stop-data-information>). To start I'm just going to pull their population totals and demographic breakdown information and enter it manually, so that I can repeat and confirm their conclusion that Black individuals are stopped more frequently than White individuals. Only race was included in the demographic information, not ethnicity (unlike in the discretionary stop data), so there are no baselines for White or Black Hispanic individuals and therefore are excluded from this first analysis.

Let  $N_R$  denote the number of individuals of race  $R$  in the Boulder policed population. I assume that every such individual has an equal probability of being stopped by BPD, denoted  $p_R$ . This is absolutely a "first-order" assumption and is highly suspect; even given race, an individual's location, behavior, appearance, etc. all likely effect the probability of being stopped by BPD. For each  $R$ , the value  $N_R$  is set equal to the estimate provided by the city of boulder

From the BPD discretionary stop data I calculate  $y_R$ , the number of stops which occurred, where the stopped individual was of race  $R$ . This is simply the number of occurrences of each race category in the `race` column of the tidied data. Note that a single individual may have been stopped multiple times, and in that case would appear as two or more rows in the tidied data. I do not account for multiple occurrences in this, or any model in this analysis. Below are the total number of stops and demographic information for White ('W'), Black ('B'), Asian ('A'), and indigenous ('I') individuals:

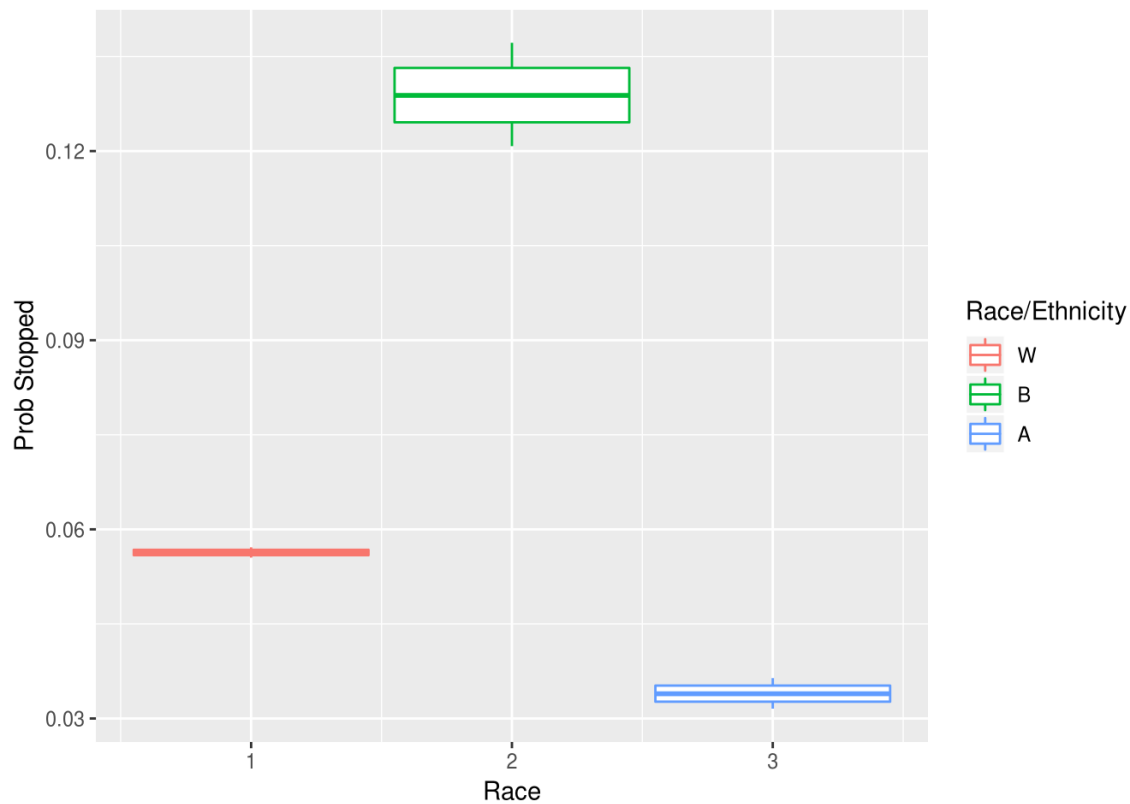
##	Race, Total Stops ( $y_R$ ), Percent Stops, Percent of Pop.,
## 1	A 310 0.04 0.06
## 2	B 353 0.04 0.02
## 3	I 36 0.00 0.01
## 4	W 7425 0.90 0.91
##	Total Pop. ( $N_R$ )
## 1	9154.942
## 2	2743.798
## 3	1114.862
## 4	131818.746

