US Census Demographic Data

Tableau Data visualization: Nanodegree Project Report

Introduction

This report details the process and outcomes of Project 1 for the Data Visualization with Tableau Nanodegree. Using a dataset of US Census demographics, this project explores key demographic trends across the country. The following sections will elaborate on the methodology used to prepare the data, the analytical questions addressed, and the visual findings presented through Tableau dashboards and a final story.

About the dataset

The dataset used in this project is sourced from Kaggle [1]. It is an expanded version of an earlier dataset that focused only on New York City. This extended version includes demographic and socioeconomic data for the entire United States, enabling broader and more meaningful analyses at both county and state levels.

The dataset contains variables related to population, income, poverty, housing, transportation, and other demographic indicators, making it well-suited for exploratory analysis and visualization.

The dataset file used is acs2015-county-data.csv, it was loaded into Tableau as a text file data source. The dataset includes 37 fields and 3220 rows. Below table is a breakdown through the fields pulled out from documentation [1]:

Column	Data Type	Description
CensusId	Integer	Unique identifier for each census entry
State	String	U.S. state name
County	String	County name within the state
TotalPop	Integer	Total population of the county
Men	Integer	Male population count
Women	Integer	Female population count
Hispanic	Float	Percentage of population identifying as Hispanic
White	Float	Percentage of population identifying as White
Black	Float	Percentage of population identifying as Black

Column	Data Type	Description
Native	Float	Percentage of population identifying as Native American
Asian	Float	Percentage of population identifying as Asian
Pacific	Float	Percentage of population identifying as Pacific Islander
Citizen	Integer	Number of U.S. citizens in the county
Income	Integer	Median household income
IncomeErr	Integer	Margin of error for household income
IncomePerCap	Integer	Per capita income
IncomePerCapErr	Integer	Margin of error for per capita income
Poverty	Float	Percentage of population below poverty line
ChildPoverty	Float	Percentage of children below poverty line
Professional	Float	Percentage employed in professional/management jobs
Service	Float	Percentage employed in service jobs
Office	Float	Percentage employed in office/administrative jobs
Construction	Float	Percentage employed in construction/extraction jobs
Production	Float	Percentage employed in production/transport jobs
Drive	Float	Percentage commuting by driving alone
Carpool	Float	Percentage commuting by carpool
Transit	Float	Percentage commuting by public transit
Walk	Float	Percentage commuting by walking
OtherTransp	Float	Percentage commuting by other means
WorkAtHome	Float	Percentage working from home
MeanCommute	Float	Mean commute time (minutes)
Employed	Integer	Number of employed individuals
PrivateWork	Float	Percentage employed in private sector
PublicWork	Float	Percentage employed in public sector
SelfEmployed	Float	Percentage self-employed
FamilyWork	Float	Percentage working in family businesses
Unemployment	Float	Unemployment rate

Data Assessing and cleaning

Before conducting the analysis, the dataset was assessed for **quality and structure**. Tableau's **Data Interpreter** was used to detect headers, remove extraneous formatting, and prepare the dataset for analysis.

