ISSN: 2473-4901
v11 n6354

Developing Florida Digital Divide Index: A Comprehensive Analysis of Internet Accessibility and Socio-economic Characteristics

Tushya Vemuri
tushya09@gmail.com
Southeastern Grocers
Jacksonville, FL, USA

Karthikeyan Umapathy
k.umapathy@unf.edu
School of Computing, University of North Florida
Jacksonville, FL, USA

Abstract

Along with the technological advancement and infiltration of Internet-based devices into our daily lives, Digital Divide research domain has evolved to focus on social development issues. Purdue researchers have developed the Digital Divide Index to measure digital access and use gaps for U.S. counties. Digital Divide Index goes beyond the access gap to focus on identifying communities disconnected from social ties and economic opportunities of the 21st century. However, the Digital Divide Index currently is calculated only at the county level. We have developed the Florida Digital Divide Index at the Zip Code level. We collected relevant datasets from the Census Bureau and the Ookla speed test. We applied Random Forest modeling to the index scores and gathered data variables to identify top importance features. The findings from the machine learning model were used to develop interactive dashboards to explore Florida zip codes with digital divide index scores.

Keywords: Digital Divide, broadband access, socio-economic, random forest, dashboards

 ISSN: 2473-4901
v11 n6354

Developing Florida Digital Divide Index: A Comprehensive Analysis of Internet Accessibility and Socio-economic Characteristics

Tushya Vemuri and Karthikeyan Umapathy

1. INTRODUCTION

President Biden remarked that “for today’s economy to work for everyone, internet access is just as important as electricity, or water, or other basic services” during the Broadband Equity Access and Deployment (BEAD) program announcement at the White House (Mason & Renshaw, 2023, para. 4). Given the vitality of connectivity, digital accessibility casts a long shadow across communities and demographics (Powell, Bryne, & Dailey, 2010). Broadband access stands as the invisible barrier that separates the technologically privileged from the underprivileged, marking a distinction far deeper than just access to hardware. The digital divide encapsulates disparities in the ability to engage with the digital world, access information, and utilize technology for advancement.

The digital divide is more than a gap that can be overcome by providing equipment or access to a service. Rather, the digital divide is a social development issue that needs to be addressed by integrating information and communication technologies into impacted communities (Warschauer, 2004). In today’s society, being connected and digitally literate means having access to education, job opportunities, healthcare information, and social networks. It’s about managing your finances online, applying for jobs, completing your education, or even starting a business. For some, these opportunities are just a click away. But for others, barriers like lack of internet access, affordability, and digital skills stand in the way.

Despite the United States being a global hub for technological advancements and innovation, significant disparities exist in digital access across different regions and communities. Roberto Gallardo from the Purdue University Center for Regional Development has developed the Digital Divide Index (DDI) to rank and identify counties in the United States with the highest digital divide (Gallardo, 2016, 2024). The DDI paints a concerning picture of digital accessibility disparities, emphasizing the uneven distribution of digital resources and connectivity. This discrepancy underscores the importance of conducting focused studies on the digital divide in different parts of the USA. By identifying regions with high DDI scores, researchers and policymakers can better understand the underlying causes of digital disparities and implement targeted interventions. These studies are crucial for ensuring that the benefits of technology and digital access are equitably shared, supporting educational opportunities, economic development, and social inclusion across all American communities.

The DDI doesn’t just highlight these gaps; it also points us toward solutions. Breaking down the digital divide into measurable components shows us where to direct resources and efforts. Whether providing internet access to remote areas, making technology more affordable, or offering digital literacy programs, the DDI guides policymakers, educators, and community leaders in making informed decisions to ensure everyone can benefit from the digital revolution.

However, DDI values are provided at the county level. Florida Philanthropic Network (FPN) is an organization focused on addressing issues with the 2020 Census and planning for the 2030 Census. The Census Bureau has revealed that the census count for Florida has an estimated net coverage error of -3.48%, which means around 749,529 people in Florida were undercounted in the 2020 census, as the recorded Florida population is 21,538,187 (America-Counts-Staff, 2022). FPN engaged the Florida Data Science for Social Good (FL-DSSG) team to develop data-driven strategies aimed at mitigating the undercount observed in the 2020 Census. The collaboration seeks to inform and enhance methods for ensuring that the 2030 Census data collection process avoids similar disparities in representation. As the Census Bureau predominantly utilizes online data collection methods, FPN sought digital divide measurement at the zip code level. Analyzing the digital divide at the zip code level provides a fine-grained understanding of areas at risk of digital access disparities that can directly impact Census response rates. As county-level data may obscure localized challenges, zip code level insights would allow researchers and FPN to identify

ISSN: 2473-4901
v11 n6354

communities at higher risk of undercount due to limited internet access.

