Data Science Final Project

Goal: Find a data set that has some impact on you or your community. Use that data to create visualizations in p5 and then make conclusions about the data. Acknowledge any possible bias or errors in the data and how another study could address them.

Requirements (see the rubric for more details):
1) Find a multivariate data set that applies to you, your community, or a community that you care about.
2) Look through the data to filter and clean the data and check that it seems credible.
3) Create a single variable data visualization in p5 (circle graph, bargraph)
4) Create a 2 variable data visualization in p5 (scatterplot, line graph, bubble plot)
5) Come up with a conclusion or argument based on your data visualizations
6) Acknowledge any possible errors or bias in the data or the conclusion that you made.
7) Create a brief (5 min) presentation to show your class what you have completed.

Some place to find data sets:
These are some good places to find relevant data sets but you can get them from anywhere that seems credible.
- Data on NYC: https://opendata.cityofnewyork.us/
- US census: https://www.census.gov/data/datasets.html
- All-sorts: https://www.kaggle.com/datasets/giodev11/usstates-dataset
- Random: https://www.openintro.org/data/
- Sports: https://utopia.ut.edu/c.php?g=887297&p=6377997

Rubric: Final Data Science Project

	Exceeding	Meeting	Approaching	No evidence
Data Set	Data set is reliable, has several categories of data, and is meaningful for the student or their community.	Data set is reliable, has several categories of data,	Data set is unreliable and has limited data and application.	Data set is made up
1st Visualization using p5	1st visualization uses at least one category of data in p5. Is clearly labeled and scaled. Easy to read. Shows something interesting about the data.	1st visualization uses at least one category of data in p5. Has some titles/labels. shows some fact about the data.	1st visualization uses at least one category of data in p5 to produce a graph/chart	Does not create a 1 variable data visualization.
2nd Visualization using p5 and uses at least 2 variables	2nd visualization includes at least 2 variables (bivariate) that could reasonably be related. Is clearly labeled and scaled. Easy to read. Shows something interesting about the data.	2nd visualization includes at least 2 variables that could reasonably be related. Is clearly labeled and scaled.	2nd visualization uses at least two variables of data in p5 to produce a graph/chart	Does not create a 2 variable data visualization.

Conclusions about your data, possible explanation	Conclusion is thoughtful and references the data and data visualizations. It does not overstate or assume too much. It includes some considerations of why or how to change.	Conclusion is thoughtful and references the data It includes some considerations of why or how to change.	Conclusion is says does not significantly reference the data/visualizations and or has errors.	Conclusion does not reference the data.
Concerns of Bias, Errors	The reliability of the data is considered. Concerns of bias or errors are addressed as well as possible solutions.	Either bias or errors in the data are considered and addressed.	Minimal reference to potential errors/bias.	Does not address potential issues with bias/errors.
Presentation	Presentation is 5 minutes, clear, organized, shows both visualizations, thoroughly explains the conclusions as well as bias	Presentation is 5 minutes, shows both visualizations, some explanation of conclusions as well as bias	Presentation shows one visualization and some comments made on conclusions and bias.	Presentation is missing most of the components