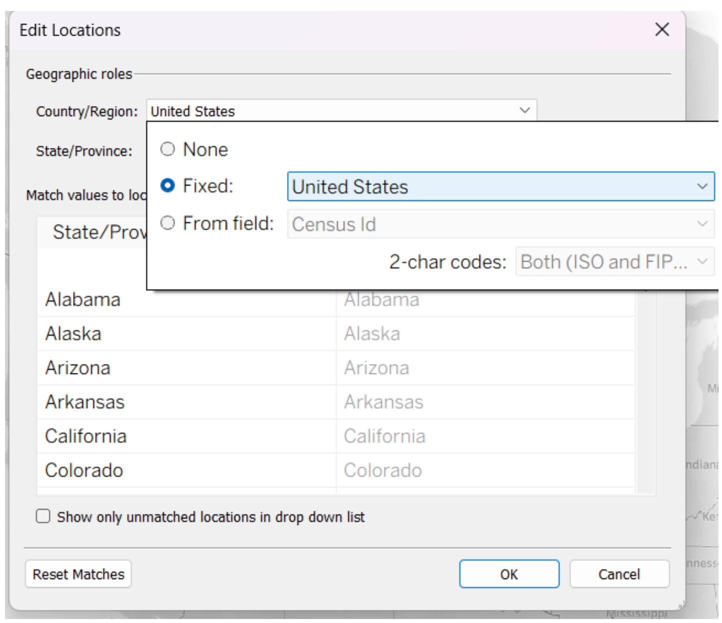
CensusId	State	County	TotalPop	Men	Women	Hispanic	White	Black	Native	Asian	Pacific	Citizen	Income	IncomeErr	IncomePer	IncomePer	Poverty	ChildPover	Profession	Service	Office
1001	Alabama	Autauga	55221	26745	28476	2.6	75.8	18.5	0.4	1	0	40725	51281	2391	24974	1080	12.9	18.6	33.2	17	24.2
1003	Alabama	Baldwin	195121	95314	99807	4.5	83.1	9.5	0.6	0.7	0	147695	50254	1263	27317	711	13.4	19.2	33.1	17.7	27.1
1005	Alabama	Barbour	26932	14497	12435	4.6	46.2	46.7	0.2	0.4	0	20714	32964	2973	16824	798	26.7	45.3	26.8	16.1	23.1
1007	Alabama	Bibb	22604	12073	10531	2.2	74.5	21.4	0.4	0.1	0	17495	38678	3995	18431	1618	16.8	27.9	21.5	17.9	17.8
1009	Alabama	Blount	57710	28512	29198	8.6	87.9	1.5	0.3	0.1	0	42345	45813	3141	20532	708	16.7	27.2	28.5	14.1	23.9
1011	Alabama	Bullock	10678	5660	5018	4.4	22.2	70.7	1.2	0.2	0	8057	31938	5884	17580	2055	24.6	38.4	18.8	15	19.7
1013	Alabama	Butler	20354	9502	10852	1.2	53.3	43.8	0.1	0.4	0	15581	32229	1793	18390	714	25.4	39.2	27.5	16.6	21.9
1015	Alabama	Calhoun	116648	56274	60374	3.5	73	20.3	0.2	0.9	0	88612	41703	925	21374	489	20.5	31.6	27.3	17.7	24.2
1017	Alabama	Chambers	34079	16258	17821	0.4	57.3	40.3	0.2	0.8	0	26462	34177	2949	21071	1366	21.6	37.2	23.3	14.5	26.3
1019	Alabama	Cherokee	26008	12975	13033	1.5	91.7	4.8	0.6	0.3	0	20600	36296	1710	21811	1556	19.2	30.1	29.3	16	19.5
1021	Alabama	Chilton	43819	21619	22200	7.6	80.5	10.2	0.4	0.3	0	31728	41627	2025	21399	1446	19.1	30.6	27.2	14.3	23.7
1023	Alabama	Choctaw	13395	6382	7013	0.4	55.9	42.9	0	0	0	10568	33536	2231	20755	1380	23.4	34.4	24.9	15.4	19.6
1025	Alabama	Clarke	25070	11834	13236	0.3	53.4	45.3	0	0.4	0.4	19258	32011	3088	20064	1469	24.7	29.8	21.1	15	25.5
1027	Alabama	Clay	13537	6671	6866	3.2	79.9	14.4	0.7	0	0	10312	35327	4517	18905	1067	16.7	22.5	21.5	13.3	20.8
1029	Alabama	Cleburne	15002	7334	7668	2.3	92.5	2.9	0.2	0.4	0	11367	38056	3883	20151	1357	17	26.3	28.9	11.7	23.1
1031	Alabama	Coffee	50884	25174	25710	6.4	71.5	17.2	0.8	1.2	0	37575	46729	2066	24936	854	17.1	26.4	32.2	18.6	22.7
1033	Alabama	Colbert	54444	26303	28141	2.4	78.9	15.6	0.6	0.3	0.1	42075	40576	2536	22546	917	17.4	26.1	27.3	15.5	25.7
1035	Alabama	Conecuh	12865	6176	6689	1.6	51	44.7	0.3	0	0	9930	24900	3001	15968	1369	33.8	41.5	19.4	21	20.4
1037	Alabama	Coosa	11027	5579	5448	2.1	65.2	30.7	0.1	0	0	8803	31212	2971	17519	961	19.3	27.5	18	19.8	21.9
1039	Alabama	Covington	37886	18339	19547	1.5	83.3	13	0.5	0.3	0.1	29358	36444	1673	20977	723	20.2	31.2	26.2	15.6	23.6
1041	Alabama	Crenshaw	13938	6863	7075	1.7	70.5	23.5	0.8	1.8	0	10591	36022	3041	20585	1229	16.7	19.2	27.6	11.6	23.8
1043	Alabama	Cullman	80965	40081	40884	4.3	92.2	1.1	0.4	0.6	0	61222	38971	1526	20881	727	18.5	25	28.5	15.6	23.9
1045	Alabama	Dale	49866	24708	25158	6	69.8	19.4	0.5	1.2	0	37331	45028	2413	22566	911	19.1	26.2	27.1	17.2	24.8
1047	Alabama	Dallas	42154	19450	22704	0.3	28.6	69.2	0.2	0.3	0	31231	27306	2075	17808	788	35	53.8	26.5	19.5	19.3
1049	Alabama	DeKalb	71068	35474	35594	14	80.9	1.8	1.1	0.3	0.1	49341	38192	1360	18411	588	18.7	25.2	24.2	15.7	21.3
1051	Alabama	Elmore	80763	39362	41401	2.8	73.6	21	0.2	0.5	0	61172	53555	1980	24381	941	13.6	20.5	33.3	16	25.2
1053	Alabama	Escambia	37935	19524	18411	1.2	60.7	33.4	3.2	0.1	0	29399	32330	2209	17045	848	24.2	31.8	25.4	21.8	23.1
1055	Alabama	Etowah	103766	50207	53559	3.6	78.5	15.4	0.3	0.7	0.1	78881	39220	1976	20500	518	19.7	31.7	28.7	18.5	21.3

Additional checks (data types, missing values) were manually reviewed to ensure consistency. The dataset does not show missing values and data types that were automatically set are quite accurate.

However, IDs are usually categorical identifiers, not numerical values. So the data type is set into String.

To use the Map visualization, the states were defined as Unknown as shown in the below screenshot:

The solution was to click on the 52 unknown and change the location to United states.



Now tableau identifies all the states properly.

Data Visualization

In this section, each of the main questions is explained and answered using Tableau visualizations, with attention to design choices.

Q1: Which states have the best transportation?

This is a subjective question, as “best transportation” can be defined from different perspectives. Does it mean the **highest percentage of people using public transit**? Or does it mean the **shortest average commute times**?

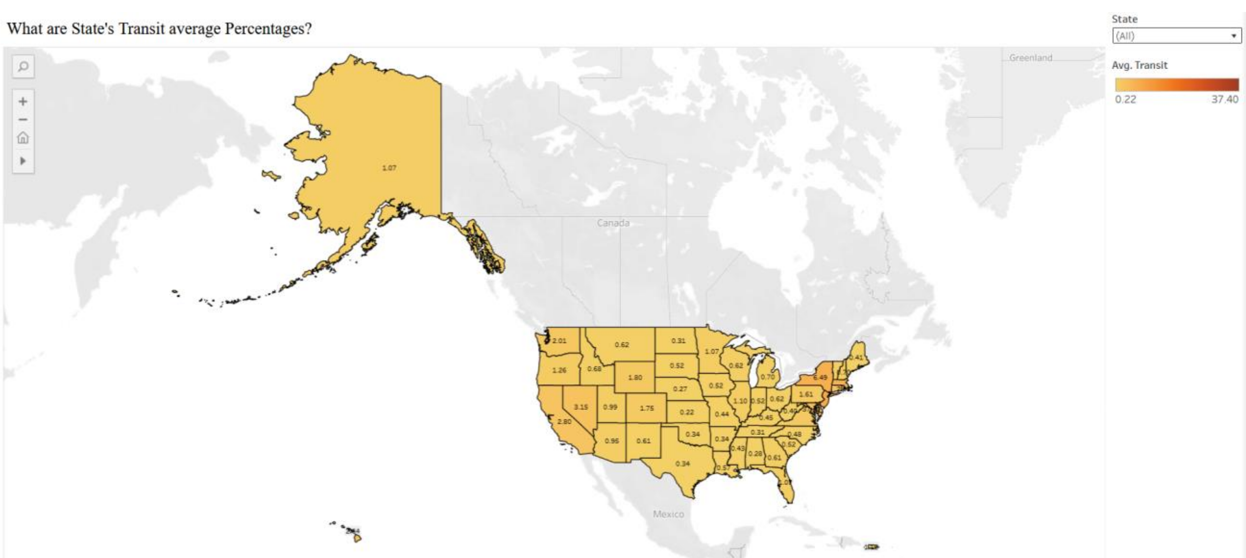
To approach this, the dataset was aggregated from the county level to the state level using a hierarchy of *State* → *County*. This allowed for analysis at state-level perspective.

