HEALTHCARE ACCESSIBILITY IN PITTSBURGH NEIGHBORHOODS

Avonworth High School

Marion Haney - Owen North - Darren Hunt - Lauren Pflueger
Mike Frank - Charlie Bozada - Langley Turcsanyi

In which Pittsburgh neighborhoods is healthcare most accessible?

Definitions

Access to healthcare refers to one's proximity to healthcare offices or hospitals by either phsyical distance or travel time. Good access to healthcare may also involve health insurance and care management. We also considered life expectancy and types health insurance coverage. Bad access to healthcare is usually due to a lack of health insurance and a low income.

We defined poverty using the 2018 Federal Poverty Guidelines.

Medicaid is a healthcare program that assists families and individuals with low incomes in paying for medical costs. This is a federal program but at a state level, the coverage varies.

Resources

Our main resource was the Western Pennsylvania Regional Data Center wesbite, which allowed us to retrieve demographic data and the data that we required in order to define healthcare. The American Community Survey allowed us to find this data categorized by neighborhood.

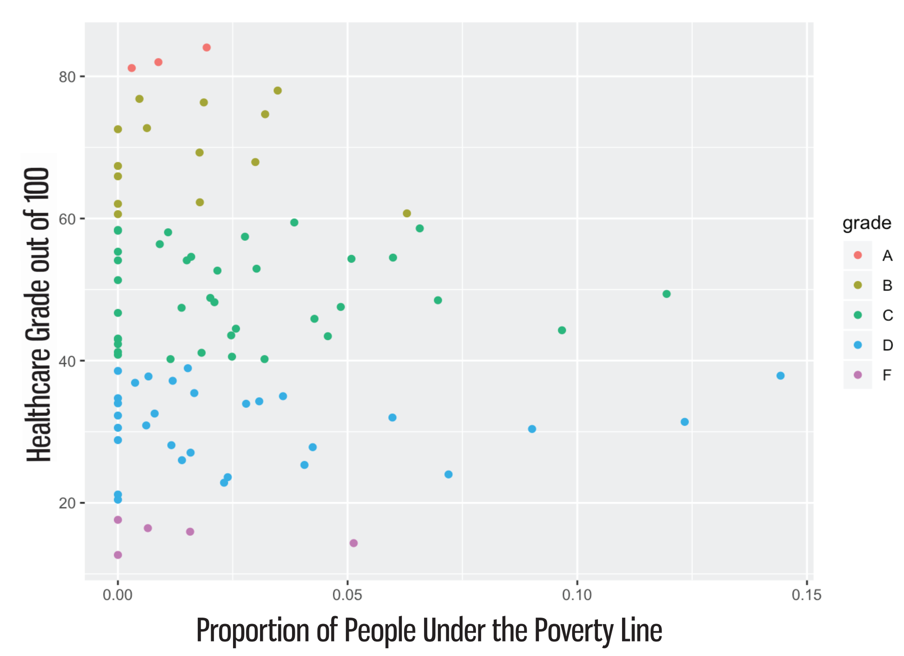
We also created our own dataset, which helped us to define healthcare, by researching the life expectancy in each neighbhorhood in Allegheny County. We accessed this information from The Robert Wood Johnson Foundation.

Challenges

- Choosing a topic that was interesting and necessary
- Cleaning data in order to simplify data analysis
- Defining a location's proximity to primary care center and hospitals
- Deciding which significant factors to analyze
- Simplifying location status by converting neighborhoods and counties to zip codes
- Creating an objective rubric with average rank that would still work when analyzed through the ANOVA test
- Manually creating a heatmap since Tableau is unable to process our information based on neighborhoods rather than counties or zip codes

SUMMARY OF RESULTS

Poverty Rate vs Healthcare Grade



ANOVA Output

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
grade_factor	3	0.0009153	0.0003051	0.3403	0.7963
Residuals	85	0.07622	0.0008967	NA	NA

Table: Analysis of Variance Model

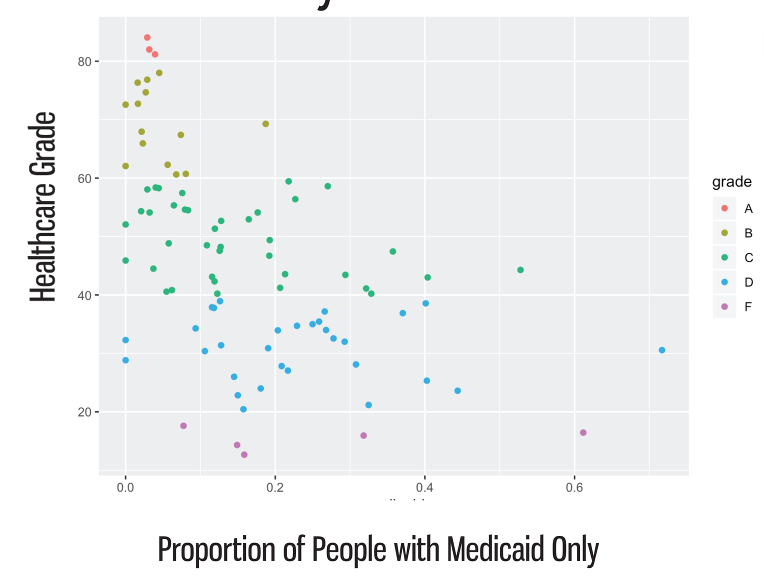
The F statistic produces a probability greater than 0.05, proving insignifance.

Confidence Intervals

	2.5 %	97.5 %
(Intercept)	-0.007633	0.04562
grade_factorB	-0.02514	0.03134
grade_factorC	-0.02049	0.03592
grade_factorD	-0.04927	0.0377

Each interval contains 0, further confirming the insignificance of our ANOVA test. In other words, we are 95% confident that the average poverty rate of each healthcare grade are not significantly different.

Medicaid-Only vs Healthcare Grade



Medicaid-Only ANOVA Test Confidence Intervals

	2.5 %	97.5 %
(Intercept)	-0.112	0.1785
grade_factorB	-0.1474	0.1727
grade_factorC	-0.03313	0.2686
grade_factorD	0.04601	0.3506
grade_factorF	0.04597	0.4134

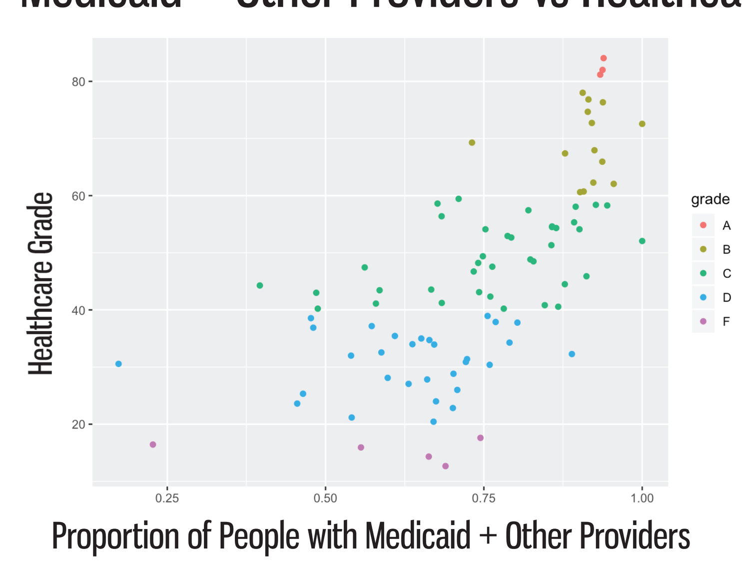
This ANOVA test tells us that neighborhoods with a grade of F are supported by only Medicaid slightly more than neighborhoods with a grade of D.

ANOVA Output

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.03328	0.07304	0.4556	0.6498
grade_factorB	0.01264	0.08049	0.1571	0.8755
grade_factorC	0.1177	0.07587	1.552	0.1245
grade_factorD	0.1983	0.07661	2.589	0.01132
grade_factorF	0.2297	0.09239	2.486	0.01488

Since these values are the only ones less than 0.05, the D & F grades are the only average Medicaid-only proportions that are significant.

Medicaid + Other Providers vs Healthcare Grade



Medicaid + Other Providers ANOVA Test Confidence Intervals

	2.5 %	97.5 %
(Intercept)	0.786	1.087
grade_factorB	-0.1919	0.1402
grade_factorC	-0.3277	-0.01459
grade_factorD	-0.4587	-0.1426
grade_factorF	-0.5513	-0.17

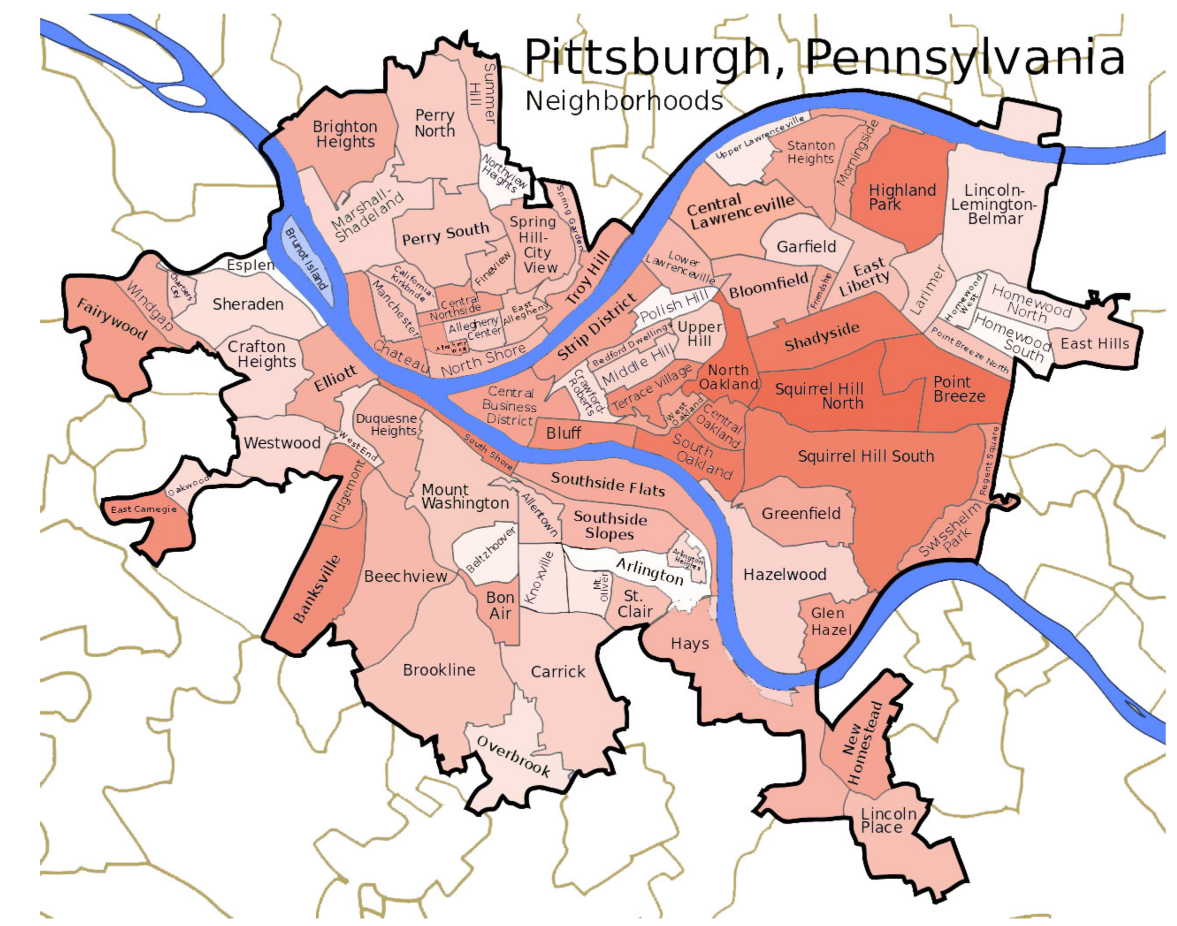
Since the confidence interval for C is greater than D, which is less negative than F, we can conclude the average proportion of people with Medicaid + other providers for C is greater than D, which is greater than F.

ANOVA Output

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.9367	0.0758	12.36	1.116e-20
grade_factorB	-0.02586	0.08353	-0.3096	0.7576
grade_factorC	-0.1711	0.07874	-2.174	0.03252
grade_factorD	-0.3006	0.0795	-3.781	0.0002891
grade_factorF	-0.3607	0.09588	-3.762	0.0003095

Since these probabilities are all less than 0.05, the average proportion of people with Medicaid + other providers are significantly different between C, D and F grades.

Our heatmap displays the range of healthcare grades in each neighborhood. The neighborhoods with the darkest shades of red have the best access to healthcare, while the lightest have the worst access to healthcare. It is evident that the neighborhoods around Oakland, Squirrel Hill, and Shadyside have the best access as a result of their proximity to several UPMC and other medical locations.



Our Rubric

We crafted our own rubric so we could assign a grade based on the data from several sets that we were able to find from online resources, as well as data from our own datasets that we put together.

The Top 5

Squirrel Hill North	84.06	A
Point Breeze	82	A
North Oakland	81.17	A
Highland Park	78	B
South Oakland	76.83	B

The Bottom 5

Polish Hill	17.61	F
Northview Heights	16.44	F
Homewood South	15.94	F
Arlington	14.33	F
Esplen	12.67	F

Our Grading Scale

A	81-100
B	61-80
C	41-60
D	21-40
F	0-20

This is a sample of a few neighborhoods and their grades based on our scale.

Conclusion

The main factors that can be used to assess a neighborhood's access to healthcare includes average life expectancy, proportion of people insured, and proximity to hospitals and primary care facilities. It was interesting to see that the neighborhoods with the best healthcare access were close to eachother geographically, while the ones with the worst were also near eachother. We discovered that the neighborhoods with the best access to healthcare are Squirrel Hill, Central Oakland, Point Breeze and some of their surrounding areas. The worst neighborhoods for healthcare access are Homewood, Northview Heights, Polish Hill, Arlington and Esplen. The neighborhoods with a D grade, based on our rubric, had a higher average proportion of people with Medicaid-Only as their insurance than the neighborhoods with a F grade. We also found that neighborhoods with a C grade have a higher average proportion of Medicaid + other providers as their insurance than neighborhoods with a D grade. F grades also have a lower average proportion of Medicaid + other providers than D grades. From our research and data analysis, we can also confirm that the access to healthcare in Allegheny County is not affected by poverty levels. The correlation between these factors are insignificant which tells us that they are independent of one another. A recommendation based on our findings could be offering tax subsidies or other incentives for non-government healthcare providers to expand their client base for neighborhoods with a lower grade.

Are There Identifiable Factors That Effect the Location of Gun Violence in the City of Pittsburgh?

INTRODUCTION

As we all know, our country has been riddled with gun violence, but it is hard for many of us to believe that this would ever affect our communities in Pittsburgh. Unfortunately, gun violence is incredibly common, causing us to ponder why it is common. According to Forbes Magazine, since economic fluctuation changes human behavior, when there is an economic downturn there is a noticeable increase in gun violence. Since economic downturns aren't predictable, we wanted to identify other factors that we predicted would have a similar effect on gun violence: employment rate and median income by neighborhood in the City of Pittsburgh.

PROCESS

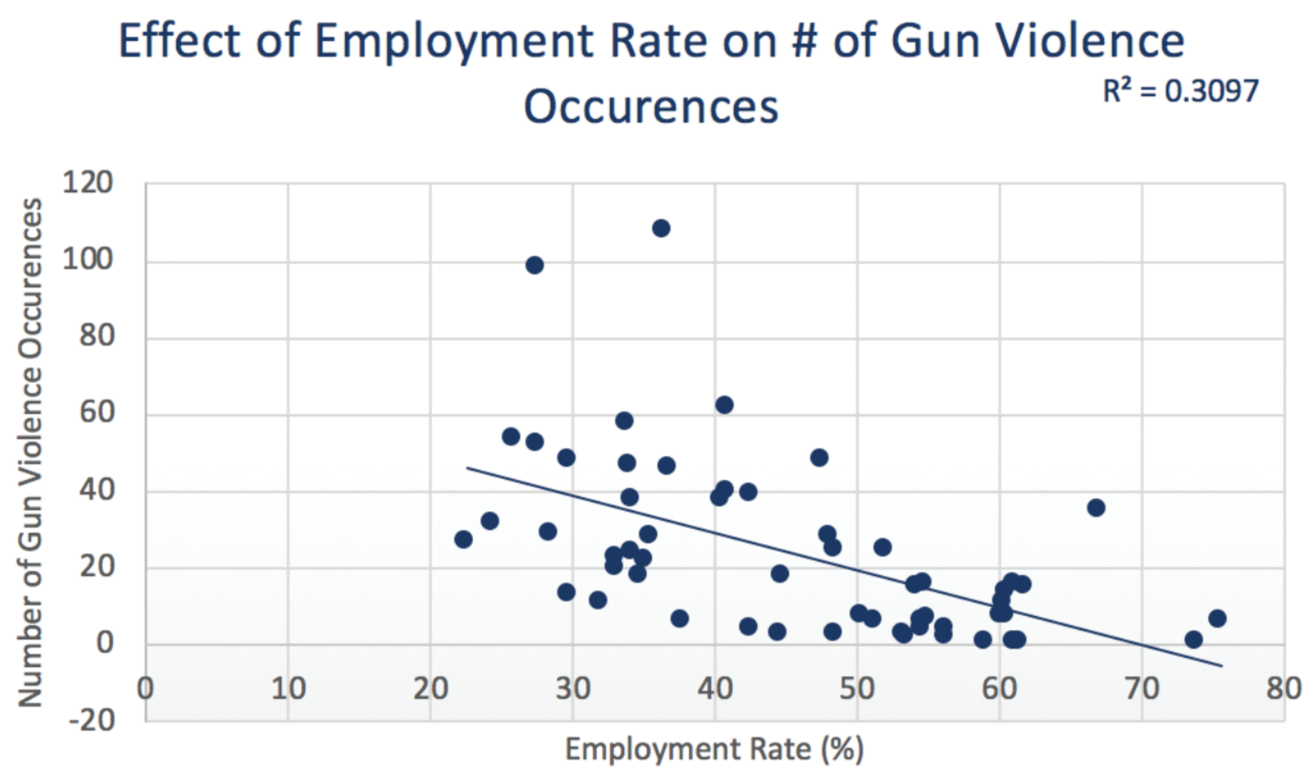
After reading the Pittsburgh Post Gazette report that gun violence has been decreasing in the City of Pittsburgh for the past five years and officers are solving more homicides than ever before, it felt necessary to dig into what is contributing to the numbers of gun violence incidents decreasing. This led us to compile the following data sets:

