Pittsburgh Allderdice HS
Data Jam Competition
2020

Religious institutions and Health

Maxim Yaskolko
Nathan Feinstein
Jonah Saitz
Sam King
Eli Holst
Amos Chapman
Maxwell Seeley
Eliza Cham

Background

- **Religious Institutions** are organizations of individuals who **share religious** and **philosophical beliefs.**
  - Often manifested as a physical location for communities to gather/practice.
- Religious participation has always been a **prominent** component of **society** with **noticeable effects** on the **people**.
- Recent polls have shown that more than **80% of Americans** identify as **religious** while more than **60%** claim to be members of a **congregation**.
- Proponents of religious institutions argue that churchgoers enjoy **increased charitableness** and **better physical health.**

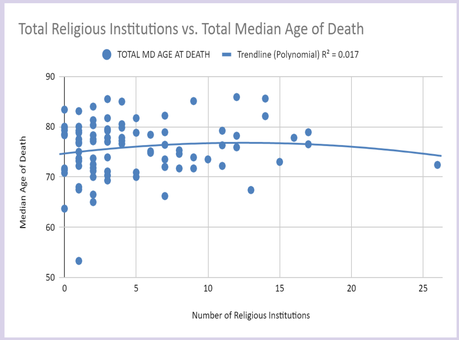
Purpose

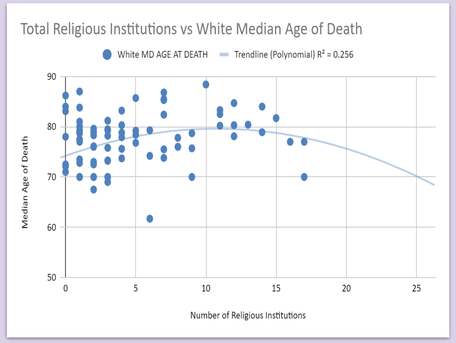
- We aim to determine whether the **number of religious institutions in a neighborhood** is associated with the **neighborhood's health**. We hypothesize that neighborhoods with greater numbers of institutions will have a higher median age of death and lower obesity rate.

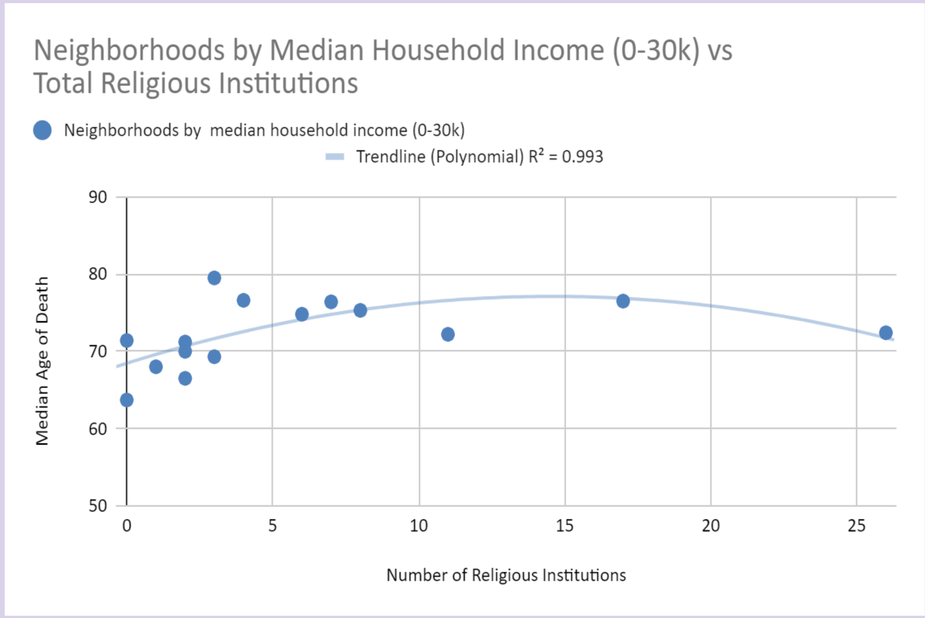
Data

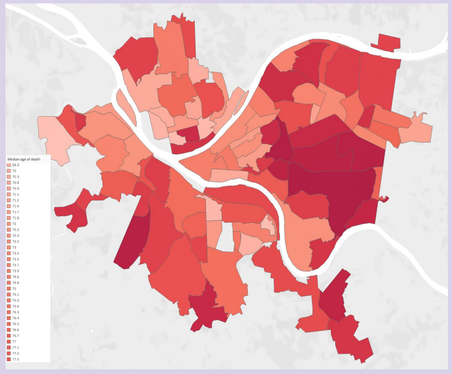
- Median age of death data was provided for **88 of the 90 neighborhoods** in Pittsburgh.
- Religious Institution data included **address**, zip code, and **religion**/denomination (neighborhood location was found manually)
- Out of the **~800** original institutions, **~450** were used after **filtering** repeats and institutions not in a neighborhood
  - Sources: Western PA Regional Data Center, Salatomatic, Maven Search,

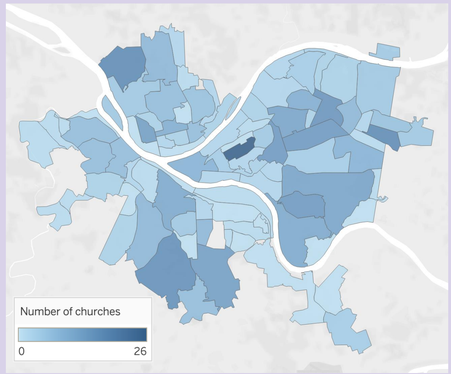
Results











Analysis

- The first graph shows the **total** religious institutions has a **very weak polynomial** association with the **total** median age of death **(r^2 = .017)**. This means 1.7% of the variation in the median age of death can be explained by the number of institutions in each neighborhood. Although a **polynomial** association is rarely used, it was chosen as it showed a **significant increase in r^2** in all graphs
- However, when we **controlled by race**, the correlation **increased** to a **weak** association **(r^2 = .256)**
- Furthermore, when **controlling for income**, the correlation increased significantly to a **very strong** association at the **lower income** range **(r^2 = .993)** but changed to a very **weak** positive, logarithmic correlation at the **higher income** range **(r^2 = .227)**

Challenges

- Data regarding **religious institutions** is **not** usually well **documented** and thus is difficult to obtain. The data that we did obtain had the **address** rather than the neighborhood so we **manually** had to **find** the **neighborhood** for ~800 institutions.
- People are **not confined** to one neighborhood; they often **use institutions** located in **other neighborhoods**.
- All religions **don't** have the **same influence** but they are **generally focused** in certain **geographic areas** so controlling for type of religion was not a promising option

Conclusion

- Religious Institutions should **continue to be built** and **supported,** especially in **low income neighborhoods** as they noticeably **increase** median age of death until **up to 15** are present in a neighborhood.
- In the future, we would like to control by type of religion, and increase our scope beyond Pittsburgh, as well as use other health metrics to analyze the effect.