As indicated in the Annual Report, there does appear to be some mismatch between stop rates and demographic representation. But just from this table it's not fully clear if these differences are meaningful, or if they just represent noise. To quantify the uncertainty in the above totals, I perform a basic Bayesian analysis. Given the above  $y_R$  and  $N_R$ , as well as the previous assumptions, we have that  $y_R$  is binomially distributed with parameters  $N_R$  and  $p_R$ . Placing a uniform prior on each of the  $p_R$  then gives a Beta-distributed posterior density over the  $p_R$ :

$$P[p_R | y_R, N_R] \propto p_R^{1+y_R} (1 - p_R)^{1+N_R-y_R}$$

Below are boxplots for the posterior over each of the  $p_R$ :

```
## Using quantiles as value column: use value.var to override.
```



Since the posterior densities exhibit substantial amounts of separation it's clear that there are *serious* racial differences between the probabilities of experiencing a discretionary stop. Asian and White individuals seem to have a similarly, low probability of being stopped, however Black individuals are *substantially* more likely to have been stopped (relative to the population demographics). Let's now dig in on what is contributing to this discrepancy.

## Most Biased Search Reason?

Every stop in the dataset lists the reason the officer initiated the stop. Some of the given reasons seem especially likely to be exhibit differences across races (eg. one stop reason is that the officer was "suspicious"). I therefore repeated the above analysis for subsets of the data corresponding to each different stop reason. I then ranked each stop reason by the number of "extra individuals stopped for that reason". This was calculated as the difference  $(\bar{p}_B - \bar{p}_W)N_B$ , where  $\bar{p}_B$  is the median probability that BPD stopped a Black individual (respectively, White individual). When this index is positive and large it indicates a potential source of the difference in top-level stop rates. Since different populations might be experience different stop reasons (eg. commuters are more likely than residents to be stopped for speeding), I performed separate analyses on the resident and non-resident populations, as well as on both populations together:

```
## [1] "Black-White Bias for all individuals:"
```

```
##              reason    bw.bias
## 1    municipal.violation 78.5473878
## 2          traffic.speeding 42.1522988
## 3    traffic.reckless.careless 22.8549956
## 4          equipment.violation 22.2949612
## 5 traffic.right.of.way.violation 22.2313367
## 6              suspicious 16.9806761
## 7              state.violation  6.2157071
## 8    traffic.reddi.observed.pc  5.8407590
## 9    traffic.parking.violation  2.3882599
## 10             noise.violation  1.5178949
## 11             disturbance  0.9975681
## 12             welfare.check -0.2304922
```

```
## [1] "Black-White Bias for resident stops:"
```

```
##              reason    bw.bias
## 1    municipal.violation 13.6309960
## 2    equipment.violation  8.9655858
## 3          traffic.speeding 7.6798996
## 4    traffic.reckless.careless 6.6324664
## 5 traffic.right.of.way.violation 6.4662227
## 6              suspicious  5.3619983
## 7              state.violation 2.9895774
## 8    traffic.parking.violation 2.5755342
## 9             noise.violation 1.5595054
## 10             welfare.check  1.1640818
## 11    traffic.reddi.observed.pc 0.5396494
## 12             disturbance  0.3666836
```

```
## [1] "Black-White Bias for non-resident stops:"
```

```
##              reason    bw.bias
## 1      municipal.violation 62.8205047
## 2      traffic.speeding 34.2093637
## 3      traffic.reckless.careless 16.9810439
## 4 traffic.right.of.way.violation 16.5651207
## 5      equipment.violation 13.2128490
## 6      suspicious 11.4415417
## 7      traffic.reddi.observed.pc 5.9864338
## 8      state.violation 3.9467150
## 9      disturbance 1.3097782
## 10     noise.violation 0.6371474
## 11     traffic.parking.violation 0.4915556
## 12     welfare.check -0.6740324
```

Almost every stop reason in the dataset exhibits a large and significant anti-Black bias. The `bw.bias` column above lists the discrepancy in stop rates (in units of “stopped individuals”). The bias is largest for non-resident stops, although the leading stop types appear to be similar for both categories (ie. municipal violations and speeding or other traffic violations). The relatively high bias in the “suspicious” stop reason is also telling. That Black individuals viewed with elevated suspicion seems further underlined by the fact that the only time BPD are more likely to stop a White individual is when performing a non-resident “welfare check”. In total, over 220 “extra” Black individuals are policed per year (ie. 220 folks are subject to discretionary stops who would not have been if Black individuals were policed at the same rate as White individuals).

