

Longitudinal Analysis of Disparities in Marijuana Arrests in the District of Columbia

Rajeev Agrawal

3/17/2021

1. Introduction

Several studies consistently find that young whites use marijuana at higher rates than young Blacks or Hispanics [1][2][3]. Yet, in large cities and counties throughout the United States, young Blacks are arrested and jailed for marijuana possession at much higher rates than young whites. In 2014, Washington, D.C. residents approved Initiative 71, thereby legalizing the recreational use of marijuana. However, racial disparity remains a big issue in Washington, D.C.’s marijuana arrests. In 2017, Congresswoman Eleanor Holmes Norton pointed out, “91 percent of all cannabis-related arrests in D.C. were of Black residents, even though Blacks and whites use marijuana at the same rate.” A report [4] by the Drug Policy Alliance indicated that arrest rates for possession in Washington D.C. dropped about 99 percent for both Black people and white people between 2013 and 2016. But still, it is reported that African Americans were arrested for possession at four times the rate of white Americans. In fact, the racial disparity between African Americans and white Americans that exists in marijuana arrest rates is likely even more pronounced than usually reported when we consider the fact that most Hispanic arrests were likely counted as “white” arrests, meaning that the white arrest rate was artificially inflated.

1.1 Research Questions

The current study investigates the extent of age and racial disparity in Washington, D.C.’s marijuana arrests by using longitudinal analysis. In addition, the study asks: how is racial disparity in marijuana arrests linked to a systematic racial bias?

1.2 Data

This study primarily uses the D.C. crime incident data [5] made available by the D.C. police department. The data for marijuana arrests between 2012 and 2019 can be found on the Metropolitan Police Department (MPD) website in a csv format. The data represents individuals arrested with a marijuana charge, regardless of whether there was a more serious secondary charge. If an arrestee was charged with multiple marijuana charges, the arrest is only counted once under the more serious charge type (Manufacture/Cultivation > Distri-

bution > Possession with Intent to Distribute > Possession > Public Consumption). The category of “Manufacture or Cultivation” was added in the 2019 data and is not utilized in prior years. MPD collects race and ethnicity data according to the United States Census Bureau standards. Hispanic, which was previously categorized under the Race field prior to August 2015, is now captured under Ethnicity. All records prior to August 2015 have been updated to “Unknown (Race), Hispanic (Ethnicity).” Arrestee age is calculated based on the number of days between the self-reported or verified date of birth (DOB) of the arrestee and the date of the arrest; however, DOB data may not be accurate if self-reported or if the arrestee refused to provide it. Due to the sensitive nature of juvenile data and to protect the arrestee’s confidentiality, any arrest records for defendants under the age of 18 have been coded as “NA” for fields such as Age, Defendant Race, Defendant Ethnicity and Defendant Sex. MPD Adult Arrests (2013-2019) data includes reports entered into the MPD’s Record Management System (Cobalt) whereby an adult arrest occurred between January 1, 2013 - December 31, 2019. Totals are based on the most serious arrest charge (i.e., each row denotes one arrest, not one charge), and one person may be booked on more than one arrest charge. Separate csv files are available on the MPD website for different years ranging from 2013 to 2019.

There are 21 fields in the MPD Marijuana Arrests dataset with 12221 records, while there are 26 fields in each of the 7 MPD Adult Arrests datasets with 210740 total records. Some of the fields in the MPD datasets include:

- **Categorical variables**

- *Marijuana_Arrest Type*: Type of marijuana arrest; for example, for public consumption, possession, distribution, cultivation etc.
- *Adult/Juvenile*: If the arrestee is an adult or a juvenile.
- *CCN (Criminal Complaint Number)*: A number assigned to the criminal complaint against which the arrest was made.
- *Arrest Year*: The year when the arrest was made.

- *Arrest Date*: The date when the arrest was made.
- *Arrest Hour*: The hour of arrest.
- *Arrest Number*: A number assigned by the police for the arrest made.
- *Defendant Race*: Race of the arrestee.
- *Defendant Ethnicity*: Ethnicity of the arrestee; for example, Hispanic, non-Hispanic or unknown.
- *Defendant Sex*: Sex of the arrestee.
- *Defendant PSA*: Police service area where the defendant lives.
- *Defendant District*: Police district where the defendant lives.
- *Arrest Category*: Type of arrest; for example, narcotics, simple assault, damage to property, theft, traffic violations, robbery, homicide etc.
- *Arrest Location PSA*: Police service area where the arrest was made.
- *Arrest Location District*: Police district where the arrest was made.
- *Offense Location PSA*: Police service area where the offense was committed.
- *Offense Location District*: Police district where the offense was committed.

- **Quantitative variables**

- *Arrest Time*: Time when the arrest was made.
- *Age*: Age of the arrestee.
- *Arrest Latitude*: The latitude location where the arrest was made.
- *Arrest Longitude*: The longitude location where the arrest was made.
- *Offense Latitude*: The latitude location where the offense was committed.
- *Offense Longitude*: The longitude location where the offense was committed.

- **Other**

- *Charge Description*: Description of the reason(s) for the arrest.

Data regarding the relative ethno-racial composition by year in the District of Columbia is sourced from the US Census Bureau [6]. It is found in the MS excel format (Fig. 1).

