What affects income more? Education Level or Sunny Weather

By Muhammad Areeb

5th December 2024

DATA 363

To Dr. Jaehui Lim

Contents

Figures

Introduction

Education and income are the two most fundamental metrics by which an individual's standing is evaluated on the socio-economic pyramid, and their relationship is widely studied across multiple disciplines. The general assumption is that higher education leads to higher income and prosperity as a good level of education leads to better or wider prospects and consequently should pay better. However, this relationship isn't always as straightforward as it seems since multiple variables like economy, job market, laws, or even location are in play.

Another interesting metric to explore is the overall climate of a region. It is commonly assumed that most people prefer sunny weather to colder frigid ones as suggested by (Pew Research Center, 2009). This preference can lead to more jobseekers and families wanting to relocate in tropical or subtropical climates, which could boost the economy of the region and drive up the prices, which could in turn lead to higher incomes to cope with the high cost of living.

This project aims to analyze two hypothetical comparisons—whether the income of a place is largely influenced by the education level of an individual or can geographical features like more sunlight play a larger role in determining a region's economy.

Methodology

This project aims to compare two response variables, sunlight and education level against the explanatory variable—income. For this, three different datasets were collected, namely Global Horizontal Irradiation (GHI), median household income data, and percentage of bachelor's degree graduates' data. All of these datasets were collected based on county values for the United States to make a fair comparison.

Sunlight (GHI)

The first response variable *sunlight* is collected by solar radiation dataset provided by the World Bank (Global Solar Atlas, n.d). This dataset is containing the irradiation values collected when the sun strike the surface of the earth, which is primarily used to estimate photovoltaic potential, but is being used to estimate the overall amount of sunlight a region gets. This dataset is further processed and classified based on county values using a combination of ArcGIS Pro software and Python Programming language.

Education Level

The second response variable *education* is collected through the datasets provided by the National Institute on Minority Health and Health Disparities (HDPulse, 2024). This dataset provides the number and percentage of people in a given county with a bachelors' degree. Percentage is used instead of the actual number, because it gives an accurate representation of proportions, whereas number of people with bachelor's degree can conflate the arguments, because big cities will tend to have a higher population resulting in a higher number of people having with undergraduate education, even if the overall proportion of them is quite low.

Income

The explanatory variable, income is collected through the datasets provided by the National Institute on Minority Health and Health Disparities (HDPulse, 2024). This dataset provides the median hosehold income sourced by counties.

Correlation

Pearson's correlation coefficient is used, which is a statistical measure that quantifies the strength and direction of the linear relationship between two variables. This coefficient ranges from -1 to +1, where -1 indicates a perfect negative correlation and +1 indicates a perfect positive correlation. Values equal or close to zero represent no linear relationship. The Person's correlation coefficient is calculated using the formula below (Britannica, 2024).

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \hat{y})}{\sqrt{\Sigma(x_i - \bar{x})^2(y_i - \hat{y})^2}}$$

$r = Pearson's\ Correlation\ Coefficient$

$x_i = Explanatory\ Variable\ (sunlight\ or\ education\ level)$

$\bar{x} = Mean\ of\ Explanatory\ Variable$

$y_i = Response\ Variable\ (Per - Capita\ Income)$

Data Summaries and Visualization

The map shown in (Figure 1), shows the distribution of sunlight across all counties of continental United States, the units for which are watt-hour per sq. meter (Wh/m^2). It can be observed that the areas in the south tend to have higher solar radiation as compared to regions in the north due to their proximity to the equator and the orientation of sunlight hitting the surface (BBC, n.d.).