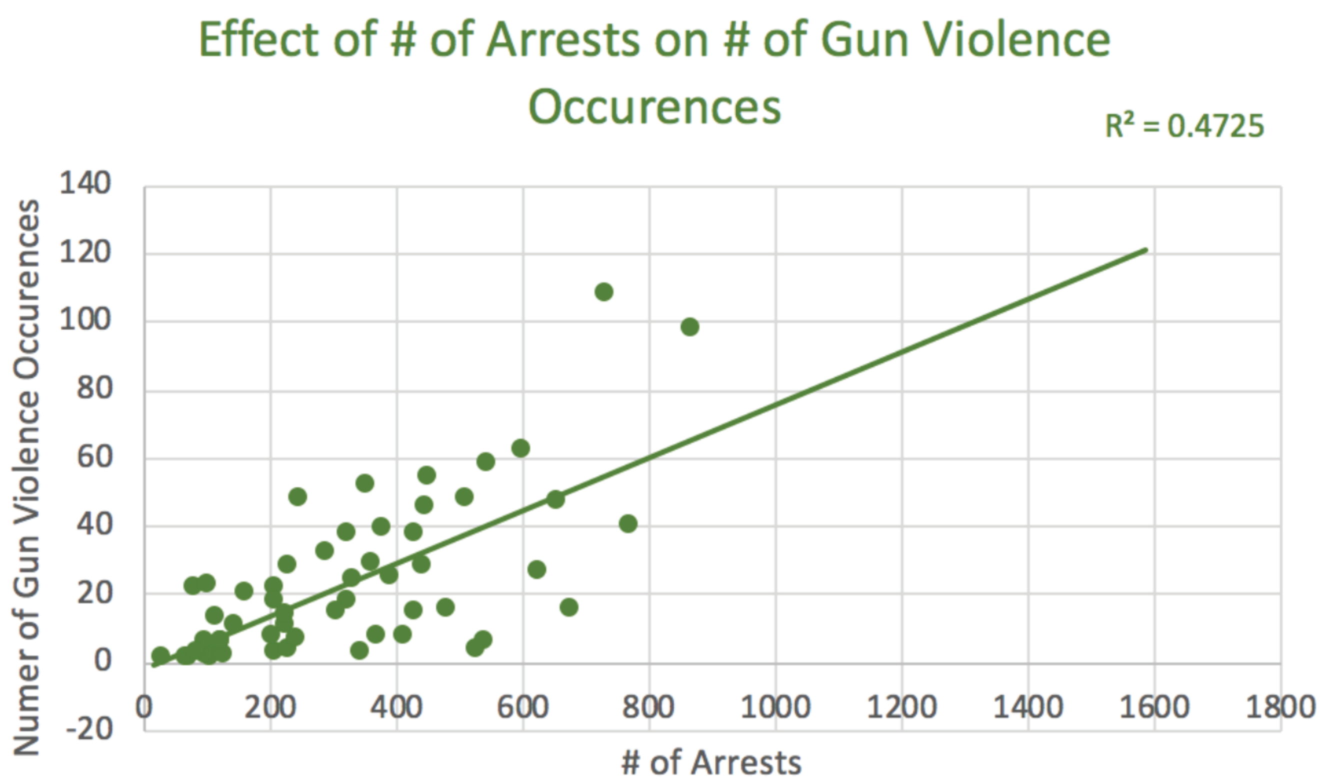
- Shooting Homicides in PA and US (2014-2017)
- Shooting Homicides in City of Oittsburgh v Allegheny County (2010-2017)
- by neighborhood v Arrests by neighborhood (City of Pittsburgh 2010-2013)
- Income by neighborhood v Arrests by neighborhood
- Employment rate by neighborhood v Arrestsby neighborhood
- Employment rate by neighborhood v Shootings by neighborhood
- Offenses v Shootings (2013)

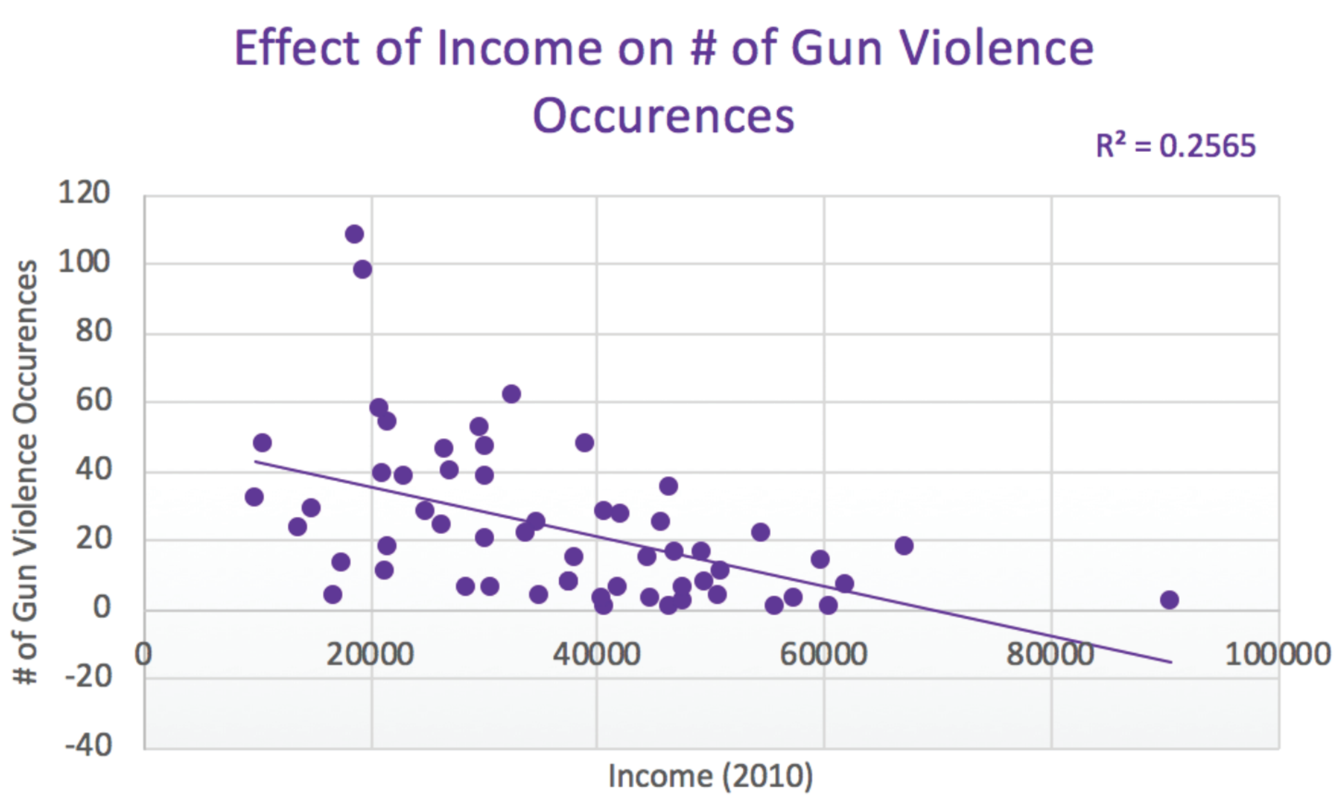
GRAPH NAME	SOURCES	CONCLUSION
Effect of Employment Rate on # of Gun Violence Occurrences (City of Pittsburgh)	WPRDC: Employment Status for the Population 2014 PBP and DHS: Gun Violence, City of Pittsburgh, 2010 to February 2019	r^2 = 0.3097 Based off of this value and the trendline, there is a positive correlation between the employment rate in each neighborhood of the City of Pittsburgh and the number of gun violence occurrences.
Effect of # of Arrests on # of Gun Violence Occurrences (City of Pittsburgh)	WPRDC: Pittsburgh Police Arrest Data 2016-2019 PBP and DHS: Gun Violence, City of Pittsburgh, 2010 to February 2019	r^2 = 0.4725 Based off of this correlation value, the number of gun violence occurrences does increase in areas where more arrests take place. These may correlation because some of the arrests are due to gun violence related crimes, or it could be due to chance.
Effect of Income on # of Gun Violence Occurrences (City of Pittsburgh)	Data USA: Income by Location 2013-2016 PBP and DHS: Gun Violence, City of Pittsburgh, 2010 to February 2019	r^2 = 0.397 This is a slightly positive correlation between median income in each neighborhood of the city and the number of gun violence occurrences in the corresponding locations. The trendline for this relationship shows a slightly negative correlation, but that may be due to the small size or range of the data set.
Shooting Homicides 2010-2017: Allegheny County Breakdown	Pittsburgh Post Gazette: Homicide: Allegheny County medical examiner and area police reports 2010-2017 PBP and DHS: Homicides in the City of Pittsburgh, 2010-2018	Looking at the breakdown of inside city limits of Pittsburgh and outside, the number of homicides by gun violence has been decreasing over the past few years in the City of Pittsburgh. The graph shows a gradual incline in shooting homicides

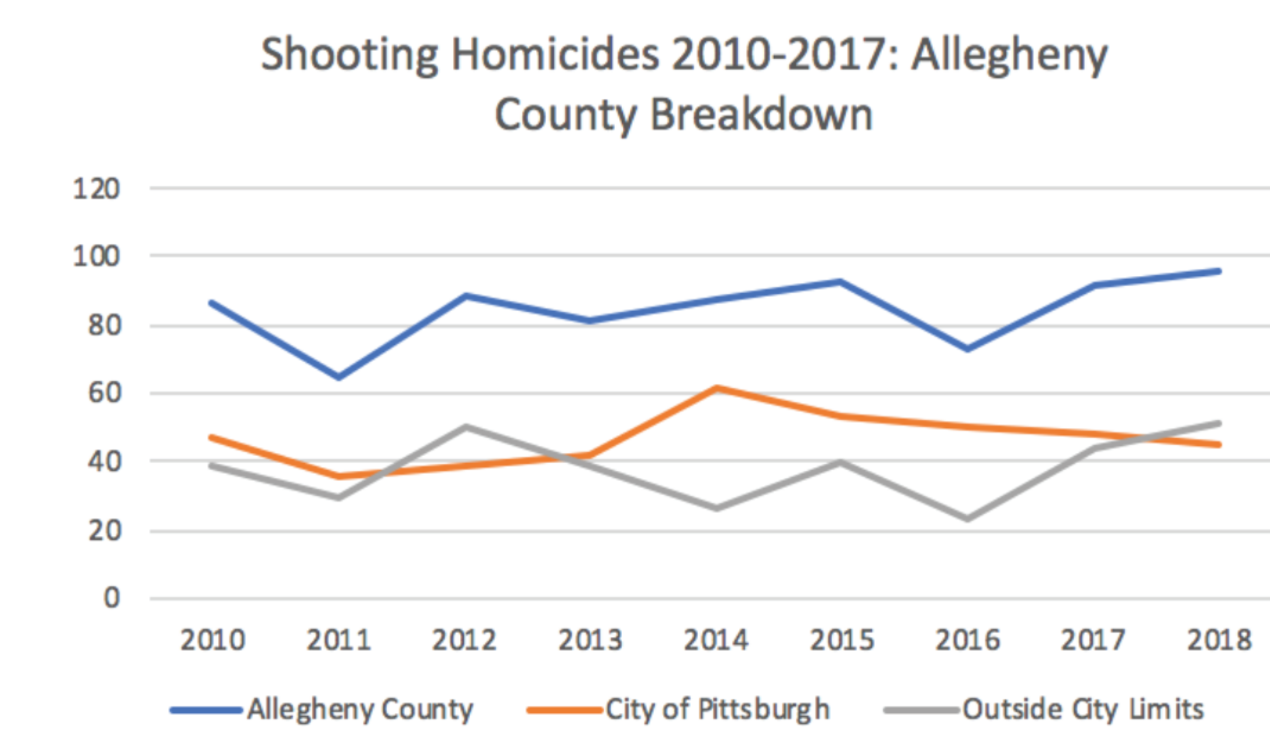
CONCLUSION

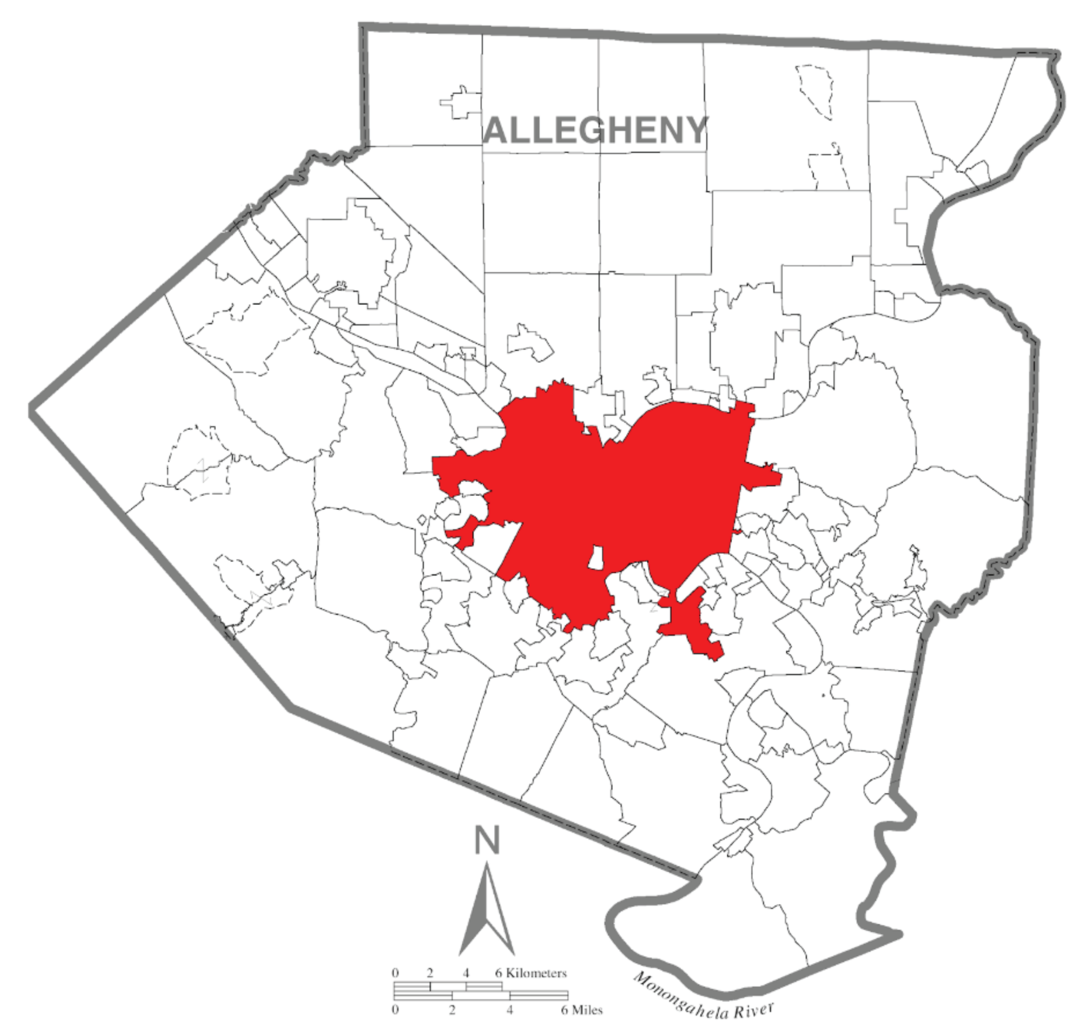
While more people may own guns, there may also be a larger awareness of safety training. Because our project focussed on the socio-economic aspects of gun violence, we are also interested in whether the 2008 financial crisis affected gun violence. To do this, we would examine the same factors (employment and income) but instead look specifically around the stock market crash.Ultimately, gun safety and restriction laws have one of the largest known influences on gun deaths. States with the most of these laws such as Massachusetts, Connecticut, and California also have low death rates (Boston University School of Public Health). Gun violence is most likely a result of the culmination of many factors (laws, regulations, socio-economic status). We hope that future progress in these areas will instigate a decline is gun violence.











Basile, Thomas J. "The Best Gun Control Measure Is Always A Job." Forbes, Forbes Magazine, 27 Feb. 2013, www.forbes.com/sites/thomasbasile/2013/02/27/the-best-gun-control-measure-is-always-a-job/#7d11143b2118.

Rabinowitz, Kate, and Youjin Shin. "The Great Recession's Great Hangover." The Washington Post, WP Company, 7 Sept. 2018, www.washingtonpost.com/graphics/2018/business/great-recession-10-years-out/?utm_term=.7d0eb5360150.

Schoen, John W. "States with Strict Gun Laws Have Fewer Firearms Deaths. Here's How Your State Stacks Up." CNBC, CNBC, 27 Feb. 2018, www.cnbc.com/2018/02/27/states-with-strict-gun-laws-have-fewer-firearms-deaths-heres-how-your-state-stacks-up.html.

By The Ellis School

Technology vs. Test Scores

Jacob Black, Handuo Chen, Devin Demnyan, Ian Ellis, George Kohan, Julian Palmgren, Zachary Somma, and Michael Ulizzi

Introduction

Research Question

Does an increase in advance technology within public schools positively increase the standardized test scores of students?

Background Information

Many of the schools within Alleghany County differ in their spending towards technology for students. Some schools are established with advance technology like Promethean boards, individual computers, or individual iPads.

We observed that certain schools like Montour excel in their advancements of test scores while providing their students with computers of their own.

We also observed schools such as Sto-Rox that were not performing well in test scores and did not provide their students with any type of technology other than a shared computer classroom.

