

Final Project

Group 7: Saujanya Acharya and Rona Hu

2023-12-13

Geospatial Network Analysis of U.S. Universities' Competitiveness

Abstract

This project aims to conduct a comprehensive geospatial and competitive analysis of universities listed in the U.S. News school rankings. By leveraging Google Maps data, we intend to create a matrix that encapsulates the driving distances between each pair of universities, treating the institutions as nodes within a network. The values within the matrix will represent the shortest driving distances, transforming the network's edges from binary indicators to quantifiable distances. This project will provide valuable insights for prospective students, educators, and policymakers into the competitive landscape of higher education institutions in the United States. It will also highlight the potential regional dynamics that could influence university admission policies and strategies.

Introduction

In recent years, the landscape of higher education in the United States has undergone a significant transformation, marked by a heightened level of competitiveness among universities. This transformation has been driven by several factors, including globalization, advancements in technology, and evolving student preferences. As a result, universities are no longer confined to competing solely within their regions; they are now vying for students on a national and even global scale. This paper aims to delve into the intricate dynamics of this competitiveness by examining three key aspects: the geographic proximity of universities, their rankings, and their admission standards.

The American higher education system has long been regarded as one of the most prestigious and sought-after in the world. With a diverse range of institutions, from Ivy League universities to state colleges and private liberal arts schools, students have numerous options when pursuing higher education. However, as the demand for quality education has grown both domestically and internationally, universities have found themselves engaged in fierce competition to attract the brightest minds.

In this evolving landscape, geographic proximity has emerged as a crucial factor influencing

student choices and university strategies. Traditionally, universities competed primarily with institutions within their immediate geographic vicinity. However, advancements in transportation and communication technologies have diminished the barriers of distance, allowing universities to compete on a broader scale. This paper will explore how the geographic distribution of universities across the United States impacts their competitive positions.

Additionally, university rankings have gained increasing importance in recent years. Organizations such as U.S. News & World Report provide annual rankings of colleges and universities, which are closely scrutinized by students, parents, educators, and policymakers. These rankings often play a pivotal role in shaping institutional reputations and attracting prospective students. Hence, understanding how rankings correlate with other factors, such as geographic proximity and admission standards, is essential to comprehending the competitive landscape of higher education.

Furthermore, admission standards, particularly the emphasis on standardized test scores like the SAT (Scholastic Assessment Test), are critical in shaping a university's student body. Different institutions place varying degrees of importance on these standardized tests during the admission process. This paper will investigate whether universities' rankings correlate with their admission standards and if regions with a higher density of universities exhibit different admission criteria compared to less competitive regions.

To undertake this comprehensive analysis, a multi-faceted approach will be employed, encompassing data collection and analysis across these three critical dimensions: geographic proximity, rankings, and admission standards.

Methods

Step 1: Acquiring Qs University Ranking from 2024 data

This initial step involves web scraping the 2024 QS World University Rankings data and locally storing it as a CSV file. Manually collecting the data reduces the need for repetitive web scraping, ensuring efficiency. We use `RSelenium` to automate web scraping from the QS World University Rankings website. The data includes the names of universities in North America.

