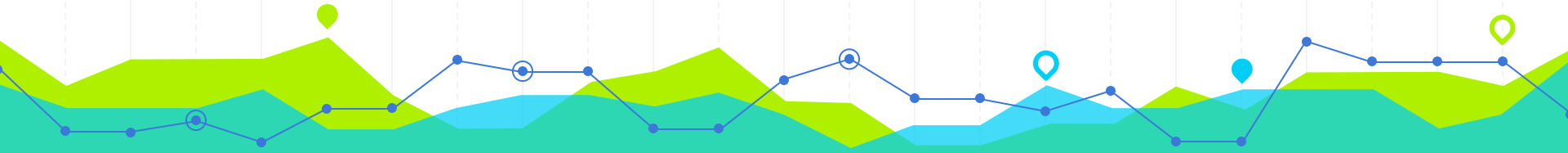
Analyzing High School Completion Rates Across Baltimore Neighborhoods

Group 5: Noah Medina, Trace Terrell, Maanya Bajaj, Emily Perez-Rodriguez, Bella Xia

PRESENTATION OUTLINE

1. Background
2. Data Cleaning and Coding with R
3. Our Visualization and Conclusions



1 BACKGROUND

STARTING INTERESTS

- Graduation rates correlated with property tax, household income, rent affordability, home ownership, vacant properties, etc.
- Other public health issues in Baltimore

GRADUATION RATES IN BALTIMORE

High school completion rate: the percentage of students in high school who attain a high school diploma within four years

- In Maryland, the high school completion rate was 87.2% in 2021. In comparison, Baltimore's was 69.2% in 2021.
- Baltimore's graduation rate has been consistently low for decades despite changes in funding, leadership, and plans for academic equity initiatives.

[https://www.marylandpublicschools.org/stateboard/Documents/2022/0322/AdjustedCohortGraduationRateAdvancedPlacementSATPostsecondaryEnrollment.pdf]

Stringfield, Samuel C., and Mary E. Yakimowski-Srebnick. "Promise, Progress, Problems, and Paradoxes of Three Phases of Accountability: A Longitudinal Case Study of the Baltimore City Public Schools." *American Educational Research Journal*, vol. 42, no. 1, 2005, pp. 43–75. *JSTOR*, http://www.jstor.org/stable/3699455. Accessed 19 Jan. 2023.

BALTIMORE'S HOUSING CRISIS

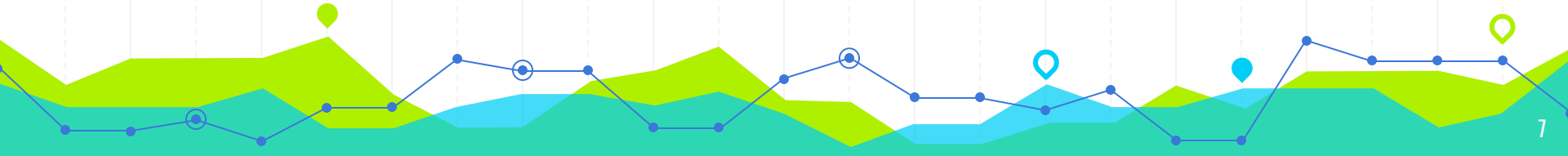
In February 2022, there were almost 15,000 vacant properties in Baltimore, with most in predominantly black and African American neighborhoods. We theorized that the high schools serving these and surrounding neighborhoods would have lower graduation rates because there would be less funding from property taxes.

https://dhcd.baltimorecity.gov/sites/default/files/Vacants%20Q&A%20website.pdf

ON THE SEARCH FOR DATA

The first dataset we encountered was a collection of statistics provided by Baltimore City Public Schools. This data offered a comprehensive overview into enrollment and demographic data in each school.

However, we soon discovered that Baltimore has a particular public high school system where student enrollment is not based on vicinity but instead through collective testing. We also learned that property taxes are not a direct funding source for high schools, which meant that we had to step back and take a new approach.