Example of a 1:1 classroom

Data

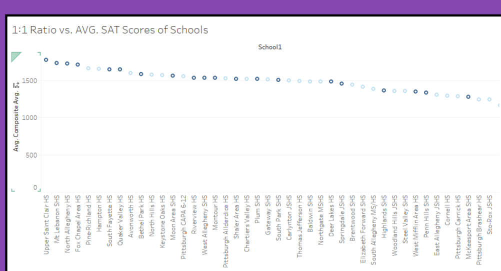
The data used was an average of SAT scores within schools in the year 2016. These scores were based on the old 2400 points over the new 1600 points used today. The schools highlighted in red are schools that are partly in Alleghany county but not considered part of the county.

AUN	LEA	School Nbr.	School Name	N Tested (More than 11)	Reading Average	Math Average	Writing Average	Composite Average
103020753	Avonworth SD	5199	Avonworth HS	77	535	547	513	1595
103021102	Baldwin-Whitehall SD	50	Baldwin SHS	236	503	504	478	1485
103021252	Bethel Park SD	62	Bethel Park HS	268	534	533	512	1579
102027451	Pittsburgh SD	6915	Pittsburgh Brashear HS	195	417	432	394	1242
103021453	Brentwood Borough SD	70	Brentwood SHS	45	499	472	472	1443
103021603	Carlynton SD	79	Carlynton JSHS	60	505	505	486	1496
102027451	Pittsburgh SD	412	Pittsburgh Carrick HS	112	428	448	406	1282
103021752	Chartiers Valley SD	6706	Chartiers Valley HS	186	509	514	495	1518
102020001	City CHS	7727	City CHS	18	439	453	421	1313
103021903	Clairton City SD	8094	Clairton MS/HS	19	388	417	346	1151
103022103	Cornell SD	8087	Cornell HS	27	440	436	413	1289
103022253	Deer Lakes SD	513	Deer Lakes HS	130	502	500	478	1480
103022803	East Allegheny SD	8340	East Allegheny JSHS	75	444	442	420	1306
103023153	Elizabeth Forward SD	144	Elizabeth Forward SHS	122	482	487	442	1410
103023912	Fox Chapel Area SD	156	Fox Chapel Area HS	291	569	576	562	1708
103024102	Gateway SD	170	Gateway SHS	204	514	509	487	1510
103024603	Hampton Township SD	5190	Hampton HS	228	558	563	533	1655
103024753	Highlands SD	5153	Highlands SHS	112	459	471	433	1363
103025002	Keystone Oaks SD	5112	Keystone Oaks HS	108	523	535	510	1568
103026002	McKeesport Area SD	6105	McKeesport Area SHS	126	436	428	414	1277
103026303	Montour SD	5017	Montour HS	170	508	514	507	1530
103026343	Moon Area SD	4951	Moon Area SHS	246	524	537	502	1563
103026402	Mt Lebanon SD	254	Mt Lebanon SHS	318	573	578	577	1729
103026852	North Allegheny SD	8305	North Allegheny HS	596	574	588	559	1722
103026873	Northgate SD	53	Northgate MSHS	43	506	509	469	1484
103026902	North Hills SD	7101	North Hills SHS	233	523	540	511	1574
103027352	Penn Hills SD	309	Penn Hills SHS	183	462	445	428	1335
107657103	Penn-Trafford SD	6648	Penn Trafford HS	262	505	525	489	1518
102027451	Pittsburgh SD	416	Pittsburgh Perry HS	62	395	402	368	1165
103021003	Pine-Richland SD	315	Pine-Richland HS	336	550	572	538	1660
102027451	Pittsburgh SD	8106	Pittsburgh CAPA 6-12	131	529	517	512	1557
103027503	Plum Borough SD	435	Plum SHS	241	512	517	488	1517
103027753	Quaker Valley SD	448	Quaker Valley HS	122	544	560	538	1643
103028203	Riverview SD	6928	Riverview HS	53	521	509	502	1532
103028302	Shaler Area SD	460	Shaler Area HS	256	515	509	495	1519
103028653	South Allegheny SD	440	South Allegheny MS/HS	67	475	466	445	1386
103028703	South Fayette Township SD	470	South Fayette HS	173	539	568	538	1645
103028753	South Park SD	467	South Park SHS	119	507	523	472	1502

Visuals

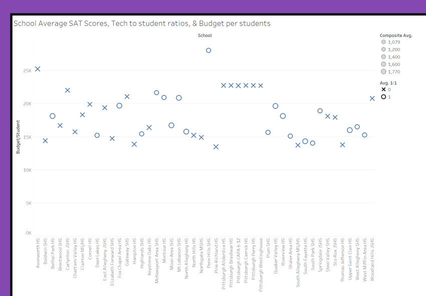
Visualization 1:

This graph simply shows the composite average of SAT scores compared to the status of a school having a 1:1 classroom. This graph shows that there is a slight correlation to these two factors which was found to be .378 correlation on a -1 to +1 scale. Since the correlation was weak we decided to understand what could be causing the issue.



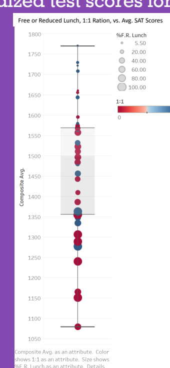
Visualization 2:

The data surrounding average SAT scores, tech to student ratios, and budget per student is scattered; but this graph shows that there must be another factor in order to pin-point the source of the problem. We choose to incorporate this data in order to show the process in which we took towards the conclusion we established.



Visualization 3:

The data surrounding average SAT scores, tech to student ratios, and free or reduced lunch populations show that schools with a high percentage of free or reduced lunch students are not performing well on tests and do not have a 1:1 classroom. This proves our conclusion that income of families is the main source affecting standardized test scores for students.



Analysis & Conclusions

Analysis

While processing the data we found that a 1:1 classroom, a computer to each student, slightly increases the test scores. We then decided to analyze the budget and percentage of students eligible for free or reduced lunches. We went this route in order to centralize the main reason for these schools low test scores.

Conclusion

Looking at all the data we were able to come up with the conclusion to our problem. We found that the SAT scores were affected by the income of students. Schools with a high percentage of free or reduced lunch populations, typically contain lower income families. Those same schools also were the schools with lower SAT scores.

Further Direction

We decided that the solution is extremely difficult to solve. Perhaps students that stem from lower income families could be given special help by the State and/or school districts by directing money towards tutoring programs. This could potentially assist the students within schools that are struggling. Hopefully those students would increase their SAT scores.

Challenges

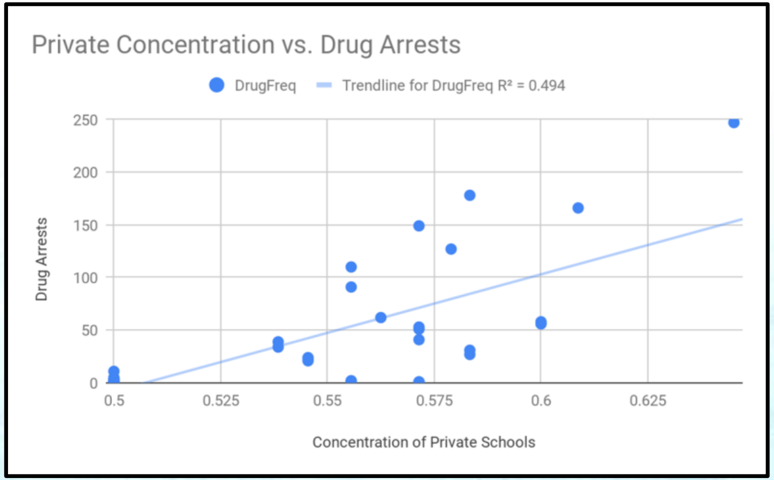
Throughout the process of our collection of data and analysis of that data there were quiet a lot of problems. The main issue was trying to obtain data on the ratio and type of technologies between schools. Many schools would not reply to our inquires for information. Since that data was difficult to obtain in the time we had, we found other public sources of information which were more general than what would have received from a district source and used that to analyze our main problem.

References

1. https://www.niche.com/k12/search/best-school-districts/c/allegheny-county-pa/
2. https://www.post-gazette.com/news/education/2015/08/21/What-will-students-in-Pittsburgh-s-south-suburbs-find-when-they-get-to-class/stories/201508210085
3. https://www.point2homes.com

Correlation Between Drug Activity and Concentration of Private Schools

by Jiwoo Cheon, Teresa Huang, Steven Pan, Gloria Ye, and Josh Zhou



The Question:

Does the concentration of private schools in an area correlate with the number of drug-related arrests in the same area?

Hypothesis and Reasoning:

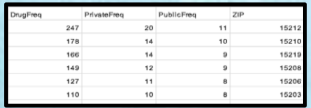
We expect to find a negative correlation between private school concentration and drug arrests because more private schools could indicate an area of greater affluence, and therefore lower crime rates.

Data Set Contents:

1. Pittsburgh Police Arrest Data: Contains details of an arrested person (age, gender, race) as well as details of their crime (location, time, offense)

PK	CCR	AGE	GENDER	RACE	ARRESTTI	ARRESTLO	OFFENSES	INCIDENT	INCIDENT	INCIDENT	INCIDENT	COUNCIL_	PUBLIC_W	X	Y

1. Allegheny County Public School Locations: Contains the building name, school district, and address of a school
1. Allegheny County Private School Locations: Contains the building name and address of the school



DrugFreq	PrivateFreq	PublicFreq	ZIP
247	20	11	15212
178	14	10	15210
166	14	9	15219
149	12	9	15208
127	11	8	15206
110	10	8	15203

Recommendations:

- Private schools should adopt more effective drug education programs.
- Private schools should also look into the culture of underage drug use at their facilities, and devote more funds into preventing them.

Procedure:

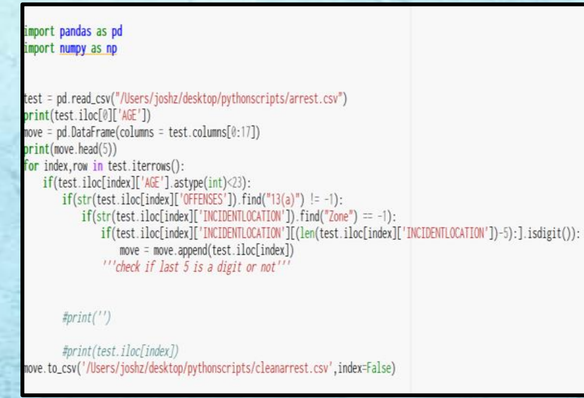
1. We collected our data and formulated a hypothesis. (To protect against confounding factors such as population density, we used the ratio of private schools to public schools.)
2. We cleaned our data using a Python program.
3. We created graphs with the data, including our r^2-value so we could examine correlation, grouping data points by zip code

Results:

- Private school concentration has a moderately positive correlation with drug arrests (r^2 = 0.494). This data leads us to believe that drugs are actually more prevalent in areas with more private schools.

Challenges and Roadblocks:

- Many data points were corrupted
- We started out with thousands of data entries, and filtered them through a Python code.
  - We filtered arrest data for those under 23 (to collect entries most influenced by schools in the area) and those related to drugs
  - We also used Python to find the zip codes that matched up between data sets --both between public and private schools, and the ratio list with arrest data.
- Many data points were unusable because they did not match with a zip code in the other data set.
  - Our data set ended up very small, even though it had started in the thousands, so we are not able to extrapolate or make bold claims about our data.



```python
import pandas as pd
import numpy as np


test = pd.read_csv("/Users/joshz/desktop/pythonscripts/arrest.csv")
print(test.iloc[0]['AGE'])
move = pd.DataFrame(columns = test.columns[0:17])
print(move.head(5))
for index,row in test.iterrows():
    if(test.iloc[index]['AGE'].astype(int)<23):
        if(str(test.iloc[index]['OFFENSES']).find("13(a)") != -1):
            if(str(test.iloc[index]['INCIDENTLOCATION']).find("Zone") == -1):
                if(test.iloc[index]['INCIDENTLOCATION'][(len(test.iloc[index]['INCIDENTLOCATION'])-5):].isdigit()):
                    move = move.append(test.iloc[index])
                '''check if last 5 is a digit or not'''


        #print('')

        #print(test.iloc[index])
move.to_csv('/Users/joshz/desktop/pythonscripts/cleanarrest.csv',index=False)
```

Brick & Mortar Decline vs. Pittsburgh Economy

Maddy Allen, Manny Athwal, Stefani Fotovich, Seth Mascellino, Maddie Minsinger, Abby Minzer, Dior Pemberton, Mac Polny, and Mitchell Ward

Introduction

Research Question

- To what extent do property closings affect property value in Allegheny county?

Background Information

- With online stores such as Amazon on the rise, more brick-and-mortar stores have closed as a result of decreased sales
  - A.K.A. The "Retail Apocalypse" – J.C. Penny, Macy's, Sears, Kmart, etc.
  - Around 12,000 stores nationally
- Due to these closings, property values are being impacted heavily in the surrounding areas

Century 3 Mall

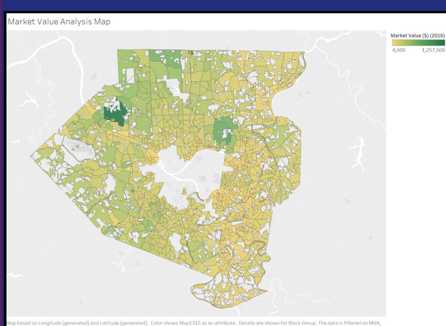
This is an example of retail mall that has closed in Allegheny county recently. Is this a result of less brick & mortar sales, or something else entirely?

Data

The Market Value Analysis dataset contains land values for Census tract block groups. The Census Bureau QuickFacts dataset contains total retail sales & retail sales per capita for Allegheny county townships, boroughs, & municipalities. The datasets were compiled into an Excel sheet as shown below.

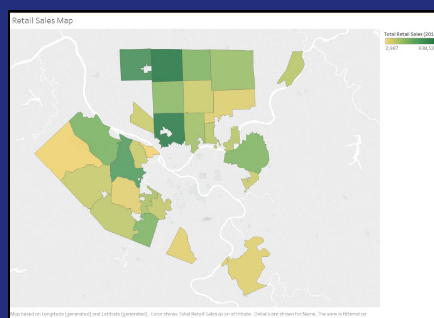
Census Tract	Township	Total Retail Sales (2012)	Total retail sales per capita (2012)	Average Msp (2016)
4200315216	Collier	54243	$7,298	$235,941.50
4200323000	Elizabeth	61183	$4,603	$95,265.00
4200325904	Findlay	9559	$1,849	$182,145.00
4200332328	Hampton	120712	$6,542	$220,160.00
4200332832	Harrison	173938	$16,616	$99,875.00
4200336808	Indiana	63426	$8,678	$179,843.40
4200339312	Kennedy	130431	$16,711	$167,507.40
4200345900	McCandless	320294	$11,147	$239,834.47
4200347696	Marshall	662014	$91,249	$293,275.00
4200350784	Moon	371934	$15,006	$74,342.94
4200351696	Mount Lebanon	142193	$4,296	$233,941.96
4200355016	North Fayette	134030	$9,534	$153,555.00
4200356384	O'Hara	174986	$20,726	$229,808.44
4200356392	Ohio	116859	$20,834	$322,500.00
4200359032	Penn Hills	346262	$8,185	$65,096.35
4200360272	Pine	838524	$70,595	$386,515.75
4200364528	Richland	357000	$31,343	$237,660.00
4200365352	Robinson	559385	$41,335	$199,830.56
4200366264	Ross	794635	$25,525	$157,805.32
4200368388	Scott	216602	$12,729	$258,665.71
4200369584	Shaler	249843	$8,682	$151,604.17
4200372160	South Fayette	160521	$10,759	$212,259.86
4200372400	South Park	53043	$3,928	$144,805.00
4200374648	Stowe	2987	$470	$44,221.43
4200379274	Upper St. Clair	361867	$18,751	$266,450.00
4200382800	West Deer	275103	$23,208	$134,741.75
4200385184	Wilkins	117781	$18,525	$74,250.00

Visuals



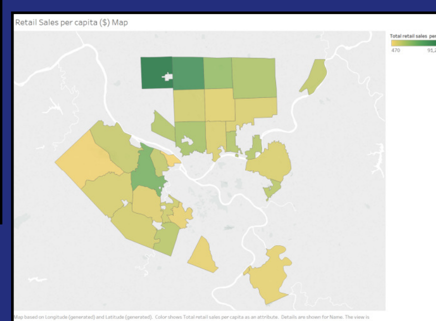
Visualization 1:

This map shows the market land values for Census tract block groups across Allegheny county from 2016. The highest market value was in Sewickley Heights, while the lowest was in West Mifflin



Visualization 2:

This map shows the retail sales per Allegheny county townships in dollars for 2012. The highest amount of sales was in Pine Township while the lowest was in Stowe Township.



Visualization 3:

This map shows the retail sales per capita per Allegheny county townships in dollars for 2012. The highest sales per capita was in Marshall Township while the lowest was in Stowe Township.

Analysis & Conclusions

Analysis

- Visualizations 1, 2, & 3 look like they might correlate to each other at first glance, but correlation coefficients say otherwise.
- Total retail sales & Retail sales per capita resulted in an R-value of .77864
- Total retail sales per capita & Market land value resulted in an R-value of .53497
- Total retail sales & Market land value resulted in an R-value of .38092

Conclusion

- The R-value between total retail sales & retail sales per capita shows a strong positive correlation, which makes sense & acts as a control data set.
- The R-value between total retail sales & market land value shows a moderate positive correlation.
- The R-value between total retail sales & market land value shows a slight positive correlation.

Further Direction

- Does the amount of retail sales per store chain relate to whether it will close?
- Can this data & its correlations compare to retail sales & land value on a national scale?
- Do the amount of online retail sales determine whether brick & mortar stores will close?
- If the retail store closings are compared to other factors, will the results be different?

Challenges

- We had trouble narrowing down the topic that we wanted to analyze.
- We struggled to find the data that involved the actual numbers of store closings in Allegheny County.
- Because of this, we then struggled to find the link between the data that we could access.