Step 2 : Getting geocodes from google maps api

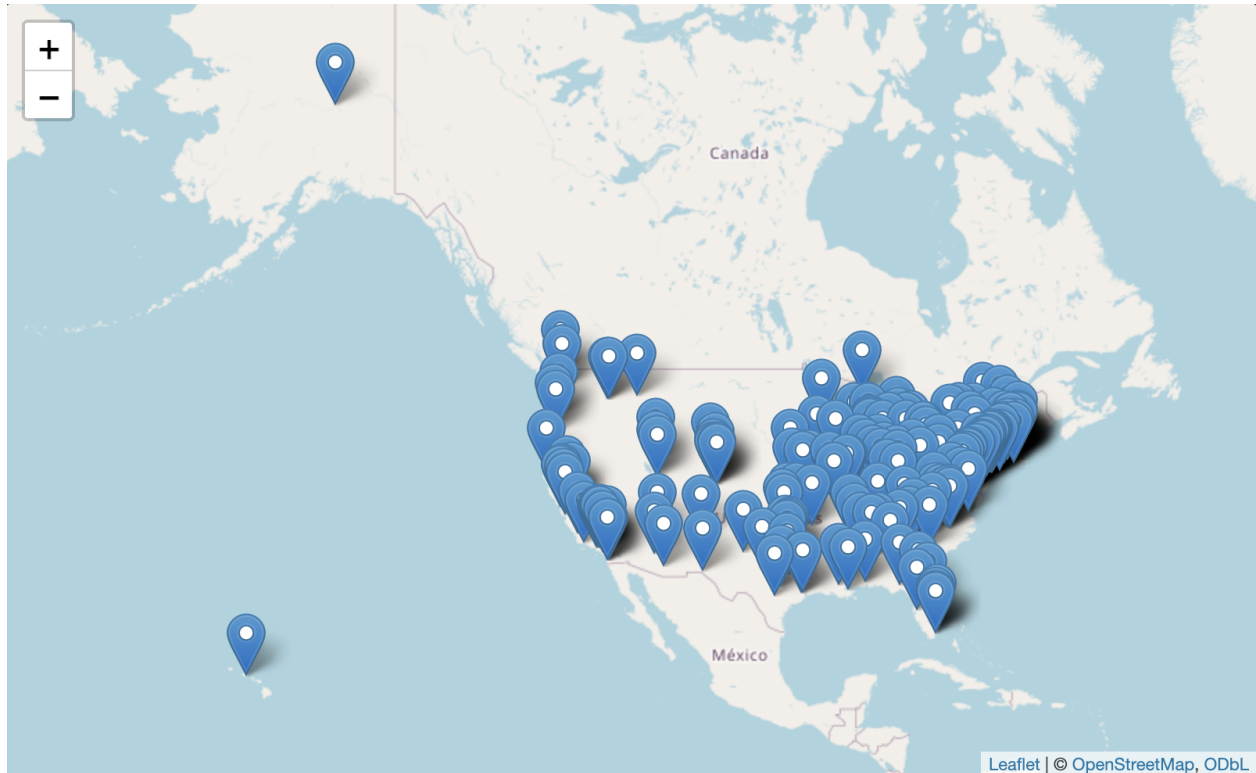
This step involves obtaining geographic coordinates (latitude and longitude) for each university by using the Google Maps API. These coordinates will be crucial for calculating driving distances between universities.

Step 3: Cleaning Google Maps Coordinate data

Here, the retrieved coordinates are processed to ensure consistency and accuracy. The process involves handling missing data and making sure the coordinates are in a suitable format for further analysis.