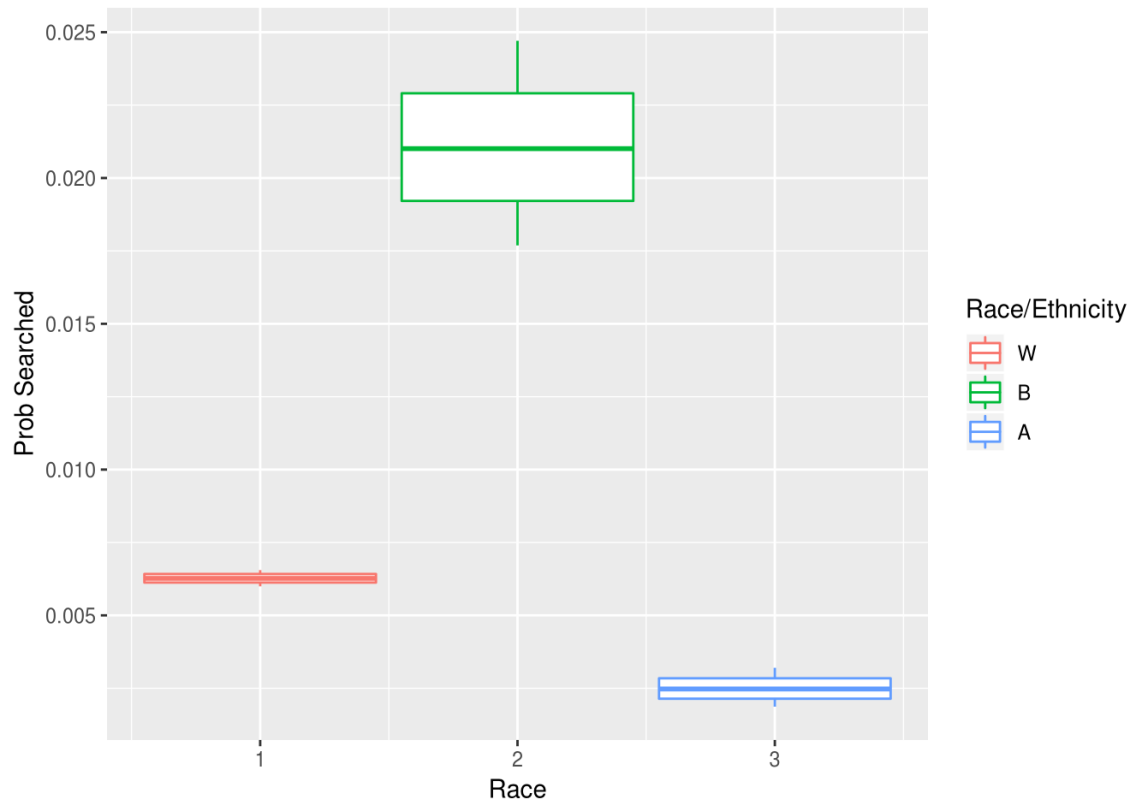
## Outcome Testing Search Results

A problem with benchmark analyses like the above is that racial differences in stop rates may reflect racial differences in the underlying crime rates, rather than racial bias per se. Alternatively, someone might take issue with some of the other assumptions made above. For example, it could be argued that the discrepancy in stop rates may be caused by different levels of police exposure due to the spatial distribution of populations, rather than racial animus in the individual police officers. One way that we can test whether these claims are consistent with the data is by performing an outcome test.

