DATA VISUALIZATION 44613 – Dr. Case

MODULE 7: FINAL PROJECT

Topic: High School Student Academics in the United States

Naiema Elsaadi

Section 1. Introduction

For this project, I put together a storyboard that analyzes High School student performances based on several factors. Through this analysis I looked at GPAs, Standardized Test Scores, Courses, Family Income, and States of the students and visually presented the trends in the data and mapped out the distribution of academic performances across the entire United States. From this analysis, we can determine the demographic, geographic, and academic factors that are correlated with Student Performances in academics.

All the project's files are available in GitHub:
https://github.com/NaiemaElsaadi/Final_Project_Data_Visualization

Section 2. Data Description

- The domain of the dataset is Education.
- This Data file is a CSV file that was found and downloaded from Kaggle and uploaded to Excel.

school_scores.xlsx

- This data source is a source that contains information about High School Students and their academic, geographic, and demographic information. This includes GPA, Standardized Test Scores, State, and Family Income. This source was found on Kaggle, and it is called "US School Scores". The Excel file was downloaded from this site:

  https://www.kaggle.com/datasets/mexwell/us-school-scores/data

- This dataset has 99 columns and 578 rows.

Section 3. Data Cleaning Strategies

- No cleaning was needed for this dataset.

Section 4. Clean Dataset

school_scores.csv (359.71 kB)

Detail Compact Column 99 of 99 columns

# Year Year	State.Code State Code	State.Name State name	Total.Math total math	Total.Test-takers total test takers	Total.Verbal total verbal	Academic Subjec...	
2005 2015	AL 2% AK 2% Other (555) 96%	Alabama 2% Alaska 2% Other (555) 96%	383 619	134 242k	401 612	3.43 3.96	1
2005	AL	Alabama	559	3985	567	3.92	2
2005	AK	Alaska	519	3996	523	3.76	1
2005	AZ	Arizona	530	18184	526	3.85	2
2005	AR	Arkansas	552	1600	563	3.9	2
2005	CA	California	522	186552	504	3.76	1
2005	CO	Colorado	560	11990	560	3.88	2
2005	CT	Connecticut	517	34313	517	3.66	2
2005	DE	Delaware	502	6257	503	3.71	1
2005	DC	District Of Columbia	478	3622	490	3.54	1