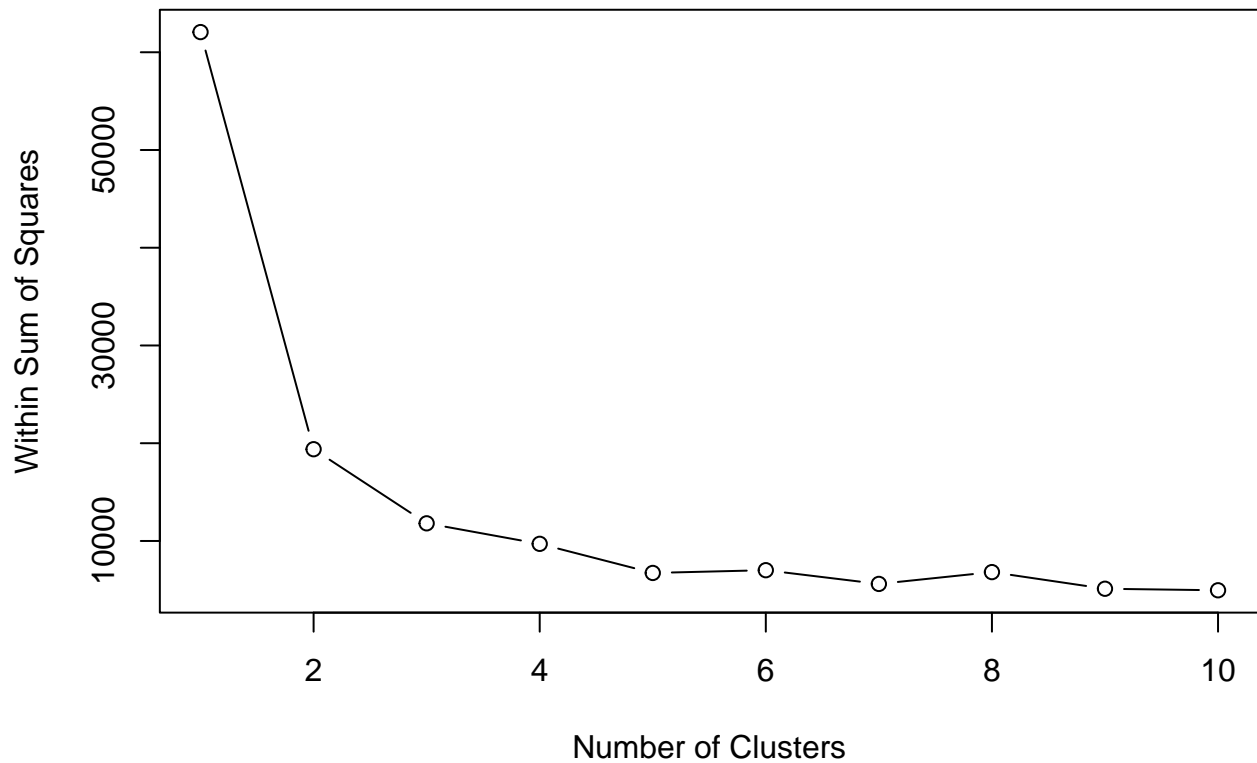
Step 4 : Showing universities in a map

In this step, we create a map visualization using the **Leaflet** library to display the universities' locations based on the latitude and longitude coordinates obtained from Google Maps. This provides a visual representation of where the universities are situated.



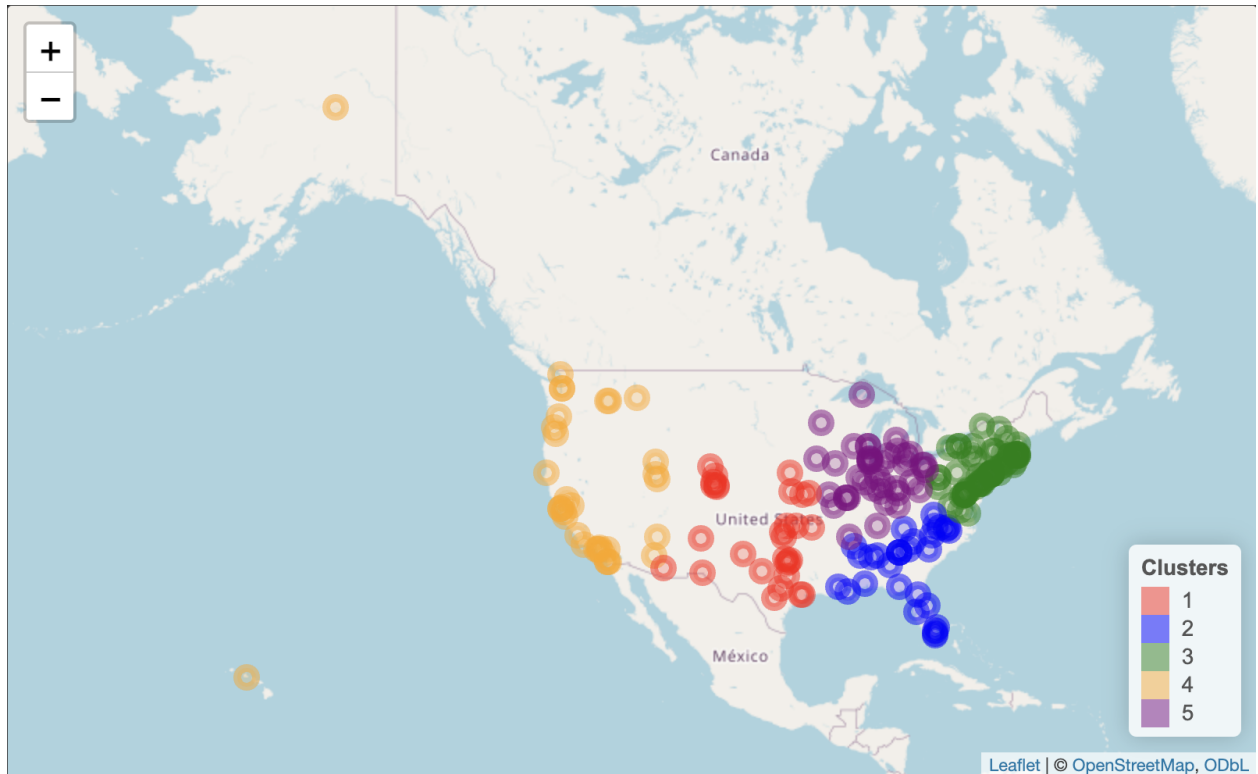
Step 5: Clustering universities based on their latitudes and longitudes

We perform clustering on the universities based on their geographic coordinates using the k-means algorithm. This step groups universities that are close to each other in terms of location into clusters, helping to identify geographical patterns in university distribution.