Florida's diverse cultural and topographical landscape presents a unique case study for exploring the digital divide issue. With bustling metropolitan hubs and secluded rural locales, the state is a microcosm of the nation's wider digital disparities. We expand on Gallardo's Digital Divide Index (DDI) work and offer a quantifiable look into DDI at Zip Code levels in Florida, shining a light on the areas where the digital age is but a distant echo and those where it resonates clearly. This paper delves into the fine-tuned DDI's findings to paint a comprehensive picture of Florida's digital landscape. By mapping out the contours of connectivity and access, we aim to provide a foundation to build more inclusive digital strategies, ensuring that all Floridians can confidently navigate the digital future.

2. BACKGROUND – DIGITAL DIVIDE INDEX

The digital divide is a sociotechnical phenomenon that has attracted public policy and information systems researchers' attention. Vassilakopoulou and Hustad (2023) conducted a systematic literature review of information systems research on the digital divide on articles published from 2010 to 2020. One of the findings identified by researchers is the lack of studies that focused on innovative approaches to address the grand challenge of multi-faceted dimensions of the digital divide and drawing insights into bridging the digital divide. The Digital Divide Index (DDI), a nationwide measurement of the digital divide, is one such effort. Roberto Gallardo developed the Digital Divide Index (DDI) from the Purdue University Center for Regional Development (Gallardo, 2024). DDI score ranges from 0 (low) to 100 (highest digital divide). The DDI measurement uses the Ookla speed test and Census data. The DDI score comprises infrastructure/adoption (INFA) and the socio-economic (SE) dimension scores. Infrastructure dimension score is calculated based on broadband infrastructure and adoption variables. Socio-economic dimension scores is calculated based on variables known to impact technology adoption. INFA and SE scores are combined to calculate the overall DDI score for each county in the United States. Figure 1 displays a map of DDI score across US.

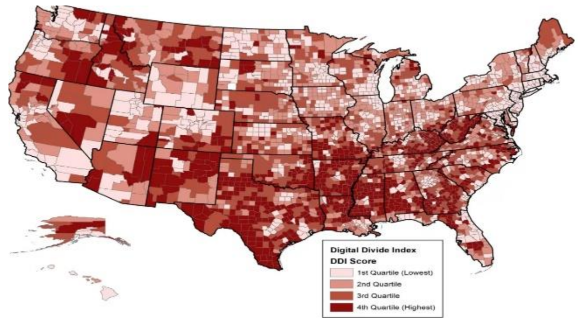

Figure 1. Digital Divide Index across the United States.

2.1. Literature Review

Literature on the digital divide has evolved through increasingly granular levels of analysis, starting with national-level comparisons and progressing into sub-national and behavioral dimensions, reflecting growing awareness of intra-country disparities. Initial research efforts focused on disparities within the United States (NTIA, 1999), while international comparisons emerged with efforts such as Carrocher and Ordanini (2002). The conceptualization of a "global digital divide" gained traction through works by Norris (2001), and Dasgupta, Lall, and Wheeler (2005), which lead to continent-wide analyses such as Fuchs and Horak's (2008) study of Africa.

Sub-national analyses have grown in prominence, reflecting a shift toward localized understandings of digital inequality. Pioneering studies in the U.S. (Atkinson & Coduri, 2002) and China (Jin & Xiong, 2002) paved the way for regional assessments across India, Europe, Australia, and beyond. These studies often rely on first-level metrics—such as access and infrastructure—though some, like Korovkin, Park, and Kaganer (2023), have begun to explore second-level divides in urban contexts. The increasing frequency of sub-national research post-2010 suggests a growing recognition of intra-national disparities and the need for targeted policy interventions.

Despite the proliferation of sub-national studies, there is a notable absence of research that operationalizes the digital divide at the ZIP code level. This gap is particularly consequential given the increasing policy emphasis on hyper-local interventions and the need to identify digital exclusion within neighborhoods and communities, limiting their utility for targeted resource allocation and community-based programming.