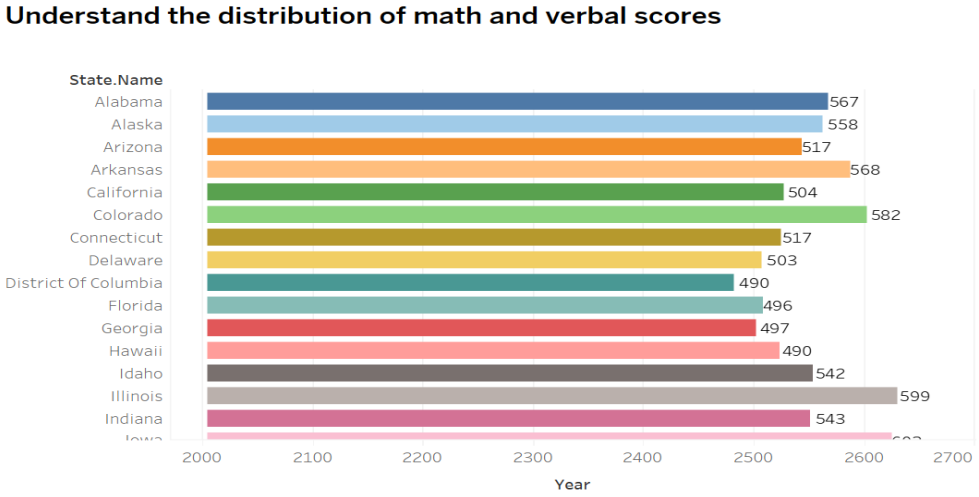
Year	State.Cod	State.Nam	Total.Mat	Total.Test	Total.Verb	Academic	Academic	Academic	Academic	Academic	Academic	Academic	Academic	Academic	Academic	Academic	Academic	Family Inc	F
2005	AL	Alabama	559	3985	567	3.92	2.2	3.53	3.9	3.54	2.6	3.41	4	3.52	3.9	3.59	3.9	513	
2005	AK	Alaska	519	3996	523	3.76	1.9	3.35	3.9	3.34	2.1	3.06	3.5	3.25	3.2	3.39	3.4	492	
2005	AZ	Arizona	530	18184	526	3.85	2.1	3.45	3.9	3.41	2.6	3.25	3.9	3.43	3.4	3.55	3.3	498	
2005	AR	Arkansas	552	1600	563	3.9	2.2	3.61	4	3.64	2.6	3.46	4.1	3.55	3.7	3.67	3.6	513	
2005	CA	California	522	186552	504	3.76	1.8	3.32	3.8	3.29	2.8	3.05	3.7	3.2	3.2	3.38	3.3	477	
2005	CO	Colorado	560	11990	560	3.88	2.2	3.49	4	3.41	3.1	3.33	3.9	3.43	3.7	3.56	3.7	533	
2005	CT	Connectic	517	34313	517	3.66	2.1	3.13	3.9	3.03	3.1	3	3.8	3.07	3.5	3.18	3.6	463	
2005	DE	Delaware	502	6257	503	3.71	1.8	3.21	3.9	3.18	2.7	3.07	3.8	3.19	3.6	3.3	3.6	449	
2005	DC	District Of	478	3622	490	3.54	1.8	3.03	3.8	3.04	2.7	2.91	3.7	2.99	3.3	3.11	3.4	391	
2005	FL	Florida	498	93505	498	3.77	1.8	3.29	3.8	3.3	2.4	3.07	3.8	3.27	3.5	3.39	3.5	471	
2005	GA	Georgia	496	59842	497	3.8	1.8	3.3	3.8	3.31	2.4	3.12	3.9	3.22	3.4	3.38	3.6	459	
2005	HI	Hawaii	516	7878	490	3.77	2	3.21	3.9	3.31	2.5	2.98	3.8	3.18	3.5	3.33	3.8	487	
2005	ID	Idaho	542	3506	544	3.84	2	3.5	3.9	3.48	2.4	3.31	3.8	3.45	3.4	3.57	3.4	520	
2005	IL	Illinois	606	12970	594	3.85	2.1	3.55	4	3.5	3.4	3.42	4	3.48	3.7	3.57	3.6	569	
2005	IN	Indiana	508	41553	504	3.77	2.2	3.25	3.9	3.17	2.8	3.03	3.8	3.19	3.4	3.37	3.3	478	
2005	IA	Iowa	608	1671	596	3.92	2.8	3.75	4.1	3.68	3.5	3.6	4	3.69	3.8	3.78	3.7	585	
2005	KS	Kansas	588	2667	585	3.92	2.5	3.62	4	3.57	3	3.49	3.9	3.59	3.6	3.68	3.7	558	
2005	KY	Kentucky	559	4666	561	3.87	2.1	3.57	3.9	3.57	2.9	3.46	4	3.55	3.6	3.65	3.5	520	
2005	LA	Louisiana	562	3290	565	3.87	1.8	3.46	4	3.46	2.8	3.34	4.1	3.42	3.8	3.52	3.6	497	

Section 5. Visualization Tools

- I chose Tableau for each goal, because it was the visualization tool that I worked with the most and was the one I was the most comfortable with. It also had all the applications that I needed to complete my data visualization.
- The detailed work for this project is available on Tableau Public: https://public.tableau.com/views/Elsaadi_Final_project/Goal6?:language=en-US&:display_count=n&:origin=viz_share_link
- All the project's files are available in GitHub: https://github.com/NaiemaElsaadi/Final_Project_Data_Visualization

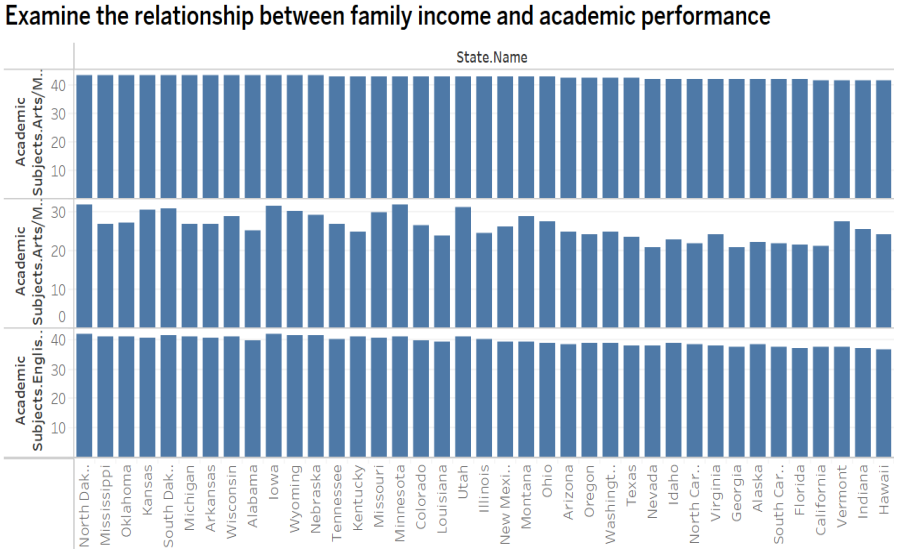
Section 6. Visualizations and Stories

- **Goal 1: Visualize the Distribution of Math and Verbal scores.**
  - For this goal I used a Side-by-Side Bar Chart



