

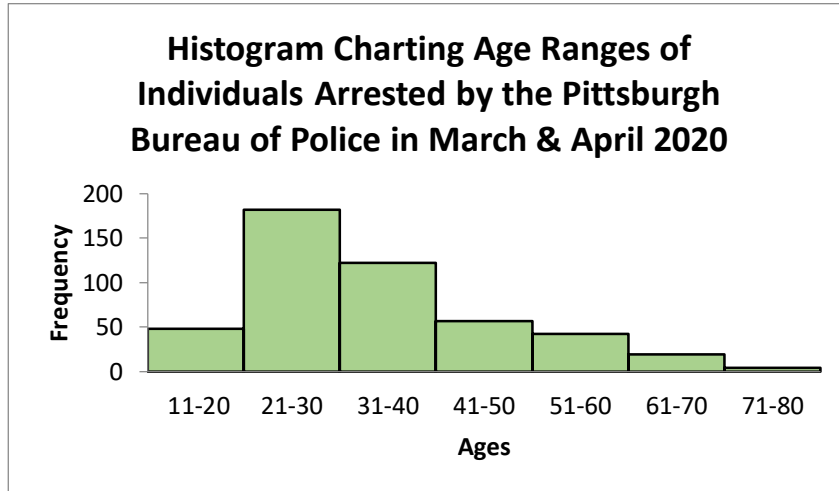
Policing in the Time of Coronavirus

A Comparative Look at Arrests Made by the Pittsburgh Bureau of Police in 3/9/2020-4/22/2020 versus 3/9/2019-4/22/2019

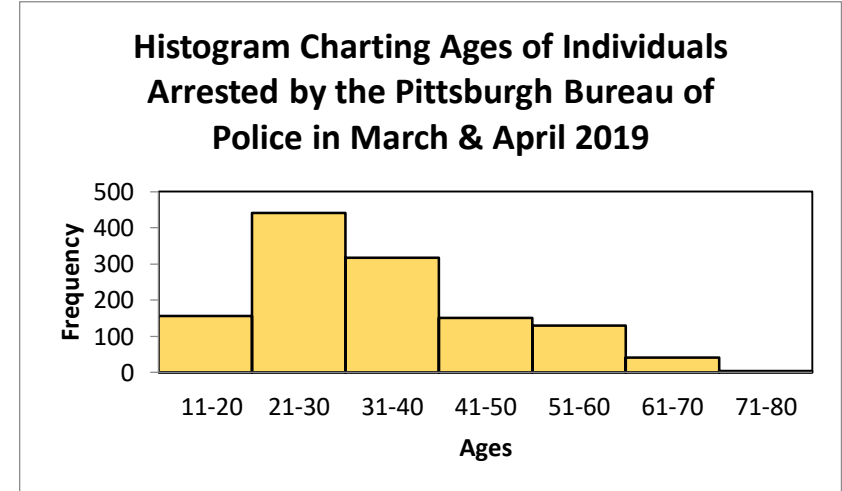
Core inquiry questions: While citizens of Pittsburgh are being asked to stay at home as much as possible and public services are being scaled back, how has policing changed in the city? Where is the Pittsburgh Bureau of Police focusing its resources, and who is being policed? Do the demographics and zip codes of the people being arrested during the stay-at-home order reflect the overall population of Pittsburgh, and how do they compare to the demographics and zip codes of those arrested in spring 2019? Finally, do the most common offenses that lead to arrests right now match up with the most common offenses in a comparable, pre-coronavirus timeframe?

Sample population 1: Arrests made by the Pittsburgh Bureau of Police (PBP) between March 9, 2020 and April 22, 2020 (n = 474)

Sample population 2: Arrests made by PBP between March 9, 2019 and April 22, 2019 (n = 1,250)



MEAN AGE	34.13319239
MEDIAN AGE	31
STANDARD DEV.	12.86720052
MINIMUM AGE	12
MAX AGE	73
LOWER FENCE	1
Q1	25
Q3	41
IQR	16
UPPER FENCE	65



MEAN AGE	33.8389694
MEDIAN AGE	31
STANDARD DEV.	12.95395036
MINIMUM AGE	11
MAX AGE	78
LOWER FENCE	-1.5
Q1	24
Q3	41
IQR	17
UPPER FENCE	66.5

Analysis of variance in ages of offenders in 2019 versus 2020:

The histograms and descriptive statistics tables above do not show much variance in age between the two sample populations. Both datasets have a median offender age of 31, a mean age close to 34, and similar spread across all ages of offenders (standard deviations are both about 12.9). The max age is notably higher in 2020, but calculating the IQR and upper/lower fences helped to determine that this is an outlier. An F test of ages in random samples of 30 from each of the two sample populations helped to determine that there is no significant variance in age of offenders between the two sample populations.