	A	B	C	D	E	F	G	H	I	J	K	L	M
2	Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for District of Columbia: April 1, 2010 to July 1, 2019												
3	Sex, Race, and Hispanic Origin	April 1, 2010		Population Estimate (as of July 1)									
4		Census	Estimates Base	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
14	White	262,139	262,167	264,625	275,058	285,587	294,771	302,863	312,267	320,245	328,038	334,390	340,498
15	Black or African American	319,177	319,192	319,969	323,370	327,137	332,496	335,384	337,835	339,218	339,654	339,385	337,049
16	American Indian and Alaska Native	8,084	8,084	8,117	8,355	8,611	9,107	9,300	9,463	9,626	9,687	9,733	9,862
17	Asian	27,369	27,371	27,720	28,951	30,302	31,814	32,929	34,650	36,213	37,661	38,628	39,586
18	Native Hawaiian and Other Pacific Islander	1,514	1,514	1,554	1,621	1,681	1,746	1,799	1,854	1,860	1,866	1,867	1,871
19	NOT HISPANIC	546,914	546,999	549,379	560,051	571,844	583,817	592,787	602,819	611,170	618,643	623,556	626,272
20	One Race:												
21	White	210,389	210,397	211,946	218,971	226,093	231,955	237,317	243,699	249,422	255,398	260,081	264,400
22	Black or African American	302,051	302,066	302,598	304,905	307,938	312,337	314,555	316,378	317,187	316,992	316,103	313,290
23	American Indian and Alaska Native	1,342	1,342	1,345	1,326	1,401	1,447	1,396	1,362	1,414	1,398	1,418	1,413
24	Asian	21,151	21,152	21,333	22,126	23,064	24,144	25,136	26,576	27,818	29,009	29,808	30,541
25	Native Hawaiian and Other Pacific Islander	220	220	230	229	237	252	291	301	327	346	345	339

Figure 1: Snapshot of the excel file sourced from US Census Bureau with D.C.'s annual population estimates by race and ethnicity.

1.3 Literature Review

The research article by C. L. Firth et al. [7] studies the changes in adult marijuana arrest rates and disparities in rates for African Americans in Washington State (WA) after legalization of possession of small amounts of marijuana for 21+ year olds in December 2012, and after the marijuana retail market opening in July 2014. The ACLU research report, A Tale of Two Countries: Racially Targeted Arrests in the Era of Marijuana Reform [8], details marijuana arrests from 2010 to 2018 and examines racial disparities at the national, state, and county levels. Updating their previous report, The War on Marijuana in Black and White, that examined arrests from 2000 to 2010, the latest ACLU report reveals that more than six million arrests occurred between 2010 and 2018, and Black people are still more likely to be arrested for marijuana possession than white people in every state, including those that have legalized marijuana.

2. Initial Hypotheses

- *Hypothesis 1:* Young adults (ages 18 – 35) are more likely to get arrested for marijuana offenses than older adults in Washington, D.C.
- *Hypothesis 2:* There is ethno-racial disparity in marijuana arrests in Washington, D.C.
- *Hypothesis 3:* After the marijuana use was legalized in Washington, D.C. in 2015, the total number of marijuana arrests was reduced; however, ethno-racial disparity in marijuana arrests persists.

- *Hypothesis 4*: There is an association between ethno-racial disparity in marijuana arrests in Washington, D.C. and a systematic racial bias.

3. Exploratory Data Analysis

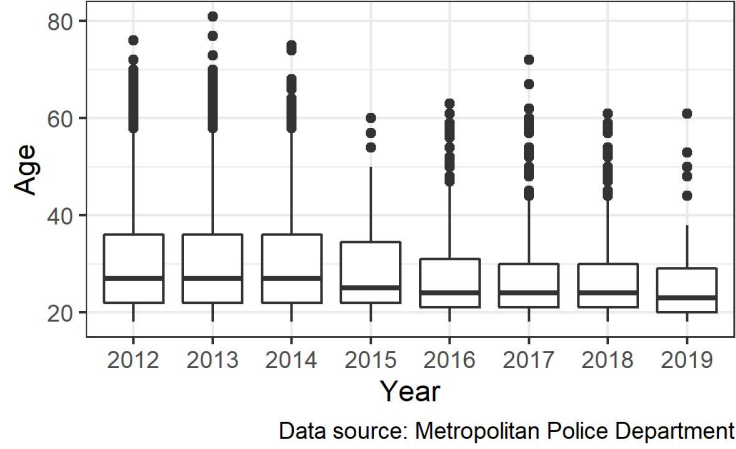
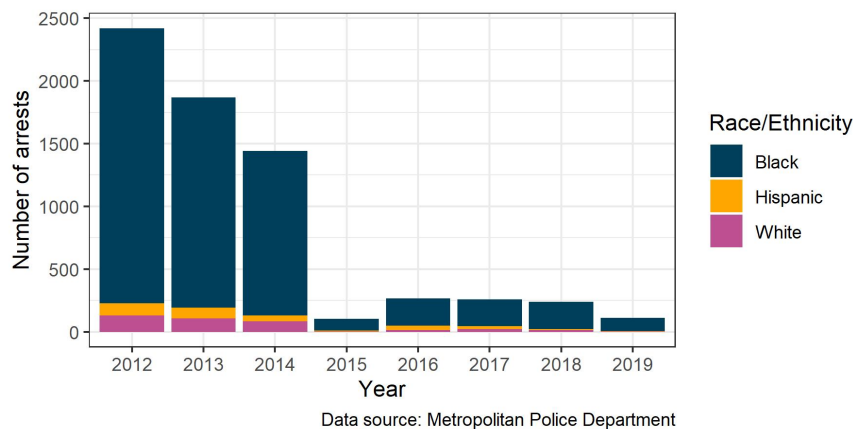