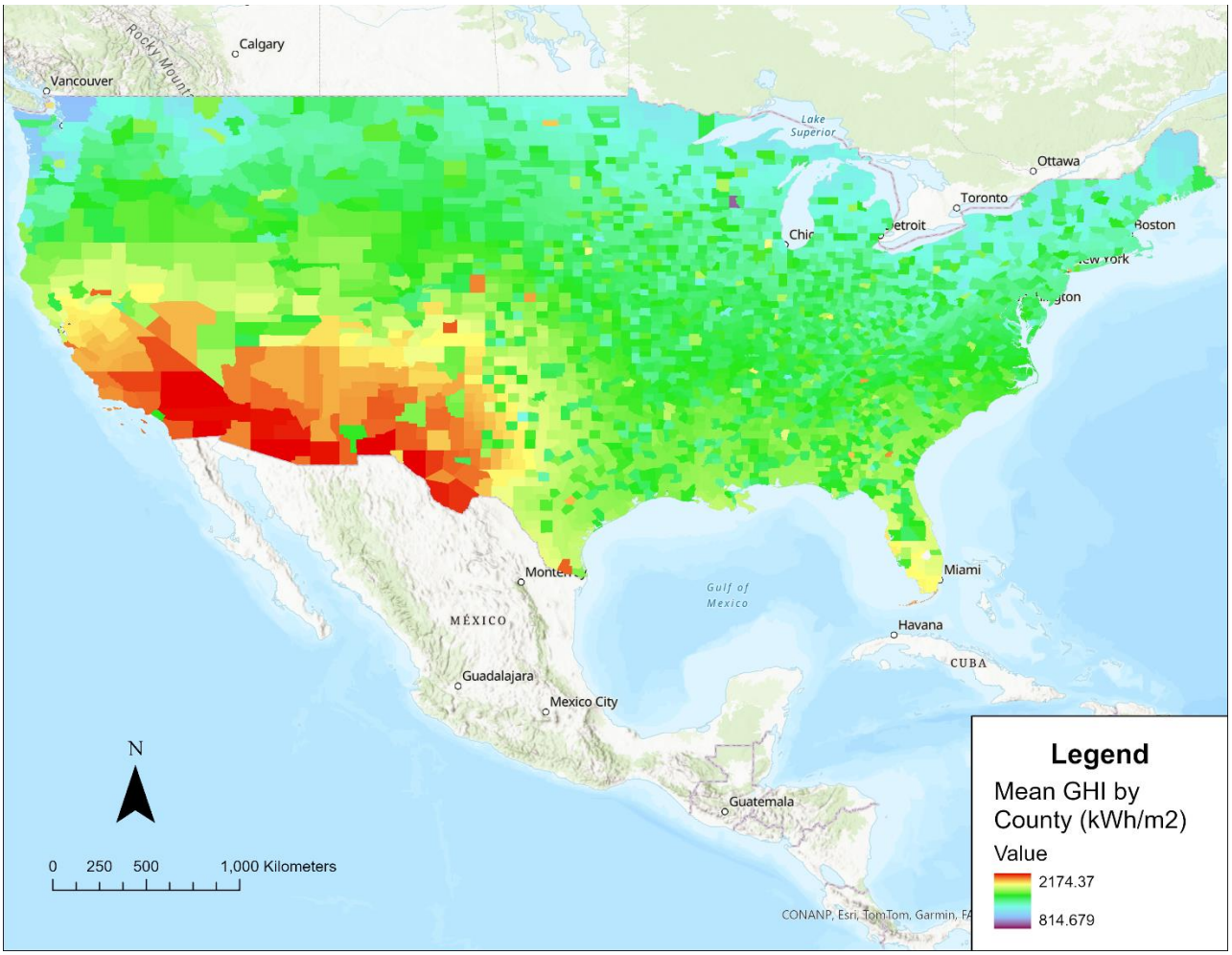

Figure 1: Solar Radiation by County

As shown in (Figure 2), the distribution of education performance is quite uniform around the entire region. It can be however noted that counties that occupy big cities tend to have a higher percentage of college graduates compared to rural or suburban counties, which is a common trend seen in the United States.

This tends to happen because of many factors like higher educational opportunities, better pay incentives, and better infrastructure in cities as compared to rural areas (Wood. Richard, 2023; Zahl-Thanem et. Al, 2024).