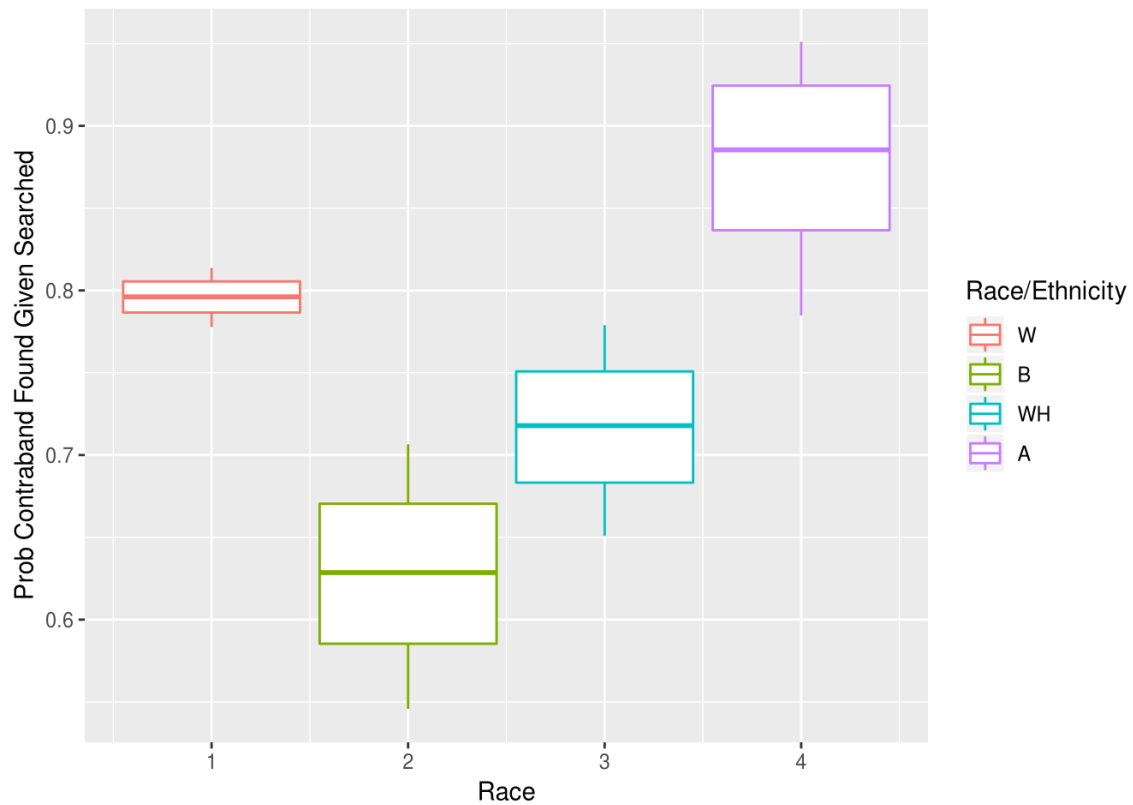
Outcome tests originated in the economics literature to test whether loan officers were discriminating against Black applicants. The idea (in its original use) was to look at whether Black individuals who *did* receive home loans defaulted at a lower rate than their White counterparts. If the data showed that this was the case (and it did), then it indicated that loan officers were holding Black applicants to a higher standard than White applicants. We can apply a similar logic (<https://Sharad.com/papers/frisky.pdf>) here by looking at the rates at which discretionary police searches turn up contraband (the “hit rate”). If Black individuals who are searched are *less* likely to have contraband than White individuals who are searched, it suggests that police are searching Black individuals with a lower threshold of evidence, ie. that racial animus is probably a factor in their decision to search. This test is not perfect (<https://openpolicing.stanford.edu/findings/>), but the case where this test fails still rules out the possibility that heightened policing of Black individuals reflects higher criminality on their part, and suggests that any overpolicing detected in the benchmark test is not warranted by crime trends.

We can apply the same basic modeling framework as in the benchmark test to perform the outcome test. Now, however,  $N_R$  represents the total searches performed on individuals of race  $R$  and  $y_R$  is the number of searches which turned up contraband. Let’s plot the posterior density for both the probability that an individual is searched, and that a search turns up contraband:

```
## Using quantiles as value column: use value.var to override.
```



## Using quantiles as value column: use value.var to override.





Black individuals are the most likely to be searched, but the least likely to carry contraband (among those searched). The discrepancy in outcome holds for both residents and non-residents, and for consent and non-consent searches (for brevity those results are not plotted, but code can be made available on request). Similar to the previous analysis, we can also estimate the number of "extra searches" of Black individuals (ie. the number of searches of non-contraband possessing Black individuals which would not have been performed if Black and White individuals were searched at similar rates):

```
## [1] "Unnecessary searches of Black individuals:"
```

```
## [1] 56.94529
```