SCHOOL SUCCESS FACTORS IN ALLEGHENY COUNTY SCHOOL DISTRICTS

WHAT OUTSIDE FACTORS AFFECT A SCHOOL DISTRICT'S SUCCESS?

AVONWORTH HIGH SCHOOL
DARREN HUNT - ADDISON DEXTER
LUC AZEN - CHARLIE BOZADA
LAUREN PFLUEGER - MIKE FRANK

DEFINITIONS

- **Success of a school** = **SAT scores** and **SPP scores**.
- **SPP score** = **School Performance Profile** provided by the state that grades all schools throughout PA.
- **Community factors** = any confounding variable that may have **impact on the success of a school district that is not directly associ- ated with the school itself**.
- **Millage rate** = the number of **tax dollars assessed for each $1,000 of property value**.
- **Cohort** = **size of the graduating class** in the school.

RESOURCES

- Data.gov
- US Bureau of Labor Statistics
- WPRDC.org
- Pennsylvania Department of Education
- 2014 Community Need Index
- DataUSA.io

ASSUMPTIONS

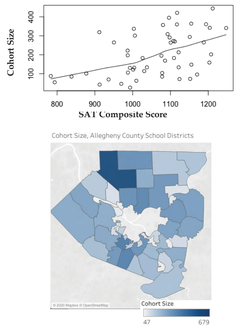
Our prediction was that a successful community would lead to a suc- cessful high school. For example, a district with a high median income or low death reate would result in an increased success rate of students.

CHALLENGES

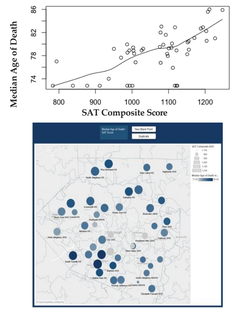
Learning how to **program using R**, including learning the **functions**, learning the outpu[illegible]ring out how to do **pairs function** [illegible] values

- Creating **heatmaps** that effectively depict our results
-
- Finding **datasets** based on school districts
- Figuring out how to convey our results about **Pittsburgh city schools**
- Deciding what **results** were most essential to summarize our work and findings
- Working with **different years** for different data sets

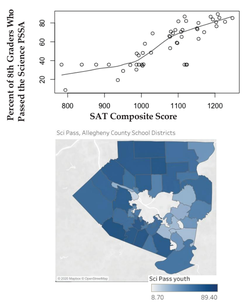
SUMMARY OF RESULTS



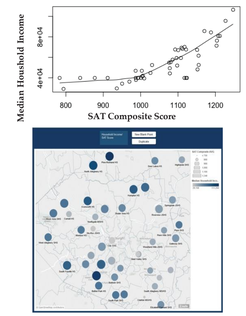
The class size or **cohort is positively correlated** with **SAT** scores with an r value of **0.4885**. The map shows **North Allegheny** school district and **Shaler Area** school district as two of the highest cohort sizes, while **Clariton** has the lowest, depicted in near-white. We included this as a significant factor because we **believed** that the **size of classes** at a school would have an **impact** on the **success** of the students, but they **did not have a significant effect**.



The **median age of death** was **surprisingly cor- related** with a value of **0.6413**. This map shows the **connection** between **SAT composite** scores and the **median age of death**. The **size** of the dots for each district **represents SAT scrores**, while the **color gradient** of the dots **shows the median age**. A district with a small, light col- ored dot, for instance, has a lower SAT compos- ite and low median age of death.

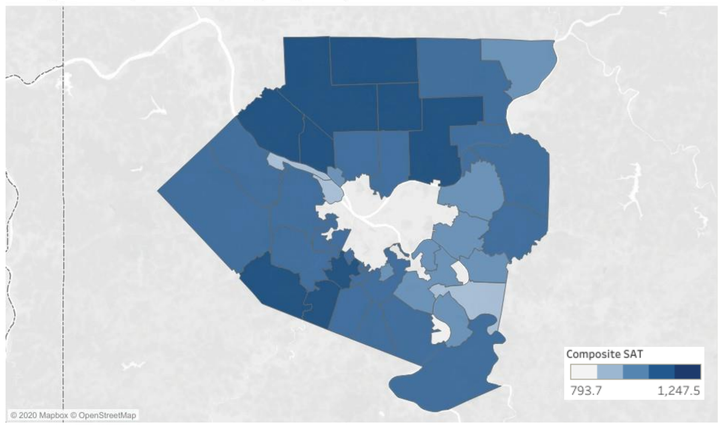


All three standards assessed by the PSSAs, **ELA, Math, and Science**, had r values of **0.8135, 0.7935, and 0.8169** respectively. We chose to represent the correlation between the PSSA Science passing percentages and SAT scores **because science is not covered by the SAT**. Many of the school districts had high passing percentages, with **districts to the East of the city being visibly lower**.



Median income for the district is correlated at a value of **0.7903**, the **second highest correla- tion** of all of the factors that we came up with. We believe that this is one of the **best indica- tors of the wealth** of a district. The **size** of the dots, similarly to median age of death, **corre- sponds to the SAT composite** scores, while the **color** corresponds to the **median income** for the district.