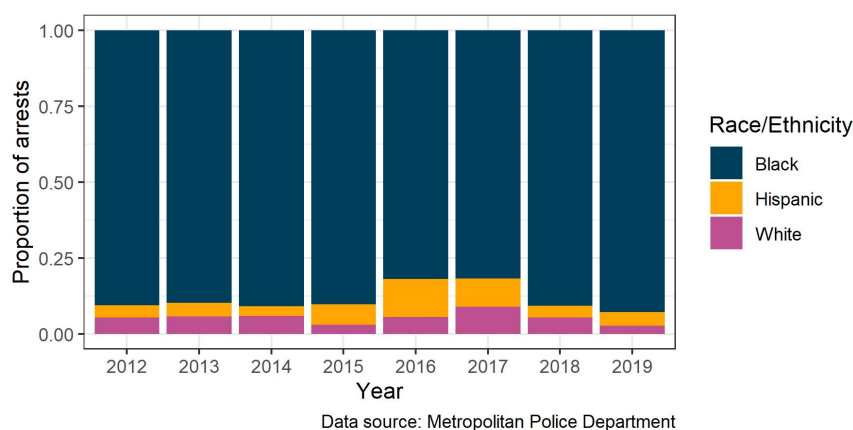
Figure 2: D.C.'s marijuana arrestees' age distribution over the years

As evident from Fig. 2, the average median age of D.C.'s marijuana arrestees is 27 years and most arrestees are below 40 years of age. From Fig. 3(a), it can be seen that the number of arrests for marijuana possession/public consumption came down significantly after 2014, when marijuana consumption for adults was legalized in D.C. However, as evident from Fig. 3(b) and 3(c), the proportion of marijuana-related arrests and arrest rates for African Americans, respectively, has not declined.

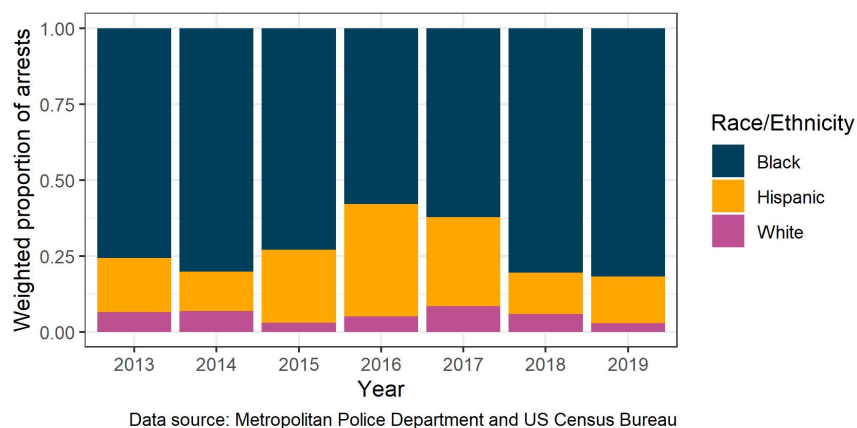
In order to explore the question of systematic racial bias against African Americans in marijuana arrests, we use the arrests made by the D.C. police for serious crimes, such as homicide, robbery and burglary, as our baseline. It is assumed that there is much lower likelihood of any ethno-racial bias in arrests made by police when it comes to serious crimes, such as homicide, robbery and burglary, which is why it has been used as the baseline in our current study to compare disparity in marijuana-related arrests with. From Fig. 4(a), it is evident that the number of arrests for serious crimes, such as homicide, robbery and burglary, has been slowly decreasing over the years. However, we can see from Fig. 4(b),



(a)

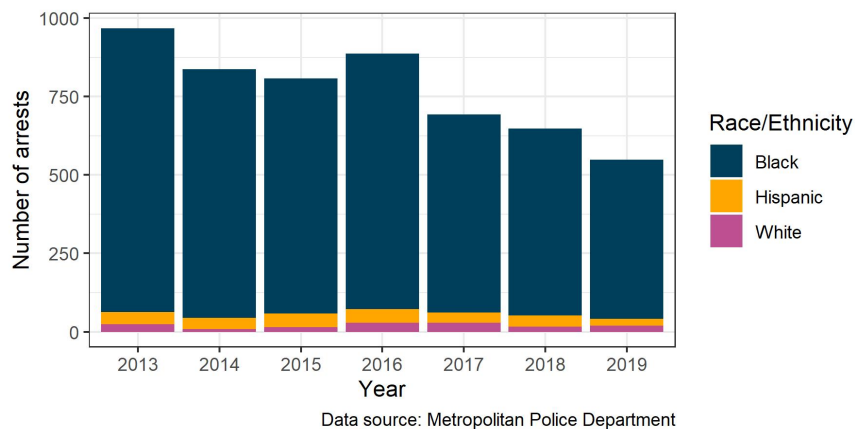


(b)