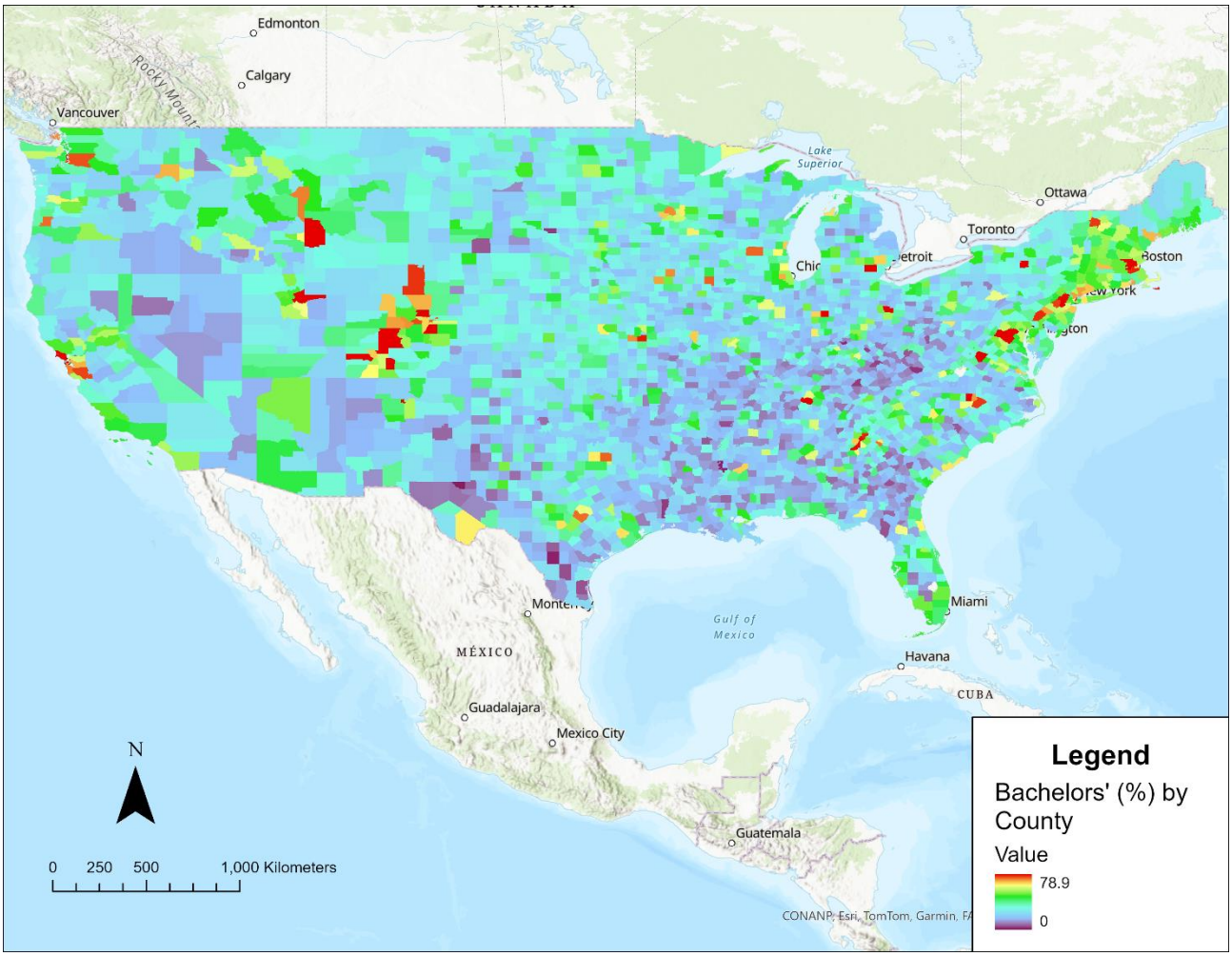

Figure 2: Bachelor's percentage by County

(Figure 3) shows the distribution of median household income or the 50th percentile income of each respective county. Similar to (Figure 2), the distribution of is relatively uniform with areas around big cities tending to have higher income as compared to rural areas. However, this assumption might be incorrect in suburban areas, since a lot of high income families tend to live in suburban areas away from the cities' congestion and high property prices (Urban Design Lab, 2022).

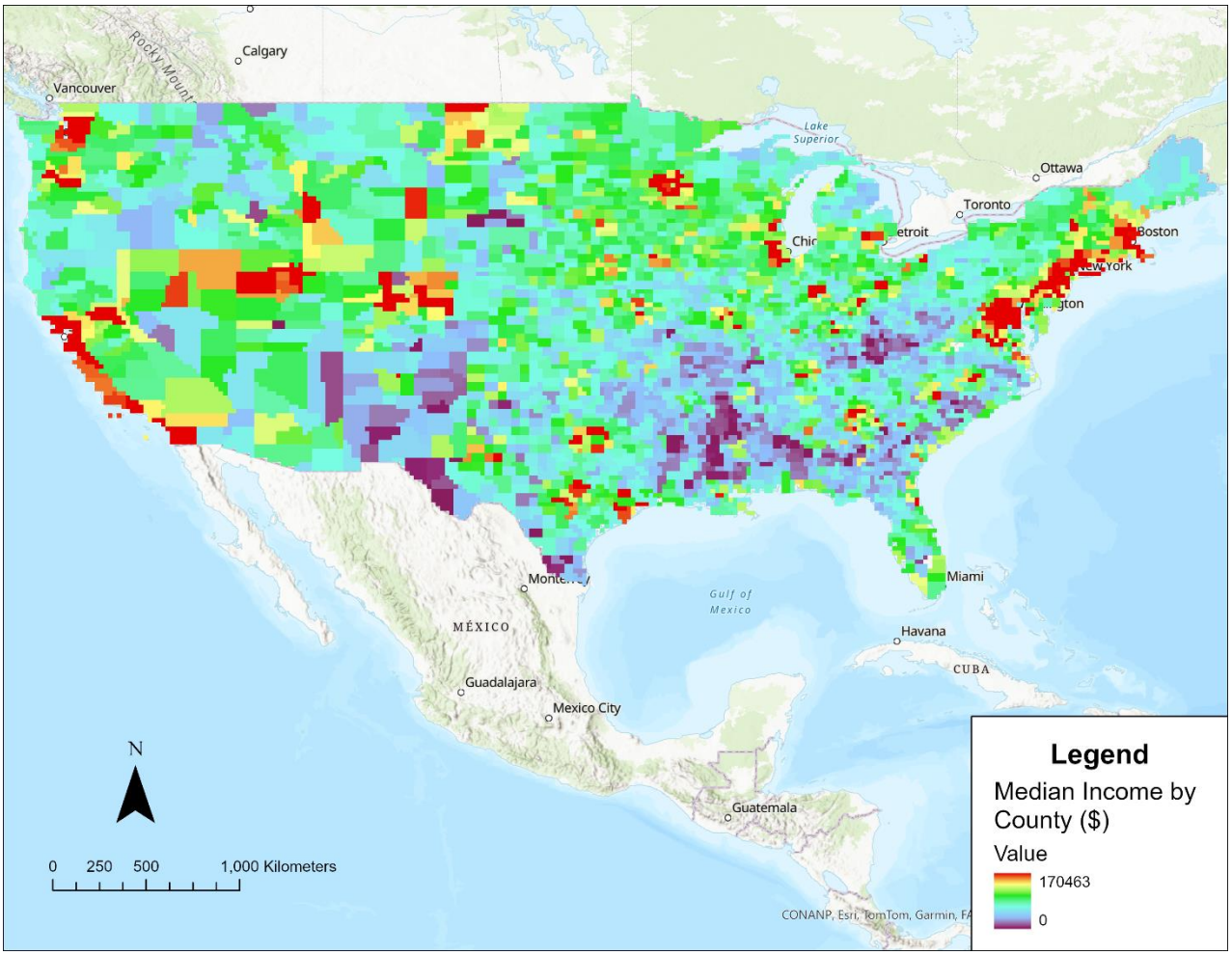

Figure 3: Median Household Income by County

The correlation model as shown in (Figure 4) between sunlight and income gives a coefficient of -0.0008, which is very close to zero indicating no relation between the variables.