Analysis of Variance in Age of Offenders in March-April 2019 versus March-April 2020				
AGE 2020 (RANDOM SAMPLE)	AGE 2019 (RANDOM SAMPLE)			
42	21			
22	21		F-Test Two-Sample for Variances	
27	27			
22	21			
25	63		Mean	AGE 2020 32.96666667 AGE 2019 33.83333333
22	46		Variance	208.5850575 182.3505747
17	45		Observations	30 30
26	56		df	29 29
44	35		F	1.143868385
37	38		P(F<=f) one-tail	0.35991502
28	25		F Critical one-tail	1.860811435
24	27			
52	19		Conclusion: Because the F Critical value is higher than the F value, we retain the null hypothesis that variance in age of those arrested is basically equal between 2019 and 2020.	
55	40			
61	26			
20	25			
14	53			
14	13			
55	29			
24	58			
27	25			
25	57			
47	37			
35	31			
26	24			
69	30			
45	47			
30	23			
26	22			
28	31			

Frequency Table for Race/Ethnicity of Those Arrested by PGH Police in March & April 2020		
<i>Race/Ethnicity</i>	<i>Frequency</i>	<i>%</i>
Asian	1	0.20%
Black	329	69%
Hispanic	4	1%
Other	4	1%
Unknown	4	1%
White	132	28%
Total	474	100%

Frequency Table for Race/Ethnicity of Those Arrested by PGH Police March & April 2019		
<i>Race/Ethnicity</i>	<i>Frequency</i>	<i>%</i>
Asian	4	0.32%
Black	787	62.96%
Hispanic	10	0.80%
Other	14	1.12%
Unknown	15	1.20%
White	420	34%
Total	1250	100%

Frequency Table for Gender 2020		
<i>Gender</i>	<i>Frequency</i>	<i>%</i>
Female	121	26%
Male	352	74%
Unknown	1	0.20%
Total	474	100%

Frequency Table for Gender 2019		
<i>Gender</i>	<i>Frequency</i>	<i>%</i>
Female	314	25%
Male	931	74%
Unknown	5	0.40%
Total	1250	100%

Frequency Table for Zip code 2020		
<i>Zip Code</i>	<i>Frequency</i>	<i>%</i>
PA 15226	12	3%
PA 15221	25	5%
PA 15220	11	2%
PA 15219	111	24%
PA 15214	15	3%
PA 15212	48	10%
PA 15210	43	9%
PA 15208	13	3%
PA 15206	28	6%
PA 15203	24	5%
PA 15201	12	3%
Total with known zip	458	

Frequency Table for Zip code 2019		
<i>Zip Code</i>	<i>Frequency</i>	<i>%</i>
PA 15226	18	1%
PA 15221	44	4%
PA 15220	18	1%
PA 15219	396	33%
PA 15214	66	5%
PA 15212	140	12%
PA 15210	71	6%
PA 15208	68	6%
PA 15206	78	6%
PA 15203	69	6%
PA 15201	10	1%
Total with known zip	1204	

Hypothesis testing to determine the significance of differences in proportions highlighted in the frequency tables:

The frequency tables for race show that the proportion of individuals arrested who are white is 6% lower in 2020 than in 2019, and the proportion of individuals arrested who are black is 6% higher in 2020 than in 2019. The two screenshots below show the results of two hypothesis tests to determine whether the differences in these demographic proportions are statistically significant (since the % change is the same for both white and black individuals, I only ran these tests once using the data for black individuals). Both tests returned p values smaller than the significance level of 0.05, which indicates that there would be a very low chance resampling with this great of a difference in proportions (0.064). This means we can reject the null hypothesis and conclude that there is indeed a significant increase in the proportion of black individuals arrested in March-April 2020 compared to March-April 2019. I ran similar tests to determine the significance of the 9% proportional difference in arrests made in the 15219 zip code from 2019 to 2020, which also resulted in p values < 0.05.