Average SAT Composite Scores, Allegheny County School Districts



r Correlations Without Outliers	SPP Score	SAT Score
Cohort Size	0.5735	0.4885
Population	0.2988	0.2459
Median Household Income	0.8030	0.7903
Millage Rate	0.4480	0.3544
% of 8th Graders Passing ELA PSSA	0.7963	0.8135
% of 8th Graders Passing Math PSSA	0.8044	0.7935
% of 8th Graders Passing Science PSSA	0.8059	0.8169
% of Population in Poverty	0.6001	0.4622
% of Population under 200% of Poverty Line	0.5551	0.4220
% of Families with Single Parents	0.5356	0.3601
% of Population (16-64) Unemployed or Not in Labor Force	0.4076	0.2637
% of Vacant Housing Units	0.5231	0.4180
% of Households with No Vehicle	0.6132	0.4713
% of Pop 25 or Older without Highschool Diploma	0.5034	0.3589
Foreclosures in District since 2009	0.2089	0.0866
Median Age of Death in SD	0.6069	0.6413
Total Deaths in SD	0.0444	0.0099

CONCLUSION

The **most influential factors** that we found in our analysis were **Middle School (8th grade) PSSA Scores** and **Median Household Income**. They both had high r values of about **0.805** and **0.79** respectively in their **linear regressions compared to SAT scores**. Some other major factors that we found include **median age of death**, percent- age of the district's **population in poverty**, percentage of **households** in the community **with no vehicle**, and **cohort size**, otherwise considered class size. We found that this last factor, **cohort size, was correlated to SAT score much more than expected**; this was especially **surprising** because the correlation of **overall district popula- tion** to SAT score had an r value of only **0.2459**, and we had **believed that cohort would be dependent on population**. Besides the high correlations, other factors we tested, **foreclosures** in the school district since 2009 and the total **number of deaths in the district**, were the **least correlated** with SAT score.

The Return on Investment of College

Bethel Park High School - Team 2

Alex Chitsazzadeh, Anthony Hudson, and Allison Kurtz

Research Question

How does receiving a degree in a field of work from a more or less expensive university affect one's starting and average pay during their career?

Data Sources

The two primary datasets we used were sourced from https://www.payscale.com/college-salary-report and https://collegescorecard.ed.gov. We then programmed a python script to merge specific aspects of both data sets together. Mainly, the College Scorecard data provided statistics such as: Admission rate, in-state tuition, and out-state tuition. Then, Payscale provided statistics on the salaries of majors early on and later in their career based on what college or university they graduated from.

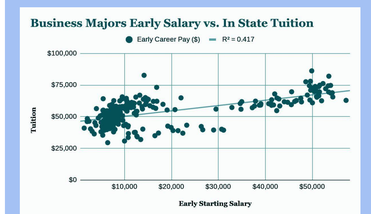
Challenges

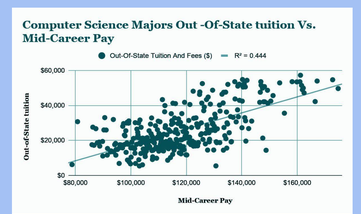
One of the challenges we had with this project was finding appropriate data sets. For example, at first, we found a data set that was too general to be used, however we later found one more tailored to our question. One issue we had was that not every college or university provided all the necessary information, therefore we were not able to truly account for every college or university.

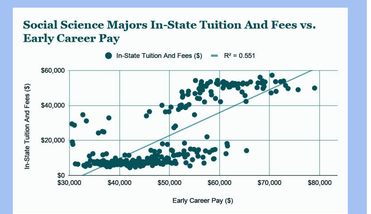
Data Set Examples

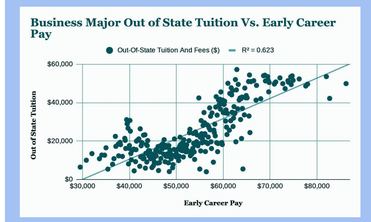
Rank	Institution Name	Early Career Pay ($)	Mid-Career Pay ($)	In-State Tuition And Fees ($)	Out-Of-State Tuition And Fees ($)	Latitude	Longitude	Admission Rate	State Postcode
1	Stanford University	$73,600	$123,700	$49,617	$49,617	37.429434	-122.167359	0.0473	CA
2	Colgate University	$60,700	$111,800	$53,960	$53,960	42.819475	-75.539346	0.2314	NY
3	Dartmouth College	$67,900	$110,000	$53,368	$53,368	43.704115	-72.288946	0.1046	NH
4	The New School	$55,300	$107,000	$47,276	$47,276	40.735608	-73.997114	0.6151	NY
5	Rhode Island School of Design	$55,400	$104,900	$49,370	$49,370	41.825935	-71.407592	0.3193	RI
6	Brown University	$56,200	$104,600	$53,419	$53,419	41.82617	-71.40385	0.0849	RI
7	Georgia Institute of Technology-Main Campus	$56,200	$103,900	$12,418	$33,014	33.77242	-84.394632	0.234	GA
8	University of Notre Dame	$54,300	$103,800	$51,505	$51,505	41.703058	-86.239959	0.1992	IN
9	Cornell University	$55,200	$103,700	$52,653	$52,653	42.4472	-76.483084	0.1267	NY
10	Yale University	$59,200	$102,600	$51,400	$51,400	41.311158	-72.926658	0.0693	CT
11	Art Center College of Design	$53,000	$101,600	$42,008	$42,008	34.150031	-118.185102	0.7038	CA
12	Columbia University in the City of New York	$60,200	$101,100	$57,208	$57,208	40.808286	-73.961885	0.0651	NY
13	New York University	$53,500	$101,000	$50,464	$50,464	40.729452	-73.997264	0.2766	NY
14	Fashion Institute of Technology	$51,800	$100,800	$5,463	$14,663	40.74731	-73.994781	0.4704	NY
15	University of Virginia-Main Campus	$52,400	$100,500	$16,953	$47,273	38.035985	-78.503428	0.2735	VA
16	California College of the Arts	$59,900	$100,000	$47,266	$47,266	37.787485	-122.399451	0.8369	CA
17	Otis College of Art and Design	$54,500	$99,600	$45,620	$45,620	33.956907	-118.416515	0.9125	CA
18	The Juilliard School	$54,900	$99,500	$43,170	$43,170	40.773725	-73.982913	0.0592	NY
19	University of California-Santa Cruz	$53,000	$99,400	$14,020	$42,034	36.995504	-122.058931	0.5094	CA
20	Pratt Institute-Main	$51,800	$99,900	$50,038	$50,038	40.691297	-73.96431	0.568	NY