References

1. http://www.pasda.psu.edu/uci/SearchResults.aspx?originator=%20&Keyword=crop%20estimates&searchType=keyword&condition=AND
2. https://www.census.gov/quickfacts/fact/map/alleghenycountypennsylvania/RTN131212
3. https://data.wprdc.org/dataset/market-value-analysis-allegheny-county-economic-development\
4. https://pittsburgh.cbslocal.com/2018/04/12/century-iii-mall-sheriffs-sale-posted-notice/

TJ Faber
Ethan James
Pavan Otthi

Impacts of Forestation on Illegal Waste Dumping

Alex Pizov
Alex Vlad
North Allegheny HS

INTRODUCTION:

How does the forestation in a neighborhood impact the number of illegal waste dumping sites in the neighborhood?

BACKGROUND:

Allegheny County has 486 illegal dump sites, 175 more than any other county in Pennsylvania. This figure does not even include the city of Pittsburgh. (Pittsburgh Post Gazette)

Illegal waste has suffocated children, bred insects, contaminated groundwater, discouraged development and cost millions of dollars in cleanup. (North Carolina Department of Environmental Quality)

A cost-effective way to curb illegal waste dumping is in the best interests on an entire community.

DATA SETS USED:

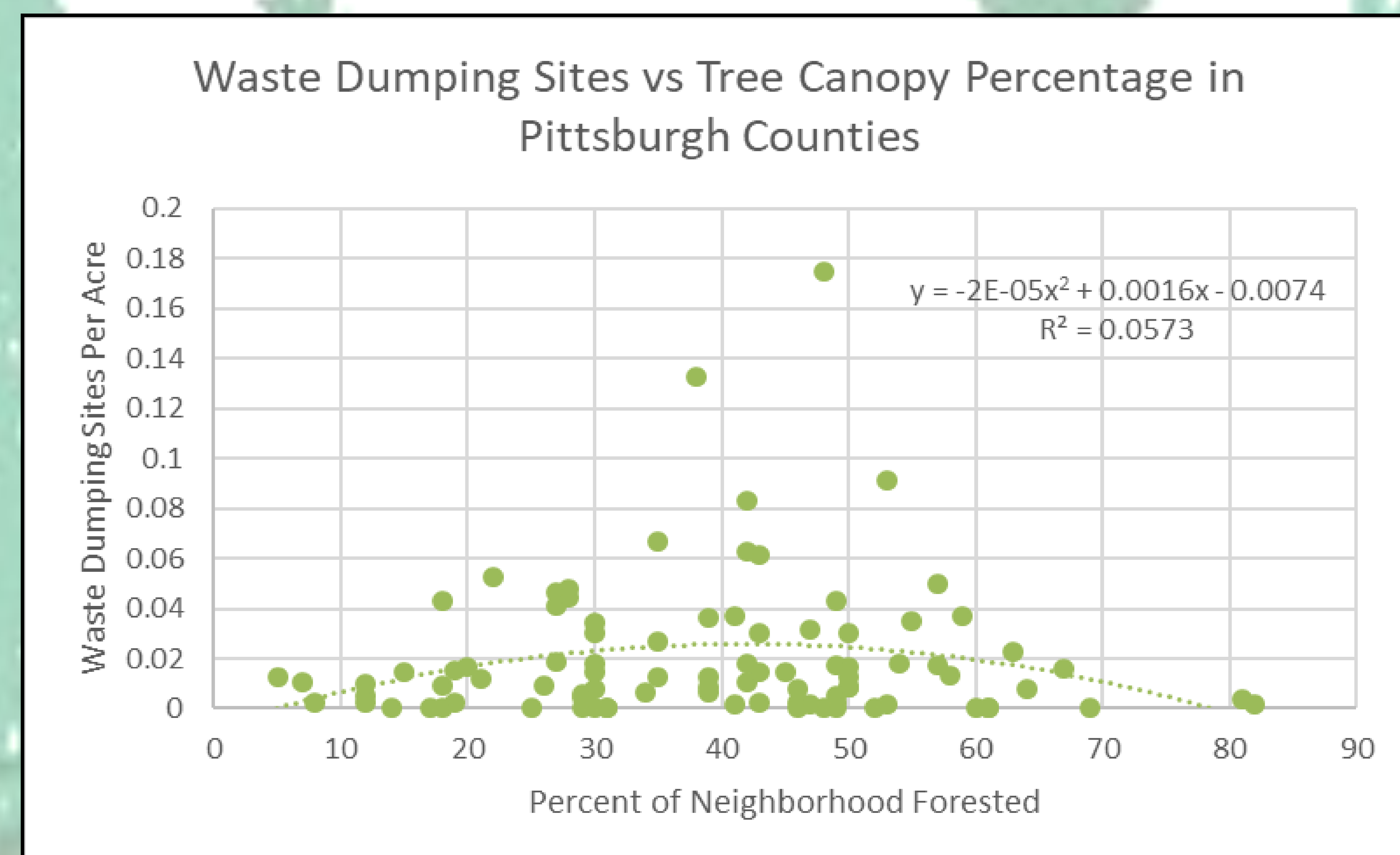
Pittsburgh data works collects data on the canopy cover by neighborhood as well as the location of each waste dumping site by neighborhood. A list of Pittsburgh's neighborhoods was obtained from the City of Pittsburgh's Map Room.

site_name	Status	City	Neighborhood	estimated_	location_description	latitude	longitude	qty
St. Martin	Surveyed	Pittsburgh	Allentown	0.5		40.42222	-79.9902	1
Brosville St	Surveyed	Pittsburgh	Allentown	3		40.4237	-79.9866	1
McCain Str	Surveyed	Pittsburgh	Allentown	1		40.42427	-79.9902	1
Ceres Way	Surveyed	Pittsburgh	Allentown	0.5		40.42327	-79.9925	1
Eureka Str	Completed	Pittsburgh	Allentown	0.1		40.42372	-79.9948	1
222 Walte	Completed	Pittsburgh	Allentown	5	dump is in deteriorated garage	40.42022	-79.9948	1
Grimes and	Completed	Pittsburgh	Allentown	0.3	residential street, across the stree	40.41544	-79.993	1
Parkwood	Surveyed	Pittsburgh	Arlington	3		40.41068	-79.9658	1
Parkwood	Surveyed	Pittsburgh	Arlington	4		40.41145	-79.9671	1
Parkwood	Completed	Pittsburgh	Arlington	2.5		40.41397	-79.9773	1
Mountain .	Completed	Pittsburgh	Arlington	2		40.4147	-79.9779	1
Jonquil Str	Surveyed	Pittsburgh	Arlington	0.5		40.4161	-79.9806	1
Bassler Str	Surveyed	Pittsburgh	Arlington	0.5		40.414	-79.9646	1
Azul Street	Surveyed	Pittsburgh	Arlington	1.5		40.41648	-79.9657	1
Rothman S	Surveyed	Pittsburgh	Arlington	2		40.41384	-79.9731	1
Parkwood	Partially Cl	Pittsburgh	Arlington	1.5		40.41354	-79.9761	1
Medhurst S	Surveyed	Pittsburgh	Banksville	1		40.41339	-80.0368	1
Barnett W	Completed	Pittsburgh	Bedford Dwelli	1		40.45159	-79.9692	1
Andick Wa	Surveyed	Pittsburgh	Beechview	2		40.41422	-80.0169	1
Napoleon	Surveyed	Pittsburgh	Beechview	2		40.41223	-80.0293	1

TYPICAL VALUES:

The Pittsburgh tree canopy cover dataset is simply a list of the percent of tree canopy cover per Pittsburgh neighborhood. The waste dumping dataset includes data on the location of each dump, materials dumped, the number of tons of material dumped, and the neighborhood it occurred in. (See above)

FIGURE 1:



RESULTS:

The dataset analyzed, waste dumping sites vs tree canopy percentage, was modelled with a quadratic equation of power 2. This equation was deemed to be the best fit for the dataset because it had the highest r^2 value out of any major correlation. Unfortunately, the correlation was still extremely low at $r^2 = 0.05$, meaning that only 5% of the data was well explained by the model. The statistical analysis for this same dataset, however, resulted in a p-value of 0.02, which indicates statistical significance. Both of these together indicate that while there is a correlation between tree canopy percentage and waste dump sites, there are other much more significant factors that affect waste dumping.

CONCLUSION:

The results of this analysis were inconclusive. Although we did discover statistical significance for the forestation of a neighborhood and the number of waste dumping sites per acre, the model we tried did not appropriately represent all the data. We believe this is due to regression to the mean of tree canopy cover per neighborhood, and is not a result of a correlation.

CHALLENGES:

Data cleaning the set of all Pittsburgh neighborhoods proved to be more difficult than expected due to the discrepancy between the waste dumping dataset's neighborhoods, those on the forestation dataset and those listed in the City of Pittsburgh's official map room registry. In the end, unless a neighborhood's name was nearly identical across all three datasets, it was excluded from the calculations.

In addition, we felt tree dataset we were originally using (see above) did not accurately reflect the number of trees in any given county. The list of trees we obtained were only those that are publicly maintained and does not reflect the whole forested area

FURTHER DIRECTION:

What are the costs associated with cleaning up illegal waste dumps?

Are neighborhoods composed primarily of people of color more likely to contain dump sites?

Do quality waste disposal programs reduce dumping incidences?

SOURCES:

https://www.post-gazette.com/news/environment/2018/07/21/hundreds-illegal-dumps-survey-allegheny-county-pittsburgh-CleanWays/stories/201807200145

https://deq.nc.gov/about/divisions/waste-management/waste-management-permit-guidance/solid-waste-section/illegal-dumping

https://issuu.com/treepittsburgh/docs/final_pittsburgh_urban_forest_management_plan_augu

https://data.wprdc.org/dataset/pgh

https://data.wprdc.org/dataset/allegheny-county-illegal-dump-sites

Alcohol-Related Car Crashes and Ridesharing

Cassandra Moats, Rachel Moret, Eliane Rectenwald, Helen Tan, Vivian Chen, Kristin Liang, Alexis Hagerty, & Lindsay Worrall
Oakland Catholic High School

Problems

Does ridesharing have an impact on alcohol-related DUIs and alcohol-related car crashes? Is this impact positive? How can we increase the positive impact that ridesharing has?

Top Destinations for Uber in Pittsburgh

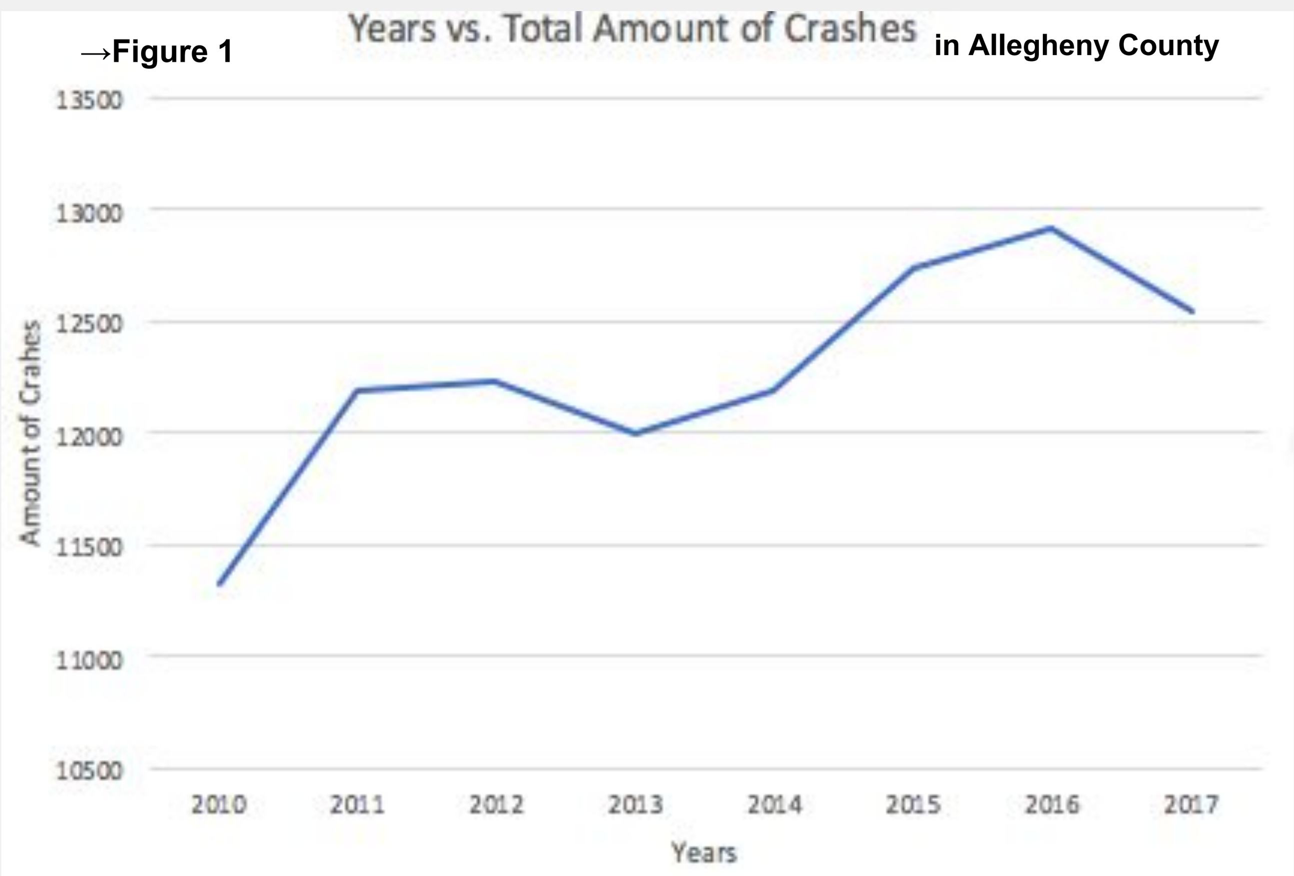
1. Pittsburgh International Airport
2. Rivers Casino
3. Tequila Cowboy
4. PPG Paints Arena
5. Heinz Field
6. Mario's Southside Saloon
7. Wyndham Grand Pittsburgh Downtown
8. David L. Lawrence Convention Center
9. Carnegie Mellon University
10. Carson City Saloon
11. McFadden's

Highlighted selections are drinking establishments. These being 7/11 of the top Uber destinations in Pittsburgh show that people who are more likely to be inebriated are taking Uber.

Source: Pittsburgh Magazine, December 2018

Importance

In 2016, 28% of fatal crashes were related to alcohol-impaired driving, resulting in a grand total of 10,497 deaths in the US. Fortunately, in recent years, there have been more options for avoiding driving under the influence. As a result, fatal accidents have steadily decreased. In 2014, Uber was introduced to Pittsburgh. Uber, along with Lyft, has been a popular option to for impaired would-be-drivers to use to ensure safe travels. With the increase of Uber and Lyft in major cities, our project will examine if the presence of these services decreases the number of traffic incidents associated with alcohol, which theoretically would lead to a safer driving environment.