In sum, literature underscores a critical

ISSN: 2473-4901
v11 n6354

methodological and empirical gap: the absence of ZIP code-level digital divide measurement. Addressing this gap requires integrated models that combine multidimensional indicators with geospatial precision, enabling researchers and policymakers to more effectively diagnose and address digital inequities at the community level.

Carrocher and Ordanini proposed a model for measuring the digital divide for 10 countries. For calculating digital divide, authors utilized 36 indicator variables and applied principal components analysis for aggregating the variables into synthetic index of digitalization. The digital divide measurement framework was used to highlight opportunities and risks for business managers working in the digital economy environment. The research described in this paper differs in the geographic unit level of analysis and variables used for calculating digital divide index.

3. IMPLEMENTATION METHODOLOGY

After reviewing DDI the methodology outlined in (Gallardo, 2024), the Florida Digital Divide Index (FL-DDI) was calculated to incorporate the latest U.S. Census data. This approach allowed for a contemporary assessment of the digital divide, capturing nuances in internet access, digital literacy, and technology usage across different zip codes in the Florida region.

The methodology from (Gallardo, 2024) provides a robust framework for analyzing the digital landscape, considering variables such as Households with no computer, Households without an internet subscription, average download speed, average upload speed, population percentage of 65 years and above, population percent 25 years and above with Less Than High School (LTHS), Percentage of population with disability, Percentage of population below poverty rate. Using U.S. Census data enhances the DDI's reliability by grounding it in comprehensive and systematically collected information.

All the attributes listed in Table 1 in Appendix A have been sourced from the 2022 American Community Survey (ACS) 5-year census data sources. This comprehensive dataset allows for a detailed and scaled analysis, providing a broad yet nuanced snapshot of demographic, economic, and technological factors across various regions. The extended duration of the data collection ensures that the attributes reflect sustained trends and patterns, making them highly reliable for in-depth analysis in studies such as the Digital Divide Index (DDI). This approach ensures that the attributes encompass a wide range of variables, from socio-economic status to technology access, which is crucial for accurately assessing the scope and impact of the digital divide in different communities.

3.1. Data Collection

We explored the U.S. Census Bureau's website to identify data sources relevant to the Digital Divide Index (DDI). We explored a range of Census data profiles to identify relevant datasets and determine the availability of Florida-specific data. This approach was taken to gather a comprehensive dataset encompassing a range of attributes relevant to understanding the facets of the digital divide across Florida. Appendix A contains a Table that outlines the specific attributes selected for analysis and the corresponding profiles from which these data were extracted.

3.2. FL-DDI Calculation

The calculation of the Florida Digital Divide Index (FL-DDI) involved a detailed process of calculating scores for factors like infrastructure access (INFA) and social equity (SE). This methodological approach aims to comprehensively understand the digital divide in specific areas, incorporating both quantitative and qualitative aspects of digital access and literacy. Here's a step-by-step breakdown of how the FL-DDI is calculated and the additional process of standardizing and scaling data for further analysis.

Step 1: Standardizing Internet Speed Data: Initially, the average download (avg_d_mbps_wt) and upload (avg_u_mbps_wt) internet speeds are extracted from the Ookla speed test website (Ookla, 2024). Ookla Speed Test is a widely recognized tool for assessing the performance of internet connections globally. It measures download and upload speeds to provide users with a clear view of their internet service performance.

Download speed, measured by the Ookla Speed Test, refers to the rate at which data is transferred from the internet to a user's device. Broadband download speed is typically expressed in megabits per second (Mbps). Higher download speeds allow for smoother streaming of high-definition videos, faster loading of webpages, and more efficient downloads of large files.

ISSN: 2473-4901
v11 n6354

Upload speed measures how quickly data is sent from a user's device to the internet. The upload speed is also measured in Mbps. Upload speeds are crucial for sending large amounts of data, such as video conferencing, uploading large files to a server, or live streaming.

A comprehensive approach has been followed to retrieve and analyze broadband speed data across different geographical areas, specifically focusing on Florida's zip codes. The goal is to extract average upload and download speeds for each zip code, using weighted averages where the weights are the number of tests conducted. A Python script (see Appendix B) has been implemented to extract data that involves the following steps.