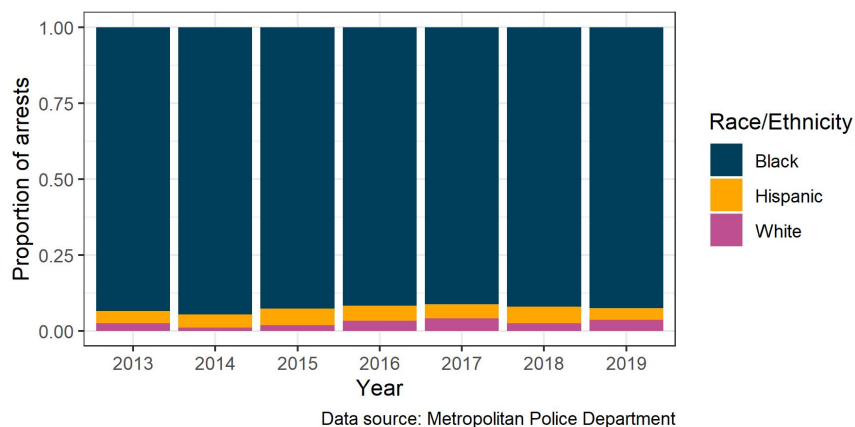


(c)

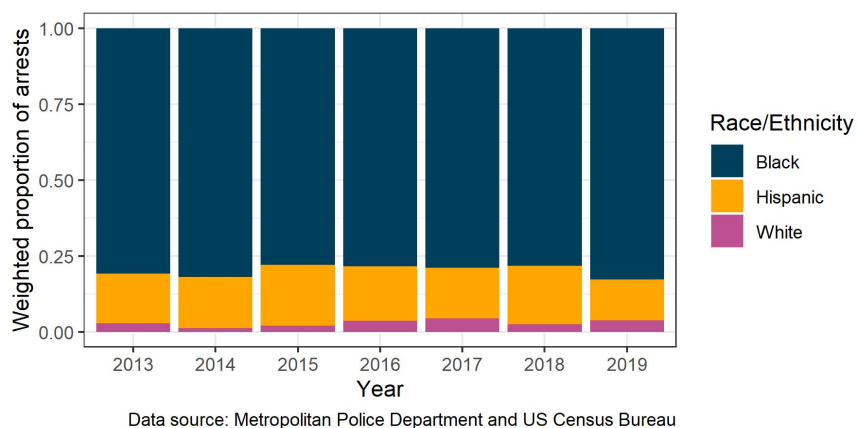
Figure 3: (a) Number of marijuana arrests for possession/consumption in D.C. for different ethno-racial groups over the years. (b) Relative ethno-racial composition of marijuana arrests for possession/consumption in D.C. over the years. (c) Relative ethno-racial composition of marijuana arrests for possession/consumption in D.C. weighted by respective ethno-racial populations over the years.



(a)



(b)



(c)

Figure 4: (a) Number of arrests for homicide, robbery and burglary in D.C. for different ethno-racial groups over the years. (b) Relative ethno-racial composition of arrests for homicide, robbery and burglary in D.C. over the years. (c) Relative ethno-racial composition of arrests for homicide, robbery and burglary in D.C. weighted by respective ethno-racial populations over the years.

that African Americans make up disproportionately higher proportion of arrests involving serious crimes such as homicide, robbery and burglary. Similarly, Fig. 4(c), which compares arrest rates for serious crimes per 100k residents for different ethno-racial groups, shows that serious crime related arrest rates are much higher for Blacks.

Further, to test *Hypothesis 4*, we define our dependent variable as the relative Black-nonBlack disparity in arrest rates in D.C. for marijuana possession/public consumption. Similarly, our independent variable is the ratio of the Black arrest rate to the nonBlack arrest rate related to D.C.'s serious crimes. We fit a linear regression model (Fig. 5) to get the correlation coefficient between the two variables for years 2013 - 2019 of about 0.49 (95% confidence interval: -0.28 to 1) with one-sided p-value of 0.134.

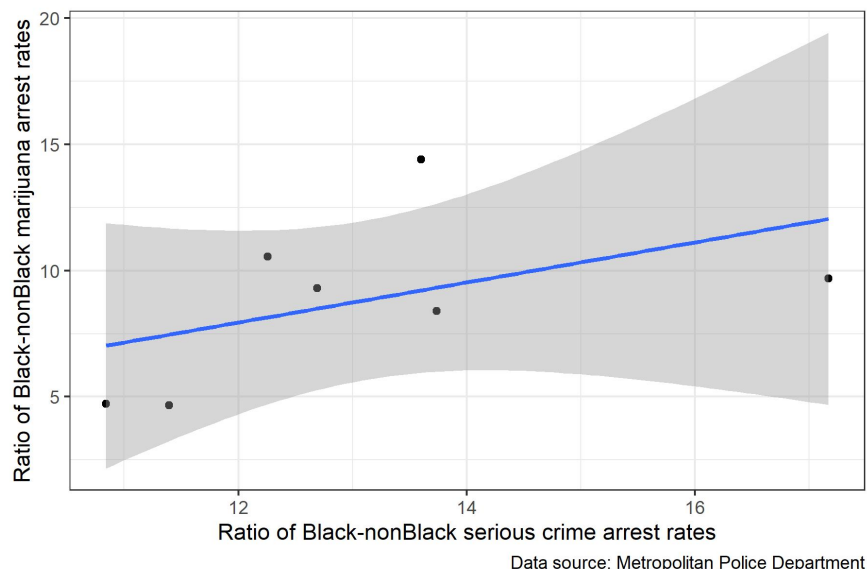


Figure 5: Relative Black-nonBlack arrest rates in D.C. for serious crime vs. Relative Black-nonBlack arrest rates in D.C. for marijuana possession/public consumption.