Step 6: Showing the above cluster in map

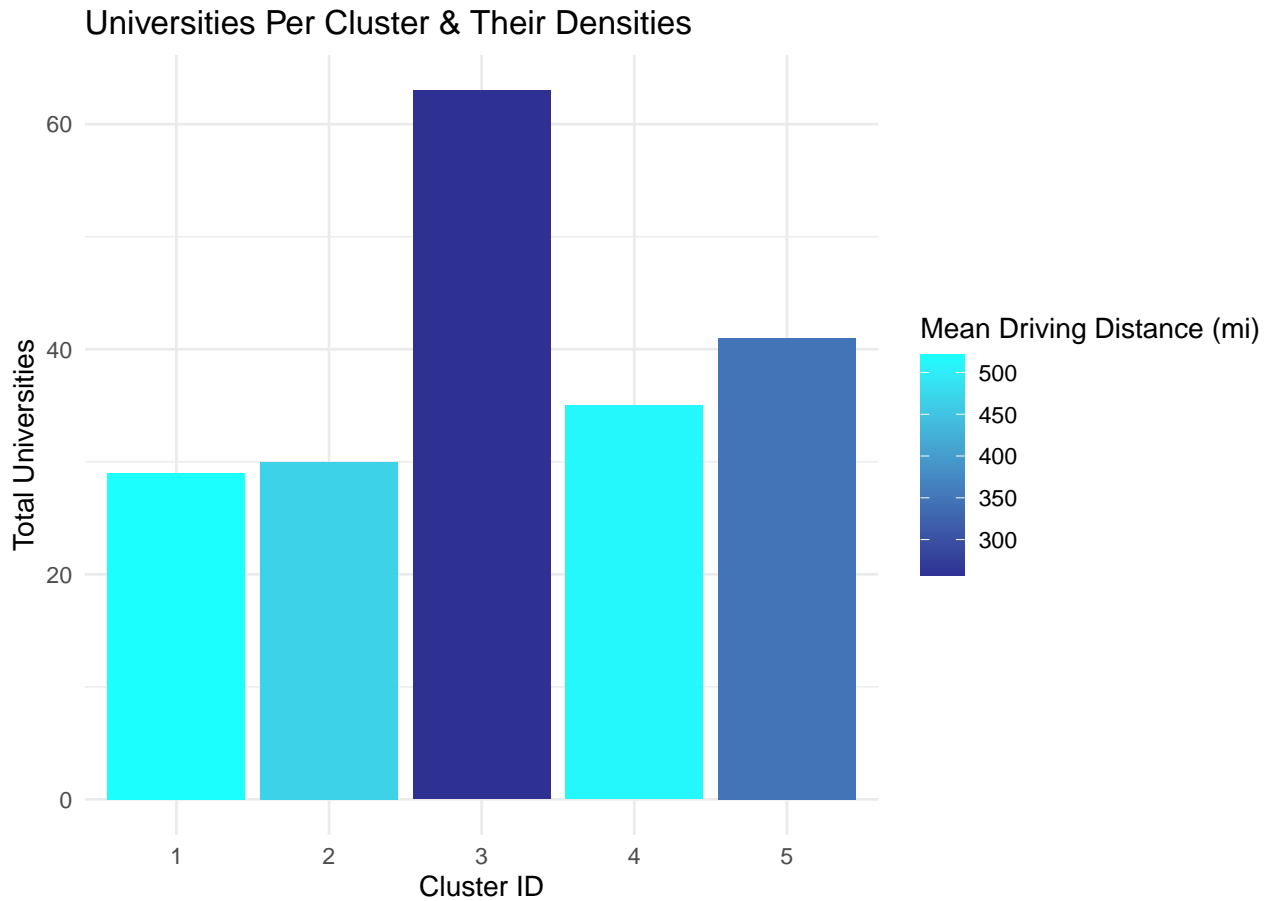
This step visualizes the clusters we created in the previous step on a map using different colors to represent each cluster. It helps we visualize how universities are grouped based on their geographic proximity. We chose the k value of 5 based on 2 factors, a) the analysis of elbow plot, b) the geospatial spread of the US.