Transit Usage: What are State's Transit average Percentages?

To explore transit usage, a map visualization was created showing the aggregated average percentage of commuters using public transit by state. A state-level filter was applied, and an orange color palette was chosen, as orange symbolizes movement and human activity. State borders were emphasized with bold outlines at 100% opacity to clearly distinguish boundaries and the total population variable was included as a detail. This chart type is well-suited for the data since it allows for intuitive geographic comparisons.

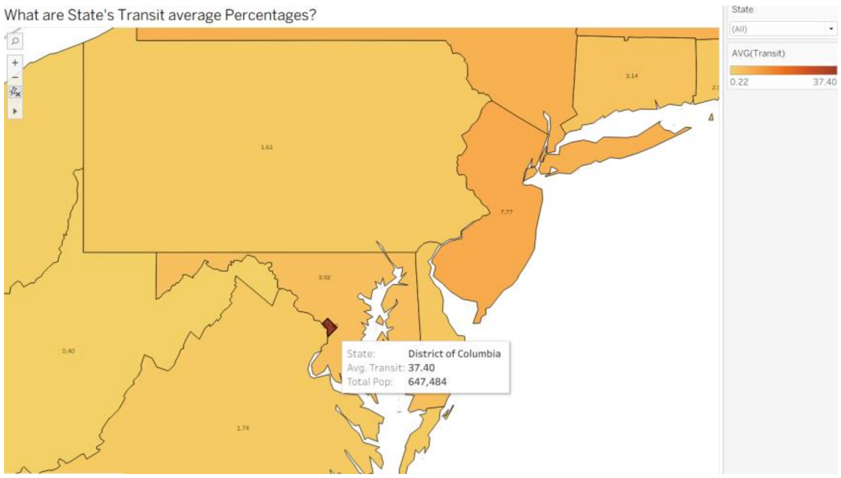
[Visualization 1: Link: Map Average Transit Across states]

The map highlights how transit usage varies across the U.S. Most states show relatively low percentages, reflected in lighter shades of orange



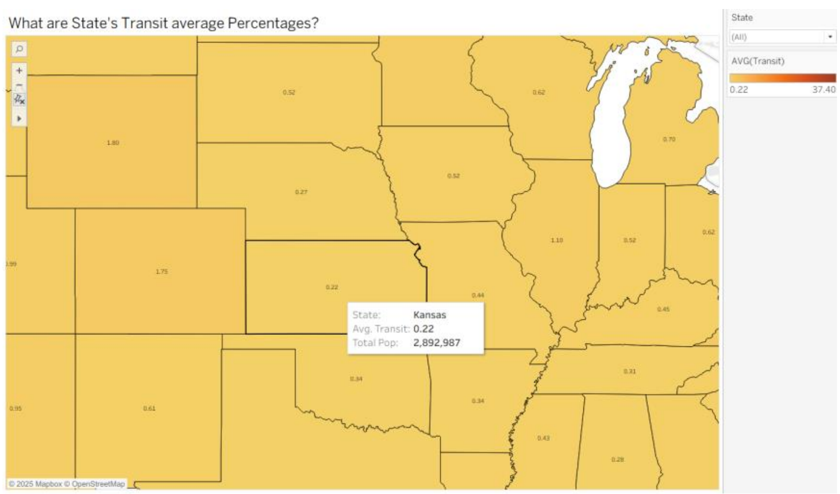
Highest Transit Use:

The District of Columbia stands out with the highest transit percentage (37.4%), shown in the darkest shade. This is logical given its smaller geographic area and dense urban population, where public transit is widely accessible and heavily relied upon.



Lowest Transit Use:

Kansas records the lowest average, with only 0.22% of commuters using transit..



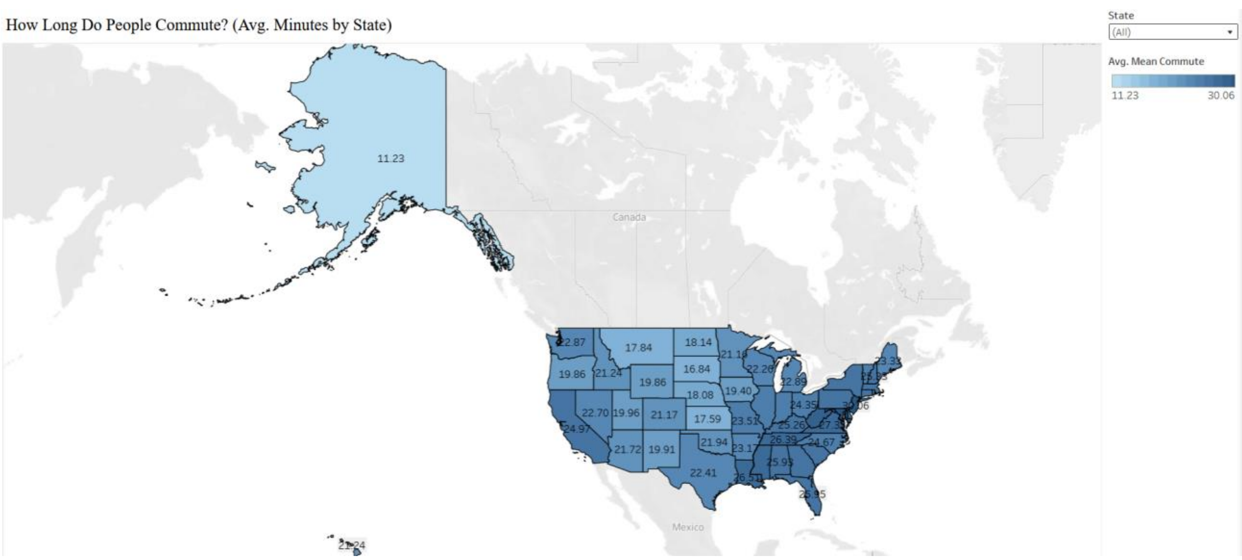
If we define "best transportation" as the **highest reliance on public transit**, then the District of Columbia is the strongest example, with 37.4% of commuters depending on transit.

Commute Duration: How long do people commute in the States (Avg. Minutes)?