4. Data-driven Hypothesis

From the exploratory data analysis of the D.C. crime incident data, there is evidence in support of *Hypotheses 1, 2* and *3*. For *Hypothesis 4*, since the one-sided p-value is greater than 0.05, the data provides inconclusive evidence of there being a positive correlation between relative Black-nonBlack disparity in marijuana arrest rates in D.C. and relative Black-nonBlack

disparity in arrest rates related to D.C.'s serious crimes. Therefore, we should test *Hypothesis 4* using some other approach that can provide us with a more conclusive answer. We may also use other predictor variables such as racial segregation, family background, education, fatal drug overdose and arrest location.

5. Discussion

If the claim made by various studies [1][2][3] that young Blacks use marijuana at equal if not lower rates than young whites is assumed to be true, then a case for the presence of conscious or subconscious racial bias can be made if a positive correlation can be found between the relative Black-nonBlack disparity in D.C.'s marijuana arrest rates and D.C.'s serious crime arrest rates, respectively. However, a correlation coefficient only measures the strength of the relationship between the relative movements of two variables. Racial bias is based on perception, which is relatively static and may not move swiftly with small changes in the ratio of Black-nonBlack arrest rates for serious crime. One possible explanation for racial bias is that, because Blacks have disproportionately higher involvement in serious crimes as compared to other ethno-racial groups, there is a negative public perception of the group as whole. This negative public image has also seeped into institutions such as the police department, which is why police might have a higher inclination to arrest African Americans for even less serious crimes such as marijuana possession/public consumption.

The perception that Blacks make up almost 90% of the marijuana arrests made in Washington, D.C. is supported by the data. It is also found that when Hispanics and whites are treated as separate ethno-racial groups, which is done in the current study, the gulf between African Americans and white Americans regarding marijuana arrest rates is much more pronounced than usually reported. Another interesting outcome is that even though marijuana legalization may help to reduce existing disparities in marijuana-related arrests for African Americans in the long run, in the near term, the exploratory data analysis suggests no decrease in existing disparities.

6. References

1. Substance Abuse and Mental Health Services Administration. Overview of Findings from the 2003 National Survey on Drug Use and Health (Office of Applied Studies, NSDUH Series H-24, DHHS Publication No. SMA 04-3963) Rockville, MD: 2004a.
2. Johnson RM, Fairman B, Gilreath T, Xuan Z, Rothman EF, Parnham T, Furr-Holden. Past 15-year trends in adolescent marijuana use: Differences by race/ethnicity and sex. *CD Drug Alcohol Depend.* 2015 Oct 1; 155():8-15.
3. Pacek LR, Malcolm RJ, Martins SS. Race/ethnicity differences between alcohol, marijuana, and co-occurring alcohol and marijuana use disorders and their association with public health and social problems using a national sample. *Am J Addict.* 2012 Sep-Oct; 21(5):435-44.
4. Drug Policy Alliance. Report: From Prohibition to Progress: A Status Report on Marijuana Legalization. 2018. Available at https://drugpolicy.org/sites/default/files/dpa_marijuana_legalization_report_feb14_2018_0.pdf [accessed Feb. 2021].
5. D.C. Crime incident data from the Metropolitan Police Department. Available at <https://mpdc.dc.gov/page/statistics-and-data> [accessed Feb. 2021]
6. Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for District of Columbia: April 1, 2010 to July 1, 2019 (SC-EST2019-SR11H-11). U.S. Census Bureau, Population Division. June 2020. Available at <https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html> [accessed Feb. 2021]
7. Firth, Caislin & Maher, Julie & Dilley, Julia & Darnell, Adam & Lovrich, Nicholas. Did marijuana legalization in Washington State reduce racial disparities in adult marijuana arrests?. *Substance Use & Misuse.* 2019. 54. 1-6. 10.1080/10826084.2019.1593007.
8. American Civil Liberties Union. Report: A Tale of Two Countries: Racially Targeted Arrests in the Era of Marijuana Reform. 2020. Available at <https://www.aclu.org/report/tale-two-countries-racially-targeted-arrests-era-marijuana-reform> [accessed Feb. 2021].