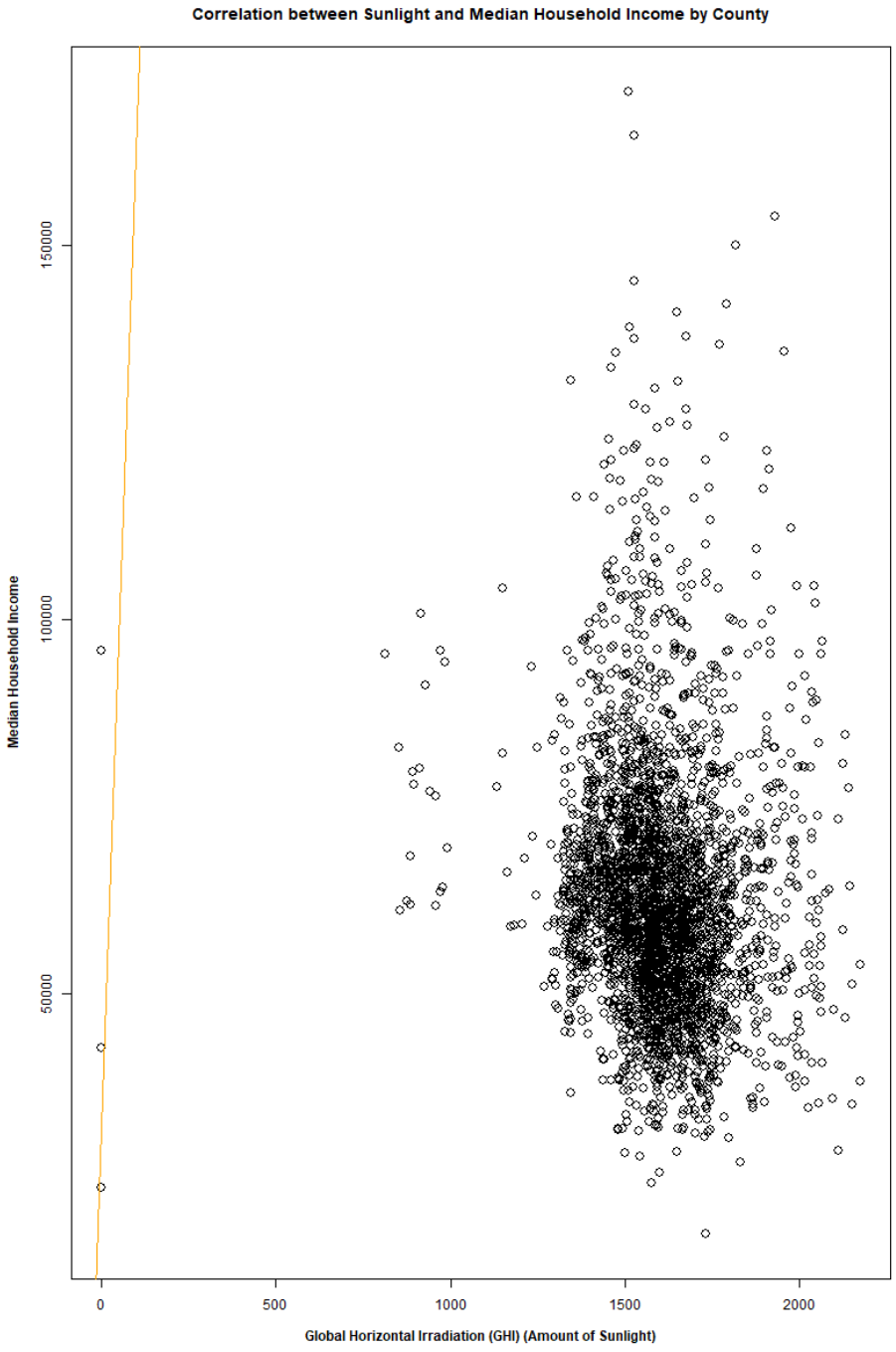

Figure 4: Correlation betwen sunlight and income

The correlation model for education level and income as shown in (Figure 5), results in a coefficient equal to 0.73, which indicates a relatively strong positive relationship between education and income