A second **map visualization** was created showing the aggregated average commute time (in minutes) by state. A blue color palette was chosen, as blue conveys the negative experience of long, slow, or frustrating commutes. State borders were emphasized with bold outlines at 100% opacity, and the total population variable was included as a detail. This chart type is well-suited for the data since it allows for intuitive geographic comparisons.

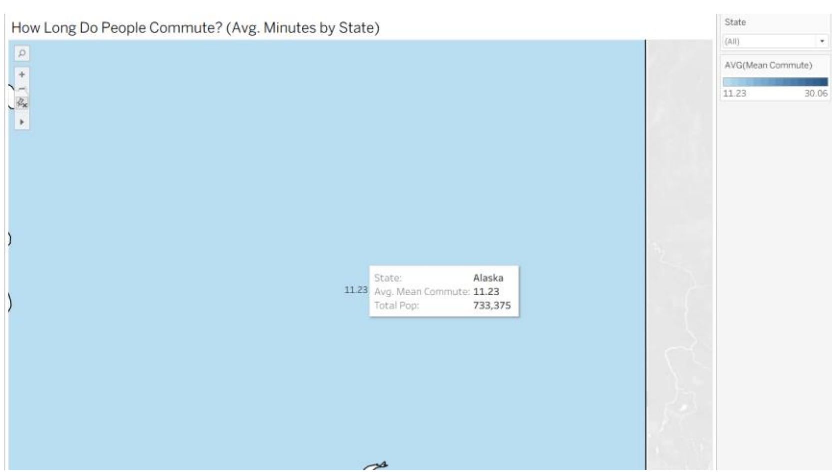
[Visualization 2: Map how long do people commute Avg Minutes?]

The map highlights variations in commute times across the U.S. Lighter shades represent shorter commutes, while darker shades represent longer commutes.



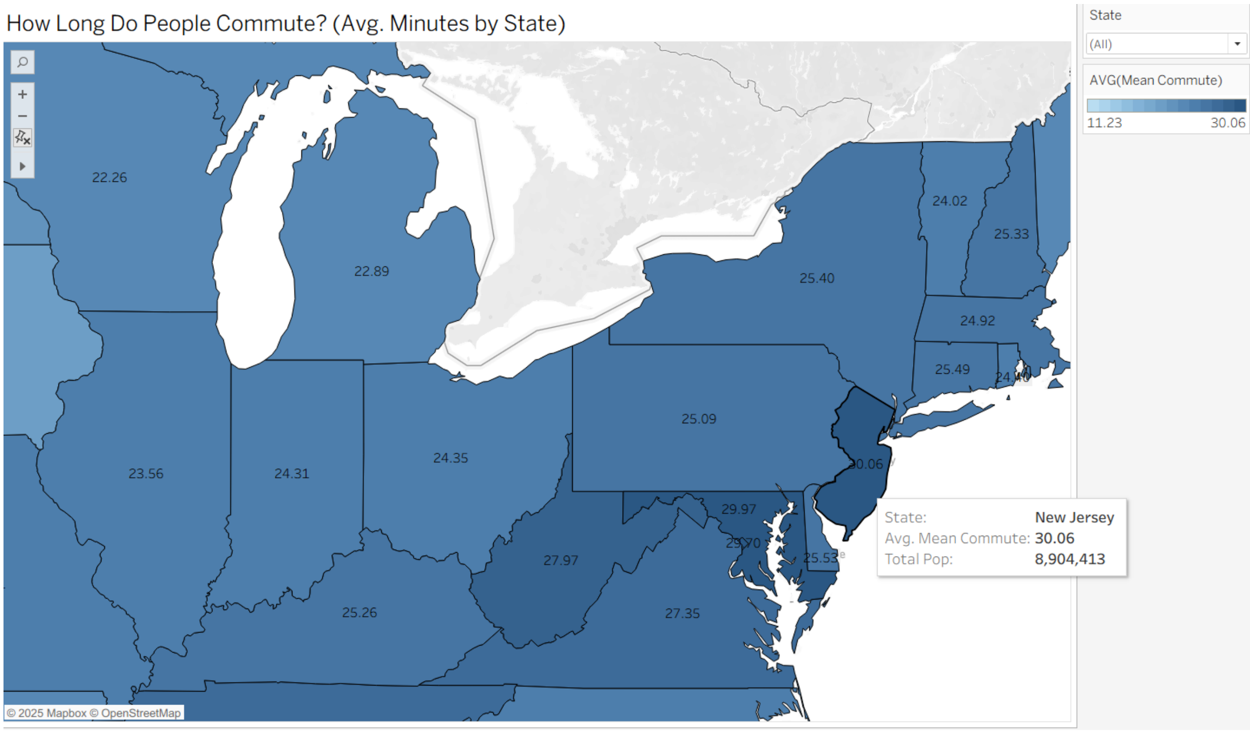
Shortest Commute:

Alaska has the lowest average commute time at 11.23 minutes.



Longest Commute:

New Jersey has the highest average commute time at 30.06 minutes.



So, which states have the best transportation?

- By **transit usage**, the District of Columbia ranks highest, with 37.4% of commuters relying on public transportation.
- By **commute time**, Alaska stands out, with the shortest average commute of just 11.23 minutes.

Therefore, **Alaska can be considered the state with the best transportation overall**, as it combines accessibility with significantly shorter commutes than the District of Columbia.

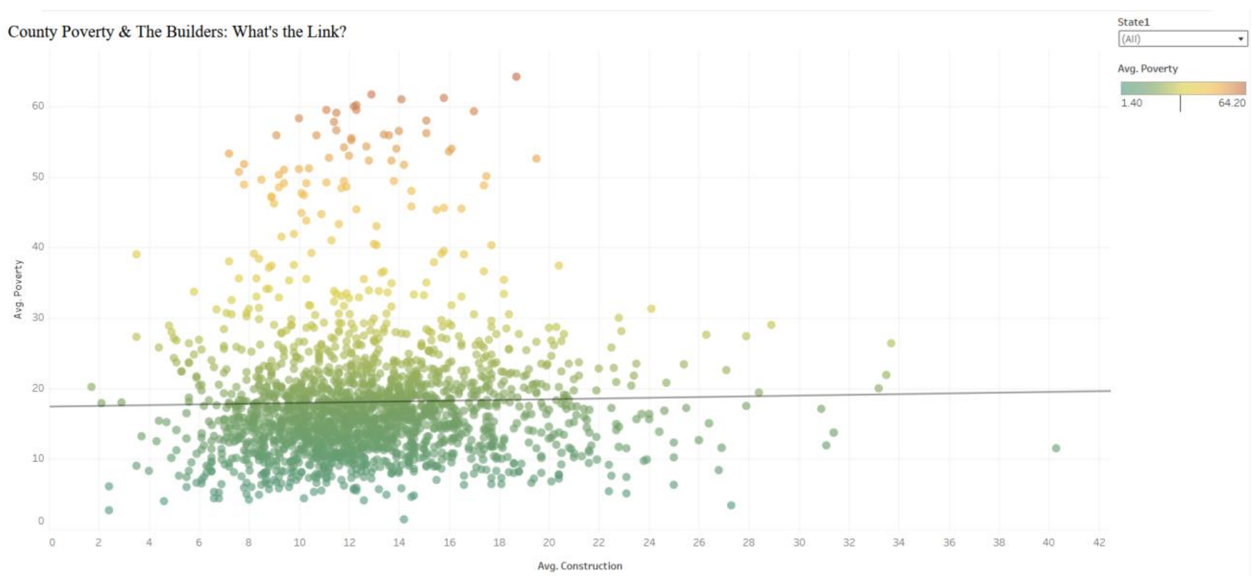
Q2: How do income and poverty look across America?