7. Appendix

Main R code chunk from the R Markdown file.

```
source("../code/tidyData1.R")
source("../code/tidyData2.R")
source("../code/tidyData3.R")

#Numerical Summary - Median Age
Median_age <- median(mar_possess$Age)

#Box plots of yearly arrests as a function of age
mar_possess %>%
  ggplot() + geom_boxplot(mapping = aes(Arrest.Year, Age)) +
  labs(caption = "Data source: Metropolitan Police Department",
       x = "Year", y = "Age") +
  theme_bw() +
  guides(fill=guide_legend(title="Race/Ethnicity")) -> p0
ggsave("../output/plot0.jpeg", width = 4, height = 2.5)

#Colors for the bar charts
ecols <- c("#003f5c", "#ffa600", "#bc5090")

#Number of yearly marijuana arrests for each racio-ethnic group
mar_possess %>%
  ggplot() +
  geom_bar(aes(Arrest.Year, fill = Race.Ethnicity),
           position = "stack") +
  scale_fill_manual(values = ecols) +
  labs(caption = "Data source: Metropolitan Police Department",
       x = "Year", y = "Number of arrests") +
  theme_bw() +
  guides(fill=guide_legend(title="Race/Ethnicity")) -> p1
ggsave("../output/plot1.jpeg", width = 6, height = 3)

#Proportion of yearly marijuana arrests for each racio-ethnic group
mar_possess %>%
  ggplot() +
  geom_bar(aes(Arrest.Year, fill = Race.Ethnicity),
           position = "fill") +
  scale_fill_manual(values = ecols) +
  labs(caption = "Data source: Metropolitan Police Department",
       x = "Year", y = "Proportion of arrests") +
  theme_bw() +
  guides(fill=guide_legend(title="Race/Ethnicity")) -> p2
ggsave("../output/plot2.jpeg", width = 6, height = 3)

#Correct for the population proportion using JOIN
mar_possess <- inner_join(mar_possess, popul_dc, by = c("Race.Ethnicity", "Arrest.Year" = "Year"))
mar_possess <- mar_possess %>%
  group_by(Arrest.Year, Race.Ethnicity) %>%
  mutate(Weight = 1e+05/Population, Arrest.Rate = Count.Arrests*Weight) %>%
  ungroup()

mar_possess %>%
  ggplot() +
  geom_bar(aes(Arrest.Year, Weight, fill = Race.Ethnicity),
           position = "fill", stat = "identity") +
  scale_fill_manual(values = ecols) +
  labs(caption = "Data source: Metropolitan Police Department and US Census Bureau",
       x = "Year", y = "Weighted proportion of arrests") +
  theme_bw() +
  guides(fill=guide_legend(title="Race/Ethnicity")) -> p3
ggsave("../output/plot3.jpeg", width = 6, height = 3)

#Yearly arrests for serious crime for each racio-ethnic group
arrests_data %>%
  ggplot() +
```

```

geom_bar(aes(Arrest.Year, fill = Race.Ethnicity),
  position = "stack") +
scale_fill_manual(values = ecol) +
labs(caption = "Data source: Metropolitan Police Department",
  x = "Year", y = "Number of arrests") +
theme_bw() +
guides(fill=guide_legend(title="Race/Ethnicity")) -> p4
ggsave("../output/plot4.jpeg", width = 6, height = 3)

#Proportion of yearly arrests for serious crime for each racio-ethnic group
arrests_data %>%
  ggplot() +
  geom_bar(aes(Arrest.Year, fill = Race.Ethnicity), position = "fill") +
  scale_fill_manual(values = ecol) +
  labs(caption = "Data source: Metropolitan Police Department",
    x = "Year", y = "Proportion of arrests") +
  theme_bw() +
  guides(fill=guide_legend(title="Race/Ethnicity")) -> p5
ggsave("../output/plot5.jpeg", width = 6, height = 3)

#Correct for the population proportion using JOIN
arrests_data <- inner_join(arrests_data, popul_dc, by = c("Race.Ethnicity", "Arrest.Year" = "Year"))
arrests_data <- arrests_data %>%
  group_by(Arrest.Year, Race.Ethnicity) %>%
  mutate(Weight = 1e+05/Population, Arrest.Rate = Count.Arrests*Weight) %>%
  ungroup()

arrests_data %>%
  ggplot() +
  geom_bar(aes(Arrest.Year, Weight, fill = Race.Ethnicity),
    position = "fill", stat = "identity") +
  labs(caption = "Data source: Metropolitan Police Department and US Census Bureau",
    x = "Year", y = "Weighted proportion of arrests") +
  theme_bw() +
  guides(fill=guide_legend(title="Race/Ethnicity")) +
  scale_fill_manual(values = ecol) -> p6
ggsave("../output/plot6.jpeg", width = 6, height = 3)

#Correlation
mar_possess %>%
  group_by(Arrest.Year) %>%
  select(Arrest.Year, Race.Ethnicity, Arrest.Rate, Count.Arrests, Population) %>%
  unique() %>%
  mutate(ArrestRate.Ratio = Arrest.Rate/(1e+05*(sum(Count.Arrests)-Count.Arrests[Race.Ethnicity == "Black"])/
    (sum(Population)-Population[Race.Ethnicity == "Black"]))) %>%
  filter(Race.Ethnicity == "Black") -> mar_possess2

arrests_data %>%
  group_by(Arrest.Year) %>%
  select(Arrest.Year, Race.Ethnicity, Arrest.Rate, Count.Arrests, Population) %>%
  unique() %>%
  mutate(ArrestRate.Ratio = Arrest.Rate/(1e+05*(sum(Count.Arrests)-Count.Arrests[Race.Ethnicity == "Black"])/
    (sum(Population)-Population[Race.Ethnicity == "Black"]))) %>%
  filter(Race.Ethnicity == "Black") -> arrests_data2

inner_join(arrests_data2, mar_possess2, by = c("Arrest.Year", "Race.Ethnicity"), suffix = c(".Crime", ".Mar")) -> arrests_data2

cor.test(arrests_data2$ArrestRate.Ratio.Crime,
  arrests_data2$ArrestRate.Ratio.Mar, alternative = "greater") -> t1

arrests_data2 %>%
  ggplot(aes(ArrestRate.Ratio.Crime, ArrestRate.Ratio.Mar)) + geom_point() + geom_smooth(method = lm) +
  labs(caption = "Data source: Metropolitan Police Department",
    y = "Ratio of Black-nonBlack marijuana arrest rates",
    x = "Ratio of Black-nonBlack serious crime arrest rates") +
  theme_bw() -> p7
ggsave("../output/plot7.jpeg", width = 6, height = 4)