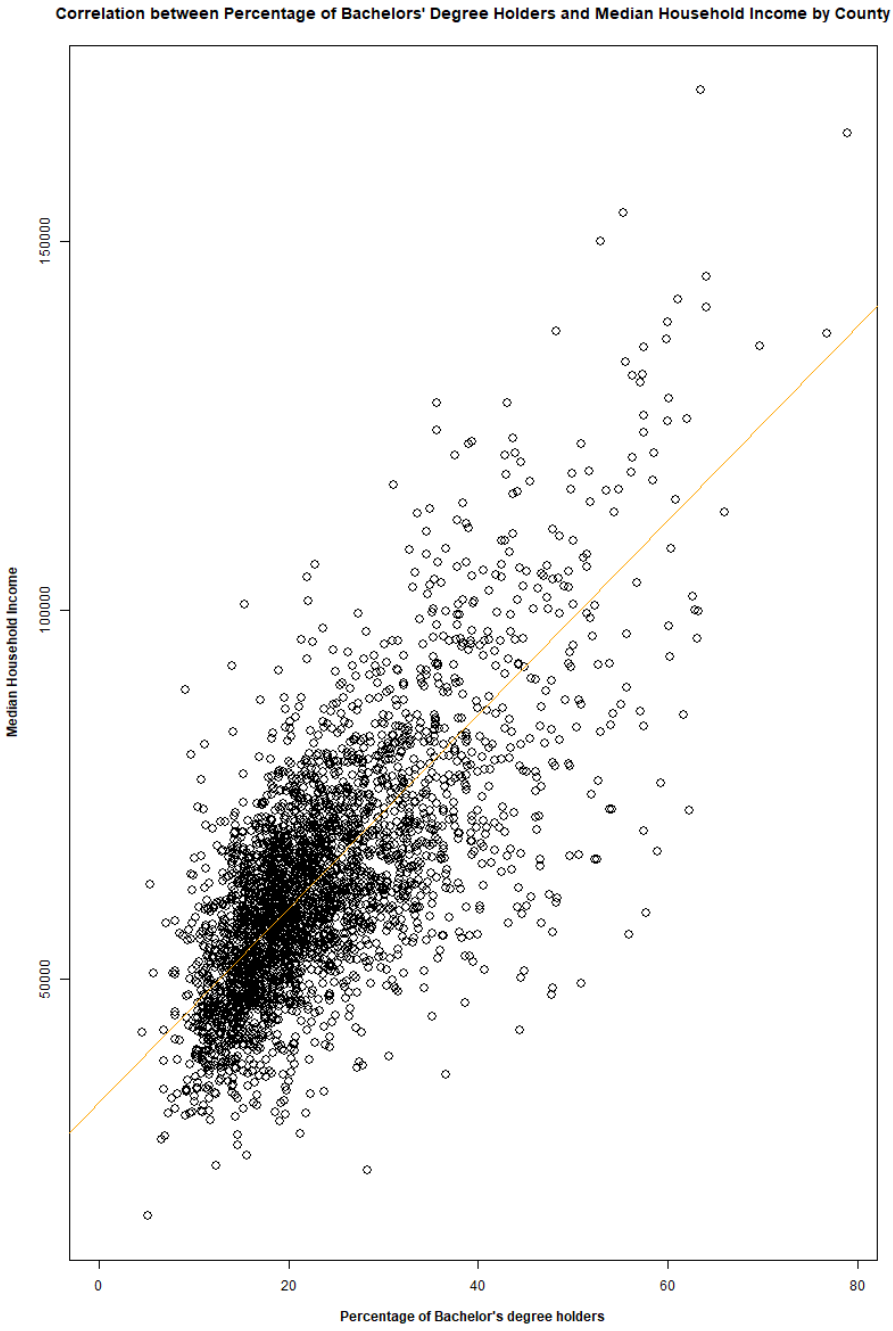

Figure 5: Correlation between education level and income

Analysis

For the purpose of further analysis, Random samples are taken from the education datasets based on county values in the state of Arizona as shown in (Figure 6). The sample sizes of these datasets are 50, 100, 500, and 5000. A hypothesis test is performed to test the accuracy of the points generated with the hypotheses as follows:

H_0

$= The\ Real\ Mean\ and\ Sample\ Means\ are\ equal\ for\ the\ percentage\ of\ bachelor's\ dataset\ (\mu = X_{50} = X_{100} = X_{500} = X_{5000} = 22.2)\ based\ on\ county\ values.$