This is a two parts question that seeks to understand the correlation between poverty, constructions and near coast located counties. The question was answered by analyzing the dataset at the county level.

County Poverty and Construction: what's the link?

To explore the relationship, a scatter plot visualization was created, mapping the average construction employment rate (percentage) on the X-axis against the average poverty rate (percentage) on the Y-axis. The Scatter Plot is the ideal choice as it shows the correlation between two continuous variables, along with a line plot to check the correlation between the two variables. County is shown using the tooltips. A temperature green-to-orange color gradient reinforced the poverty percentage.

Visualization 3: Scatter, Do counties with more construction experience more or less poverty?



The scatter strongly suggests that **construction presence is not a decisive factor** in a county's poverty level. The linear trend line is virtually flat, indicating a negligible positive correlation. On average, a county with a higher construction rate is not statistically more or less likely to have a lower poverty rate than a county with a moderate construction rate. The majority of counties form a dense cluster where:

- Construction employment ranges from approximately **8% to 18%**.
- Poverty rates are consistently between **10% and 25%**.

- It is worth mentioning that the highest poverty percentages (up to 64.20%) occur in counties with only moderate construction employment (8% to 16%). This includes territories like Puerto Rico, which have unique economic circumstances.

Near-the-coast counties experience income

This question investigates the geographic benefits of a coastal location: such as access to global trade, major finance, and specialized industries if it results in a higher median income for the counties grouped within those states.

To approach this, the dataset was categorized at the county level using a calculated dimension, Coastal Status. This dimension groups all counties into one of two categories: those belonging to an Atlantic, Pacific, or Gulf state/territory ("Coastal") and those belonging to landlocked states ("Inland"). The analysis then compared the median income per capita for each group.

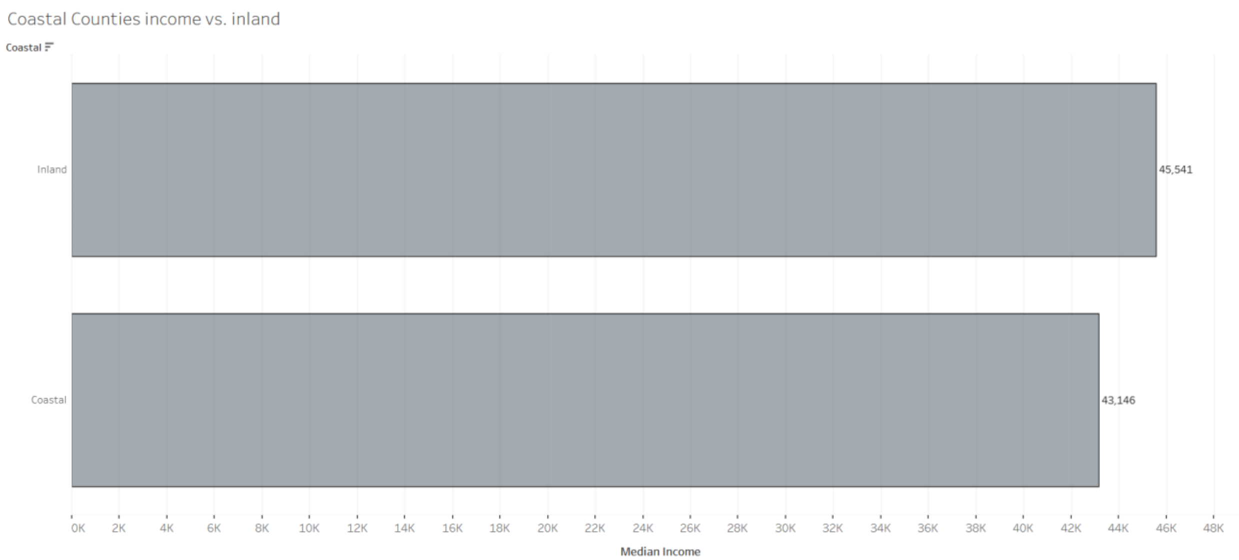
Coastal

```
IF [State1] IN (
    'California', 'Oregon', 'Washington', 'Alaska', 'Hawaii', 'Pacific',
    'Maine', 'New Hampshire', 'Massachusetts', 'Rhode Island', 'Connecticut',
 'New York', 'New Jersey', 'Delaware', 'Maryland', 'Virginia', 'North Carolina',
 'South Carolina', 'Georgia', 'Florida', 'District of Columbia', 'Atlantic',
    'Alabama', 'Mississippi', 'Louisiana', 'Texas', 'Puerto Rico', 'Gulf & Territories'
)
THEN "Coastal"
ELSE "Inland"
END
```

The calculation is valid. 1 Dependency Apply OK

A comparative bar chart was created to visualize the difference between the **Coastal** and **Inland** median income (measured in dollars). The **Comparative Bar Chart** is the most effective visualization for this question, as it allows for a direct, quantitative assessment of a single metric (median Income) across two categorical groups.

[Visualization 4: Bar plot: Near-the-coast Vs. Inland Income, which is more?]



The analysis shows that **Inland** counties have a marginally higher median income than Coastal counties.