1. The script generates a URL to download speed test data for fixed broadband services in the second quarter of 2020. This dataset is read into a GeoDataFrame named tiles.
2. Downloads and reads a shapefile of U.S. state boundaries into a GeoDataFrame named States.
3. Downloads and reads a shapefile of U.S. zip code boundaries into a GeoDataFrame named ZipCodes.
4. The script filters the States GeoDataFrame for Florida (state FIPS code '12'), ensuring the coordinate reference system (CRS) matches that of the zip code data. An inner spatial join (sjoin) is then performed to extract zip codes that fall within Florida, resulting in a GeoDataFrame florida_zipcodes. Performs an inner spatial join between the broadband speed test tiles and Florida zip codes, resulting in tiles_in_florida_zipcodes containing broadband data specifically for areas within Florida zip codes.
5. The code computes weighted average download and upload speeds (avg_d_mbps_wt and avg_u_mbps_wt) for each zip code. Weights are based on the number of tests conducted, reflecting a more accurate measure of broadband speeds experienced by users. This is done by grouping the data by zip code and using the np.average function with tests as weights.
6. The weighted average speeds are then merged with the total number of tests conducted in each zip code, resulting in the zipcode_stats DataFrame. This DataFrame is saved to a CSV file, providing a ready-to-use dataset for analysis of broadband speeds by zip code in Florida.

After extracting the average download and upload speed, the values are standardized using Z-scores. This standardization process converts the raw speed data into a format that reflects how many standard deviations each value is from the mean, facilitating comparison across different scales and distributions. This ensures that the values are in sync with other features, all in percentile.

Step 2: Handling Missing Data and Data Status Tagging: In the process of analyzing broadband speeds across Florida's zip codes, we encounter an issue common to many datasets: not all zip codes have complete data for the attributes being studied. To address this and ensure the integrity and usability of our analysis, we implement a two-way approach to manage missing values and tag data completeness.

Replacing Missing Values with Column-Wise Medians: To maintain the statistical validity of our dataset without discarding incomplete records, we opt to replace missing values with the median value of the respective attribute across all zip codes. This method is chosen because the median is less sensitive to outliers than the mean, making it a robust measure for imputing missing data. For each attribute with missing values, we calculate its median based on available data and fill in the gaps accordingly. This ensures that every zip code has a value for each attribute, allowing for comprehensive statewide analysis.

Tagging Data Status: To maintain transparency in data analysis, we introduce a "Data Status" column to our dataset. This column categorizes each zip code based on the completeness of its data.

Complete Data: This tag is assigned to zip codes where all attributes have original, non-imputed values. It indicates that the data for these zip codes is complete and has not been subjected to imputation.

Partial Data Available: This tag is applied to zip codes that require imputation for one or more attributes. It signals to users of the dataset that while the zip code is included in the analysis, some of the values have been filled in using median imputation due to the absence of original values.

This approach enhances the dataset's usability by filling in missing information and maintains data

ISSN: 2473-4901
v11 n6354

transparency by clearly indicating which records have been altered. Users can easily identify which zip codes have complete data and which have been partially imputed, enabling informed decision-making and analysis. This meticulous attention to data quality and integrity is crucial for accurately assessing broadband access and performance across Florida, providing a solid foundation for further research and policy development.

Step 3: Calculating the Infrastructure Access Score: The Infrastructure Access (INFA) score is computed by incorporating the percentages of households without a computer and without an internet subscription, each weighted at 35%, as shown in Figure 2. The standardized (Z-scored) average download and upload speeds are subtracted from this total, each weighted but negatively at 15%. This calculation reflects the positive impact of having computer access and internet subscriptions while accounting for the quality of internet access as indicated by speeds. The weightings used for calculation is same as the original DDI used by Gallardo (2024).

```
data['INFA'] = data['No computer'] * 0.35 +
data['Without an Internet subscription'] *
0.35 - data['avg_d_mbps_wt_1'] * 0.15 -
data['avg_u_mbps_wt_1'] * 0.15
```

Figure 2. Code snippet for INFRA score calculation.

Step 4: Computing Socio-economic Score: The Socio-economic (SE) score aggregates factors that reflect demographic vulnerabilities or disparities affecting digital access. It includes the percentage of the population over 65 years, the *percentage with less than a high school education, the percentage with a disability, and the percentage below the poverty* line. These components are summed directly without explicit weights as shown in Figure 3, underlining their collective impact on digital equity.

```
data['SE'] = data['65 years and over'] +
data['percent population 25 and over with less
than high school(LTHS)'] + data['Total Civilian
Noninstitutionalized Population!!With a
disability'] + data['Percent below poverty
level!!Population for whom poverty status is
determined']
```

Figure 3. Code snippet for SE score calculation.