Test for significance of 6% difference in proportion of black individuals in March-April 2019 versus March-April 2020

StatKey Randomization Test for a Difference in Proportions

Custom Data

Edit Data

Randomization method Resampling

Generate 1 Sample

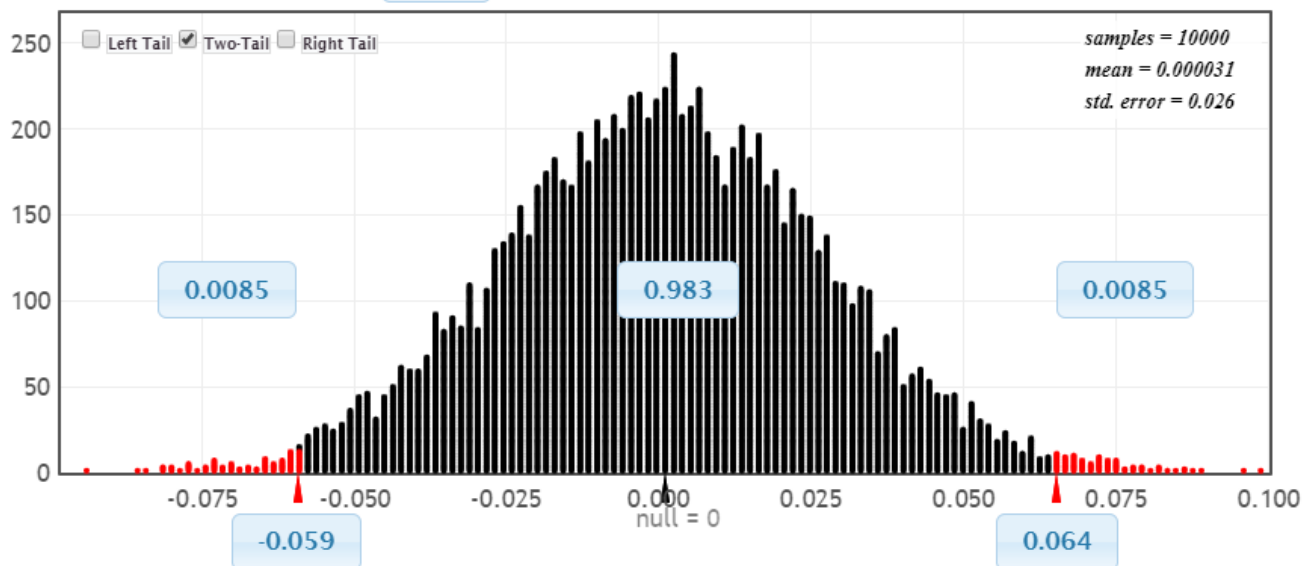
Generate 10 Samples

Generate 100 Samples

Generate 1000 Samples

Reset Plot

Randomization Dotplot of $\hat{p}_1 - \hat{p}_2$ Null Hypothesis: $p_1 = p_2$



Original Sample

Group	Count	Sample Size	Proportion
Group 1	787	1250	0.630
Group 2	329	474	0.694
Group 1-Group 2	458	n/a	-0.064

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	821	1250	0.657
Group 2	310	474	0.654
Group 1-Group 2	511	n/a	0.0028

Test for significance of difference in proportion of black individuals
in March-April 2019 versus March-April 2020

Test for significance of difference in proportion of 15219 arrests
in March-April 2019 versus March-April 2020

Z Score Calculator for 2 Population Proportions

Success!

You'll find the values for z and p below. Blue means your result is significant, red means it's not.

Sample 1 Proportion (or total number)

329

Sample 1 Size (N_1)

474

Sample 2 Proportion (or total number)

787

Sample 2 Size (N_2)

1250

Significance Level:

- 0.01
 0.05
 0.10

One-tailed or two-tailed hypothesis?:

- One-tailed
 Two-tailed

The value of z is 2.5023. The value of p is .01242. The result is significant at $p < .05$.

Z Score Calculator for 2 Population Proportions

Success!

You'll find the values for z and p below. Blue means your result is significant, red means it's not.

Sample 1 Proportion (or total number)

111

Sample 1 Size (N_1)

458

Sample 2 Proportion (or total number)

396

Sample 2 Size (N_2)

1204

Significance Level:

- 0.01
 0.05
 0.10

One-tailed or two-tailed hypothesis?:

- One-tailed
 Two-tailed

The value of z is -3.4238. The value of p is .00062. The result is significant at $p < .05$.