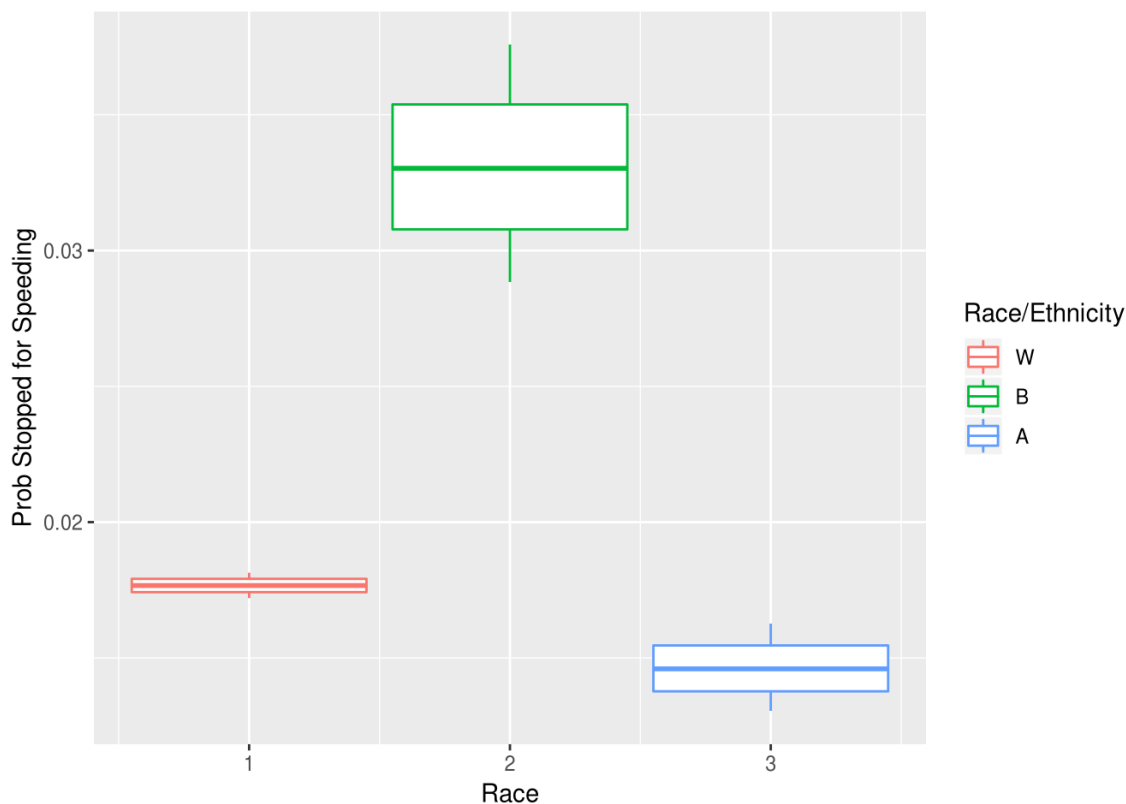
Assuming that White individuals are not more likely to carry contraband than Black individuals, the different standard of evidence with which BPD conducts searches resulted in around 56 "unnecessary" searches of Black individuals. In other words: to get equal rates of contraband discovery between Black and White individuals, BPD would need to have searched 56 fewer innocent, Black individuals.