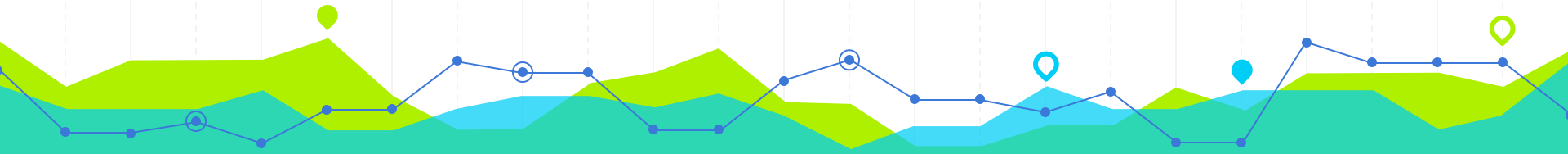
OPEN-ENDED EXPLORATION

Upon discovering the Baltimore Neighborhood Indicator Alliance (BNIA) website, we were curious as to which variables would have significant correlations with the high school completion rates within Baltimore.

We thought it likely that factors like **median household income** must have a significant correlation. We were interested in how factors like **rent affordability** or **overdose calls** might also play into the equation. Given this information, we decided we could explore multiple variables and were then able to formalize our research question.

OUR QUESTION

How do different factors correlate with high school completion rates across different neighborhoods in Baltimore, and which are the most statistically significant? Do they align with our hypotheses?



2 DATA SELECTION AND CLEANING

SELECTING AND CLEANING DATASET

Selection

Baltimore Neighborhood Indicator Alliance offers a large range of data from Census Demographics to Neighborhood Art and Culture. Additionally, there are sections with completely empty data. So we first tried to pick a number of most relevant variables to look at.

Average Healthy Food Availability Index

Copy | Excel | PDF | Print

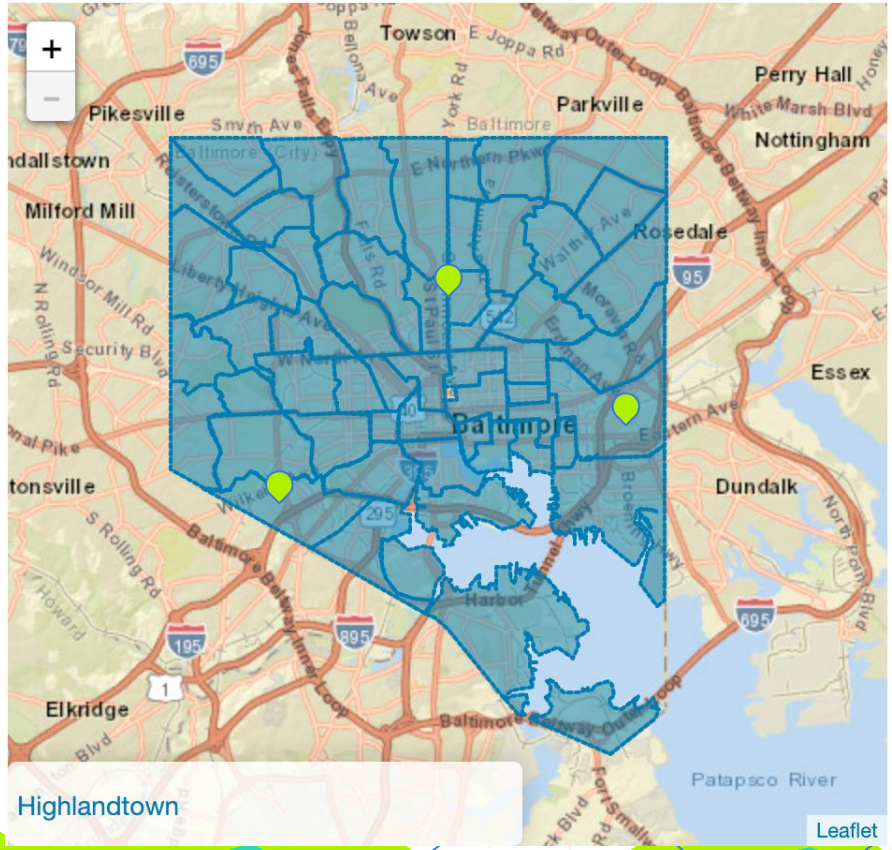
Community	2020 Data
Allendale/Irvington/S. Hilton	
Baltimore City	
Beechfield/Ten Hills/West Hills	
Belair-Edison	
Brooklyn/Curtis Bay/Hawkins Point	
Canton	
Cedonia/Frankford	
Cherry Hill	
Chinquapin Park/Belvedere	
Claremont/Armistead	
Clifton-Berea	
Cross-Country/Cheswolde	
Dickeyville/Franklintown	
Dorchester/Ashburton	
Downtown/Seton Hill	

Average Healthy Food Availability Index, for example, has a completely blank dataset unavailable to our present study.