$H_1 = The\ Real\ Mean\ and\ Sample\ Means\ are\ not\ equal$

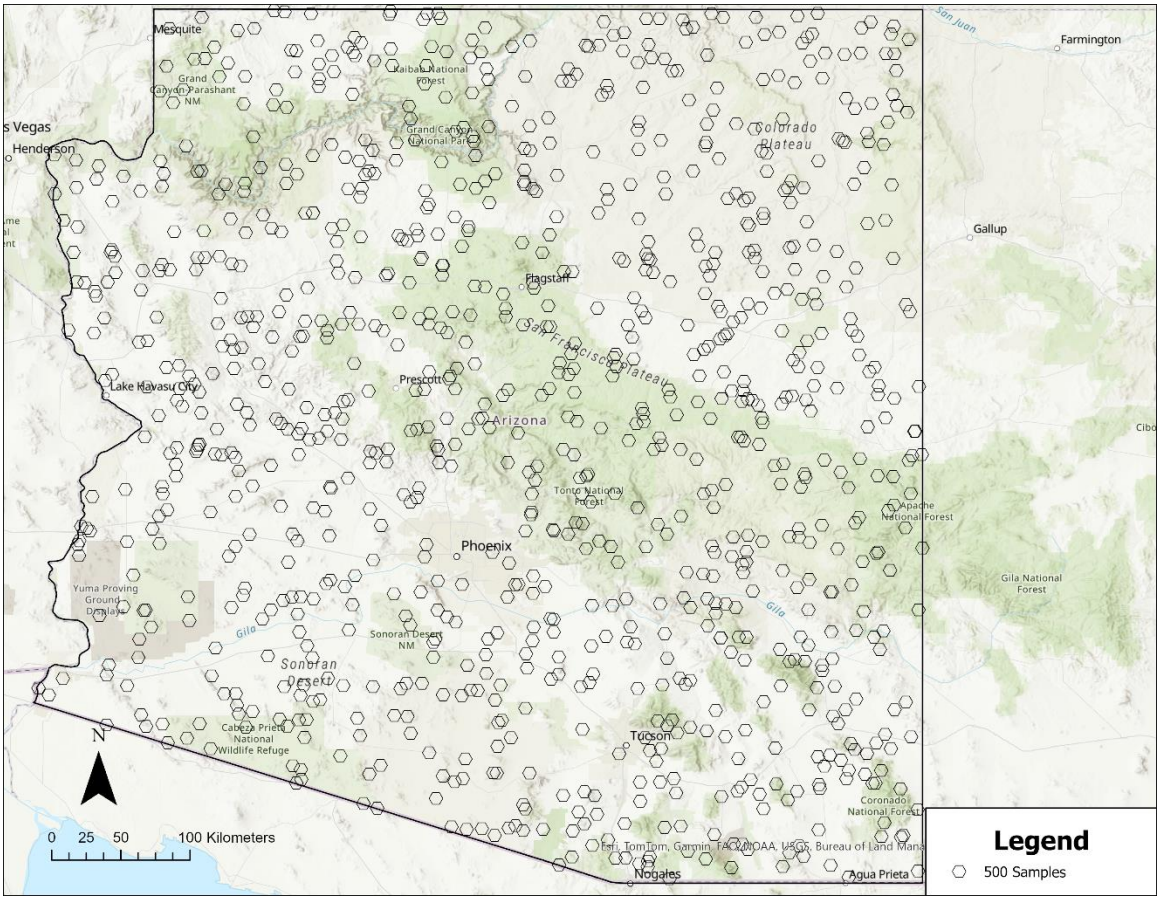

Figure 6: 500 Samples (Arizona)

The results of the test are as follows (for significance level = 0.05):

Sample Size (n)	Sample Mean	t-Statistic	p-value	Decision
50	24.558	1.57	0.123	Accept H_o
100	23.057	0.88	0.381	Accept H_o
500	24.426	4.916	1.020428e-06	Reject H_o
5000	25.5377	16.27	~0	Reject H_o

Conclusion

After performing the hypothesis test, it can be inferred that the mean is consistent for observations of sizes 50 and 100 but fails after that. This indicates that the population distribution doesn't follow a normal distribution.

For our main test, comparing the dependence of income on sunlight and education, we find that there is a stronger correlation between education and income, indicating that education as a benchmark has a much stronger influence on a person's financial status as compared to geographical features like warmth or location.