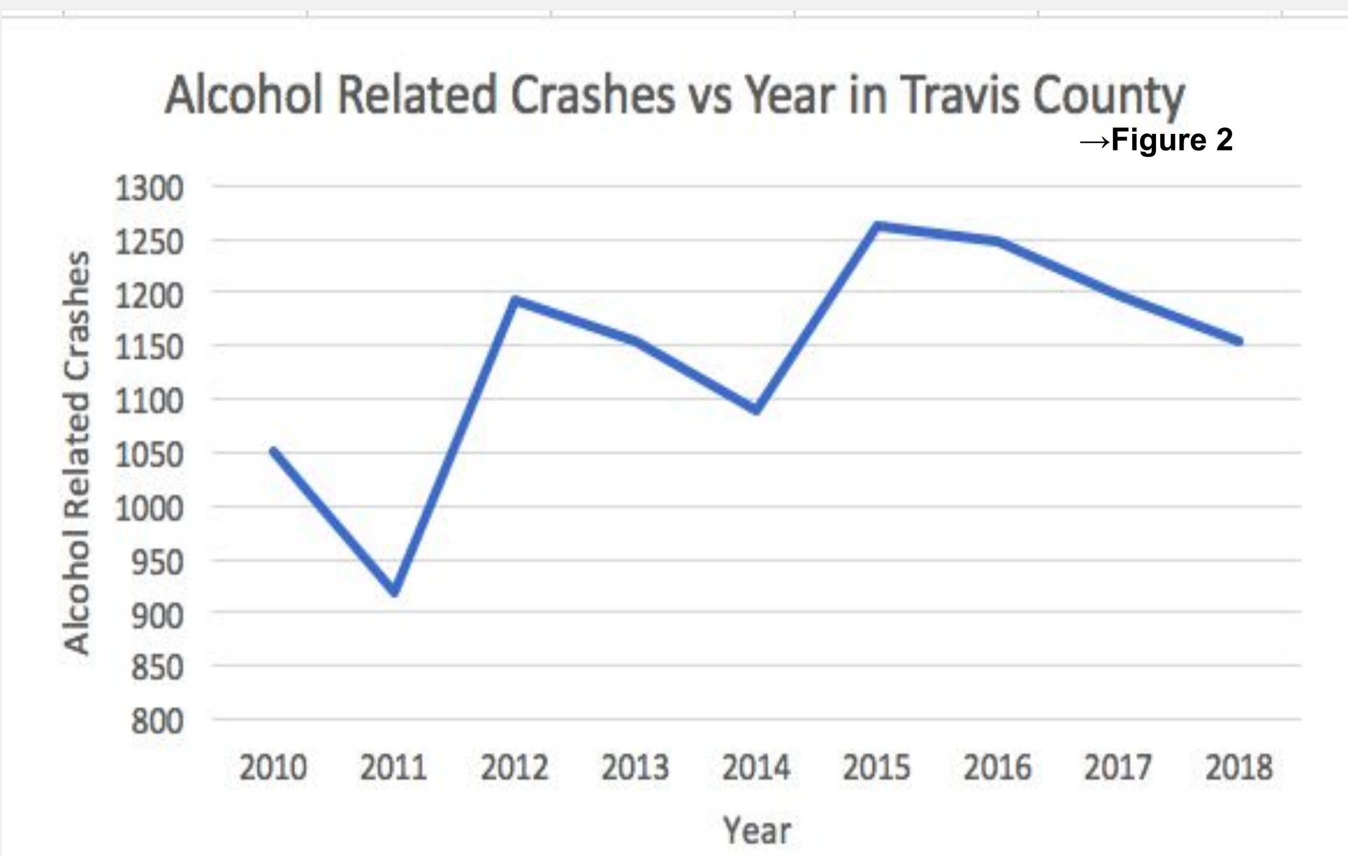
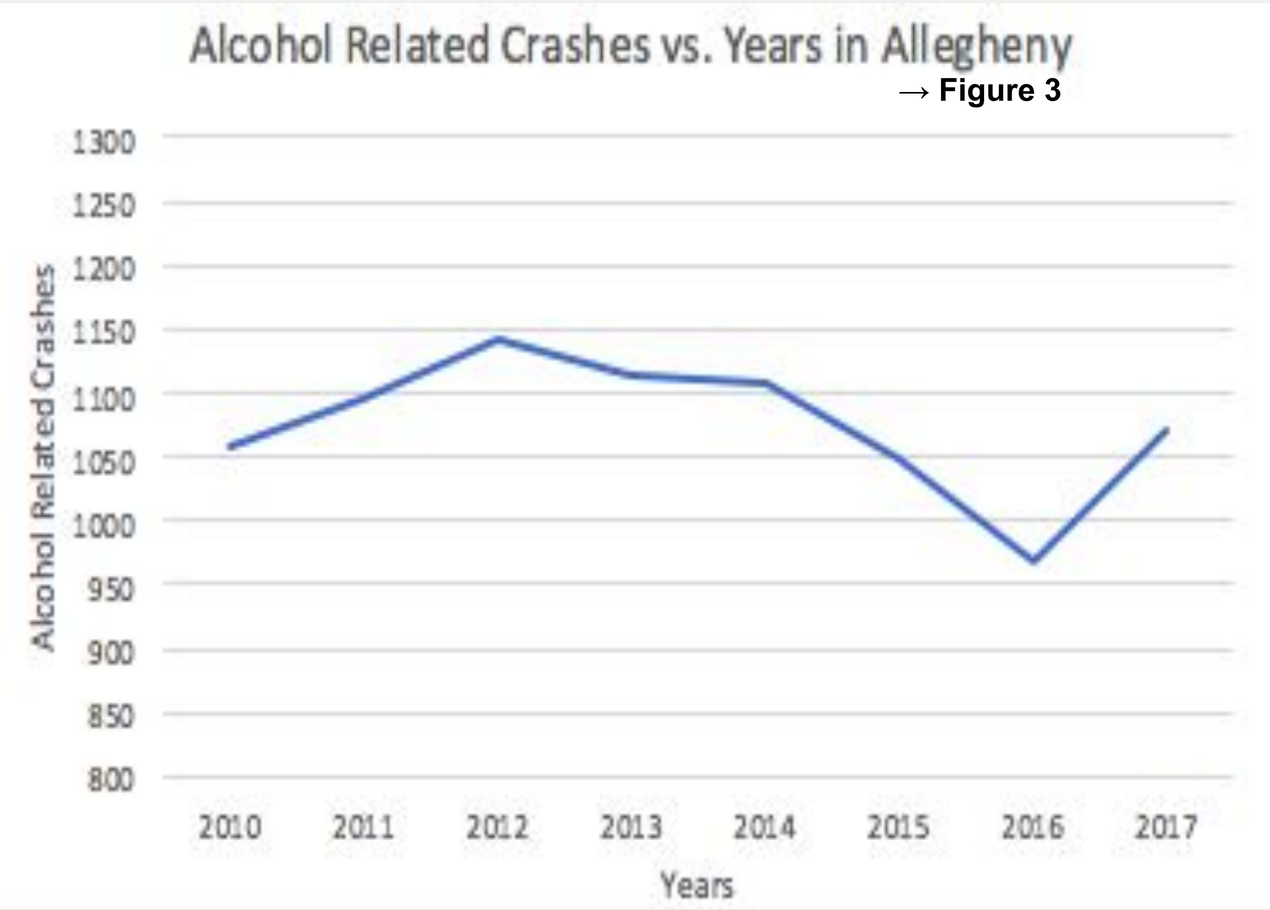


Figure 2- Drunk driving rates have fluctuated between 2010 and 2015, now beginning on a gradual descent. Uber and Lyft had been banned from Austin, TX (Travis county) from 2016-2017.



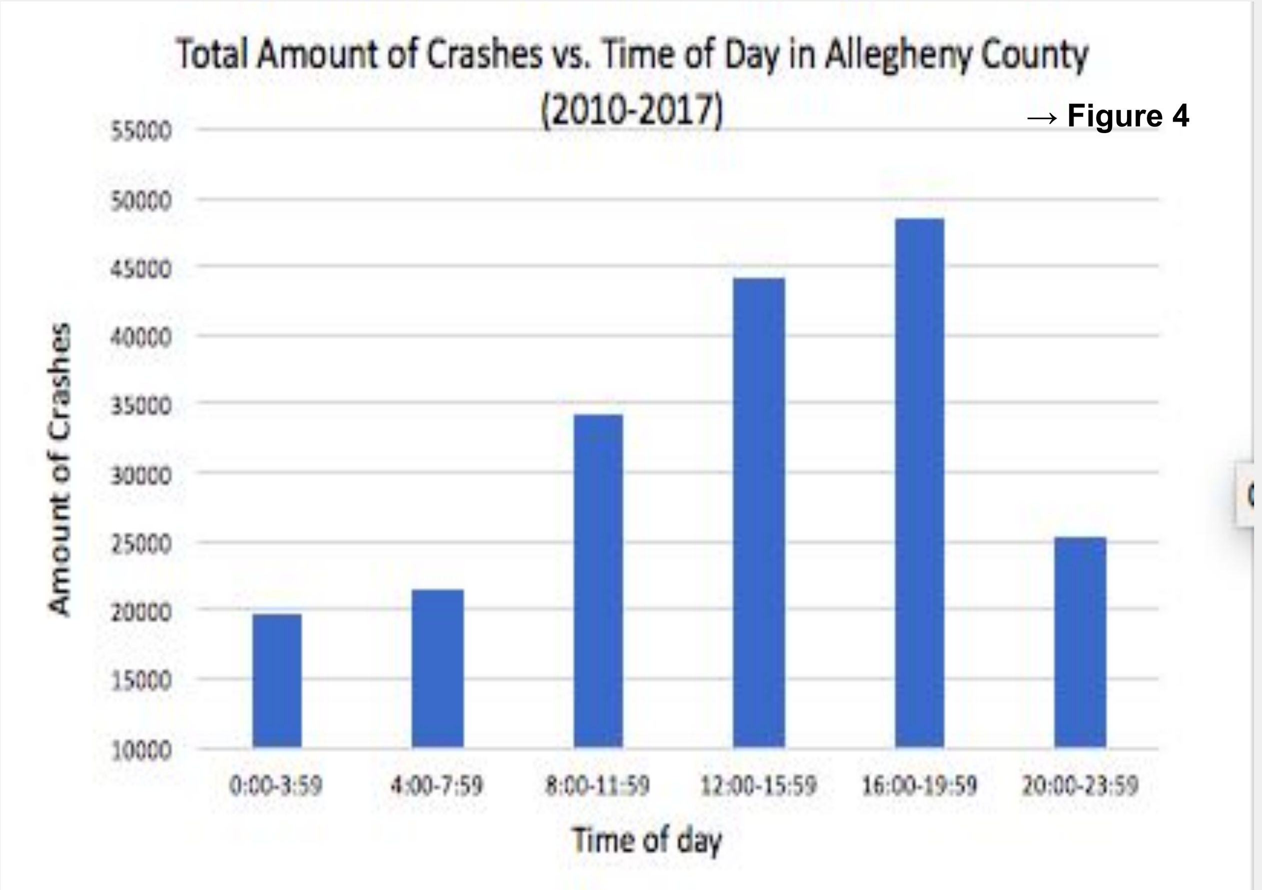


Figure 4- This graph is about the total amount of crashes from 2010 to 2017 of Allegheny County versus the time of day, which shows crashes happened, in the afternoon and evening.

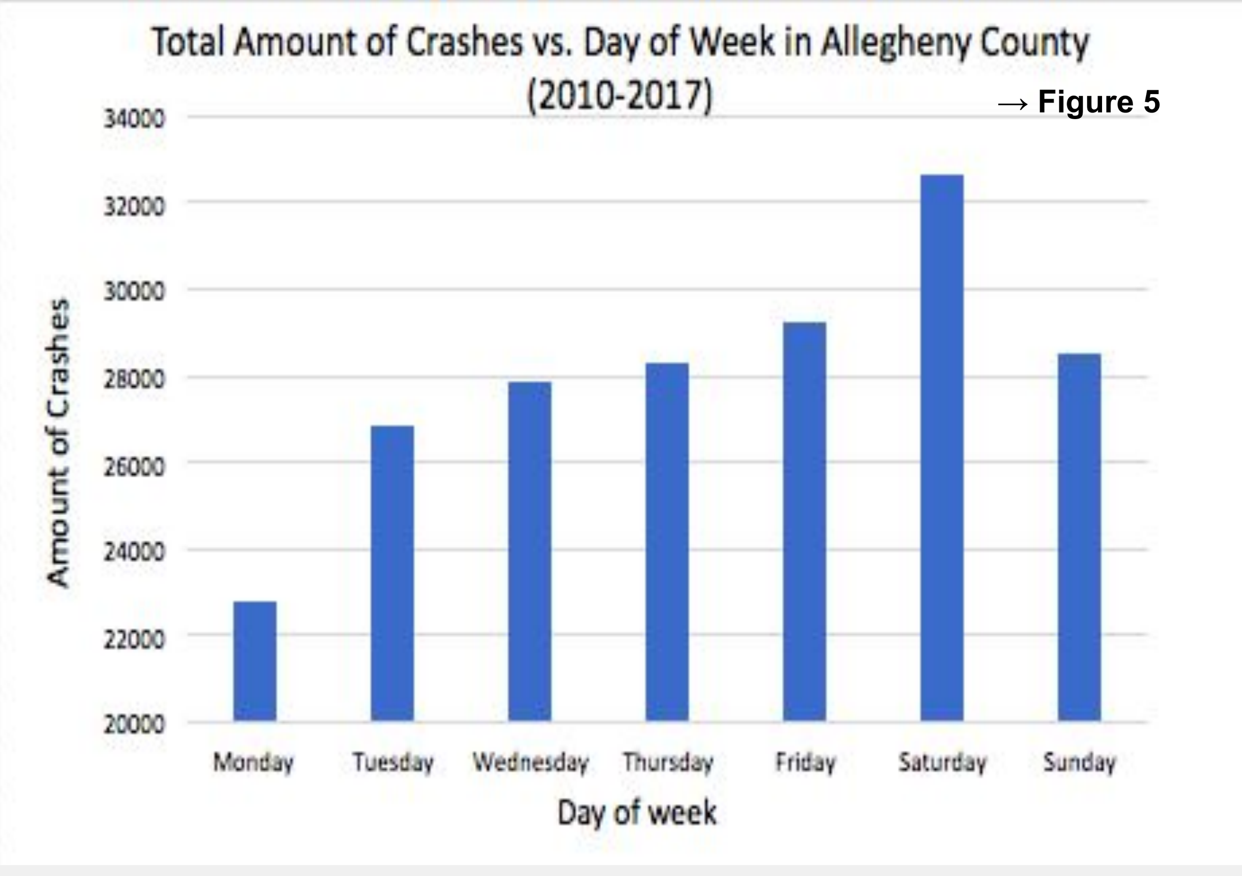


Figure 5- This graph is about the total amount of crashes from 2010 to 2017 of Allegheny County versus the day of week, it shows that crashes are more likely to happen on Fridays and Saturdays and less likely on Mondays.

Conclusion

In conclusion, the data displays an increase in DUI crashes between 2010 and 2018, and in Allegheny county the time of day/ day of the week relative to crashes. These findings show that our take on Uber or Left impacting inebriated driving and crashes is significant, although very recent. The data present the idea that these ride-sharing apps have curbed alcohol involved crashes, but only very recently in Travis county and even more minimal if not opposite results in Allegheny county. The data from Travis county was greatly affected by the removal and reintroduction of uber. Allegheny county has experienced an overall low in crashes but DUI's have maintained a pretty consistent record. We noticed while gathering data that our results could also have stemmed off from the fact that the company's name didn't pick up for a while after its arrival. The correlation between drunk driving and Uber/Lyft is noticeable in times of day in Allegheny county and Travis county (Austin) as of 2017. Overall, the data is practical to rideshare users and even authorities to see the importance of implementing safe alternatives for driving home.

Challenges Faced While Gathering Data

- Had to use third-party websites to gather crash data.
- Did not have any information directly from the ride-sharing services.
- The parameters for what constitutes a DUI in each city observed
- Police actively looking for drivers under the influence more frequently at some times of day than others.
- Tried to research cities in Oregon where Uber was banned, but crash data was only published through 2014

UBER

Is there a correlation between rates of depression, anxiety, and fatal drug overdoses in Allegheny County communities?

(Propel Andrew St. High School: Tiara Tabb, Eric Grimm, Nyla Demery, Quinton Moon, Da'Mani Washington, Destiny Travillion, Jhordan Price

INTRODUCTION:

We believe that drug abuse and mental health are both extremely important, and often neglected, issues in American society. This is particularly a problem in the Pittsburgh area and Rust Belt region. If there are less mental illness issues we think there would be less drug abuse and vice versa. Fighting the opioid/drug epidemic is one of the few issues that has bipartisan support in American politics. Mental health and/or drug abuse can affect anyone regardless of background. Maybe if a connection to opioid abuse and mental health is established, more resources can be given and action taken to solve this crisis and save lives.

METHODS:

With assistance from the Pitt students Evan and Quinn, we examined data locally from Allegheny County communities examining the rate of people who have prescribed depression and anxiety medication. This was compared to data on fatal accidental overdoses in each zip code of the county. Our data was found from the Western Pennsylvania Regional Data Center (WRPDC) along with sources used for lining up the populations of census tracts with zip codes. After the data was gathered it was analyzed by using formulas, charts, and graphs from Microsoft Excel and data maps from Tableau Public. Outliers needed to be eliminated to draw a conclusion.

CHALLENGES:

-The prescription medication data was by census tract and the overdose data was by zip code. A lot of work was done and resources used to line up census tract populations to zip codes with a site that was found.

-We temporarily used data from the wrong prescription drug category because the data sets were large and coded in a complex manner.

-Lack of Microsoft Excel knowledge or experience.

-Tableau Public was used as new data mapping resource we were all inexperienced with.

-Lack of experience, time, and sometimes motivation for our team.

DATA SETS COMPILED:

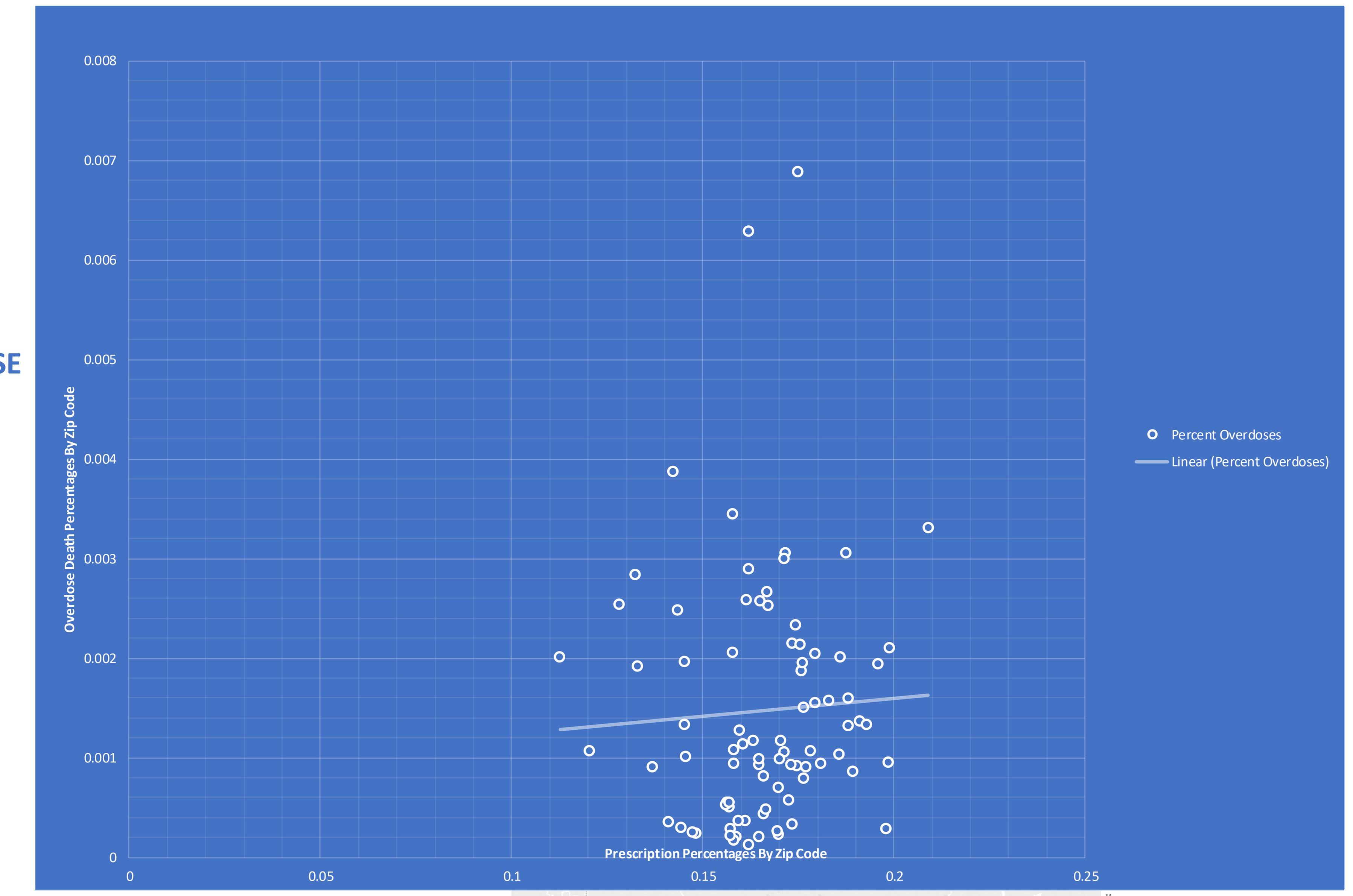
Zip Codes	Sum of Overdoses	Anx Prescriptions Per Zip Code.Column1	Anx Prescriptions Per Zip Code.Column2	Anx Prescriptions Per Zip Code.Column3	Anx Prescriptions Per Zip Code.Column4	Sum of Overdoses	Percent Overdoses	Outliers
13219	1					1		
15003	1	15003	375.0120363	62.53586368	0.166756951	1	0.002666581	
15010	1					1		
15012	1	15012	3.526315789	0.656140351	0.186069652	1		0.283582
15014	2	15014	1459.653285	278.9781022	0.19112628	2	0.001370188	
15017	4	15017	6870.063825	1186.096414	0.172647073	4	0.000582236	
15024	1	15024	4735.826181	779.8961571	0.164680064	1	0.000211156	
15025	9	15025	8301.844248	1313.065161	0.158165478	9	0.001084096	
15034	3	15034	904.5959302	189.0794574	0.209020902	3	0.003316398	
15035	3	15035	979.7365761	183.7138345	0.1875135	3	0.003062048	
15037	6	15037	5637.920307	965.5977474	0.171268428	6	0.001064222	
15044	4	15044	13977.11927	2199.04921	0.157332077	4	0.000286182	
15045	4	15045	2058	403	0.195821186	4	0.001943635	
15049	1	15049	473.8787375	94.27701447	0.198947551	1	0.002110245	
15057	4	15057	2038.326001	296.1558393	0.145293657	4	0.001962395	
15063	1	15063	496.5052632	92.3845614	0.186069652	1	0.002014077	
15065	9	15065	5723.311368	1046.848809	0.182909638	9	0.001572516	
15066	1					1		
15068	4	15068	1309.320449	224.7069758	0.171621069	4	0.00305502	
15071	1	15071	4709.646176	747.5725274	0.158732206	1	0.00021233	
15074	2					2		
15076	2	15076	579.0458002	91.43898923	0.157913224	2	0.003453958	
15084	7	15084	4650.514889	820.8868969	0.176515271	7	0.00150521	
15085	2	15085	52.19888717	8.956265115	0.171579618	2		0.038315
15088	2	15088	0	0		2		
15090	4	15090	11056.54589	1559.725408	0.141068054	4	0.000361777	
15101	5	15101	13581.55282	2186.592976	0.160997274	5	0.000368146	
15102	7	15102	16026.11139	2661.230928	0.166055936	7	0.000436787	
15104	8	15104	3221.574164	462.5204407	0.143569701	8	0.002483258	