```

tidyData1.R to tidy marijuana arrests data

```
#' @author Rajeev Agrawal
#' @description STAT612 - Final Project - R code for tidying the marijuana arrests data set
library(tidyverse)
library(mosaic)
library(lubridate)

mar_data <- read.csv("../data/MarijuanaArrestData_2012_2019.csv", header = T)

#Drop variables not used in the study
mar_data <- mar_data %>%
  select(-c(CCN, Defendant.PSA, Defendant.District, Arrest.Category, Charge.Description, Arrest.Time)) %>%
  select(-contains("Location")) %>%
  select(-contains("Adult"))

names(mar_data) #Get the names of the remaining variables

#Filter to retrieve records with only numeric ages
mar_data <- mar_data %>%
  filter(Age != "Unk", Age != "Juvenile", !is.na(Age))
mar_data$Age <- as.numeric(mar_data$Age)

#Create a new variable Race.Ethnicity
mar_data <- mar_data %>%
  mutate(Race.Ethnicity = case_when(Defendant.Ethnicity == "H" |
    Defendant.Ethnicity == "H " ~ Defendant.Ethnicity,
    TRUE ~ Defendant.Race))

unique(mar_data$Race.Ethnicity)

#Retrieve records for blacks, whites and Hispanics
mar_data <- mar_data %>%
  filter(Race.Ethnicity != "U", Race.Ethnicity != "Juvenile", Race.Ethnicity != "A ",
    Race.Ethnicity != "A", Race.Ethnicity != "P", !is.na(Race.Ethnicity))

#Make the data consistent in Race.Ethnicity column
mar_data$Race.Ethnicity[mar_data$Race.Ethnicity == "H "] <- "Hispanic"
mar_data$Race.Ethnicity[mar_data$Race.Ethnicity == "H"] <- "Hispanic"
mar_data$Race.Ethnicity[mar_data$Race.Ethnicity == "B "] <- "Black"
mar_data$Race.Ethnicity[mar_data$Race.Ethnicity == "B"] <- "Black"
mar_data$Race.Ethnicity[mar_data$Race.Ethnicity == "W "] <- "White"
mar_data$Race.Ethnicity[mar_data$Race.Ethnicity == "W"] <- "White"
unique(mar_data$Race.Ethnicity)

#Drop variables Defendant.Race and Defendant.Ethnicity
mar_data <- mar_data %>%
  select(-c(Defendant.Race, Defendant.Ethnicity))

mar_data$Arrest.Year <- as.character(mar_data$Arrest.Year)

unique(mar_data$Marijuana_Arrest.Type)

#Retrieve records for Possession or Consumption
mar_possess <- mar_data %>%
  filter(Marijuana_Arrest.Type != "Distribution",
    Marijuana_Arrest.Type != "Manufacture",
    Marijuana_Arrest.Type != "Cultivation",
    Marijuana_Arrest.Type != "Possession with intent to distribute")

unique(mar_possess$Marijuana_Arrest.Type)

#Add a new column for number of arrests by year and race
#Add another column for percentage arrests by race
#Add column for the arrest month
mar_possess <- mar_possess %>%
  group_by(Arrest.Year, Race.Ethnicity) %>%
  mutate(Count.Arrests = n()) %>%
```

```

ungroup() %>%
group_by(Arrest.Year) %>%
mutate(Percentage.Arrests = Count.Arrests*100/n()) %>%
ungroup() %>%
mutate(Arrest.Month = as.factor(month(as.POSIXlt(mar_possess$Arrest.Date,      format="%m/%d/%Y"))))

# Set ordering manually
mar_possess$Race.Ethnicity <- mar_possess$Race.Ethnicity %>%
  factor(levels = c("Black", "Hispanic", "White"))

#Remove the original data objects from the environment
rm(mar_data)