Appendix

Code

https://github.com/officialar33b/ua_363_project

Figures

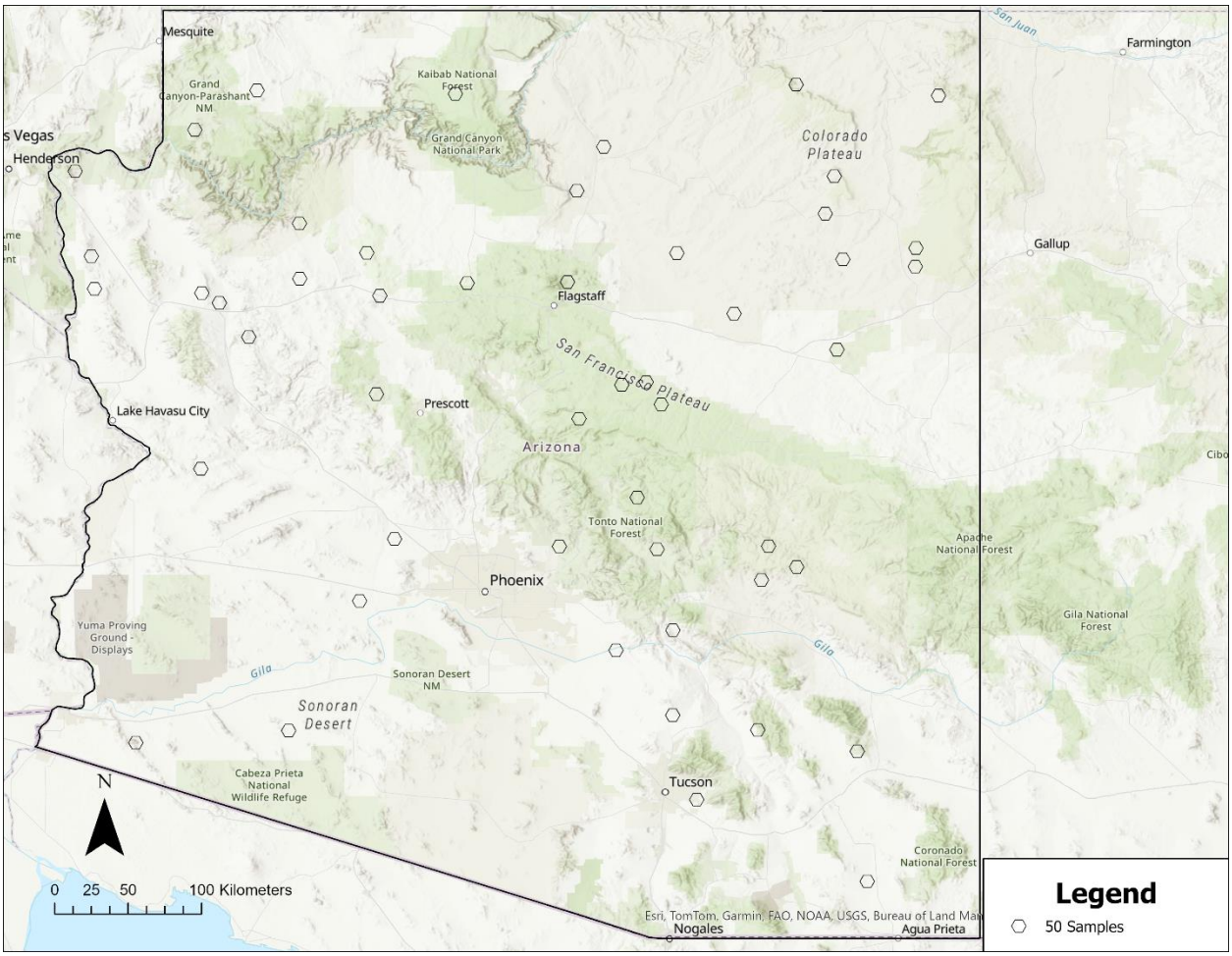

Figure 7: 50 Samples (Arizona)

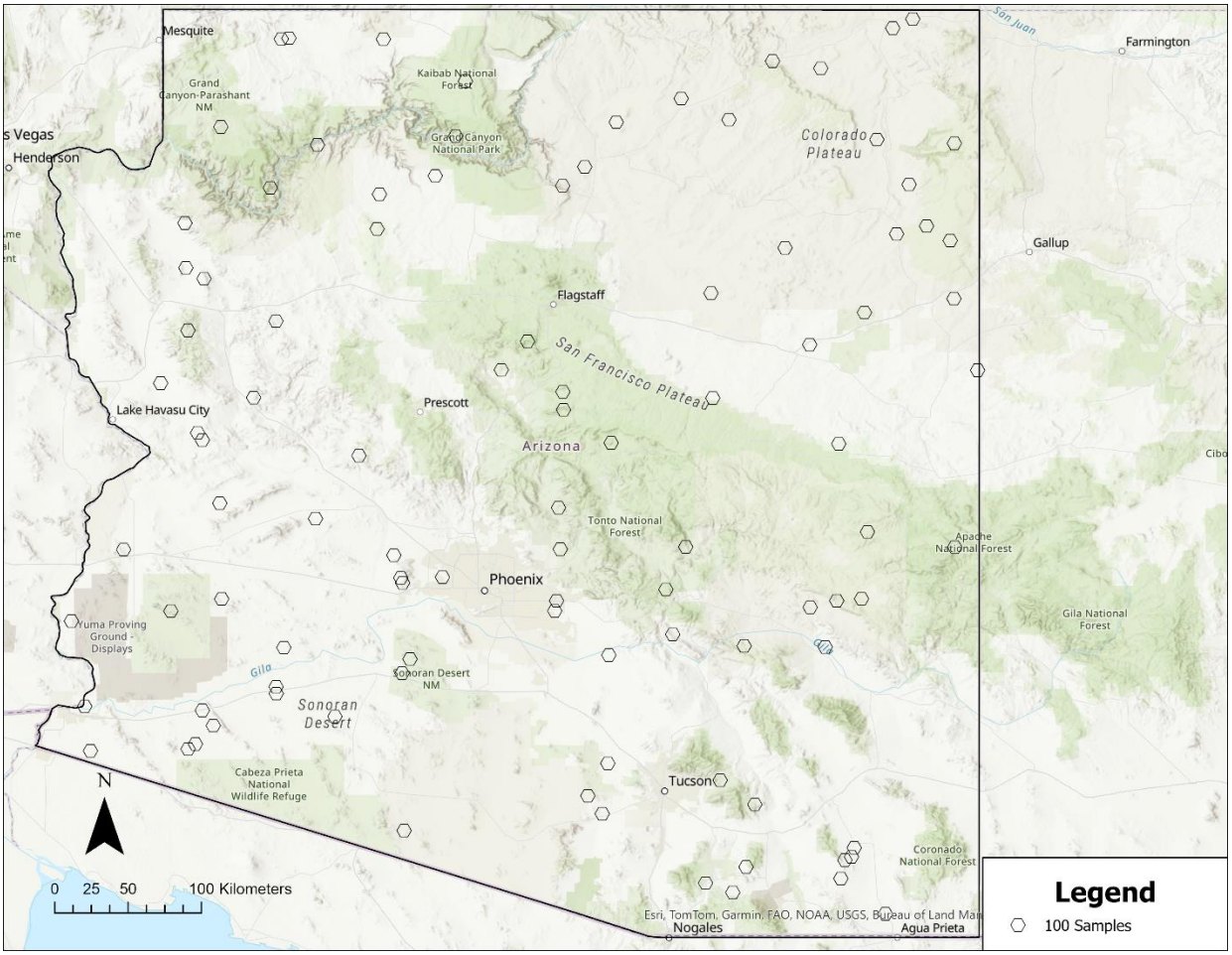

Figure 8: 100 Samples (Arizona)