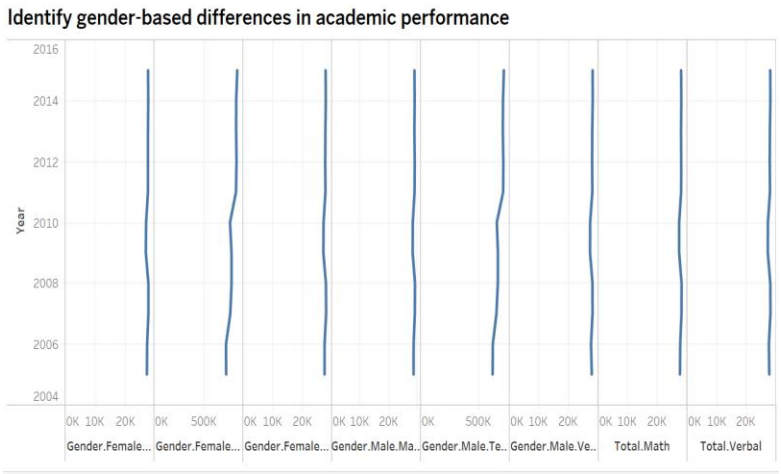
  - **Story:**
    - This chart compares the total math and verbal scores for each state and US territory, effectively showing which states score higher on these tests in comparison to each other. Looking at this, we can see that North Dakota has the highest total math and verbal scores, while the US Virgin Islands have the lowest scores.

- **Goal 2: Examine the Relationship between Family Income and Academic Performance.**
  - For this goal I used multiple Side by Side Bar Charts.



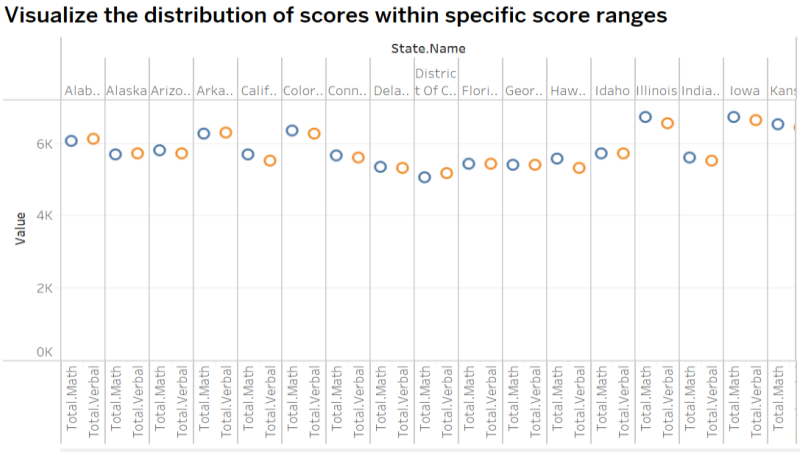
- **Story:**
  - This chart helps to show the relationship between Family Income and GPA and courses. We can see in this graph that the states that have the highest amount of High-Income test takers such as North Dakota generally have higher average GPAs as well as stay in courses for a longer time.

- **Goal 3: Identify Gender-Based Differences in Academic Performances.**
  - For this goal I used a Line Chart



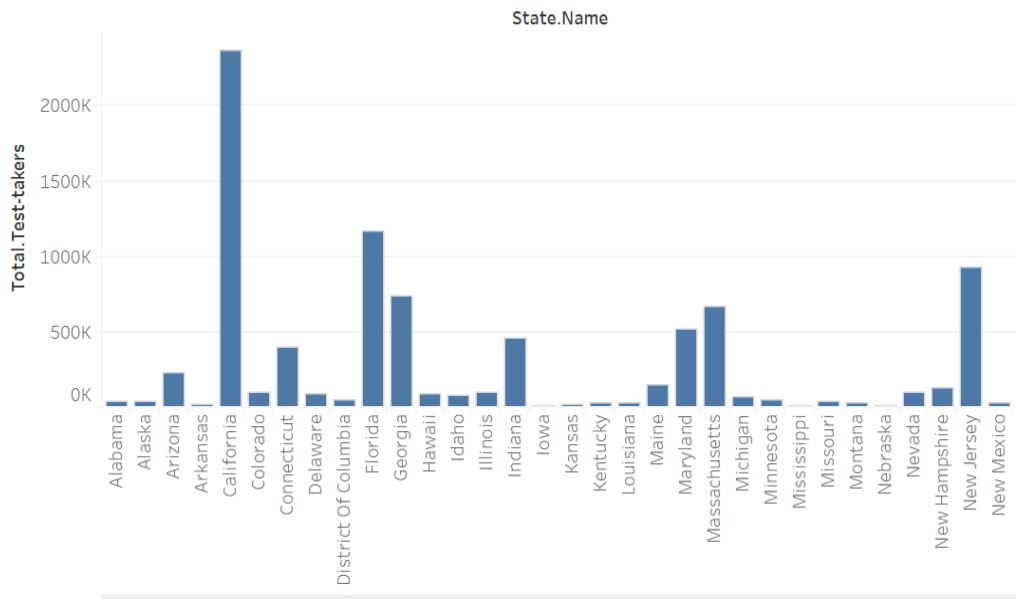
  - **Story:**
    - This graph compares the academic performance between Males and Females. Based on this chart we see that there are a lot more Female test takers than Male test takers, but the Female test takers generally perform at a slightly lower level than the Male test takers.