SELECTED VARIABLES

- Number of Overdose Calls for Service per 1000 People
- Rate of Dirty Streets and Alleys Reports per 1000 People
- Percent of Households with No Internet at Home
- Percent of Female-Headed Households with Children Under 18
- Percentage of Residential Properties that are Vacant and Abandoned
- Violent Crime Rate per 1000 Residents
- Household Income
- Percent of Family Household Living Below the Poverty Line
- Affordability Index - Rent
- Percent Population 16-64 Not in Labor Force
- Percent of Residents -Black/African American (Non-Hispanic)
- Percent of Residents - White/Caucasian (Non-Hispanic)

DATA OVERVIEW



In our present data set, the Baltimore city is separated into 56 neighborhoods. The statistics were collected in unit of each such neighborhood.

SELECTING AND CLEANING DATASET

Community	2020 Data
Allendale/Irvington/S. Hilton	73.9
Baltimore City	78.1
Beechfield/Ten Hills/West Hills	78.6
Belair-Edison	81.6
Brooklyn/Curtis Bay/Hawkins Point	69.3
Canton	
Cedonia/Frankford	77.5
Cherry Hill	72.7
Chinquapin Park/Belvedere	81.8
Claremont/Armistead	80.2

The Canton Neighborhood, in particular, lacks data for High School Completion Rate, which is the core element of our study. So we decided to focus on the remaining 55 neighborhoods.

Cleaning

In addition, within the dataset, there are also a number of neighborhoods that lack data input. Since we are unable to effectively make any prediction based on the current data, we chose not to include these neighborhoods for our data analysis.

3 DATA ANALYSIS AND VISUALIZATION

Data Exploration

In order to get an idea of the trends and visualize how each our variables correlated with high school graduation rates, we created **interactive scatter plots** and added **regression lines** to each.

Diving Deeper into our Analysis

Pearson's Correlation Coefficient ("r value") : a statistical measure of the strength and direction of the linear relationship between two variables. The closer it is to 1 or -1, the stronger the correlation.

- We chose to use this statistic to quantify the correlations within our dataset and identify the strongest ones with regards to high school completion rates

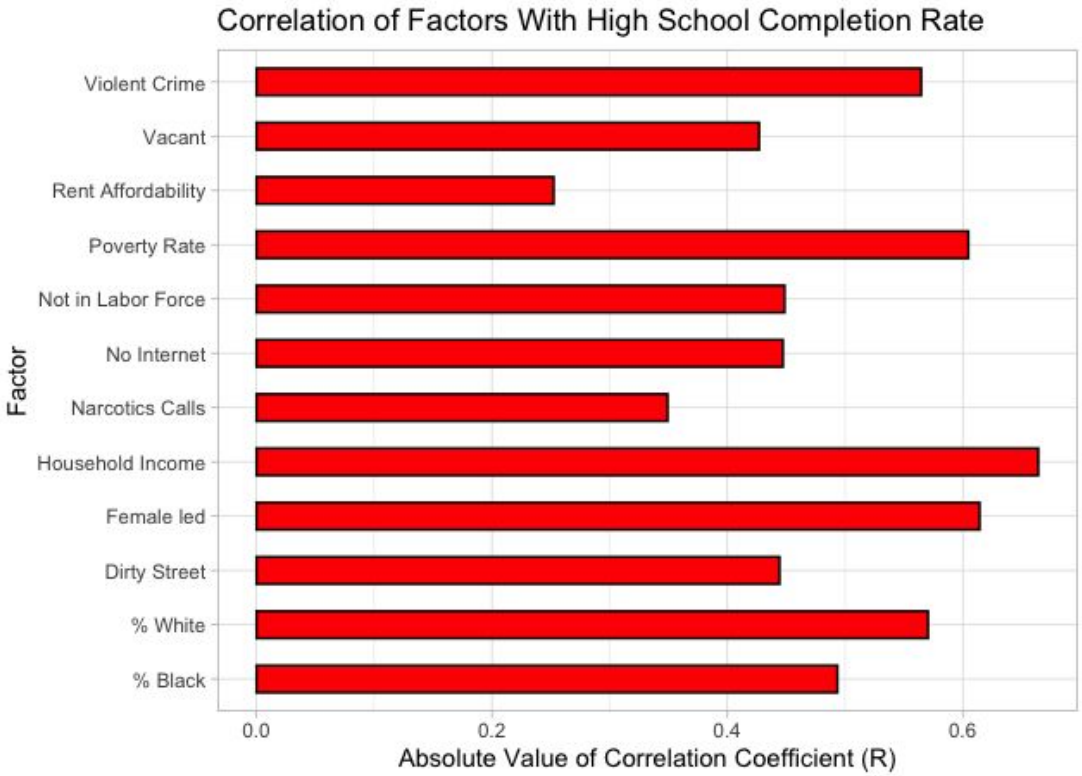
Data Visualization App

```
library(shiny)
runGitHub("graduation", "noahmedina")
```