Test for 9% difference in proportion of 15219 arrests in March-April 2019 versus March-April 2020

StatKey Randomization Test for a Difference in Proportions

Custom Data ▾

Edit Data

Randomization method **Resampling ▾**

Generate 1 Sample

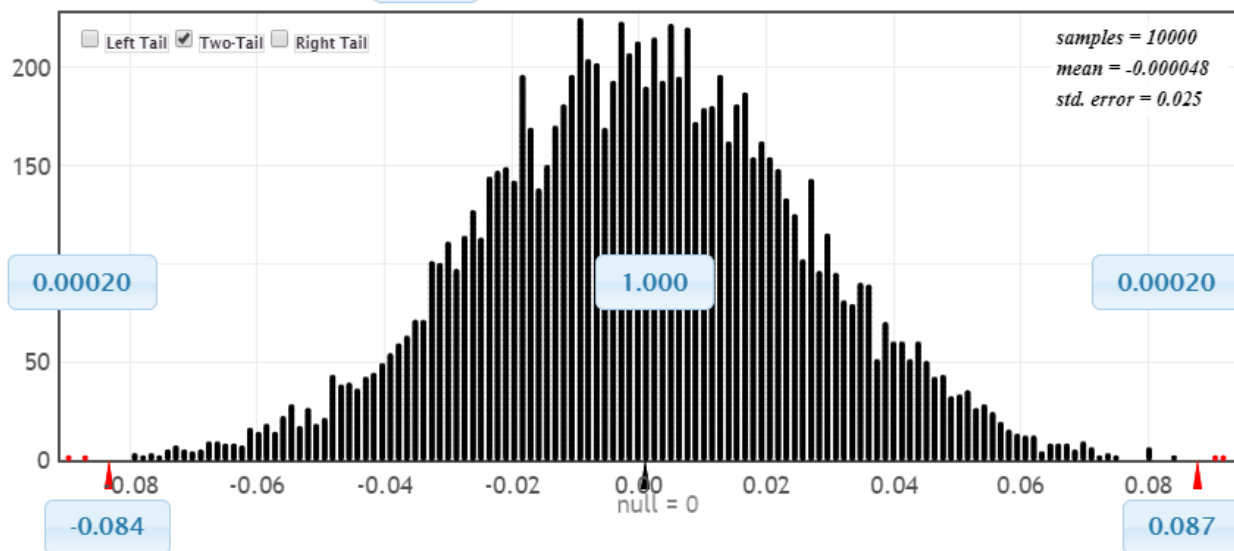
Generate 10 Samples

Generate 100 Samples

Generate 1000 Samples

Reset Plot

Randomization Dotplot of $\hat{p}_1 - \hat{p}_2$ Null Hypothesis: $p_1 = p_2$



Original Sample

Group	Count	Sample Size	Proportion
Group 1	396	1204	0.329
Group 2	111	458	0.242
Group 1-Group 2	285	n/a	0.087

Randomization Sample

Group	Count	Sample Size	Proportion
Group 1	341	1204	0.283
Group 2	122	458	0.266
Group 1-Group 2	219	n/a	0.017

Offenses represented in random samples of 30 arrests from each of the two sample populations:

Frequency Table for Randomly Sampled Offenses 2020	
<i>Offense</i>	<i>Frequency</i>
Simple Assault	6
Controlled Substance Possession	6
Driving Under the Influence	4
Bench Warrant	3
Criminal Conspiracy	3
Indirect Criminal Contempt	2
Access Device Fraud	2
Theft	3
Burglary	1
Robbery	1
Aggravated Assault	1
Marijuana Possession (Small Amount)	2

Frequency Table for Randomly Sampled Offenses 2019	
<i>Offense</i>	<i>Frequency</i>
Simple Assault	6
Controlled Substance Possession	5
Driving Under the Influence	3
Bench Warrant	5
Criminal Conspiracy	1
Failure to Appear in Court	2
Public Drunkenness	4
Theft	3
Terroristic Threats	1
Robbery	1
Aggravated Assault	1
Marijuana Possession (Small Amount)	3