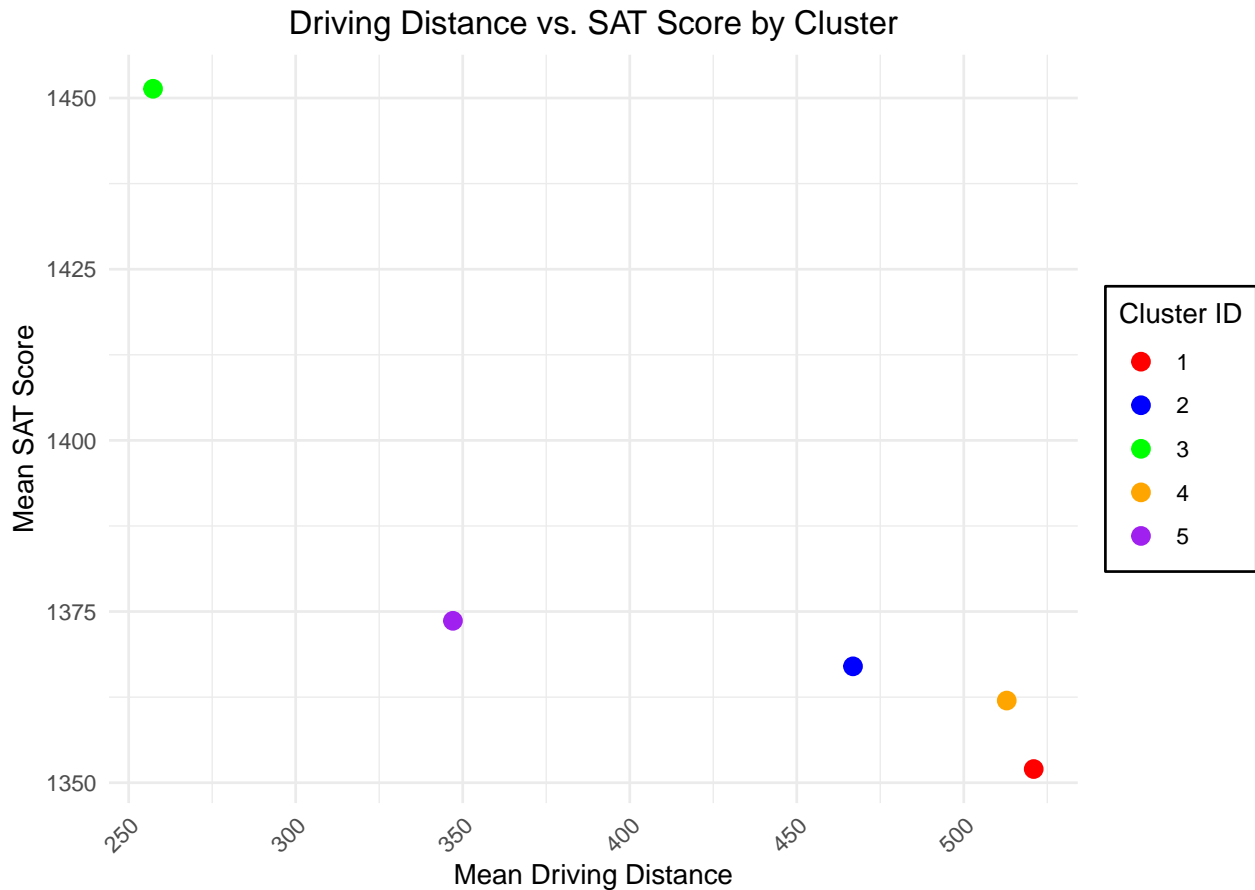
Step 7: Calculating Driving Distances and Analyzing Mean Distances:

In this phase, we calculate the driving distances between universities within each cluster. This is done using the Google Maps Distance Matrix API, considering both time and distance. We analyze the mean distances between universities in each cluster, providing insights into the geographical spread of universities within each group.