The Numbers (R Values) and Ranking

1. **Household Income** 0.66414
2. **Female led** -0.61431
3. **Poverty** -0.60472
4. **Percent white** 0.57047
5. **Violent crime** 0.56482
6. **Percent black** - 0.49350
7. **Not in Labor Force** -0.44866
8. **No Internet** -0.44723
9. **Dirty Street** -0.44436
10. **Vacant** -0.42699
11. **Narcotics Calls** -0.34925
12. **Rent Affordability** -0.25248

Our Findings

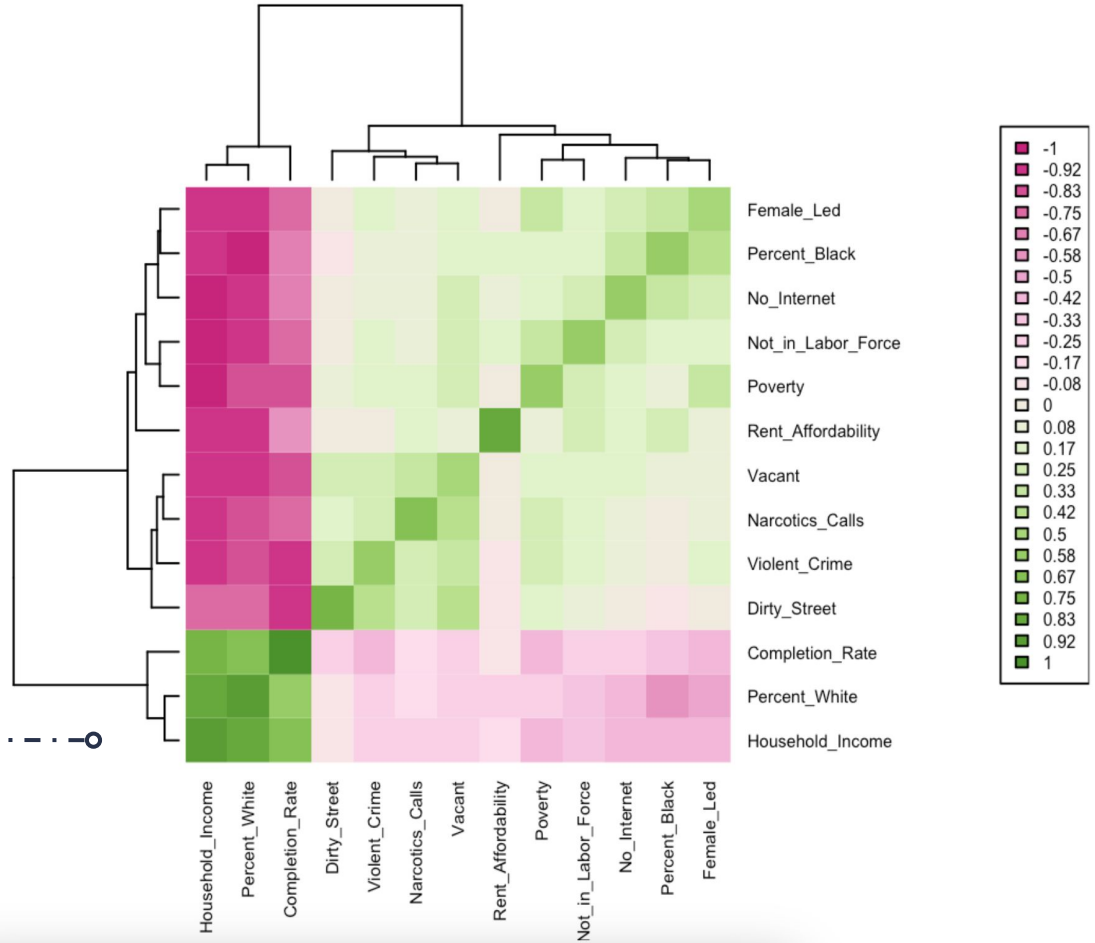


- We created a bar graph to visualize the magnitude of the correlations — the greater the absolute value, the stronger the correlation