## Outcome Tests of Speeding Stops

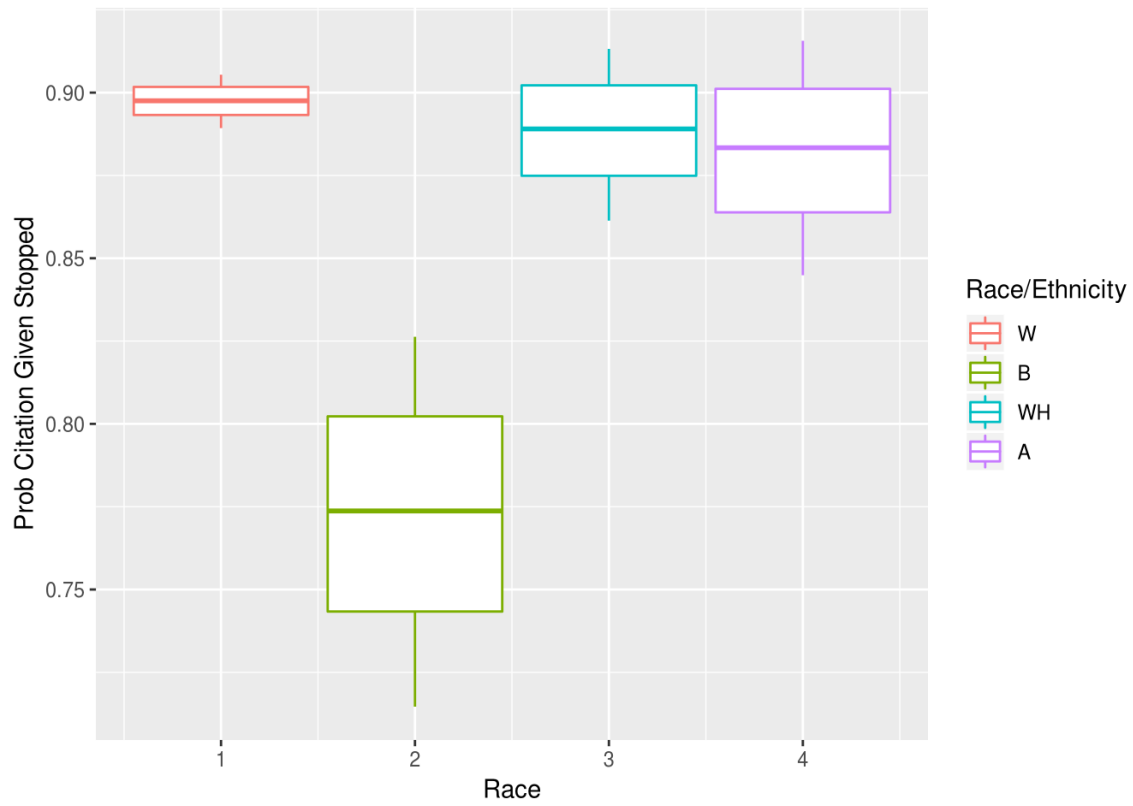
According to my top-level analysis of stop rates, one of the most biased stop reasons was speeding stops. We can apply an outcome test here as well, to assess whether this difference reflects real differences in perpetration rates. I find that, despite being cited less for speeding than other races, Black drivers were arrested more. Furthermore, of those arrested during a speeding stop, Black drivers were ultimately more likely to have the charge overturned.

An outcome test does struggle here a little, because the outcome (the issuance of a speeding citation) is also a decision at the stopping officer's discretion (versus in the case of search outcomes, where the contraband is either present or not). If Black drivers are cited less it could be the case that Boulder PD are more lenient towards Black drivers. Indeed we do see that Black drivers are more likely to receive a warning than other drivers. However it could also be the cases that Black drivers are being stopped at lower driving speeds which don't justify a citation (maybe 1-10 MPH over the limit), which would be more consistent with the other results in this analysis. A more complete dataset could clarify this ambiguity: if we had access to the recorded driving speed of the stopped drivers we could apply the test to those values, rather than just the binary cited/not cited outcome.

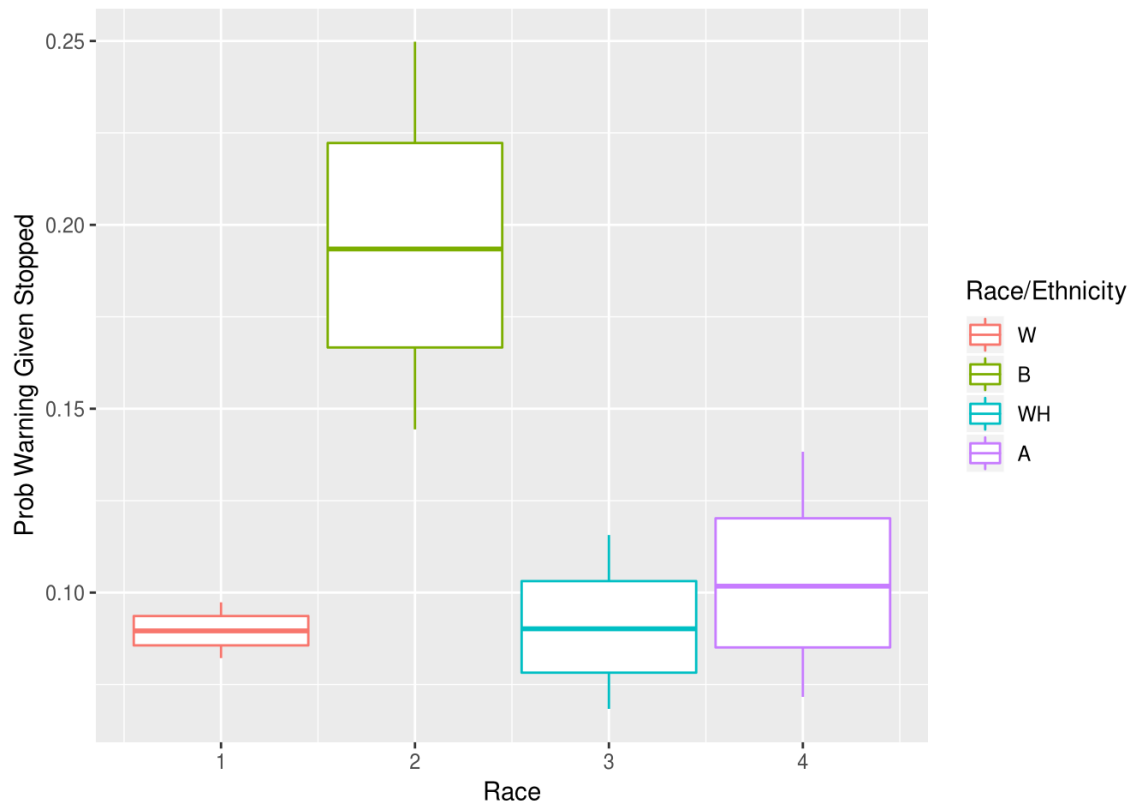
```
## Using quantiles as value column: use value.var to override.
```