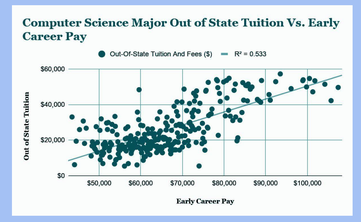
Visualizations

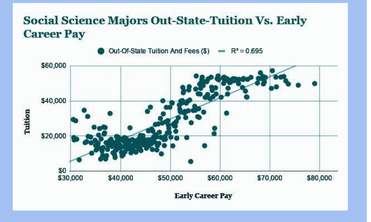












Summary

What we found is that the correlation between majors and what they could make early and mid-career veries. We believe this to be true as the various majors come with their own stipulations. We found that there was more correlation between out of state tuition and early career pay in comparison to out of state tuition and mid-career pay. We did find that most majors did have a strong correlation between out of state tuition and the early career pay.
Although we had a few outliers including: Engineering majors where the correlation was only .272, Education majors with .207 and Health Science majors with .305.

Policy

Being able to interpret all of this data will allow future college students and other individuals to conclude the optimal choice for them based on the current situation pertaining to the individual.

Coffee, Tea, or Prozac?

Nini Curcione, Zach Deschon, Jack Hric, Jay Patel, Alex Pretka, Pete Rauch, Kenneth Swimkosky, Derek Wassel

Does Starbucks contribute to lowering depression and anxiety, and therefore increase SAT scores?

They are communal places where people of all ages feel comfortable to sit, relax, work, and study. SAT scores are still the leading factor in acceptance into higher education programs, colleges, and universities. With teen depression and anxiety on the rise, a parent's best strategy to help a child is to promote the development of good mental health. Many neighborhoods are trying to increase their likability and one of the ways they are doing this is bringing in franchisees to promote community wellness. Starbucks and similar coffee shops continue to pop up at a fast rate with the hopes of providing income, increasing community wellness, and therefore decreasing anxiety and depression. This information would be relevant to the parents of high school age children and may influence their decisions about residential locations and school choices. We questioned, what if Starbucks signified less anxiety overall and that meant students would score better on their SATs?

Hypothesis

Crime and anxiety have a negative effect on students and therefore SAT scores, then the presence and availability of Starbucks stores would ease anxiety and have a positive effect.

FINDINGS

Results

- Property Value:
  - r-value = .647
  - r^2 = .4186
- Anxiety Index
  - r-value = .434
  - r^2 = .188
- School Funding
  - r-value = .515
  - r^2 = .265
- Median Income
  - r-value = .408
  - r^2 = .166
- Property Value, School Funding, Anxiety Index, Median Income, Starbuck location r^2 value = .561

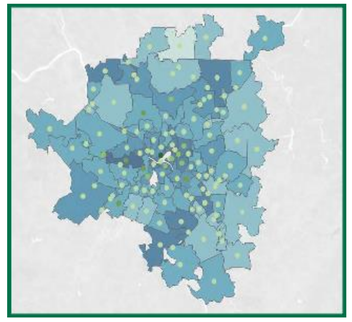
Analysis

- There is a relatively strong positive relationship between SAT scores and property value with a correlation of .647. The variable "property value" had the highest "r" value which measures how well a linear function captures the relationship between variables. Property value and SAT scores appear to tend to vary together.
- R^2, our coefficient of determination between property value and SAT scores was .4186 we see that 42% of the total variation for SAT score can be explained by the linear relationship
- Using multiple regression we found that the anxiety index was the worst predictor among property value, school funding, and median income.
- Through analysis anxiety index, property value, school funding, medium income, and Starbuck locations combined to form a good prediction equation for SAT scores.

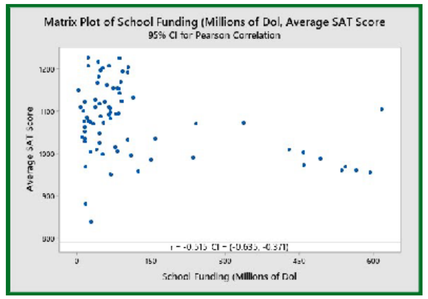
SOURCES

- Starbucks.com
- Fear of Crime: Critical Voices in an Age of Anxiety
- CityData.com
- National Institute of Mental Health

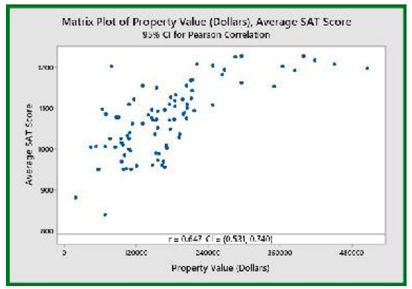
VISUALS



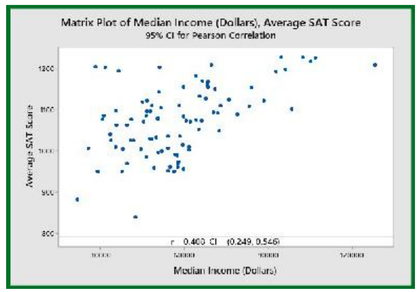
Anxiety Index: blue shading. Starbucks locations: green dots. Darker blue indicates higher levels of anxiety. The index allows us to analyze the people who are unable to afford anxiety meds



School District funding plot in relation to SAT performance. Not the best indicator of test performance.