REFERENCES:

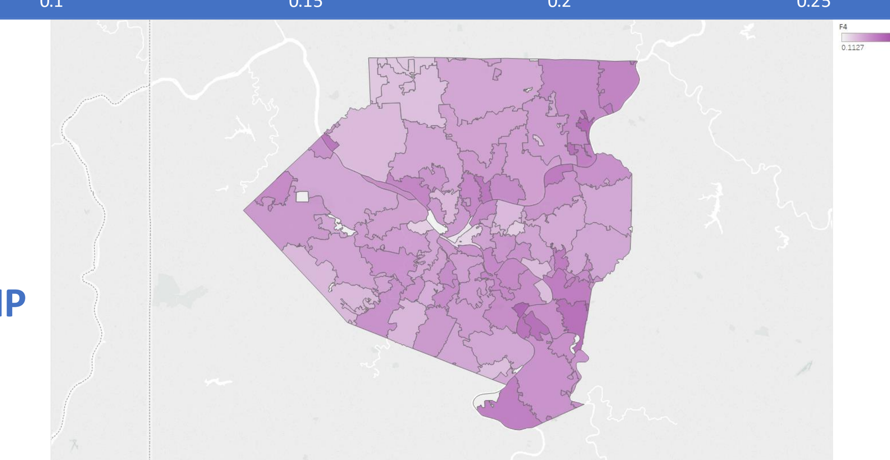
1)WPRDC (Allegheny County Fatal Accidental Overdoses)
https://data.wprdc.org/dataset/allegheny-county-fatal-accidental-overdoses
2)WPRDC (Allegheny County Anxiety Medication)
https://data.wprdc.org/dataset/anxiety
3) WPRDC (Allegheny County Depression Medication)
https://data.wprdc.org/dataset/allegheny-county-depression-medication

DATA RESULTS:

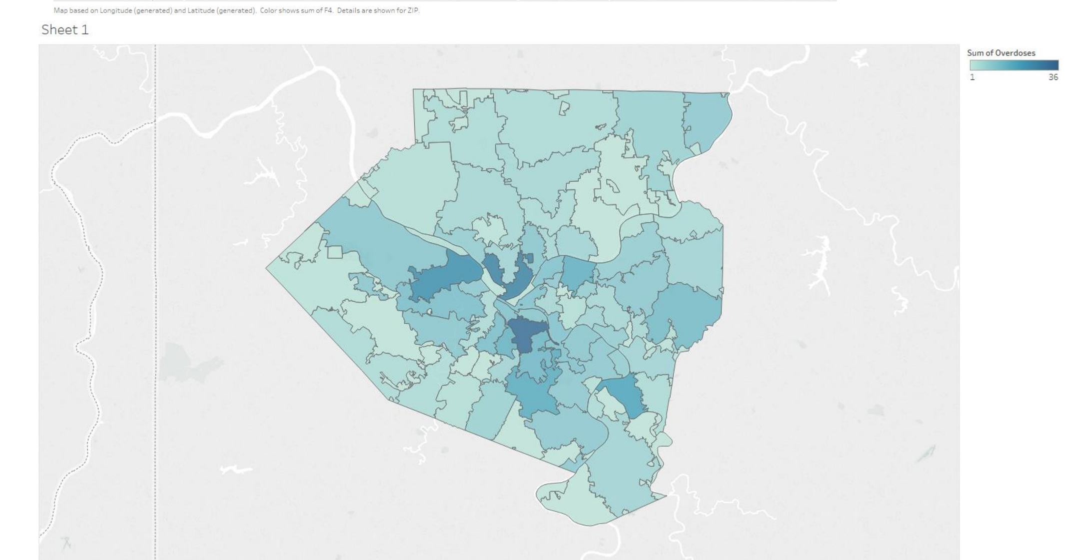
PRESCRIPTION VS. OVERDOSE PERCENTAGES BY ZIP CODE



MAP OF PRESCRIPTION PERCENTAGES BY ZIP CODE



MAP OF OVERDOSE PERCENTAGES BY ZIP CODE



ANALYSIS AND CONCLUSIONS:

From examining the graphs and maps there is a very small correlation between the rate of anxiety and depression prescriptions and overdose deaths in Allegheny County. The data analysis indicates that some communities have a "healthier" ratio between prescription and overdose rates (Fox Chapel, Upper St. Clair, etc.) than others (Mt. Oliver, McKees Rocks, etc.).

We speculated that some "Hilltop Communities" along with others may have less access to mental health/medical care than neighborhoods in the city and other locations. Communities may have disproportionate ratios in terms of being under-prescribed medication but with many sources of illegal narcotics in their neighborhoods. Citizens of some communities with more "unhealthy" ratios may be more hesitant to pursue professional mental health treatment as well because of stigmas and/or access.

Does the Presence of Full-day Vs. Half-day Kindergarten Affect PSSA Scores?

Plum Senior High School DataJam Team: Maria Czura, Gabriel Esposto, Abigail Haerr, Mackenzie Linderman, Rudra Thakkar, and Lucas Wycich

Our Data Sets

The data sets we used were the PSSA scores 2015/2016 4th grade Math and Enrollment 2010/2011 Kindergarten Half vs. Full vs. Half/Full. Our first search of datasets was very broad, covering multiple years of PSSA results and enrollment counts. When we chose a specific area of focus, we first turned toward the LEA Enrollment 2010/2011. We chose the 2010/2011 school year because the structure of the data in the Excel spreadsheet was consistent with the structure of the data in the years that followed. Choosing 2010/2011 as the targeted school year caused us to use the earliest corresponding year with the PSSA score results, which turned out to be the 2015 results. From there we focused on collecting data from the 4th grade classes because 4th grade students would be the only students who were affected by kindergarten from the 2010/2011 school year. The data sets used were the Enrollment Public LEA 2010-2011 and the 2015 PSSA School Level Data. Within the Enrollment Public LEA 2010-2011 spreadsheet, the Local Education Agencies and number of children being enrolled in kindergarten are displayed. The 2015 PSSA School Level Data showed the different schools, the percent of PSSA scores in each range (Advanced, Proficient, Basic, Below Basic), and also incorporated what district each school fell into. Some typical values for these data sets are percentages from test scores and the number of students enrolled in half-day or full-day kindergarten. Displayed to the right are two samples of our final data charts; the top one being the percentages of students that attend full day kindergarten with their PSSA scores and the bottom being the percentages of students that attended half day kindergarten with their PSSA scores.

1	Schools	Advanced	Proficient	Basic	Below Basic	Full Day		
3	Abington SD	25.6125	40.29402174	24.68369565	9.405434783	TRUE		
4	Albert Gallatin Ar	10.72375	31.64125	30.49	27.140625	TRUE		
5	Aliquippa SD	2.2175	18.845	36.3875	42.5375	TRUE	Average of all Advanced	13.56
6	Allegheny Valley	16.99772727	32.65681818	29.09772727	21.23636364	TRUE	Average of all Proficient	26.81
7	Allegheny-Clarior	11.41	30.1325	35.86	22.5875	TRUE	Average of all Basic	28.39
8	Altoona Area SD	12.58173077	37.08173077	30.67644231	19.66634615	TRUE	Average of all Below Basic	31.24
9	Ambridge Area S	15.34431818	31.55909091	28.33068182	24.76704545	TRUE		
10	Antietam SD	10.2525	25.5075	34.765	29.475	TRUE		
11	Athens Area SD	10.12333333	31.62833333	33.76333333	24.50166667	TRUE		
12	Avella Area SD	13.0125	33.7675	32.41	20.805	TRUE		
14	Bald Eagle Area	20.95	38.75263158	28.34342105	11.95657895	TRUE		
15	Baldwin-Whiteha	21.43833333	37.04333333	26.9	14.60666667	TRUE		
16	Bangor Area SD	12.82666667	34.88166667	28.83333333	23.45	TRUE		
17	Bedford Area SD	12.685	36.535	31.6175	19.1625	TRUE		
18	Belle Vernon Are	15.659375	37.475	31.3015625	15.5765625	TRUE		
19	Bellefonte Area S	21.576	36.143	25.366	16.909	TRUE		
20	Bellwood-Antis S	17.1975	36.0125	27.81	19.0125	TRUE		
21	Benton Area SD	19.3175	38.1275	28.9125	13.6375	TRUE		
22	Bentworth SD	14.89	35.655	28.8725	20.595	TRUE		
23	Berlin Brothersva	14.8675	38.305	30.315	16.51	TRUE		
24	Berwick Area SD	18.499	38.762	27.217	15.512	TRUE		
26	Bethlehem-Cente	7.65	28.8675	37.7675	25.7275	TRUE		
27	Big Beaver Falls	11.83333333	32.61833333	31.96	23.58	TRUE		

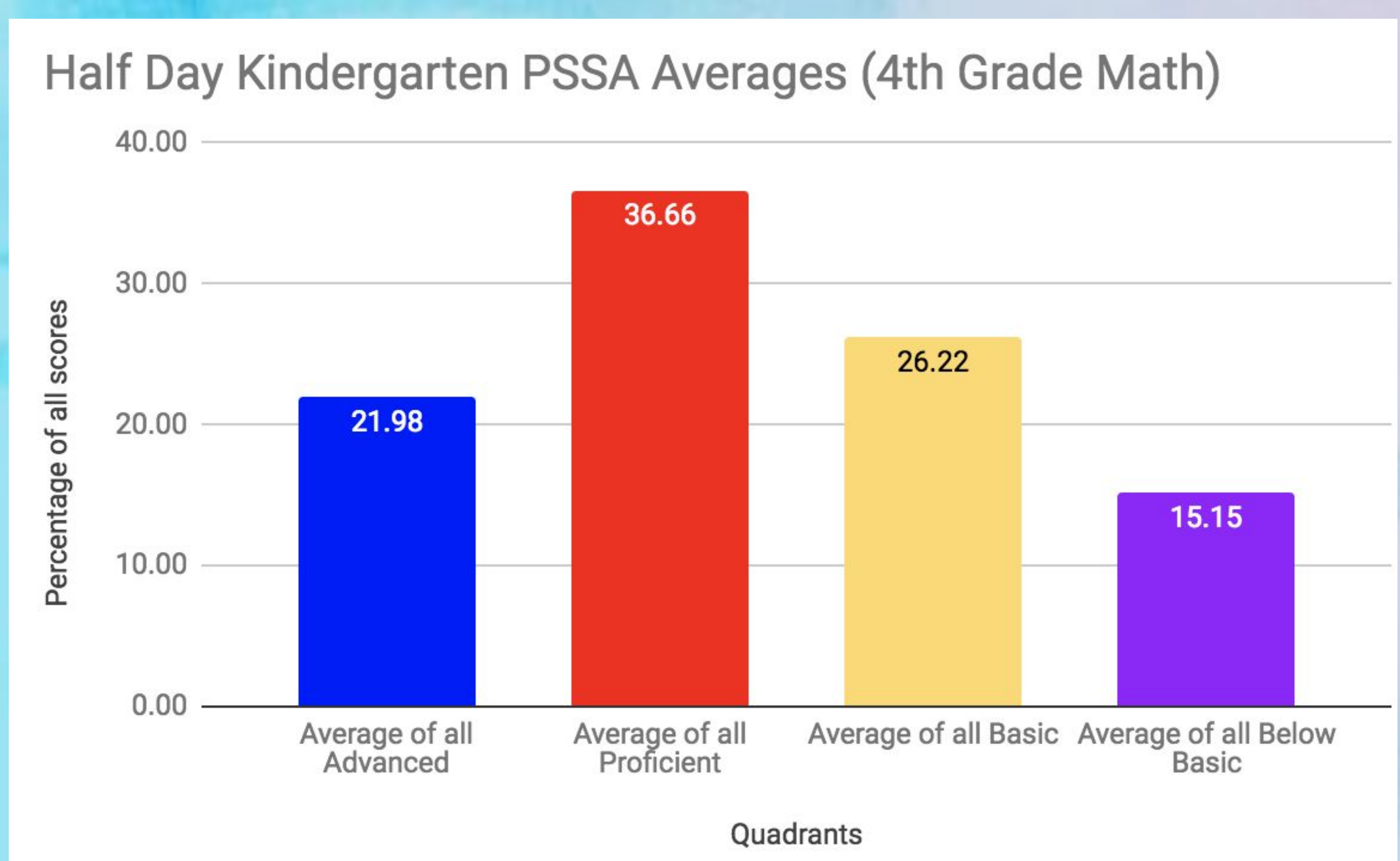
Student's PSSA Scores who attended full day kindergarten

1	Schools	Advanced	Proficient	Basic	Below Basic	Kindergarten Pr	Half Day		Average of all Advanced	21.98
2	Abington Heights	21.77435897	40.00512821	25.07051282	13.15769231	TRUE	TRUE		Average of all Proficient	36.66
3	Avon Grove SD	18.76	33.8925	28.4425	18.915	TRUE	TRUE		Average of all Basic	26.22
4	Bethel Park SD	22.59814815	39.23148148	25.78055556	12.37407407	TRUE	TRUE		Average of all Below Basi	15.15
5	Boyertown Area	22.2775	39.965	25.7375	12.0155	TRUE	TRUE			
6	Camp Hill SD	18.8725	37.1875	29.18	14.77	TRUE	TRUE			
7	Catasauqua Area	17.84	32.88	29.995	19.2925	TRUE	TRUE			
8	Central Dauphin	15.30029412	34.47147059	30.43735294	19.79823529	TRUE	TRUE			
9	Cocalico SD	20.88375	39.2825	25.835	14	TRUE	TRUE			
10	Dallas SD	25.91833333	39.565	24.95333333	9.573333333	TRUE	TRUE			
11	Derry Township S	24.37272727	36.30681818	23.78409091	15.54545455	TRUE	TRUE			
12	Garnet Valley SD	30.25166667	37.065	23.64166667	9.048333333	TRUE	TRUE			
13	Gateway SD	15.79130435	33.95434783	28.33152174	21.91304348	TRUE	TRUE			
14	Hampton Towns	34.5	36.65833333	20.32222222	8.522222222	TRUE	TRUE			
15	Jeannette City S	10.7	31.795	33.4975	23.99	TRUE	TRUE			
16	Lampeter-Strasb	24.23	35.9075	24.5425	15.315	TRUE	TRUE			
17	Leechburg Area	11.7375	32.9275	32.065	23.2525	TRUE	TRUE			
18	Mars Area SD	22.73636364	39.26818182	26.67954545	11.32272727	TRUE	TRUE			
19	Mid Valley SD	11.36	31.5575	32.695	24.39	TRUE	TRUE			
20	Moniteau SD	16.5825	37.0775	31.8275	14.53	TRUE	TRUE			
21	North Allegheny	29.38988095	35.01904762	21.86369048	13.72142857	TRUE	TRUE			