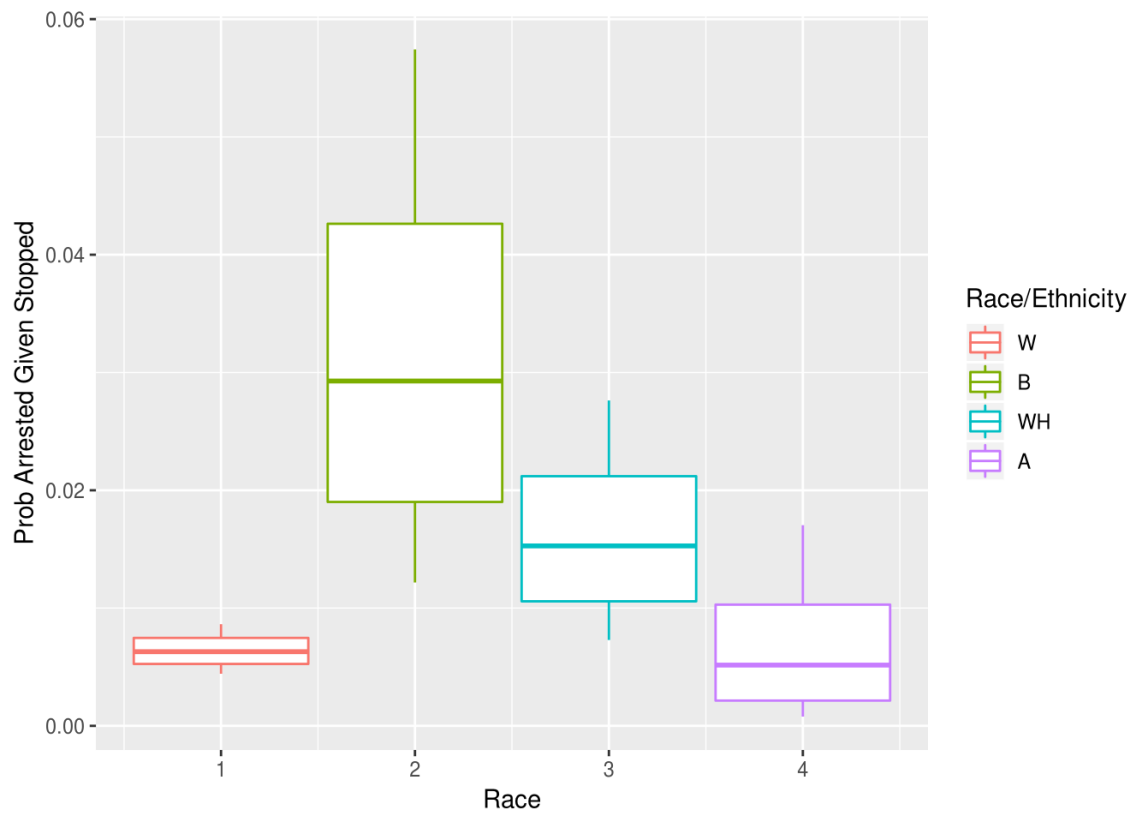
```
## Using quantiles as value column: use value.var to override.
```