Average Property Value per zip code as it relates to SAT scores in the same area



Average income per zip code as it relates to SAT scores in the same area

CONCLUSION

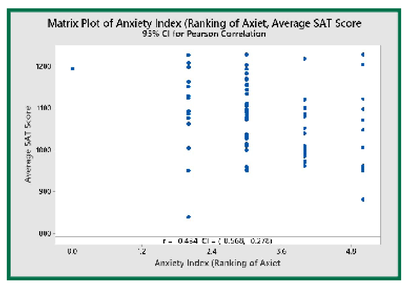
- Our initial hypotheses was centered around anxiety medication and Starbucks locations, we attempted to find correlation between low amounts of anti-anxiety medication usage and the presences of a Starbucks to help produce higher SAT scores. First using number of prescriptions per zip code; then an anxiety index about crime with the Starbucks Stores locations yielded very low correlations.
- We then considered property value, funding, and medium income which had the strongest correlations which stands to reason; those living in areas with higher income would be living on more valuable property and more funding for districts. These proved to be decent indicators of SAT Scores when combined. We would want to look into why. Perhaps students have access to more resources, more time, or some other factor. Because anxiety and Starbucks did not appear to have much influence.
- We would recommend further analysis to find and use other variables that would explain the relationship and the higher scores. In the end we are able to say that having a Starbucks nearby does not make an area less anxious nor lead to higher SAT scores.

DATA SET AND OBSERVATIONS

- A teenager's vulnerability to anxiety and depression has increased due to the stress of performing well in school and social, political, and environmental factors could affect SAT performance
- Median household income for the United States from the Bureau of Labor Statistics reported a .03% increase from December 2018 at $63,517 to January 2019 at $63,688.
- According to the U.S. Census Bureau; they released the *American Community Survey Five-Year Estimates* that the median cost of property is $217,600.
- Teenagers seem more content and alert when they have a drink from Starbucks
- A study exists, Fear of Crime, Critical Voices in an Age of Anxiety, comparing the level of crime in neighborhoods to the anxiety level of that neighborhood. It provided a numerical index of 0-5 based on the amount of crime which coincided with the zip code data
- Disparity between school districts' performance, budgets, and school funding

CHALLENGES

- Finding Starbucks locations that were not in franchises and some that were not in some areas.
- We had to create Zip Code conversions related to Districts that were covered by multiple postal codes.
- Looking through a 50 page PDFs documents on each school district's website to find their budget for the 2018-2019 school year.
- Prescription anti-anxiety data was not as useful as anticipated because there was not way to know if the people filling them lived in that zip code and attended that school district. We needed to find a way to compare the level of crime in an area to the anxiety of that same area.



Anxiety Index and SAT Scores:

Obesity Rates in Allegheny County by Income and Education Level

Oakland Catholic High School
Teresa Davison, Alexis Hagerty, Emily Gong, Frances Li, Mona Lin, Rachel Moret, Eliane Rectenwald, Lindsay Worrall, Jin Yan, Yolanda Yang

Problem

Does walkability, level of education, accessibility to grocery stores, and household income have correlations with neighborhood obesity rates? Can some of the variables (univariate or multivariate) predict the obesity rate? How can we reduce the high obesity rate in Pittsburgh?

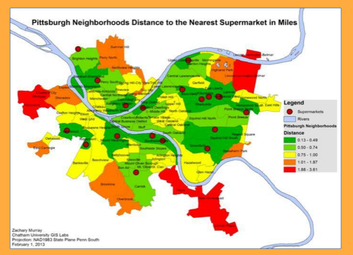
Importance

- "Mother of all diseases"[1]
- Obesity rates have doubled since 1980 and tripled for teens nationwide[2]
- Pittsburgh: 31.1 percent of adults reported being a health weight in 2016[3]

Datasets

WPRDC and US census datasets were used. All data used was compared by census tract. Obesity rates and walkability were found through the WPRDC. Income and education datasets found through US census data.

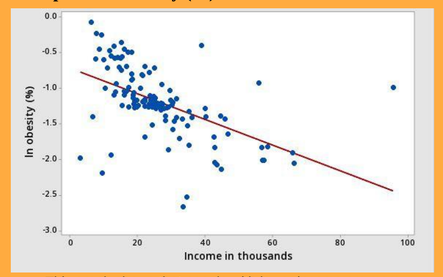
This graph shows that some areas in Pittsburgh are farther from supermarkets than other areas, meaning that it is harder for them to get food sources.
This could be a potential factor of lower obesity rate due to the fact that they can't get food as often.



Multivariate Regression Line Equation

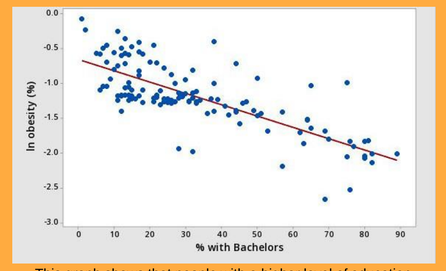
ln (obesity) = -0.7229 - 1.941 * % Bachelor's + 0.00652 * Income in thousands

Scatterplot of ln obesity (%) vs. Income in thousands



This graph shows that people with lower income are more likely to be obese than higher educated people.

Scatterplot of ln obesity (%) vs. Bachelor's or Higher



This graph shows that people with a higher level of education tend to be less obese than those who are not as educated.

This graph shows the relationship between walkability and the obesity rate.

Potential Actions

- In terms of self-improvement, people should be persuaded to prepare their own meals more and educate themselves about nutritious habits.
- Charity events such as running or walking to raise money and awareness for medical research would stimulate people's interests.
- Decrease the number of fast food restaurants in the area.
- Schools should encourage a healthy lifestyle and promote clean diets.
- Encourage more involvement in biking and running clubs

Challenges

- Census tract data concerning age-range was hard to find, along with a connection between age range and Allegheny County neighborhoods
- Information pertaining to obesity was difficult to locate because health data is usually protected.
- Pittsburgh already contains conventional solutions to obesity such as walking paths, bike lanes, yearly marathons, public parks, etc.