- **Goal 4: Visualize the Distribution of High Achieving Scores.**
  - For this goal I used a Shape Chart



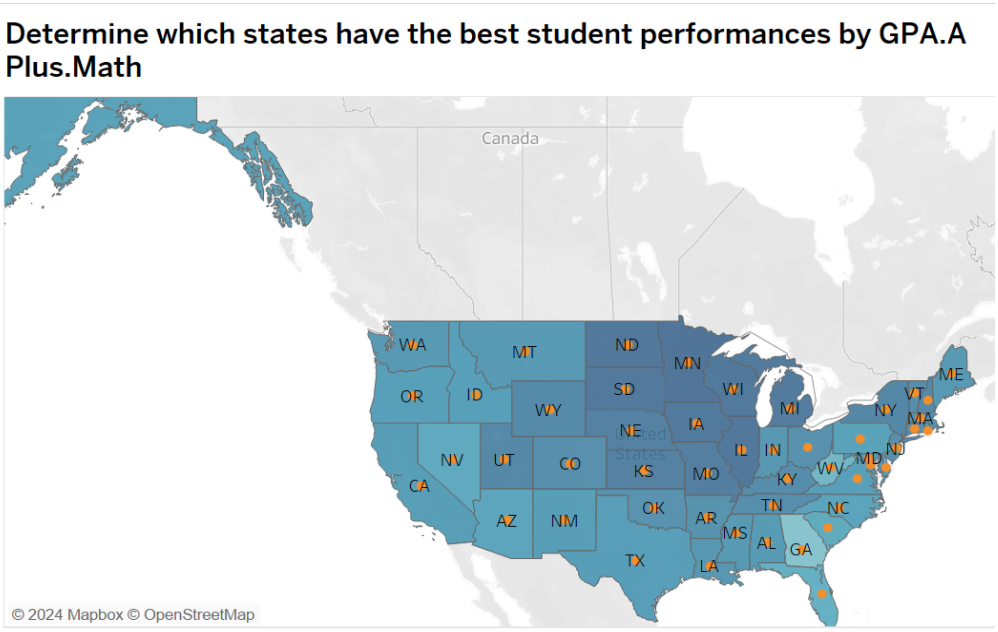
- **Story:**
  - This chart shows the distribution of students that scored between 700-800 in the Math and Verbal sections of their standardized tests for each state and territory. Based on this we see that the US territory as whole generally performed much worse than the US States. Of the US states, Florida and Georgia were off the lowest scoring states, and all the highest scoring states were all generally in the Midwest as seen with North Dakota, Illinois, and Iowa.

- **Goal 5: Determine which States have the most Test Takers each year.**
  - For this goal I used a Bar Chart



  - **Story:**
    - This graph shows the states with the most test takers. For the most part these are all correlated with the population of the states, but it provides some very important background information.

- **Goal 6: Determine which States have the best average Student GPAs**
    - For this goal I used a Heat Map of the United States



    - **Story:**
        - This heat map of the United States shows the states that have the highest average GPA. This shows us that the area of the US that has the Highest performing students is the Midwest, while we see lower averages in the south to southwest part of the United States (excluding Texas).

Section 7. Conclusions

In conclusion, by analyzing the variables from the dataset we can come up with a stronger understanding of the factors that can predict a student's academic performance. The main variables focused on in this data set and in this project are GPA, Courses, Family Income, Standardized Test Scores, and State for students. Seeing all these graphs and charts next to each other helps to show the relationship that each state has with academics.

The data visualized with these graphs show that the states that are the best performing academically are all midwestern states ranging from North Dakota to Illinois. While some of the southern states and the US territories do not have the best academic performances in comparison to the rest of the states. The Heat Map of the United States best shows this as it shows that the population of student's that have high GPAs are very prominent in the Midwest. The Shape Chart also did a good job in showing this, because the chart showed that for the most part the states that had the highest scoring students were midwestern states with a couple of exceptions. This Shape Chart also does a good job in supporting the conclusion that some of the southern states and US territories have the lowest test scores in comparison with the rest, because it shows how all of the US territories and some of the Southern states have a very low count of high performing test scores.