Student's PSSA Score who attended half day kindergarten

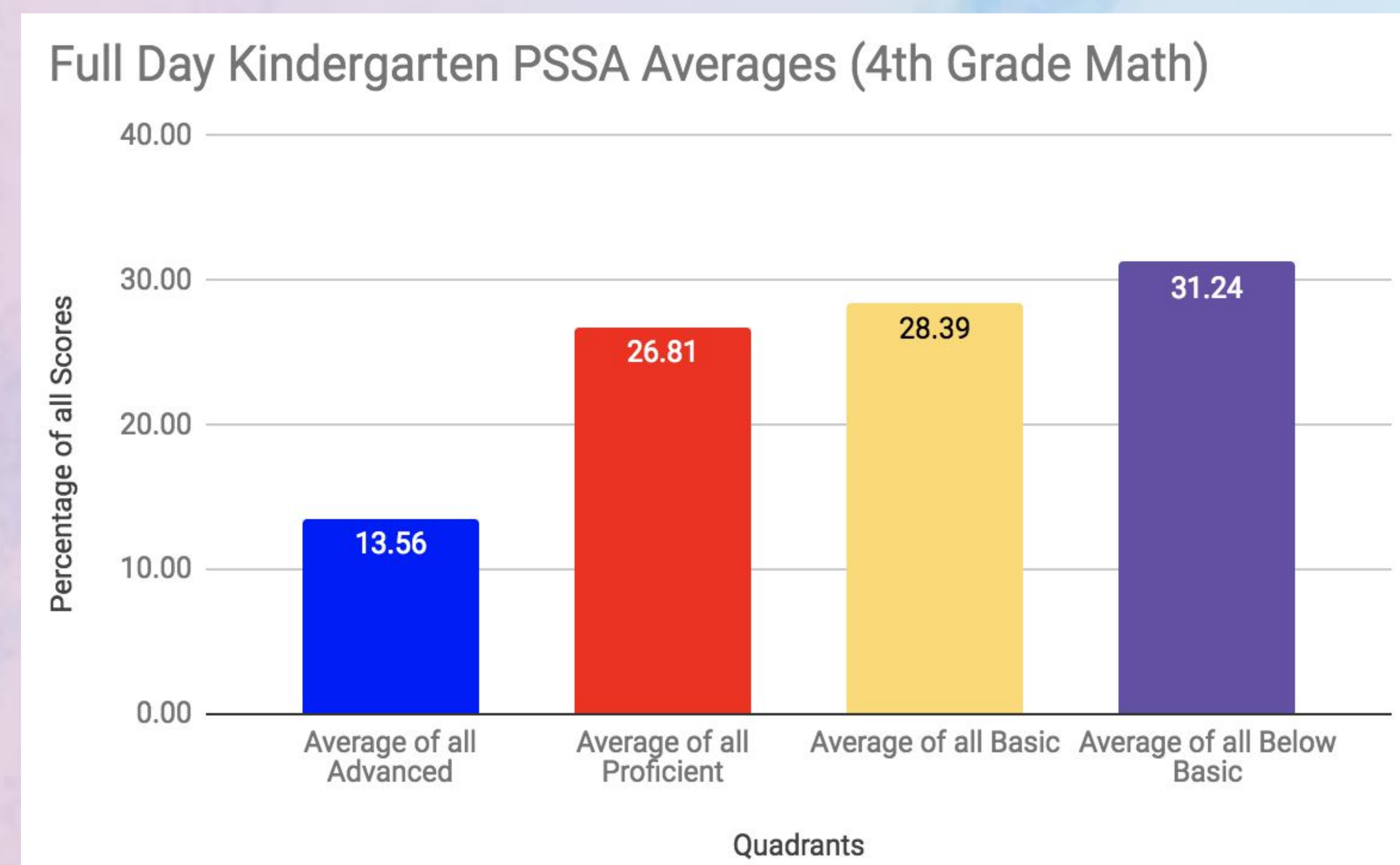
Challenges We Faced

During the time spent on our DataJam project, our team faced many challenges that we had to overcome. The first problem we encountered was finding data sets that both satisfied the question and were also formatted similarly to keep our findings consistent. Another difficulty that we faced was using our school provided Chromebooks, which do not have access to Excel. This compelled us to use Google Sheets that did not have the formatting capabilities to "conditional format" on the occurrence of like-values in two separate columns across the same row. Within our data sets, we found that every public school in Pennsylvania offers kindergarten in some form. We then had to expand our question to full-day kindergarten versus half-day kindergarten, as well as filtering out cyber and charter schools, as they open and close frequently, causing them to not have the data necessary. Although our team faced many challenges throughout our journey, we were able to work together to problem solve and use our resources to our best abilities.



Our Results/Impact and Value

After our extensive research and data collecting, our data shows that half-day kindergarten has a higher percentage of students scoring at an advanced and proficient level than full-day kindergarten does. Based on the outcome of the data collected, we believe that more research should be conducted to confirm which form of kindergarten (full-day or half-day) produces the best math PSSA scores. We came to the conclusion that outside factors such as funding, culture, as well as demographics and socioeconomics may have also affected the scores. Due to our data sets having more statistics on full-day kindergarten than half-day, it may skew the data and alter the results.



Is the Allegheny County Drug Epidemic Linked to Income?

Matias Badino, Thomas Brusilovsky, Sara Liang, Maggie Lincoln, Abigail Miller-Peterson, Hannah Seeley

Pittsburgh Allderdice High School

Introduction

The drug epidemic has been classified as a serious issue by the American public in recent years. The epidemic started largely due to doctors prescribing addictive drugs to hapless patients who continued to use them long after they were strictly necessary. Today, prescription drugs are still the gateway to addictive substances, with <u>almost 12% of patients prescribed addictive pain medicine abusing it</u>. Addiction rates are still increasing considerably every year, with the government spending considerable money in its attempt to combat the issue.

Allegheny County is one of the hardest hit areas in the United States by the epidemic, <u>ranking third nationally in overdose deaths from fentanyl</u>. Every year since 2014 <u>the number of overdose deaths in the county has increased</u>, underlining the importance of researching the crisis.

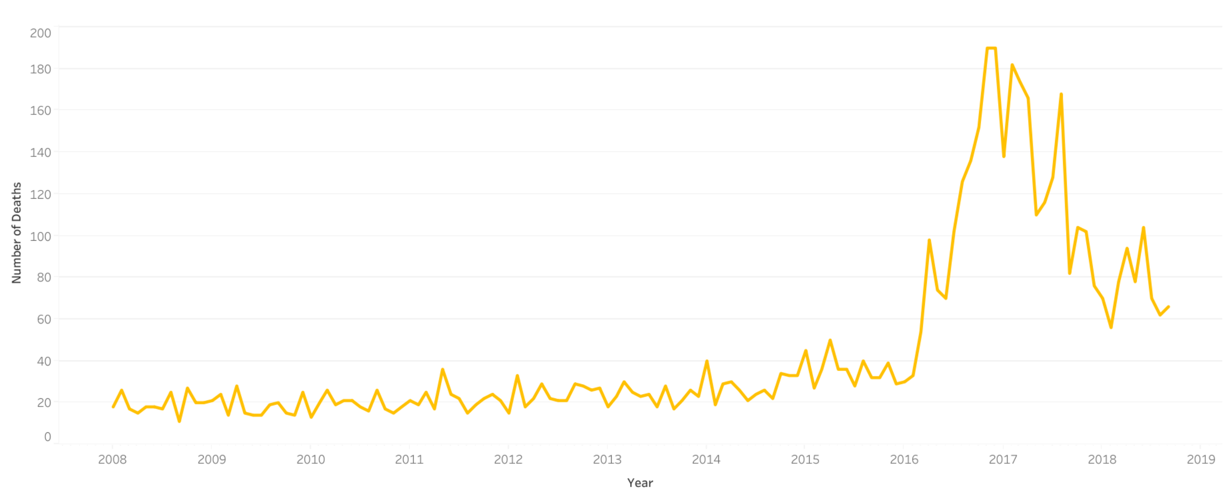


Figure 1: Number of accidental drug-related deaths in Allegheny County by year, 2008—2018.

Research Question

What relationship exists between income and the number of drug overdoses in Allegheny County?

Hypothesis

We hypothesize that there is a negative correlation between income and fatal overdoses in Allegheny County. Due to higher stressors, poorer people are more likely to resort to drug use, and therefore more likely to accidentally overdose. Additionally, those with higher incomes have more resources for overcoming addiction, and can avoid the more dangerous drug types.

Data Collection

To determine whether the two factors are linked, we obtained overdose data and median household income data in Allegheny County. Challenges we faced included finding data by ZIP code, finding data in the correct file type, and finding complete datasets.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	death_date_and_time	manner_of_death	age	sex	race	incident_zip	combined_od1	combined_od2	combined_od3	combined_od4	combined_od5	combined_od6	combined_od7	case_year
2	2016-03-13T01:40:00	Accident	33	M	W	15204	AMPH	BUTAL	CARIS	DIAZEP	DIPHEN	KETA	METHAD	2016
3	2016-03-21T04:08:00	Accident	33	F	W	15003	ALPRAZ	COCAIN	CODEI	FENTAN	HEROIN			2016
4	2016-03-15T09:28:00	Accident	34	F	W	15106	ALPRAZ	HEROIN						2016
5	2016-03-05T11:40:00	Accident	27	M	W	15102	COCAIN	FENTAN	HEROIN					2016
6	2016-03-19T12:33:00	Accident	42	M	W	15147	ALPRAZ	HEROIN						2016
7	2016-03-19T02:09:00	Accident	48	F	W	15146	FENTAN							2016
8	2016-03-04T16:43:00	Accident	56	M	B	15110	COCAIN							2016
9	2016-04-09T20:34:00	Accident	43	M	W	15220	ALPRAZ	CLONA	DIAZEP	OXYCOD				2016
10	2016-03-29T11:13:00	Accident	54	M	B	15219	COCAIN	HEROIN						2016
11	2016-04-02T10:12:00	Accident	58	M	B	15201	COCAIN							2016
12	2016-03-30T08:43:00	Accident	55	F	W	15108	ESCT	DIPHEN	HEROIN					2016
13	2016-03-24T17:40:00	Accident	49	F	W	15065	ALCOHO	DIPHEN	METHAD					2016
14	2016-04-02T15:38:00	Accident	60	M	W	15207	ESCT	HEROIN						2016
15	2016-04-01T09:15:00	Accident	36	M	W	15212	ALPRAZ	COCAIN	HEROIN	LAMO				2016
16	2016-04-06T06:57:00	Accident	28	M	W	15227	BENZOD	HEROIN						2016

Figure 2: Data from the Allegheny County Health Department on overdose deaths in the county. Contains the date and time of death, manner of death (our data used only accidental deaths), age, sex, race, ZIP code, and drugs involved.

	A	B	C	D
1	Zip	Median	Mean	Pop
4247	15202	45,961	57,053	20,840
4248	15203	39,541	56,249	9,631
4249	15204	32,525	40,651	9,611
4250	15205	47,441	59,777	22,850
4251	15206	39,661	52,125	31,493
4252	15207	35,639	47,770	13,181
4253	15208	37,759	50,215	13,271
4254	15209	54,616	63,156	12,946
4255	15210	31,997	41,291	31,279
4256	15211	44,086	61,197	12,234
4257	15212	35,831	49,810	32,013

Figure 3: Data from the U.S. Census Bureau on median household income. Includes ZIP code, income (median and mean), and population.

Visualizations

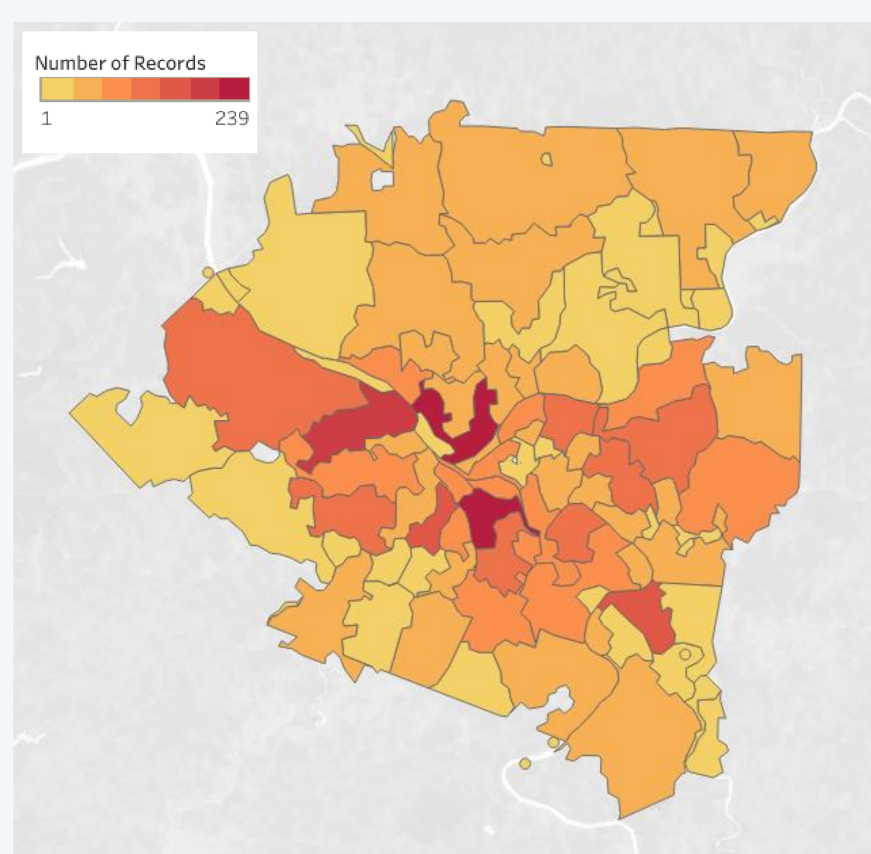


Figure 4: The number of fatal accidental overdoses by ZIP code in Allegheny County. ZIP code 15210 has the highest number of records (239), followed by ZIP code 15212 with 236.

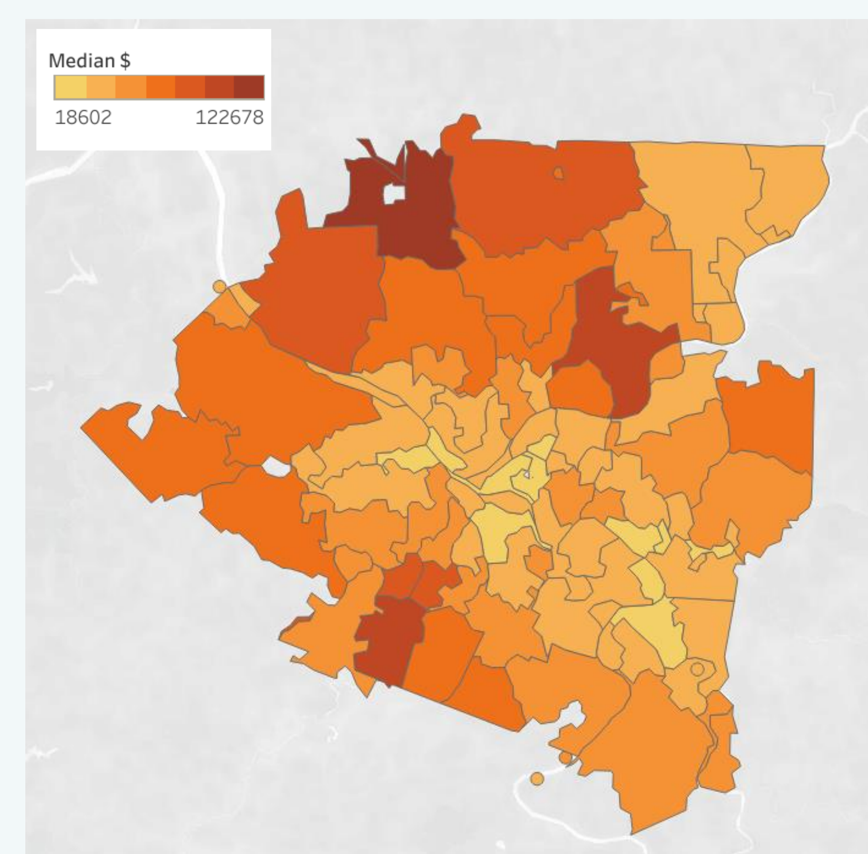


Figure 5: The median household income by ZIP code in Allegheny County. ZIP code 15090 has the highest median income ($110,788), while ZIP code 15233 has the lowest ($18,602).

Analysis

Our visualizations showed that generally, more accidental overdoses were recorded in the city, while median household income increased moving out toward the suburbs. However, to see whether a true relationship existed, we made a scatterplot from which we were able to extrapolate the correlation, along with its strength.

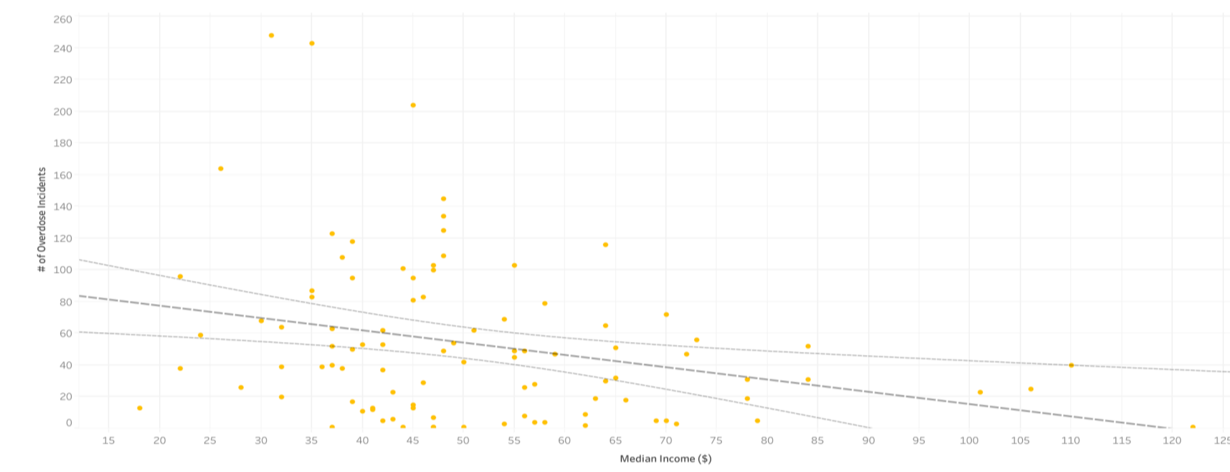


Figure 6: Number of overdose deaths vs. median income. R^2=.084, P=.0042.

Conclusions

There is a slight negative correlation between income and accidental overdoses in Allegheny County. With a p-value of .0042, <u>our model is significant</u>, but our R^2 value of .084 means that <u>it doesn't explain the variation in the data</u>. We conclude that our model shows that job-holding Americans still resort to drugs, and that <u>the opioid epidemic is affecting people of all incomes</u>. Possible errors in our analysis include that we may not have enough variation in income to see a trend, or may have used incomplete data.

<u>References</u>:

"Allegheny County Fatal Accidental Overdoses." *Western Pennsylvania Regional Data Center*. https://data.wprdc.org/dataset/allegheny-county-fatal-accidental-overdoses. Accessed 3 Mar 2019.