Summary

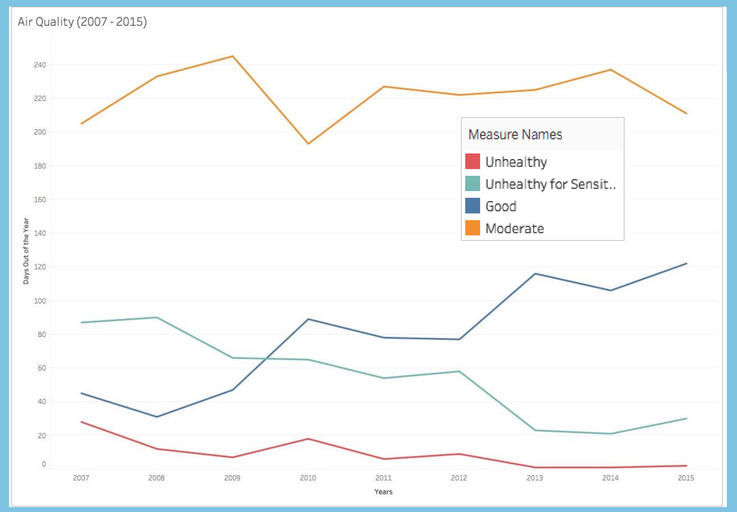
In conclusion, we found that there was a correlation between our different variables and obesity rates in the Pittsburgh area. There was a moderate correlation between obesity and income and education level while there was a much weaker relationship between walkability and obesity. We have found these connections, but coming up with a conclusive "solution" proved to be a challenge due to existing programs geared towards health. We did find out though, that education might be the best tool to keep Pittsburgh healthy.

[1] (George Eid, Allegheny Health Network's Bariatric and Metabolic Institute)
[2] (Johns Hopkins School of Medicine)
[3] (Centers for Disease Control and Prevention's Behavioral Risk Factors Surveillance System)

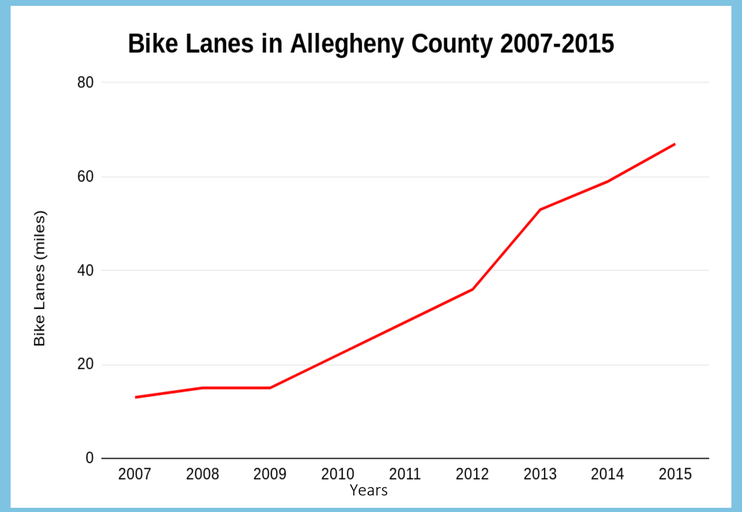
Does the Amount of Bike Lanes in a Neighborhood Affect Air Quality?

Maria Czura Abigail Haerr Mackenzie Linderman Maura Marston
Plum High School

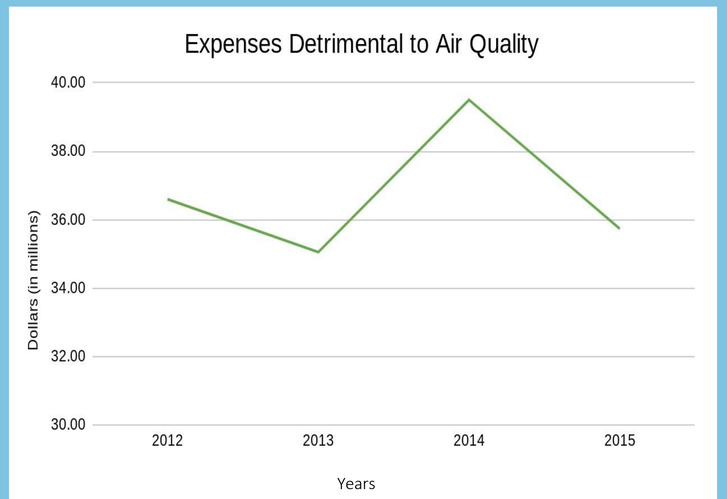
Air Quality



Bike Lanes



Expenses



Data Sets

We obtained our air quality data from the United States Environmental Protection Agency. Using the Air Quality Index Report we were able to input a year and a city, and the website would provide several how many days good, moderate, unhealthy, and unhealthy for sensitive groups. We then plotted that data on a line graph as shown to the upper left with days up to 240 on the y-axis and years 2007 – 2015 on the x-axis. We then obtained our bike lane data from the Western Pennsylvania Regional Data Center. There we were able to find the data that told us the mileage of bike lanes there were from the years 2007 to 2015.

Challenges Faced

Compare Bike Lane Data in Different Cities - The first challenge arose when attempting to gather bike lane and air quality data in cities throughout the United States. It seemed this would help draw a bike lane and air quality correlation across the country, but there is no data set that includes the mileage of bike lanes in all major American cities. As a result, individually searched bike lane data for each city was compared to the city's square footage in the 2010 United States Census. Data inconsistencies arose as the square footage had slightly different area boundaries than the bike lane data. The approach was modified by bringing in a new data set, focusing on Allegheny County air quality and bike lanes in the past years

Graphs - At Plum, high school students receive a Chrome Book to help with online schoolwork. However, these Chrome Books do not support many graphing applications that are needed to visually represent big data. Using a MacBook solved this problem. Tableau, a graphing application that cannot be used on Chrome Books, was used to create the Air Quality graph. Even so, Tableau cannot process the big data sets used to create the Bike Lane and Expenses graphs. Another change to the graphing application needed to be made. When obtaining finance data, we chose the data set labeled "Data Set name". This set described how much funding was being given to different types of organizations. Some of the funding numbers were negative and the rest were positive. To be able to graph the results, we determined that all the funding numbers must be positive. The data set was loaded onto Tableau as we attempted to take the absolute value of these numbers, however the data set was too large to fully load in Tableau. Then we tried loading this data set into Numbers. The full data set was able to load and we used "find and replace" to replace all the negative signs with "0". This gave us all positive value amounts for the money distributed to each organization. This in turn allowed us to graph the data with years on the x-axis and dollars on the y- axis. Then, utilizing Excel on a MacBook worked perfectly to create the final Bike Lane and Expenses graphs.

