An Analysis of the Socioeconomic Impact of Technology Access

Dev Dwivedy and Eric Sakk, Ph. D.

Providence Day School, Charlotte, NC 28270

Department of Computer Science, Morgan State University, Baltimore, MD 21251

Introduction

The digital divide is the boundary between communities with access to technology and people without Internet access.[1] Over the past 25 years, matters related to the digital divide have gained attention.[2] Movements such as Microsoft's Airband Initiative are looking to bridge the digital divide, but tackling this issue will not be easy.[3] Many studies have examined the disparity in information access between people affected by the digital divide, and all conclude that closing this gap in technology is essential to provide more opportunities for people without proper access to computers.[4]

Unfortunately, the disparity in information access is not the only issue that arises from the digital divide. Other studies have focused on the disparity in medical access due to the digital divide and remark that the digital divide mainly affects people of lower socioeconomic status, a seemingly obvious but important conclusion.[5]

We hypothesize that this gap in technology access should have a measurable, demonstrable effect on the economy. Our initial leanings were that, over time, increases in access to technology should have a positive net effect on economic growth and opportunity. We have found that other studies examining the economic effects of the digital divide and technology access were limited both in the United States and internationally.

For this work, we focus solely on US data by conducting a state-by-state analysis of technology access and its economic repercussions. Still, we need more reliable historical data on the topic. Literature covering the economic deficit that people without technology access face is sparse. These conclusions were considered too obvious to document thoroughly, but in this work, we demonstrate that these so-called apparent conjectures are more complex than people may believe.

Methods

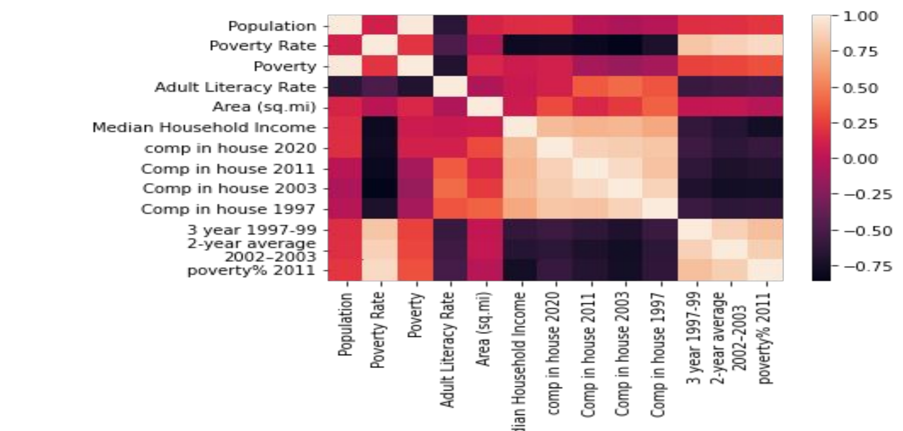
- **SciKit Learn**: A Python module used for creating Linear Regression Models, a tool to predict unknown statistics using related data.
- **Seaborn**: A Python module used for data rendering
- **TensorFlow & Keras**: Python modules used to create Neural Networks for non-linear and multivariable models
- **Matplotlib**: A Python 2D and 3D plotting library.
- **Numpy:** A Python library for numerical and array calculations
- **Pandas**: A Python module used to create the data frames that housed our datasets.
- **Linear Regression:**: Technique used to compare different statistics and verify the significance of their relationships. Linear regression determines the strength of predictors and can predict via extrapolation. The most crucial benefit of linear regression for this experiment is the ability to evaluate the importance of relationships between variables using a correlation coefficient.[6]

$$r = \frac{\sum_{i=1}^{n}(x_1 - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- **Datasets**: For this project, we compiled data for each US state over various years concerning computer access, education, and economic status. Comprehensive data for all the years in our desired range (1990-2020) due to different data sources and their various limitations. As a result, we compiled this list for the years *1994*, *1997*, *2000, 2003, 2010*, *2012, 2014*, *2018*, and *2020*. Below is a table demonstrating the sparse and scattered datasets.[7-13]

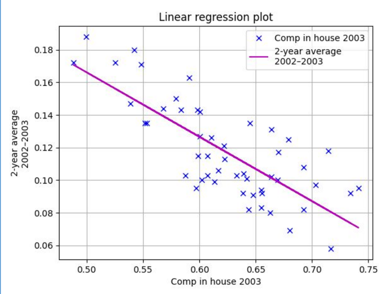
	1994	1997	2000	2001	2003	2005	2008	2010	2012	2018	2020
Computers in house	x	x	x	x	x			x	x	x	x
Adult Literacy Rate							x				x
Poverty Rate		x			x	x	x	x	x		x
Bachelor's Degree Rates											x
HS Graduation Rates					x			x			x
Median Household Income			x			x	x				x