- **Inland** Median Income: **$45,541**
- **Coastal** Median Income: **$43,146**

The Inland grouping reports a median income of **$2,395 higher** than the Coastal grouping. This finding suggests a possible **bias** in the data. While high-income metropolitan areas like New York, Boston, and Los Angeles are located on the coast, the overall "Coastal" average is diluted by:

1. The large number of lower-income **rural counties** located far from the ocean but still within a defined "coastal state" (e.g., counties in the interior of Texas or North Carolina).
2. The inclusion of **lower-income US territories** (e.g., Puerto Rico) which are classified as coastal.

The "Inland" category's median is boosted by the inclusion of wealthy, specialized inland economic hubs (e.g., major cities in the Midwest and Mountain West) that balance out its own rural, lower-income areas. Therefore, the **distribution of income** within the state categories is more complex than a simple coastal advantage.

Q3: Is there a relationship between a state's diversity (by combining different racial demographic data) and its socioeconomic well-being (e.g., median income, unemployment rate)?

This question moves beyond industry-specific factors to investigate the correlation between the demographic makeup of a state and its economic success. To dive into this investigation, we would explain two concepts: Diversity and Socioeconomic Well-being. A dashboard is used to make a full explanation of this area, this dashboard consists of the following three visualizations:

1. State diversity map
2. Top 5 earning states
3. State diversity vs. socioeconomic wellbeing.

First let's explore the two concepts in examine:

- **Socioeconomic Well-being**

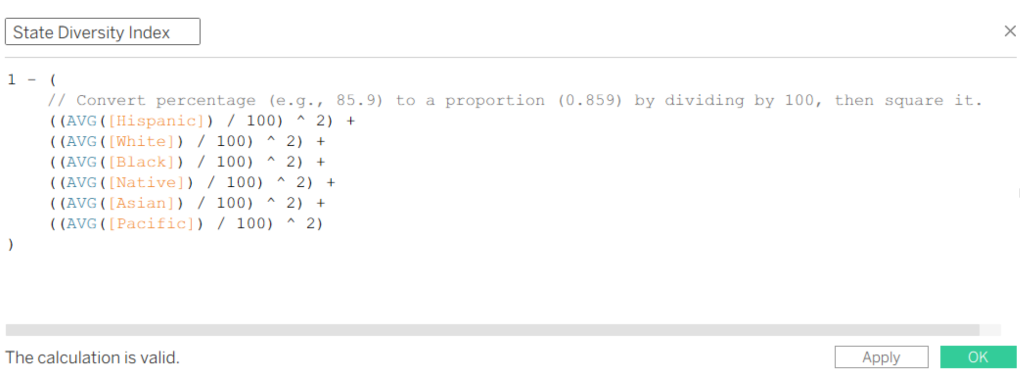
Using this dataset, this is measured either by Unemployment rate or using the median income. This metric is at the State level.

- **Diversity Index**

The diversity Index (DI) is a common method is to use a fractionalization index, which measures the probability that two randomly selected people from a state will belong to different racial or ethnic groups. [2] This metric is calculated at the State level and the following equation is how it is calculated:

$$\text{Diversity Index} = 1 - \left(\left(\frac{\text{AVG(Hispanic)}}{100}\right)^2 + \left(\frac{\text{AVG(White)}}{100}\right)^2 + \left(\frac{\text{AVG(Black)}}{100}\right)^2 + \left(\frac{\text{AVG(Native)}}{100}\right)^2 + \left(\frac{\text{AVG(Asian)}}{100}\right)^2 + \left(\frac{\text{AVG(Pacific)}}{100}\right)^2\right)$$

This is now a calculated field from the dataset as follows:

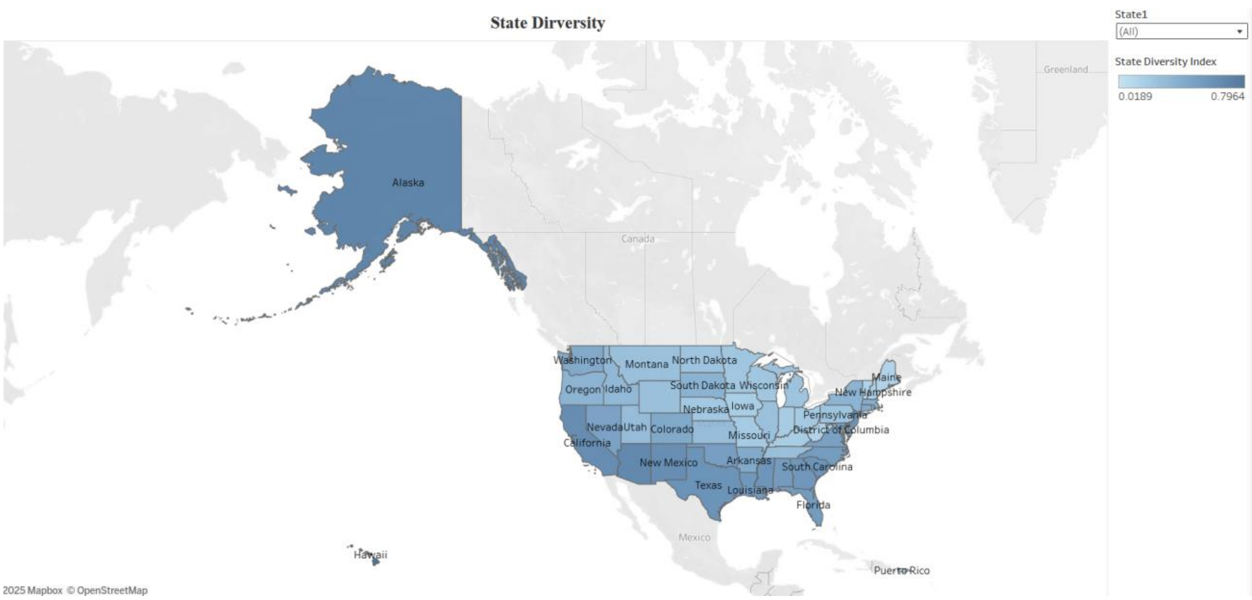


1. State Diversity Map

This visualization provides a geographic dimension to the diversity data, allowing for the identification of regional demographic trends across the United States.

[Visualization 6: Map that shows the state diversity index]

A Filled Map (Choropleth) was chosen because it is the most intuitive way to display how a variable like diversity varies across geographic boundaries. A Blue-to-Teal sequential color palette was selected to represent the Diversity Index. Darker shades represent higher diversity levels (probability of two people belonging to different groups), providing a clear visual "weight" to more diverse states.