FURTHER EXPLORATORY PLOTS

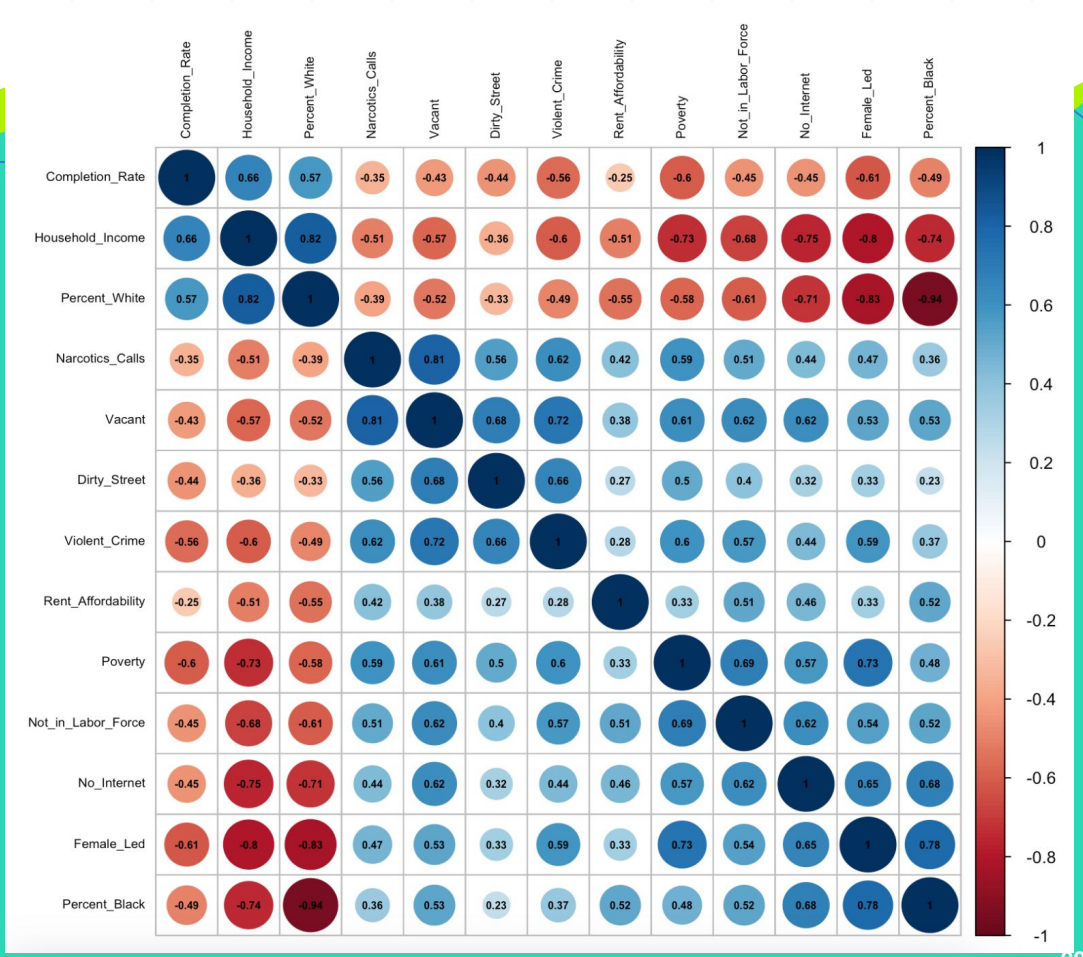
The grouping of the data on the heatmap shows that the two variables that correlated the most with Completion Rate are Percent_White and Household_Income.

This grouping echoes the histogram data on correlations where the two variables with the largest positive correlations are grouped together with Completion Rate.



FURTHER EXPLORATORY PLOTS

The Corrplot provides another visualization of the correlations between all the variables. We can see that Completion_Rate, Household_Income and Percent_White has a significantly positive correlation with each other, and at the same time very clearly negative correlation with the rest of the variables.



THANKS!

Any questions?