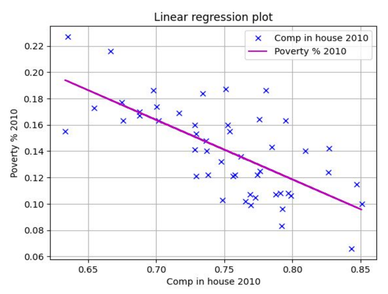
Methods

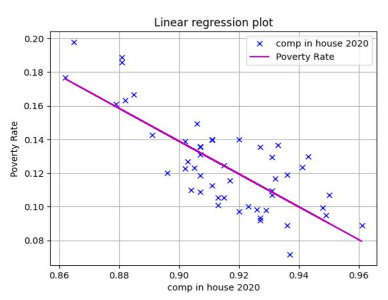


From this plot, we can conclude that the poverty rate is inversely correlated with computers in houses. This anti-correlation was consistent for each year we analyzed between 1997 and 2020. Due to the clear inverse relationship, we concluded that our initial hypothesis was correct and began to investigate further.

Results

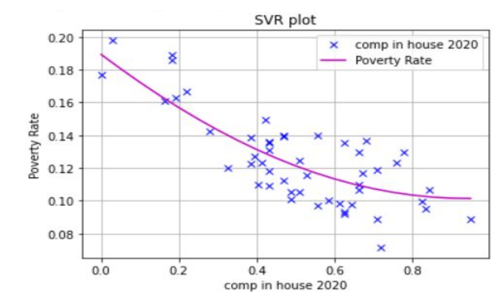


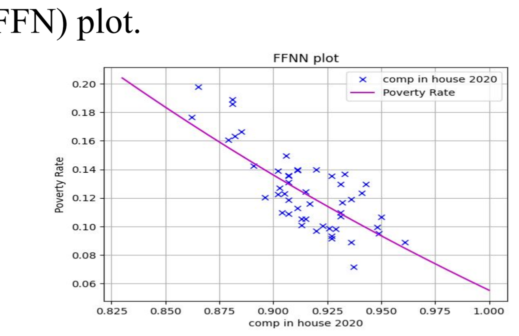




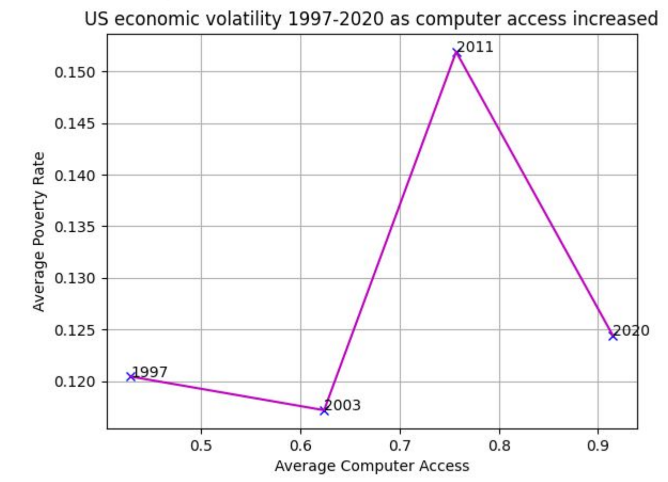
These plots had low correlation coefficients: 0.59, 0.48, and 0.60, respectively. Due to these sub-optimal values, we determined that non-linear models would be necessary to visualize this data. Multivariable linear regression models were also constructed using various combinations of the input data variables; however, those additional combinations only improved the correlation coefficient beyond that of the poverty rate and computer access combination.

After attempting linear regression models, we looked at several non-linear models, including a Support Vector Regression(SVR) plot and a Feed-Forward Neural Network (FFN) plot.





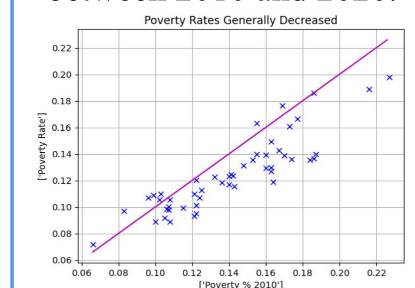
While the above linear and nonlinear models can infer and characterize data relationships, the causation between, for example, computer access increases and poverty rate decreases must still be addressed. The ideal scenario is to develop time series analysis models to predict future poverty rates based on past data observations. Such an approach would prove difficult since the available datasets are spaced at different time intervals. Furthermore, other factors can complicate such an analysis. For example, economic conditions over time are not static. For example, the plot below shows the poverty in the state of Ohio at various points in time between 2003 and 2020.