Findings



The graph above attempts to compare the densities of the universities using the number of universities in a given cluster and mean driving distances between them. The lower the driving distances, higher the densities. It can be clearly observed that cluster 3 has the highest density with the most number of universities, as well as least mean driving distances between the universities. Cluster 5 then follows cluster 3 in terms of the proximity of universities with cluster 2, 4 and 1 following there after, respectively.



Then finally, after combining the geospatial data of the top universities with the IPEDS data for respective SAT scores, the graph attempts to visualize the mean SAT scores (75th percentile) for each of the university clusters. Comparing to the graph before this, it can be easily identified that the densities of the universities have at least some correlation with the competitiveness (measured through mean SAT scores - 75th percentile as an indicator). It can be observed that cluster 3, with the highest university density has significantly higher mean SAT scores, followed by cluster 5, cluster 2, cluster 4 and 1, respectively. This correlates directly to the order or densities observed in the prior graph.

Conclusion and Discussion

This project attempts to analyze how universities, their proximities and competitiveness might be correlated. And as per the initial assumption the analysis provides some preliminary evidence on how geographical proximities affect university competitiveness and admission standards.

In the past, educational institutions primarily competed with others in their immediate geographic regions. However, in today's rapidly changing landscape of higher education, schools are no longer confined to regional competition; they are now engaged in fierce competition on both national and global scales.

In this project, we measure competitiveness by assessing how many schools are located

within a specific radius, typically a one-hour driving distance. To achieve this, we cluster schools based on their geographic proximity, allowing us to understand regional dynamics. However, it's important to note that even with this clustering approach, we must consider potential endogeneity issues. Focusing on top-ranked schools does not necessarily resolve these complexities because the endogeneity issue extends to various dimensions.

One challenge with region-based clustering is that areas with a high concentration of schools often begin with a higher population density. This inherent correlation between school and population density presents a unique challenge in our analysis.

In future work, assessing neighboring schools for fluctuations in the numbers and proportions of international students can provide additional insights. Numerous hypotheses can be rigorously tested in this framework. It is crucial to retain only those universities with nearby peers (e.g., within a 200-mile radius) for analysis.

Furthermore, exploring the changes over time, particularly by examining two distinct waves (the most recent wave and an arbitrary prior wave), can illuminate variations in SAT scores, rankings, and the composition of students (e.g., out-of-state students and international students). Analyzing variance and mean values, such as whether universities collectively advance in rankings, would be able to draw comprehensive conclusions.

In summary, this analytical approach allows for a nuanced examination of the competitive dynamics among universities, considering geographical proximity, SAT scores, and demographic compositions. By investigating variations and trends, this study holds the potential to unearth valuable insights into the shifting landscape of higher education institutions. Despite recognizing there is room for improvement in our methodology, we firmly believe this project offers valuable insights. It gives prospective students, educators, and policymakers a deeper understanding of the competitive landscape within higher education institutions in the United States. Furthermore, it sheds light on the intricate regional dynamics that have the potential to shape university admission policies and strategies in meaningful ways.

Reference

- Altbach, P. G., & Salmi, J. (2011). *The Road to Academic Excellence: The Making of World-Class Research Universities*. World Bank Publications.
- Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- Dougherty, K. J., & Natow, R. S. (2016). *Understanding the Cost Structures of Online Higher Education: Implications for Efficiency and Accessibility*. Educational Policy Institute.
- Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual review of sociology*, 40, 31-53.
- NACAC (National Association for College Admission Counseling). (2021). *College Admission Testing: A Comprehensive Review of the Literature and Policy Implications*. <https://www.nacacnet.org/globalassets/documents/publications/research/college-admission-testing-literature-review.pdf>
- Imai, K. (2018). *Quantitative social science: an introduction*. Princeton University Press.
- Kothari, C. R. (2004). *Research Methodology: Methods and Techniques*. New Age International.
- Stevens, M. L. (2007). *Creating a Class: College Admissions and the Education of Elites*. Harvard University Press.
- Treiman, D. J. (2014). *Quantitative data analysis: Doing social research to test ideas*. John Wiley & Sons.
- U.S. News & World Report. (2022). *Best Colleges Rankings*. <https://www.usnews.com/best-colleges>.
- VanDerWerf, S. (2020). *Colleges Are No Match for American Poverty*. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/colleges-are-no-match-for-american-poverty>