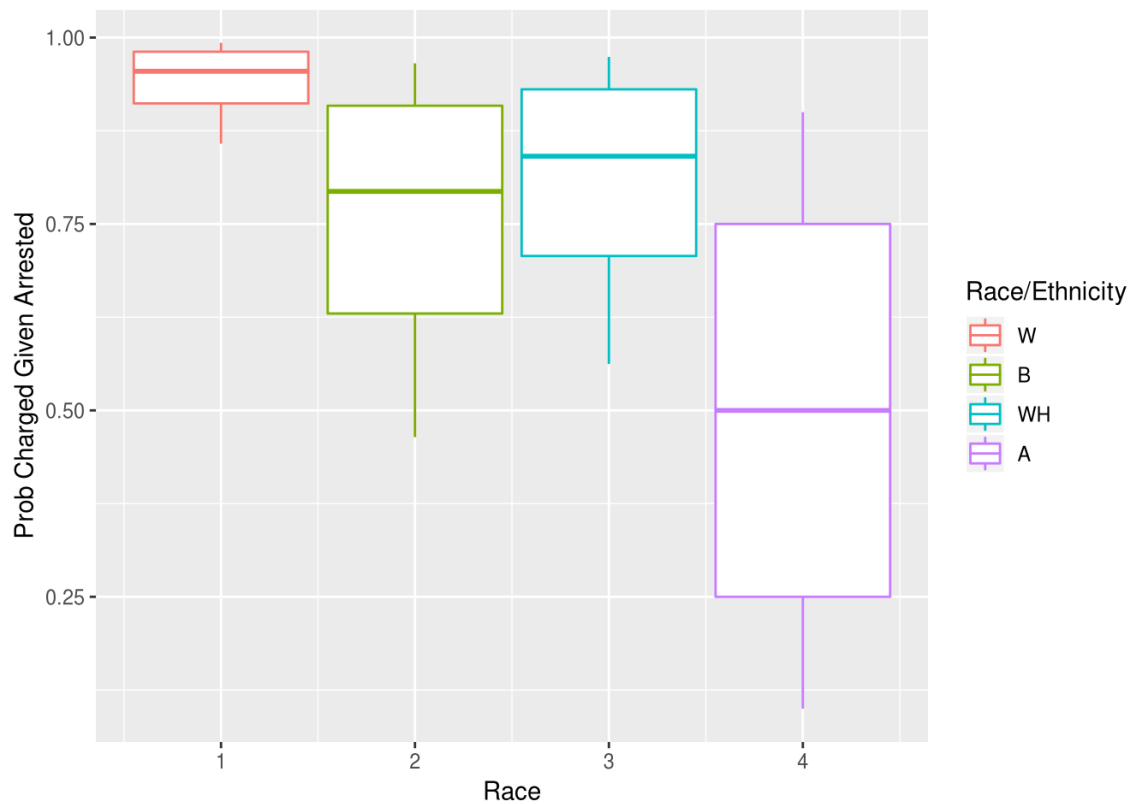
```
## Using quantiles as value column: use value.var to override.
```



```
## Using quantiles as value column: use value.var to override.
```



```
## Using quantiles as value column: use value.var to override.
```



## Conclusion

This analysis is really just a first look at the ways that BPD may be treating its policed population unequally across race. Due to limitations of the dataset we were unable to answer questions regarding police activity towards the unhoused, nor get a complete picture of what's going on with speeding stops. Nevertheless, by almost every measure considered, Black individuals in Boulder are subjected to elevated suspicion and surveillance by the police department. At the most cynical level, this represents an inefficiency in our local government. That this increased police scrutiny results in fewer productive searches and citations indicates that BPD is misallocating its time and resources towards innocent Black people. Far more importantly, however, is the real harm that this data indicates is happening. We have the right against unreasonable search and seizure, and this dataset indicates that in 2018 BPD violated that right among Black individuals who lived or traveled through our community. That the City of Boulder and its citizens are allowing this to occur is entirely unacceptable and any delay in redressing it prolongs a serious injustice.

📡 (<http://petershaffery.com/rss.xml>)

↑ (<http://petershaffery.com/blog>)