Other Factors - Pittsburgh has become more mindful of helping the Earth in the past years. Riding bikes is not the only activity that improves air quality. The Pittsburgh's Expense Report 2015-2019 includes all of Pittsburgh's spending throughout the last five years. Challenges arose when trying to simplify this data set. The data set is too large to open on a Chrome Book and needed to be accessed on a Mac Book. Also, finding topics related to air quality and searching what specific spending meant was a challenge.

Conclusion

When observing the air quality data, it can be seen that the amount of "good" days of air quality increases from 2007 to 2015. When looking at the bike lane data, it can also be seen that the amount of bike lanes increased from 2007 to 2015. Therefore, there appears to be a correlation between the air quality and the amount of bike lanes. In an attempt to view confounding variables, we graphed finance data from 2012 to 2015 in order to see if there was a significant change in funds issued to projects that would be detrimental to air quality (2012 was the earliest year we could obtain data). However, the amount of money allocated to organizations seems to be held between the range of $34 - $40 million. Following the money, we concluded that the expenses did not play a significant role in our analysis because the expenses stayed similar. Thus, while we know the air quality is obviously affected by spending going toward causes such as transportation and building construction, we are not able to conclude that there is a significant correlation between air quality and expenses because of the low variation of annual costs. In spite of this, environmentally safe technology is beginning to be used more and more. So, although funding could be kept consistent, greener technology could be utilized so there is not such a harmful effect on the air. Despite all these findings, there are also many contributors to air quality outside of public organizations that could have an effect on these numbers.

Does Concentration of Fast Food Restaurants Affect Obesity Rates?

Data Sets

--#--	ZIPCode	StAb	StCtyTract	StCty	CtyTract	TractCode1	TractCode2	Population	Housing Un...
1	17019	PA	42001030101	42001	001030101	030101	301.01	87	35
2	17316	PA	42001030101	42001	001030101	030101	301.01	821	345
3	17324	PA	42001030101	42001	001030101	030101	301.01	63	28
4	17372	PA	42001030101	42001	001030101	030101	301.01	1609	620
5	17316	PA	42001030102	42001	001030102	030102	301.02	4030	1560
6	17350	PA	42001030102	42001	001030102	030102	301.02	1719	670
7	17372	PA	42001030102	42001	001030102	030102	301.02	31	14
8	17304	PA	42001030200	42001	001030200	030200	302	714	254
9	17324	PA	42001030200	42001	001030200	030200	302	1332	542

PA Pennsyl | Show All | Find ZIP: 85258 | Find StCty: 04013 | Find StCtyTract: 04013216813

[Source]

- An online table that listed ZIP codes and the census tracts that they fell under. Because there is no direct translation between a ZIP code and a census tract, these values act as our “Translation.”
- Data was used from the columns titled “ZIPCode”, “StCtyTract”, and “Population”

	Name	Legal Name	Start Date	Street Number	Street Name	ZIP Code	Lat	Lon	Category
2	Adrian's Pizza		2014-11-07T00:0	605	Thompson Run F	15237	40.539465	-79.990764	Take Out
3	Adrian's Pizza E Rock Enterprises		2004-04-22T00:0	7824	Perry Hwy	15237	40.551219	-80.037362	Take Out
4	Allegheny Sandwich Shop		1997-02-24T00:0	414	Grant St	15219	40.43811	-79.99686	NO Dollar Menu
5	Allegheny Sandv Allegheny Sandv		2001-11-09T00:0	440	Ross St	15219	40.438514	-79.99533	NO Dollar Menu
6	Amili's Pizzeria		1999-02-26T00:0	1021	Brownsville Rd	15210	40.406082	-79.991863	Take Out
7	Angelia's Pizza	JNG Pizza LLC	2004-05-11T00:0	202	Moon Clinton Rd	15108	40.513135	-80.223406	Take Out
8	Angelia's Pizza /	Eaton Pizza Inc	2005-10-07T00:0	410	Penn Lincoln Dr	15126	40.442466	-80.235992	Take Out
9	Antney's Ice Cre	The Iceman Inc	2002-04-11T00:0	1316	Poplar St	15205	40.42747	-80.052435	Breakfast, Drink, Other
10	Arby's	Kinco Inc	1975-01-01T00:0	1617	Freeport Rd	15065	40.622125	-79.727516	Dollar Menu
11	Arby's #8	Linell Corporatio	2007-12-03T00:0	3974	Wm Penn Hwy	15146	40.437988	-79.772845	Dollar Menu
12	Arby's #42	ARG Co.	1997-07-10T00:0	1378	Banksville Rd	15216	40.414684	-80.030651	Dollar Menu
13	Arby's #164	ARG Co.	1997-07-09T00:0	3417	Forbes Ave	15213	40.439175	-79.960709	Dollar Menu
14	Arby's #836	ARG Co.	1984-06-17T00:0	1140	Washington Pike	15017	40.3679	-80.103172	Dollar Menu
15	Arby's #1203	ARG Co.	1997-07-09T00:0	2251	Noblestown Rd	15205	40.428509	-80.0531	Dollar Menu
16	Arby's #1414	ARG Co.	1997-07-09T00:0	6192	Steubenville Pik	15136	40.445808	-80.150631	Dollar Menu
17	Arby's #1691	ARG Co.	1997-07-10T00:0	5030	Wm Flynn Hwy	15044	40.606809	-79.946719	Dollar Menu
18	Arby's #1773	Arby's Restaurar	2016-01-15T00:0	19075	Perry Hwy	16046	40.6225913	-80.0533722	Dollar Menu
19	Arby's #1853	ARG Co.	1997-07-09T00:0	3700	Library Rd	15234	40.364992	-80.020333	Dollar Menu
20	Arby's #1854	ARG Co.	1997-07-10T00:0	1911	S Braddock Ave	15218	40.422721	-79.885611	Dollar Menu
21	Arby's #1858	ARG Co.	1997-07-07T00:0	6105	Saltsburg Rd	15147	40.484692	-79.819176	Dollar Menu