Conclusions and recommendations for future research

- The data supports the hypothesis that arrests are down overall in the time of coronavirus: The total number of arrests recorded by PBP in the 2020 timeframe is less than half of the total arrests made in the 2019 timeframe (474 arrests versus 1,250).
- Gender breakdown is consistent in both timeframes: Male offenders were arrested roughly three times more often than female offenders in spring 2019 and 2020.
- Age breakdown is consistent across both timeframes: Both datasets have a median offender age of 31, a mean age close to 34, and similar spread across all ages of offenders (standard deviations are very close to one another). The max age is notably higher in 2020, but calculating the IQR and upper/lower fences helped to determine that this is an outlier. An F test helped to determine that there is no significant variance in age of offenders between the two sample populations.
- The 3 most common zip codes for Pittsburgh arrests are consistent in both timeframes, but the #1 zip code for both timeframes, 15219, comprises a lower percentage of all 2020 arrests than it does in the 2019 arrests (24% versus 33%). A randomization test for difference in proportions as well as a z score calculator for 2 population proportions helped to determine that this proportional increase is statistically significant. This difference may suggest that the police zone for 15219 has been affected by the pandemic (staff illnesses, decreased in other resources, etc.), or that PBP has made a temporary change in daily patrols during the pandemic that leads to fewer arrests in 15219. Qualitative data would need to be collected directly from PBP in order to draw further conclusions about this change.
- PBP arrested black people at an extremely disproportionately high rate in both timeframes, even more so in the 2020 pandemic timeframe (69% of all individuals arrested, compared to 63% in 2019). A randomization test for difference in proportions as well as a z score calculator for 2 population proportions helped to determine that this proportional increase is statistically significant. Even though fewer people across all tracked demographic groups were arrested by PBP in the coronavirus timeframe than in spring 2019, the evidence in this study suggests that black people comprise an even

larger-than-typical majority of those arrested by PBP during the pandemic, which may give us some indication of where PBP is focusing its resources, and who is being policed. This proportion should be monitored through the end of the stay-at-home order and business closures in Pittsburgh to see if this proportion remains at 69% or increases further.

- The most common offenses from 2019 generally match up with the most common offenses in the time of coronavirus: A random sampling of 30 arrests from each sample population shows comparable proportions of arrests for simple assault, controlled substance possession, bench warrants, and thefts. Arrests related to failure to appear in court showed up more than once in the 2019 sample, but not at all in the 2020. This is unsurprising as the courts are closed.
- I hypothesized that arrests based on minor offenses would be minimal in the 2020 dataset in light of the pandemic, so I was surprised to find an arrest solely based on possessing a small amount of marijuana in the random sampling of 30 arrests from 2020. This is a pretty minor crime—particularly compared to the others in this list—because it is non-violent and does not endanger the lives of others. I went back to the full dataset for this 2020 time period to review all arrests made solely for possession a small amount of marijuana and found only 3 (there were many similar marijuana offenses tied to more serious crimes like reckless driving and/or simple assault). It's important to note that all three arrests related to minor marijuana crimes were made in the first week of the coronavirus timeline in Pittsburgh (from March 9 to March 15). In the interest of tracking potential discriminatory practices by the Pittsburgh Bureau of Police based on race, it's also important to note that, in the 26 arrests made solely for possession of a small amount of marijuana across both datasets, all but three of the offenders (88%) were black. It would also be interesting to take a closer look at arrests for minor offenses—such as possessing a small amount of marijuana—over a longer 2020 pandemic timeframe to see if these types of arrests are low compared to 2019. I do not have enough data right now to really determine if the pandemic has led to a decrease in arrests for minor offenses, but if I looked at this 2020 sample population next month with data through May 31 and there were no new arrests made for minor offenses, I think that would suggest a change in PBP's practices in response to the pandemic.

Resources:

Articles that informed the inquiry questions:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3086232

<https://triblive.com/local/pittsburgh-allegeny/pittsburghs-black-population-accounts-for-majority-of-arrests-report-says/>

<https://www.jurist.org/commentary/2010/03/pittsburgh-racial-profiling-as-law-and/>

<https://www.pennlive.com/news/2020/04/coronavirus-silver-lining-crime-and-crashes-plummet-across-pa.html>

Source for raw dataset:

<https://data.wprdc.org/dataset/arrest-data>

Online hypothesis testing tools:

http://www.lock5stat.com/StatKey/randomization_2_cat/randomization_2_cat.html

<https://www.socscistatistics.com/tests/ztest/default2.aspx>