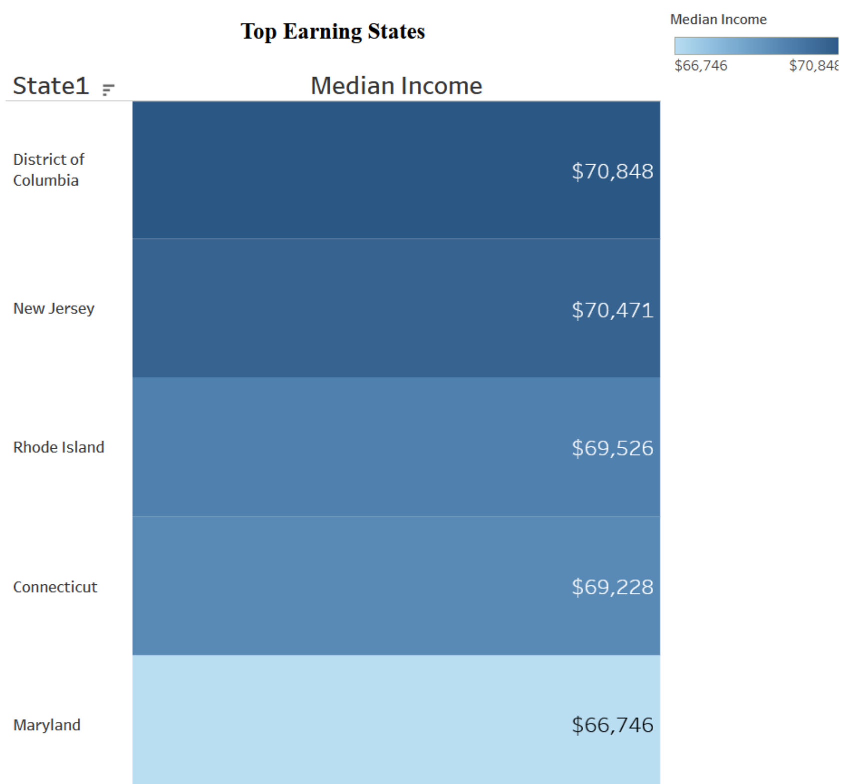
The map illustrates that ethnic diversity is not uniform across the country; it is heavily concentrated in the Western, Southwestern, and Coastal regions. High Diversity Index (DI) scores are prominent in states like California and Hawaii. whereas the more homogeneous (lower DI) states are clustered in the Midwest and Northern Plains.

2. Top 5 earning states

This sheet provides a focused look at the "Economic Leaders" to determine if the top-tier earners share a common demographic profile.

Visualization 6: Top earning states

A Highlight Table (Ranked) was chosen over a simple list to make the data "scannable." The Blue palette was used to maintain consistency with the map, ensuring that the "Diversity Index" values shown next to the income figures are visually tied to the rest of the dashboard. The font was set to higher size to make it visible.



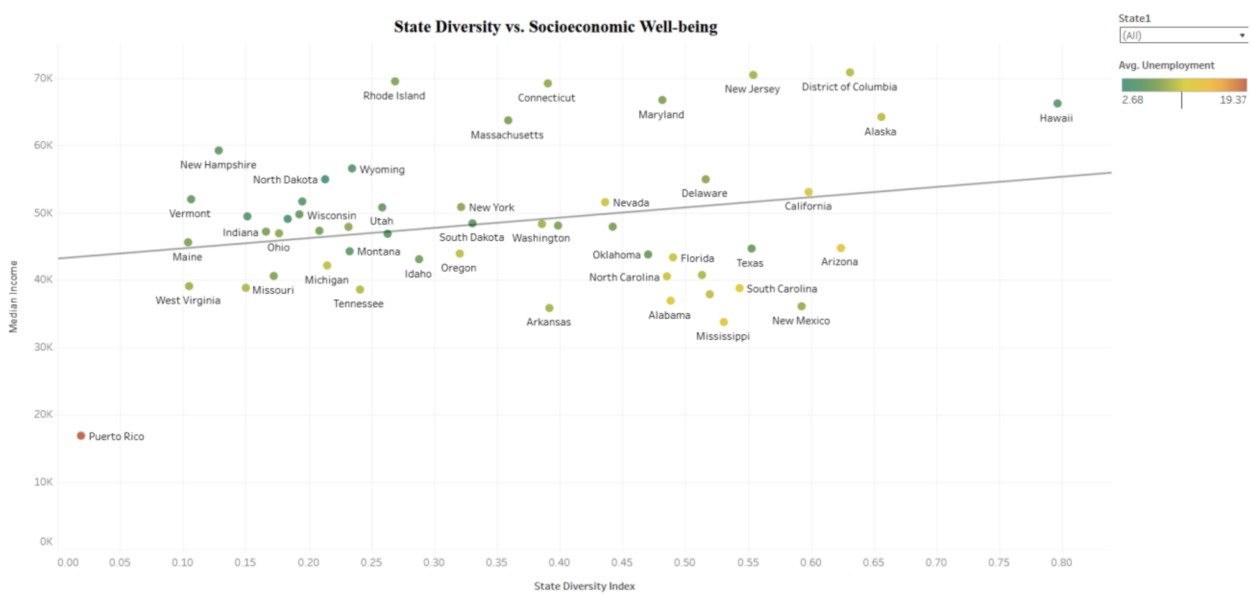
This list highlights the five states with the highest **Median Income** (measured in dollars), led by the **District of Columbia** at approximately **$70,848**. A key finding is that four out of the top five highest-earning states, New Jersey, Maryland, and Connecticut, possess moderate to high Diversity Indices (ranging from roughly 0.39 to 0.63). This insight is relevant as it provides that the nation's successful economies could be largely built within diverse demographic environments.

3. State Diversity vs. Socioeconomic Well-being

This question investigates the relationship between the **racial and ethnic diversity** of a state's population and its overall socioeconomic health, as measured by **Median Income.** The goal is to determine the correlation of highly diverse populations with economic prosperity.

The Scatter Plot is the ideal visualization, as it directly maps the value of two continuous variables, enabling the identification of clusters and the strength of the relationship via a trend line. State names were directly labeled on the scatter plot to provide immediate context, allowing the reader to quickly identify specific outlier states and cluster members. With the option to filter by state, as well a temperature green-to-orange color palette was chosen for the Unemployment Rate, with darker green representing the lowest (best) unemployment rates. This reinforces the interpretation that employment health contributes significantly to overall socioeconomic well-being.

Visualization 5: Scatter Plot: State diversity Index and the correlation with Income, highlighting Unemployment rate as average.