Original | Pivot Table 1 | Year Filter | Restaurants Built Up to 2010

[Source]

- A table of restaurants constructed in Allegheny County ZIP codes from 1968 to 2016. The restaurants that were constructed from 2011 to 2016 were filtered out because our obesity rate data was not representative of it.
- Data was used from columns C (Start Date) and F (ZIP Code).

	sname	tractfips	2000 Tract	2010 Tract	MUNICIPALITY	Municipality	City Neighborhood	2006-2010 estimate of obesity
2	Pennsylvania	42003010300	10300		PITTSBURGH	Pittsburgh	Bluff	0.246935849
3	Pennsylvania	42003020100	20100		PITTSBURGH	Pittsburgh	Central Business District	0.668012041
4	Pennsylvania	42003020300	20300		PITTSBURGH	Pittsburgh	Strip District	0.369996439
5	Pennsylvania	42003030500	30500		PITTSBURGH	Pittsburgh	Crawford Roberts	0.488690884
6	Pennsylvania	42003040200	40200		PITTSBURGH	Pittsburgh	West oakland	0.366390731
7	Pennsylvania	42003040400	40400		PITTSBURGH	Pittsburgh	North oakland	0.155867454
8	Pennsylvania	42003040500	40500		PITTSBURGH	Pittsburgh	Central Oakland	0.112780394
9	Pennsylvania	42003040600	40600		PITTSBURGH	Pittsburgh	Central Oakland	0.144556151
10	Pennsylvania	42003040900	40900		PITTSBURGH	Pittsburgh	South Oakland	0.281331416
11	Pennsylvania	42003050100	50100		PITTSBURGH	Pittsburgh	Middle Hill	0.608099152
12	Pennsylvania	42003050600	50600		PITTSBURGH	Pittsburgh	Upper Hill	0.455786185
13	Pennsylvania	42003050900	50900		PITTSBURGH	Pittsburgh	Bedford Dwellings	0.636222228
14	Pennsylvania	42003051000	51000		PITTSBURGH	Pittsburgh	Terrace Village	0.555072799
15	Pennsylvania	42003051100	51100		PITTSBURGH	Pittsburgh	Terrace Village	0.793133039
16	Pennsylvania	42003060300	60300		PITTSBURGH	Pittsburgh	Lower Lawrenceville	0.319568655
17	Pennsylvania	42003060500	60500		PITTSBURGH	Pittsburgh	Polish Hill	0.320700606
18	Pennsylvania	42003070300	70300		PITTSBURGH	Pittsburgh	Shadyside	0.128451261
19	Pennsylvania	42003070500	70500		PITTSBURGH	Pittsburgh	Shadyside	0.130383004
20	Pennsylvania	42003070600	70600		PITTSBURGH	Pittsburgh	Shadyside	0.11853283
21	Pennsylvania	42003070800	70800		PITTSBURGH	Pittsburgh	Shadyside	0.134519846

Original | Zip Codes W/ Census Tract

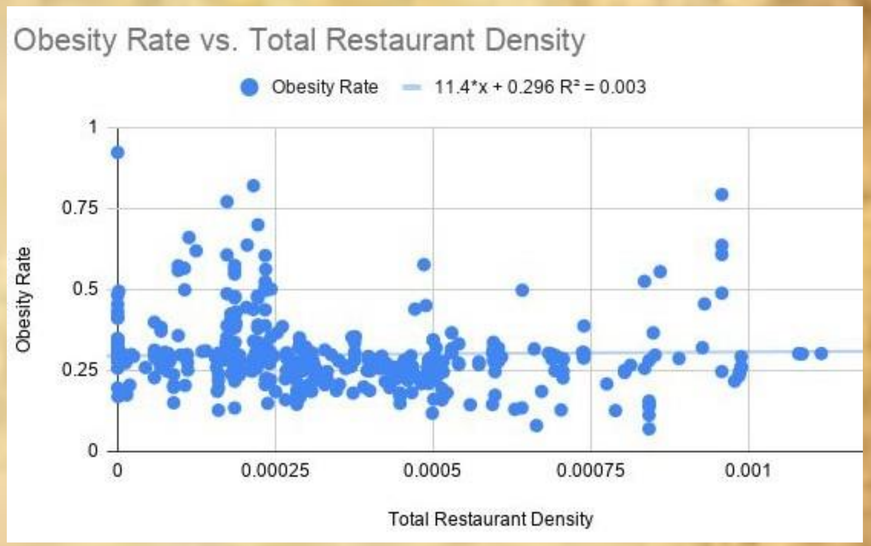
[Source]

- An obesity rate table that has obesity rates from the 2010 census data as well as their respective census tracts
- Data was used from columns C (2000 Tract) and H (2006-2010 estimate of obesity)

Challenges

The first challenge we faced was a difference in the locational data in the “Restaurants” table versus the “Obesity Rates” table. The “Obesity Rates” were organized by census tracts, whereas the “Restaurants” were organized by ZIP codes. Further research determined that there was no simple conversion between the two. To remedy this, a third “Translation” table had to be implemented. However, more often than not, there would be multiple ZIP codes that fell under a census tract, which lead to the second problem.

Once there was a way of translating between both of the spreadsheets, it was difficult to make each census tract have its own restaurant count. Each of the ZIP codes had their respective restaurant count, which was determined from the “Restaurants” table, and population number, which was copied over from our third “Translation” table. However, the populations covered the entire census tract, whereas the restaurant counts were only by ZIP code. This required searching all of the ZIP codes within a census tract, retrieving their relevant restaurants counts, and summing all of those individual restaurant counts for a grand total to one census tract.



Summary and Conclusion

It has been found that there is no correlation between the total restaurant density and obesity rates. With an increase in the independent variable, total restaurant density, there is no significant increase or decrease in the obesity rate. The data contradicts the hypothesis, which was that there would be a positive correlation between total restaurant density and obesity rate. This suggests that there are other factors that affect the obesity rates of a region, such as: diet, genetics, income, and exercise. However, the data concludes that the concentration of fast-food restaurants does not affect obesity rates in an area.