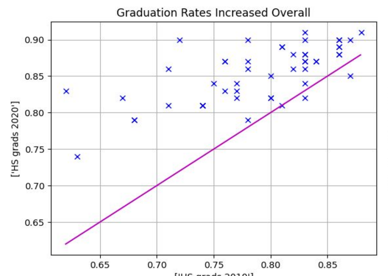


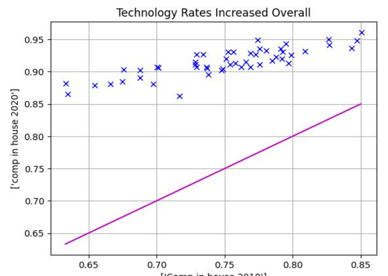
Due to this model, we realized that creating a time series model based on access to computers to predict state poverty rates would be difficult and inaccurate because several factors influence poverty rates. So, while our conjecture that providing access to technology at home influences poverty levels is potentially correct, we need more evidence to demonstrate causality. As a result of this conclusion, we gathered more data, which could help us make an accurate prediction or determination.

Results

These obstacles forced us to shift our focus, so we looked toward education statistics. First, we compared high school graduation data from 2010 with high school graduation data from 2020 (Figure 6). This display shows that high school graduation rates increased in nearly every state between 2010 and 2020.

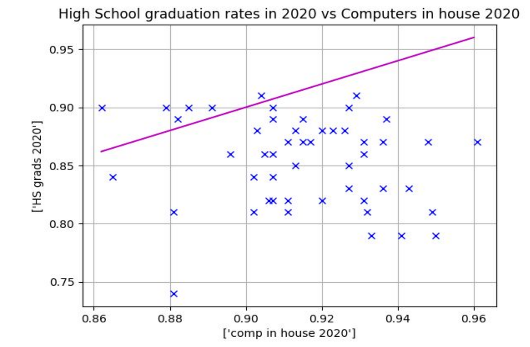


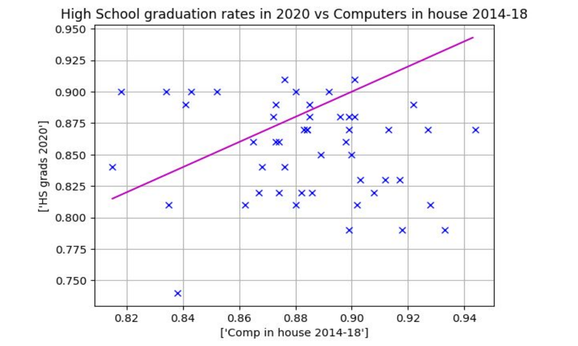




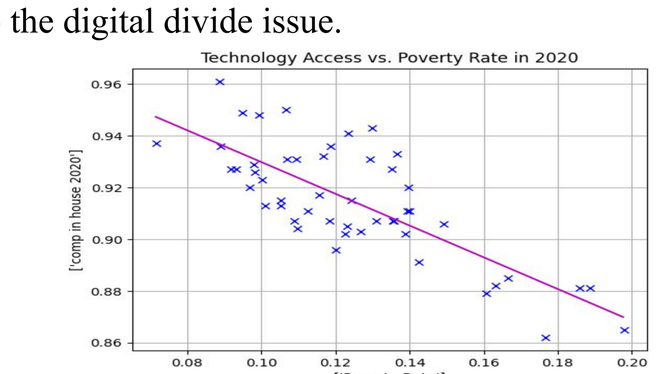
These plots show general trends in poverty rates, graduation rates, and technology access rates in the United States between 2010 and 2020. In almost every state, poverty rates decreased and both graduation rates and technology access rates increased.These conclusions are interesting for this study because our data also showed potential correlations between education and computer access data. However, we still needed to find a way to correlate the two variables directly.

The first important conclusion derived from our dataset was that high school graduation rates in a given year are unaffected by technology or median income. The model shows that technology access does not appear to correlate to high school graduation.Next, models were created to examine high school graduation dates over different years. Even though no correlation was found between graduation and technology access rates in the same year, this model could determine causality if technology access rates in previous years were correlated to high school graduation rates in later years. However, the flat behavior of the graph and low correlation coefficient indicates that high school graduation rates did not directly correlate with technology access, despite the difference in time.