The scatter shows a **weakly positive linear correlation** between the two variables. The trend line indicates that states with higher diversity tend to have a slightly higher median income. The data reveals the following findings:

- The highest-income states (those exceeding \$65,000 in Median Household Income), such as Hawaii (DI 0.77) and California (DI 0.65), are overwhelmingly found in the high-diversity cluster (upper-right quadrant). This suggests that high racial/ethnic diversity is compatible with economic performance.

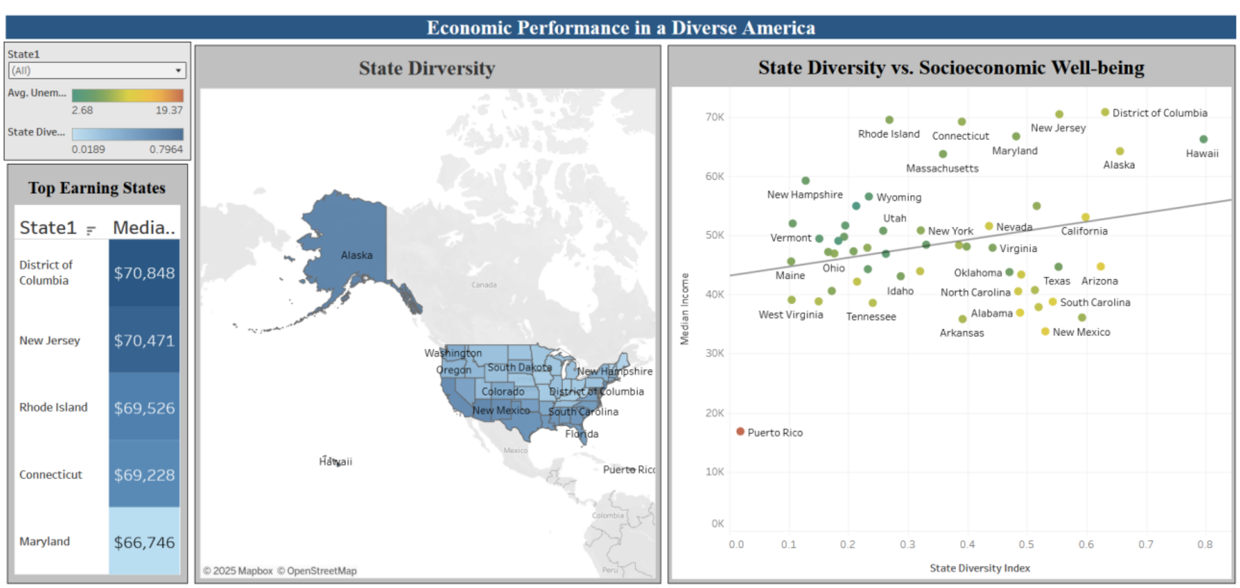
States like New Hampshire and Wyoming achieve high income with low diversity, they also boast some of the lowest unemployment rates. This contrast highlights that socioeconomic well-being is not solely driven by diversity or income but is significantly influenced by labor market factors.

Dashboard: Economic Performance in a Diverse America

The final dashboard integrates the scatter plot, map, and ranked list to provide a comprehensive answer to the relationship between diversity and socioeconomic health. The dashboard layout follows a "Macro-to-Micro" flow: the title sets the theme, the map provides the geographic "Where," the scatter plot explains the "Why" (correlation), and the ranked list provides the "Who" (specific leading states).

[Dashboard: Economic Performance in a diverse America]()

A filter of state was applied to only two sheets, the map and the scatter, given that the list filter, a top 5 filter was applied in a fixed way to only show the top 5 median income. A border was applied, with different background colors to each sheet and to the filter, legends. The choice to use a monochromatic blue scale for diversity across all sheets ensures that the user does not have to relearn the "meaning" of colors as they move between visualizations.



Conclusion

This project utilized US Census data to explore the complex intersections of geography, industry, and demography in relation to economic success. Through the systematic application of Tableau's visualization capabilities, several key conclusions have been reached:

- **Transportation as a Metric of Well-being:** While the District of Columbia leads in infrastructure utility (37.4% transit usage), **Alaska** offers a superior quality of life in terms of time, with a commute of only **11.23 minutes**. This highlights that "best" transportation is context-dependent, balancing infrastructure density with time efficiency.
- **Decoupling Construction and Poverty:** The analysis debunked the assumption that construction employment significantly mitigates poverty at the county level. The flat trend line in the scatter plot suggests that economic resilience requires a more diversified industrial base than construction alone.
- **The Inland Economic Strength:** Contrary to the "coastal advantage" hypothesis, **Inland counties** maintained a marginally higher median income (**$45,541**) than coastal ones (**$43,146**). This underscores the economic importance of specialized inland hubs and suggests that being "landlocked" is not a barrier to socioeconomic health.
- **Diversity as an Economic Asset:** The core finding of this report is the weakly positive correlation between a state's diversity and its median income. The presence of high-diversity states like **Maryland** and **DC** at the top of the income rankings indicates that multiculturalism and economic prosperity are strongly compatible in the modern American economy.

Recommendations

Based on the data-driven insights presented in this report, the following recommendations are proposed:

1. Targeted Urban Planning for Transit

States with high commute times but low transit usage (observed in various East Coast corridors) should look to the **District of Columbia's** model of high-density transit integration. Reducing the **30.06-minute average commute** in states like New Jersey should be a primary policy goal to improve worker well-being.

2. Broadening Economic Development Beyond Construction

Since construction employment showed a negligible correlation with poverty reduction, local governments in high-poverty areas (such as the identified clusters in **Puerto Rico**) should pivot toward **Professional and Service-sector** development. These sectors likely offer more sustainable pathways out of poverty than manual labor-intensive industries alone.

3. Leveraging Diversity for Regional Growth

States in the **Midwest and Northern Plains**: which currently show lower Diversity Indices—should consider policies that attract diverse talent. Since diversity correlates with higher median incomes (as seen in the **Top 5 Earning States**), fostering inclusive environments could serve as a catalyst for moving these states toward the "Upper-Right" quadrant of economic performance.

4. Further Investigation into "Inland" Wealth

It is recommended that future analyses further segment "Inland" data to identify the specific industries (e.g., Tech in Austin, Finance in Chicago) that allow these counties to outperform coastal areas. This would provide a more granular blueprint for economic success for other landlocked regions.

Sources

- [1] Dataset on Kaggle
- [2] Diversity index