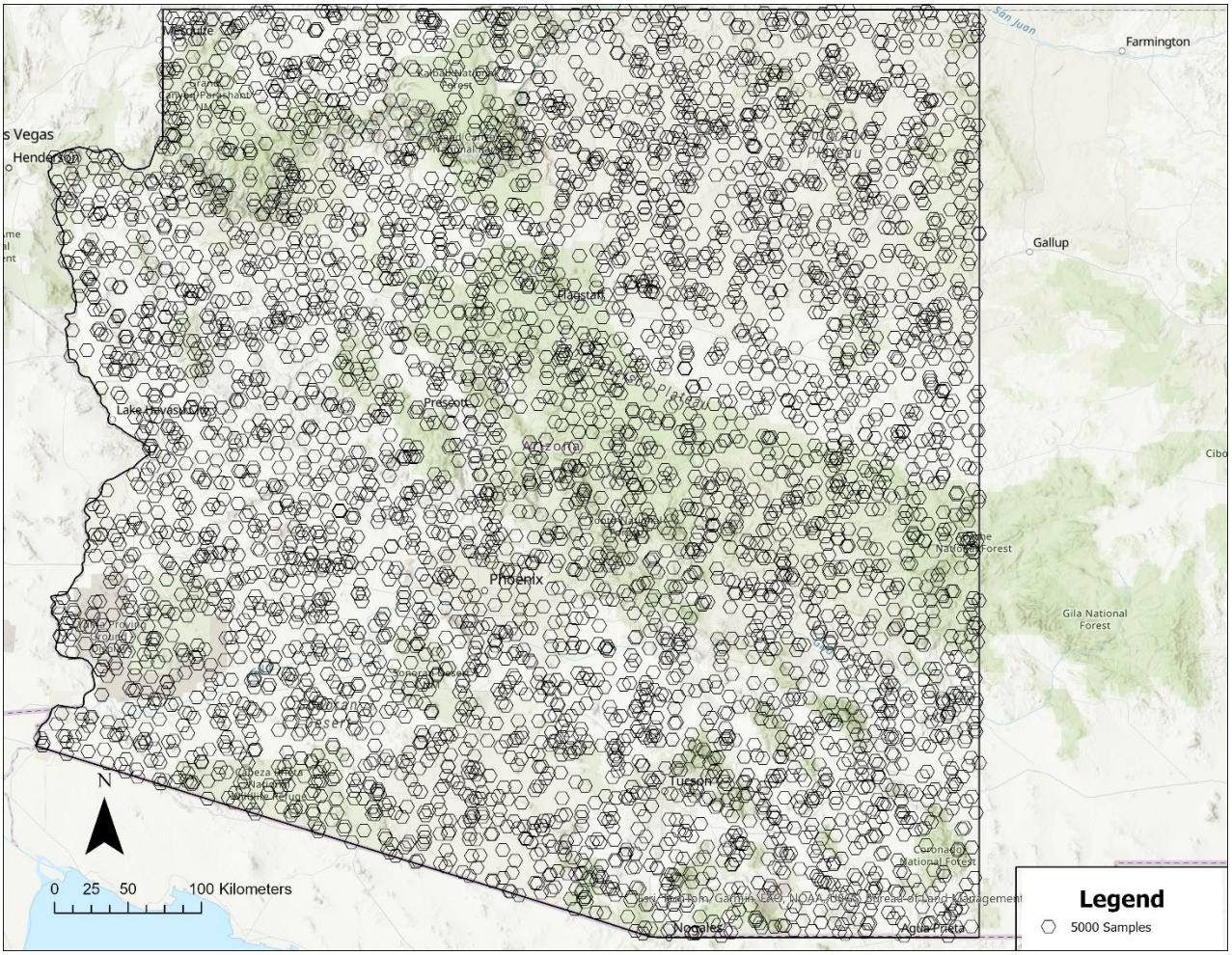

Figure 9: 5000 Samples (Arizona)

References

Alexander Zahl-Thanem, Johan Fredrik Rye, Spatial inequality in higher education: a growing urban–rural educational gap?, *European Sociological Review*, Volume 40, Issue 6, December 2024, Pages 1067–1081, https://doi.org/10.1093/esr/jcae015

BBC Bitesize. (n.d.). Climate zones and weather. https://www.bbc.co.uk/bitesize/guides/zrrjkty/revision/2

Global Solar Atlas. (n.d.). Global Solar Atlas map. https://globalsolaratlas.info/map

HD*Pulse*: An Ecosystem of Minority Health and Health Disparities Resources. National Institute on Minority Health and Health Disparities. Created 12/5/2024. Available from https://hdpulse.nimhd.nih.gov

Pew Research Center. (2009, March 18). Most like it hot. https://www.pewresearch.org/social-trends/2009/03/18/most_like_it_hot/

Richard Mark Wood. (2024, March 31). A review on Education differences in Urban and Rural Areas. https://www.interesjournals.org/articles/a-review-on-education-differences-in-urban-and-rural-areas.pdf

Stewart, K. (2024, October 23). *Pearson's correlation coefficient* | *Definition, Formula, & Facts*. Encyclopedia Britannica. https://www.britannica.com/topic/Pearsons-correlation-coefficient

Urban Design Lab. (2022, March 19). Re-distribution of Population through Suburban migration. https://urbandesignlab.in/re-distribution-of-population-through-suburban-migration/