```

tidyData2.R to tidy DC adult arrest data

```

#' @author Rajeev Agrawal
#' @description STAT612 - Final Project - R code for tidying the D.C. adult arrest data from 2013-19
library(tidyverse)
library(mosaic)
library(lubridate)

#Select only required columns from the respective data sets
arrests2013 <- read.csv("../data/Arrests2013Public.csv", header = T) %>%
  select(c(Arrest.Year, Arrest.Date, Age, Defendant.Race, Defendant.Ethnicity, Defendant.Sex, Arrest.Category))
arrests2014 <- read.csv("../data/Arrests2014Public.csv", header = T) %>%
  select(c(Arrest.Year, Arrest.Date, Age, Defendant.Race, Defendant.Ethnicity, Defendant.Sex, Arrest.Category))
arrests2015 <- read.csv("../data/Arrests2015Public.csv", header = T) %>%
  select(c(Arrest.Year, Arrest.Date, Age, Defendant.Race, Defendant.Ethnicity, Defendant.Sex, Arrest.Category))
arrests2016 <- read.csv("../data/Arrests2016Public.csv", header = T) %>%
  select(c(Arrest.Year, Arrest.Date, Age, Defendant.Race, Defendant.Ethnicity, Defendant.Sex, Arrest.Category))
arrests2017 <- read.csv("../data/Arrests2017Public.csv", header = T) %>%
  select(c(Arrest.Year, Arrest.Date, Age, Defendant.Race, Defendant.Ethnicity, Defendant.Sex, Arrest.Category))
arrests2018 <- read.csv("../data/Arrests2018Public.csv", header = T) %>%
  select(c(Arrest.Year, Arrest.Date, Age, Defendant.Race, Defendant.Ethnicity, Defendant.Sex, Arrest.Category))
arrests2019 <- read.csv("../data/Arrests2019Public.csv", header = T) %>%
  select(c(Arrest.Year, Arrest.Date, Age, Defendant.Race, Defendant.Ethnicity, Defendant.Sex, Arrest.Category))

#Add data sets vertically
arrests_data <- rbind(arrests2013, arrests2014, arrests2015, arrests2016, arrests2017, arrests2018, arrests2019)

#Remove the original data objects from the environment
rm(arrests2013, arrests2014, arrests2015, arrests2016, arrests2017, arrests2018, arrests2019)

#Create a new variable Race.Ethnicity
arrests_data <- arrests_data %>%
  mutate(Race.Ethnicity = case_when(Defendant.Ethnicity == "HISPANIC" ~ Defendant.Ethnicity, TRUE ~ Defendant.Race))

#Retrieve records for blacks, whites and Hispanics
arrests_data <- arrests_data %>%
  filter(Race.Ethnicity != "UNK", Race.Ethnicity != "ASIAN", Race.Ethnicity != "UNKNOWN", !is.na(Race.Ethnicity))

#Make data consistent in Race.Ethnicity column
arrests_data$Race.Ethnicity[arrests_data$Race.Ethnicity == "HISPANIC"] <- "Hispanic"
arrests_data$Race.Ethnicity[arrests_data$Race.Ethnicity == "BLACK"] <- "Black"
arrests_data$Race.Ethnicity[arrests_data$Race.Ethnicity == "WHITE"] <- "White"
unique(arrests_data$Race.Ethnicity)

#Drop variables Defendant.Race and Defendant.Ethnicity
arrests_data <- arrests_data %>%
  select (-c(Defendant.Race, Defendant.Ethnicity))

arrests_data$Arrest.Year <- as.character(arrests_data$Arrest.Year)

#Retrieve records for Serious/Violent Crime
arrests_data <- arrests_data %>%
  filter(Arrest.Category == "Homicide" |
         Arrest.Category == "Robbery" |

```

```

    Arrest.Category == "Burglary")

#Add a new column for number of arrests by year and race
#Add another column for percentage arrests by race
#Add column for the arrest month
arrests_data <- arrests_data %>%
  group_by(Arrest.Year, Race.Ethnicity) %>%
  mutate(Count.Arrests = n()) %>%
  ungroup() %>%
  group_by(Arrest.Year) %>%
  mutate(Percentage.Arrests = Count.Arrests*100/n()) %>%
  ungroup() %>%
  mutate(Arrest.Month = as.factor(case_when(Arrest.Year >= 2018
    ~month(as.POSIXlt(arrests_data$Arrest.Date, format = "%m/%d/%Y")),
    Arrest.Year < 2018 ~ month(as.POSIXlt(arrests_data$Arrest.Date,
      format = "%Y-%m-%d")))))

# Set ordering manually
arrests_data$Race.Ethnicity <- arrests_data$Race.Ethnicity %>%
  factor(levels = c("Black", "Hispanic", "White"))

```

tidyData3.R to tidy the population data from US Census Bureau

```

#' @author Rajeev Agrawal
#' @description STAT612 - Final Project - R code for tidying the population data from the US Census Bureau
library(tidyverse)
library(mosaic)
library(readxl)

#Write to csv file format for easy manipulation
read_xlsx("../data/PopulationDC.xlsx", sheet = "SC-EST2019-SR11H-11", range = cell_cols("G:M")) %>%
  write_csv("../data/population_dc.csv") ->
  popul_dc
read_xlsx("../data/PopulationDC.xlsx", range = cell_cols("A")) %>%
  write_csv("../data/column1.csv") ->
  column1

#Delete first few and last few rows to match the records
popul_dc <- popul_dc[-c(1:14),]
column1 <- column1[-c(1:17, 130:135),]

#Add column1 to popul_dc
popul_dc <- cbind(column1, popul_dc)

#Change name of the first variable
names(popul_dc)[1] <- "Race.Ethnicity"

names(popul_dc)
head(popul_dc)

#Filter the rows required for analysis
popul_dc <- popul_dc[c(3:4, 15),]

#Change the names of the entries in the first column
popul_dc[(1:3),1] <- c("White", "Black", "Hispanic")
unique(popul_dc$Race.Ethnicity)

#Pivot data from wide to long
popul_dc %>%
  pivot_longer(cols = starts_with("20"), names_to = "Year", values_to = "Population") ->
  popul_dc

#Remove the original data objects from the environment
rm(column1)

```