"Zip Code Characteristics: Mean and Median Household Incomes." *University of Michigan Population Studies Center*. https://www.psc.isr.umich.edu/dis/census/Features/tract2zip/. Accessed 3 Mar 2019.

Pittsburgh Allderdice HS
Data Jam Competition
April 4, 2019

Neighborhood Assets and Health

Samy Arunkumar
Will Ganger
Shane McKown
Kazuma Parkinson
Ben Cummings
Sean Graves
Dylan Olmsted
Maxim Yaskoiko

Background

- Community assets influence the **social determinants of health**.
- Community assets can **promote physical and mental health and facilitate social interactions**.
- Studies have shown that physical activity levels are affected by access to community assets like recreation centers.
- Community assets are often used to **revitalize struggling neighborhoods**.
- Many cities, including Pittsburgh, create new community assets to try to improve many aspects of their citizens' lives.

Source: Institute of Medicine

Purpose

- We aim to determine whether the **number of city-owned assets in a neighborhood** is associated with the **neighborhood's health**. We hypothesize that neighborhoods with greater numbers of assets will have a higher median age of death and lower obesity rate.

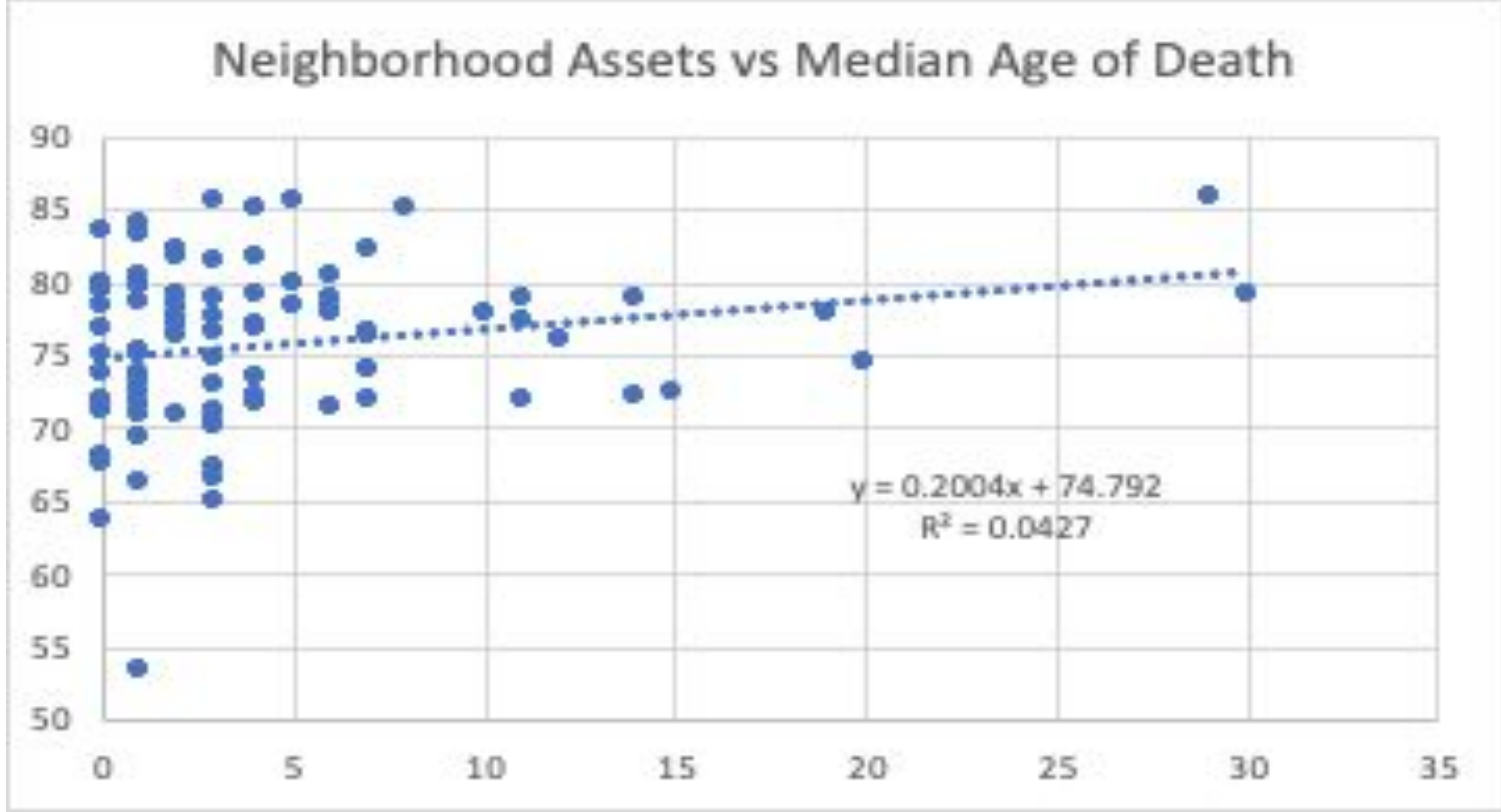
Data

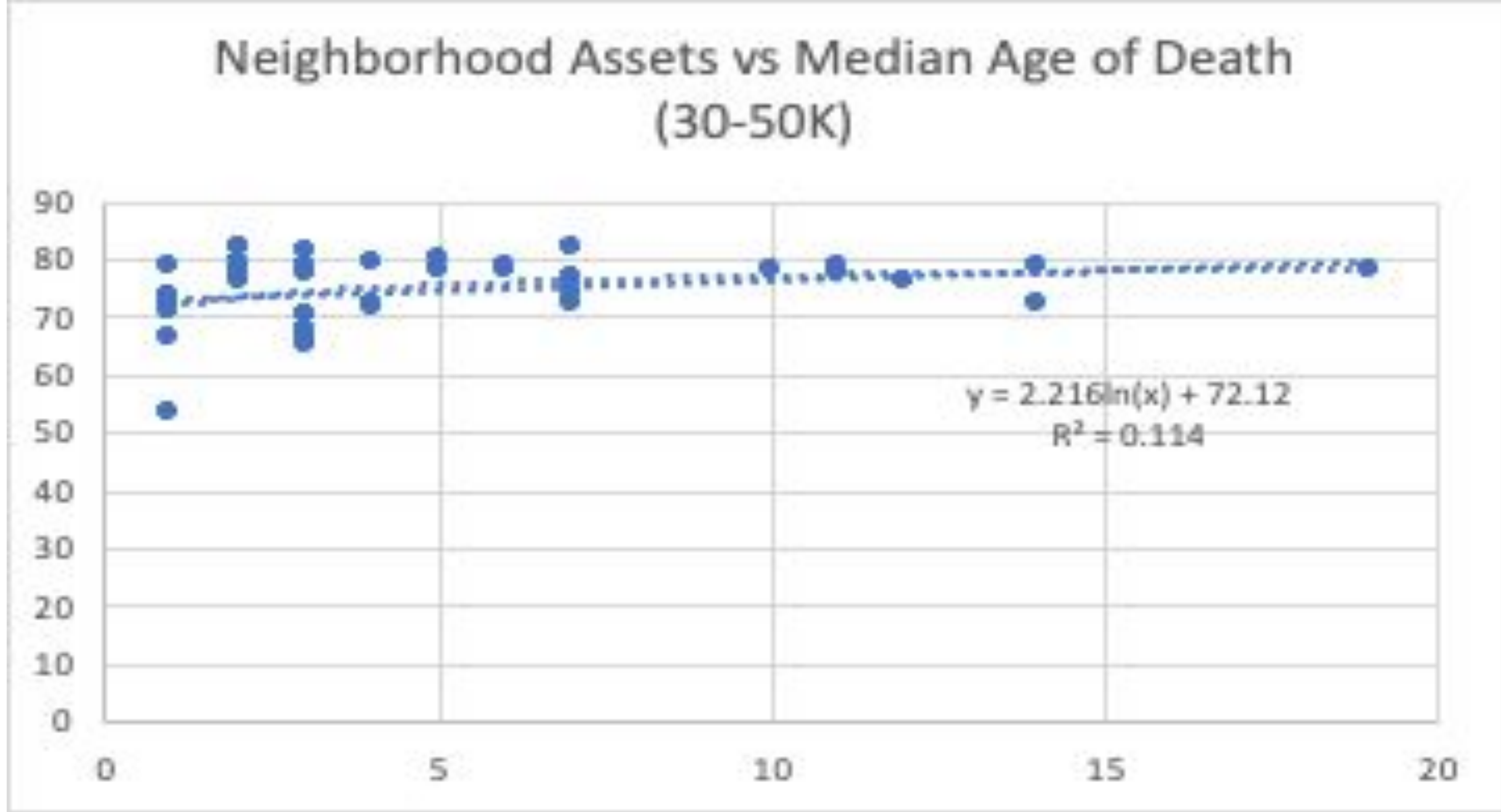
Name	Type	Neighborhood
Hazelwood Senior Center	Senior	Hazelwood

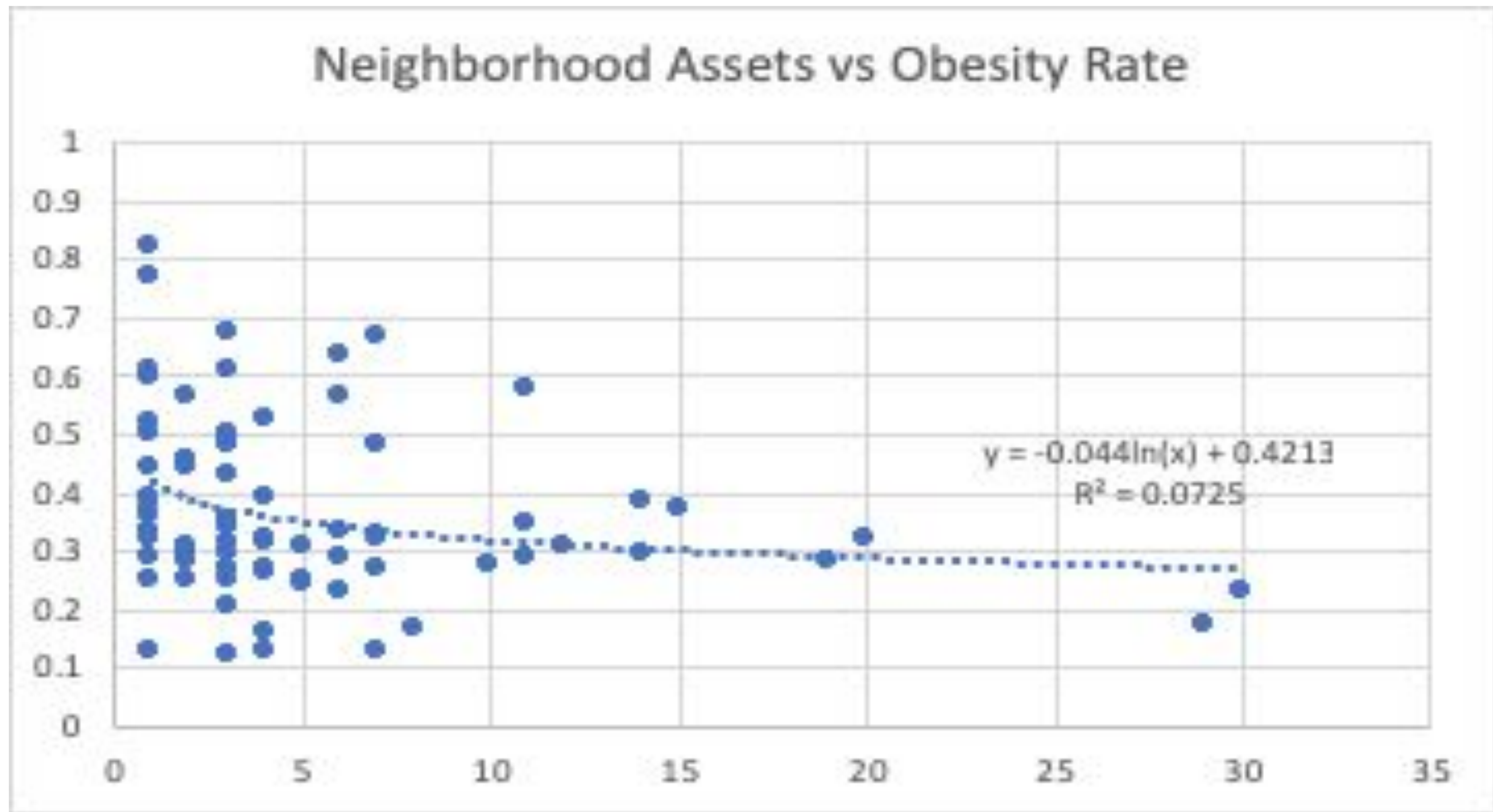
- Neighborhood asset data included over **400** rows of city facilities (e.g., community centers, offices, pools).
- Median age of death data was provided for **88 of the 90 neighborhoods** in Pittsburgh.
- Obesity rate data for all 90 neighborhoods spanned 2006-2010. Specific years were not provided so averages were used for neighborhoods with multiple data points.

Sources: Data.gov, Western PA Regional Data Center,

Results







Analysis

- The median age of death by neighborhood shows a **weak, positive correlation** with the number of neighborhood assets: **r^2 = 0.0427** (4.27% of the variability in the data can be explained by the linear association with neighborhood assets).
- When controlling for income, neighborhoods with median incomes between $30K and $50K had the strongest correlation between the number of neighborhood assets and median age of death: a **weak positive correlation with r^2 = 0.114**. However, the trend is more **logarithmic** than linear (a linear association has r^2 = 0.0669).
- The obesity rate by neighborhood shows a **weak, negative correlation** with the number of neighborhood assets: **r^2 = 0.0725**. This trend is more **logarithmic** than linear (a linear association has r^2 = 0.0569).

Challenges

- **The neighborhood assets data were limited to city-owned facilities** (Assets like the Carnegie Libraries not included).
- People are not confined to one neighborhood; they often use assets located in other neighborhoods.
- **Lurking Variables**: Health and life span are influenced by many variables (e.g., education, genetics, personal behavior). Other community factors like prevalence of gun ownership and proximity to unhealthy food options, tobacco vendors, and bars may influence health behaviors and outcomes.

Conclusion

- **Pittsburgh should continue to create new neighborhood assets.**
- Pittsburgh should determine which type of assets contribute most to health.
- In the future, we would like to add privately owned neighborhood assets to our analysis.

Does the presence of radon correlate with disease incidence?

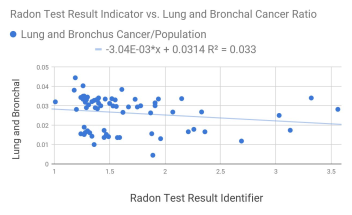
Background and Purpose

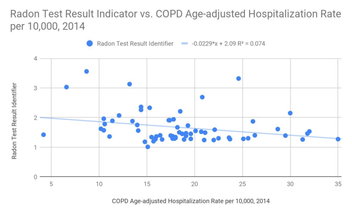
Radon is a naturally occurring radioactive gas produced when uranium, thorium, and radium break down. It is then released into the air. Radon is odorless, tasteless, and invisible. It can accumulate in places like underground mines and buildings. We thought this issue was especially relevant to us as Pennsylvanians because an estimated 40% of Pennsylvania homes have radon levels greater than the EPA guideline of 4 pCi/L.

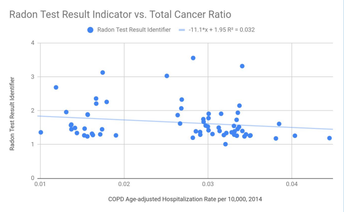
Gathering Data

Using Open Data PA, we used the data set Estimated Prevalence and New Diagnoses of HIV and HIV among Injection Drug Users by County (2012-2016) Health. We used data from the PA Department of Health Bureau of Health Statistics and Research 2012-2016 summary (pictured below) for our lung and bronchus, as well as general cancer data. In addition, we used data from the United States Census Bureau for our population data. Our radon data came from the Federated Radon Test Results 1989 to Current Environmental Protection 2016 section. Our COPD data came from the PA country health profile from the Division of Health Informatics June 2018 report. We used SAS University and google sheets for our analysis.

Data and Results







We found weak negative correlation between radon and each disease we studied.

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Measure_Value	Prevalence_HIV_Disease_Count
Measure_Value	1.00000 738560	-0.04899 <.0001 738148
Prevalence_HIV_Disease_Count	-0.04899 <.0001 738148	1.00000 738154

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Measure_Value	_percentage_total_cancer_populat
Measure_Value	1.00000 1048569	-0.00719 <.0001 1048569
_percentage_total_cancer_populat	-0.00719 <.0001 1048569	1.00000 1048575

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Measure_Value	percentage_lung_and_bronchal_pop
Measure_Value	1.00000 1048569	-0.00719 <.0001 1048569
percentage_lung_and_bronchal_pop	-0.00719 <.0001 1048569	1.00000 1048575

Using a Pearson correlation test on SAS, we were able to find statistical significance in the correlation between radon and each of the variables. All of them had negative correlations.

Analysis and Conclusion

Challenges in gathering data:

- ➔ We were not able to find data sets from the exact same time periods.
- ➔ Some disease incidences go untreated or unreported, and thus will not show up in data. In more rural counties where access to medical care is limited, this factor may have skewed results.
- ➔ Some of the radon that caused disease may have already been mitigated and thus won't show up in the data, even though it may be responsible.

Analysis:

- ➔ Surprisingly, we found weak negative correlation between radon and each of the studied diseases.
- ➔ This is an instance of Simpson's paradox because there is negative correlation between radon and lung cancer, which is in opposition to the national data.

Conclusions and recommendations:

- ➔ Our analysis does not definitively prove that radon does not cause these diseases, nor that it prevents them in any way.
- ➔ Radon testing kits are widely available. If you are worried about the presence of radon in your own home, purchase a test and based on the results, you can hire a radon mitigation contractor or visit www.radon.com for more information.

Upper St Clair - Brooke Christiansen, Mallika Matharu, Jack Clark, Dylan Jenny, Benny Pribanic, Ben Nelson, Ritvik Shah, Hersh Tripathi, and Scott Cheung