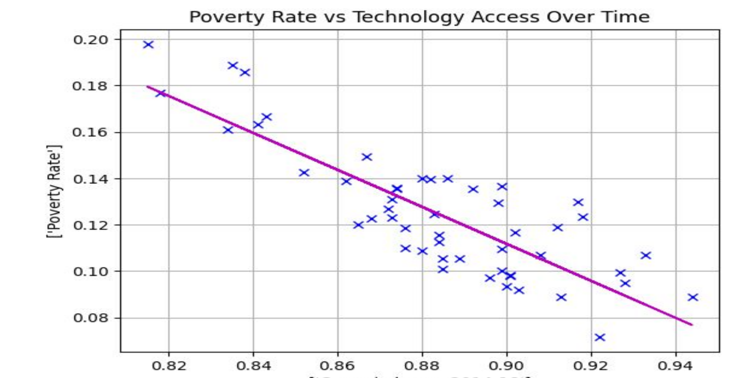




In the model below, we swapped the axes of the variables. While the correlation coefficient does not change, this graph offers a different perspective of the data, presenting access to technology(or lack thereof) due to poverty. This model shows that the poverty rate is heavily anti-correlated with computer access in 2020, demonstrating that income is central to the digital divide issue.



We came to our most significant conclusion when looking at the poverty rate in 2020 versus the technology access rate between 2014 and 2018(Above Right). The poverty rate in 2020 was strongly anti-correlated with access to technology in 2014-2018, with a correlation coefficient of -0.83. We decided to compare these data points because they helped us prove causality and were selected from a time period of relative economic stability in the US between 2014 and 2020.



Conclusions

- We hypothesized that technology access would be inversely correlated with the poverty rate in each state. Still, we found difficulties due to the need for more reliable and complete data over several years and the volatility of the US economy
- Creating a time series model based on access to computers to predict state poverty rates would be inaccurate because several factors influence the volatility of poverty rates; it would be impossible to designate technology access as the sole explanation for changes in poverty rates.
- Models were created using high school and college graduation data and adult literacy rates, but the resulting correlation coefficients were too low to support an accurate conclusion.
- Causality was demonstrated by comparing technology access in 2010 with poverty rates in 2020. This conjecture supports our hypothesis, and this work proves that technology access does have a socioeconomic impact over time. The principled analysis would effectively check the output to verify the hypothesis and predictions.

References

1. Hargittai, E. (2003). *The Digital Divide and What To Do About It*. Webuse.org; Elsevier Science.
2. Wilmon, B. (2020). *The Digital Divide*. Open.library.okstate.edu.
3. Sparks, C. (2013). *What is the "digital divide" and why is it important?* Javnost / the Public, 20(2), 27–46.
4. Ritchie, H., & Roser, M. (2017). *Technology Adoption*. Our World in Data.
5. Chang, B. L., Bakken, S., Brown, S. S., Houston, T. K., Kreps, G. L., Kukafka, R., Safran, C., & Stavri, P. Z. (2004). *Bridging the Digital Divide: Reaching Vulnerable Populations*. Journal of the American Medical Informatics Association, 11(6), 448–457.
6. Yee, G.P., Rusiman, M.S., Ismail, S., Suparman, Hamzah, F.M., & Muhammad, A.S. (2023). *K-means Clustering Analysis and Multiple Linear Regression Model on Household Income in Malaysia*. IAES International Journal of Artificial Intelligence.
7. *Citywide Public Computer Centers*. (2022, September 23). Data.gov.
8. *Common Core of Data (CCD)*. (n.d.). Nces.ed.gov.
9. *County Data*. (2015, December 22). Center on Technology, Data and Society.
10. *Public Technology Resources*. (2013, February 11). Data.gov; data.cityofchicago.org.
11. *State and County Literacy Estimates*. (n.d.). Nces.ed.gov. Retrieved February 28, 2023
12. Duffin, E. (2021, June 21). *Share of U.S. population who had completed a bachelor's degree by state 2019*. Statista.
13. Educational Attainment. (2022). *United States Census Bureau*. United States Census.
14. Turchin, P., & Korotayev, A. (2020) *The 2010 Structural-Demographic Forecast for the 2010– 2020 Decade: A Retrospective Assessment*. PLoS ONE 15(8): e0237458.
15. Blandin, A., & Herrington, C. (2022). *Family Heterogeneity, Human Capital Investment, and College Attainment*. American Economic Journal: Macroeconomics, *14*(4), 438–478.
16. Solomon, E.M., & van Klyton, A. (2020). *The Impact of Digital Technology Usage on Economic Growth in Africa*. Elsevier Ltd, 1-9.

Acknowledgements

I want to thank my mentor, Dr. Eric Sakk, for helping me and supporting me as I researched this topic. His assistance and guidance as I collected and analyzed data and drafted my paper allowed me to learn and develop an interest in the digital divide and its impact in the United States. I would also like to thank my parents for supporting me.