Step 5: Deriving FL-DDI Score: The FL-DDI is then calculated as the sum of the INFA and SE scores, as shown in Figure 4. This final metric captures a holistic view of the digital divide, integrating considerations of both the physical infrastructure and the broader socio-economic conditions that influence digital access and utilization.

```
data['DDI'] = data['INFA'] + data['SE']
```

Figure 4. Code snippet for DDI score calculation.

Step 6: Scaling for Analysis: For comparative analysis and visualization, an additional step of rescaling the Z-scores of each column (excluding non-numeric or identifier columns like “ZipCode”) to a 0-100 range is performed. This is achieved by subtracting the minimum Z-score in each column from every Z-score and dividing the result by the range of Z-scores. The scaled values are then multiplied by 100. This transformation maintains the distribution of the original data while standardizing the scale for ease of interpretation and analysis.

Through this detailed computation and scaling process, FL-DDI scores can guide researchers in identifying areas most affected by the digital divide. It can enable the prioritization of interventions and resources for those most in need and bridge the gap in digital access across the population.

4. MODELING AND DASHBOARD

FL-DDI scores plotted against zip codes in Florida: The map presented in Figure 5 visually represents the FL-DDI across Florida. Varying colors illustrate the extent of the digital divide with minimum represented with green and max represented with red. The map uses a color-coded system to indicate the severity of the digital divide across different areas:

- Red Areas: These regions exhibit higher DDI scores, suggesting a significant digital divide. Residents in these areas may face challenges due to limited internet connectivity, fewer households with computers or smart devices, and potentially lower digital literacy rates.
- Green Areas: In contrast, green areas indicate lower DDI scores. These regions will likely have better access to digital resources, including higher rates of internet subscriptions, greater availability of computers and smart devices, and possibly a more digitally literate population.

 ISSN: 2473-4901
v11 n6354

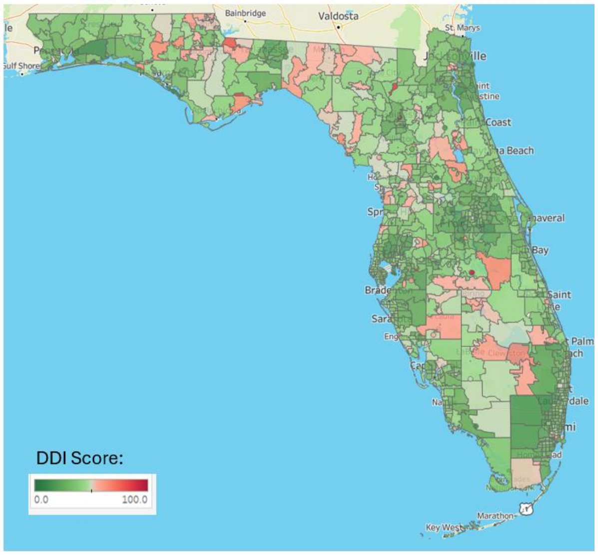

Figure 5. Digital Divide Index Scores Spread Across Florida Zip Codes.

The most severely high DDI scores are exhibited by zip codes in rural counties in Northwest and North Central regions of Florida. While, high scores are occurring for rural zip codes, each urban areas like Miami, Palm Beach, Tampa, and Jacksonville have one zip code with DDI score in range of 55 except for Orlando which have zip codes that can be considered as high DDI score. It is not surprising that zip codes with lowest DDI scores are in the urban areas with high populations. However, there were few zip codes in rural counties with low DDI scores which were further away from population hubs. These insights suggest that key variables like population and household density does not guarantee access to internet or lack off emphasizing the complexity of digital divide and the need for targeted policies to address the disparities.

4.1. Machine Learning Model Results

We utilized the Random Forest model (Breiman, 2001) to understand the feature importance of different attributes. The Random Forest model excelled in its predictions, boasting an impressive R-squared value of 0.9286, which means its predictions are closely aligned with the actual data. The feature importance scores are displayed in Figure 6.

- The "Total Civilian Noninstitutionalized Population with a disability" feature had the most significant impact, with an importance score of approximately 0.33.
- "Households without an Internet subscription" was the second most influential feature, with a score of around 0.15.
- The percentage of the "Population 65 years and over" closely followed, also with an importance score near 0.15.
- The "Percentage of Population below poverty level" had a notable contribution, with a score of around 0.13.

```
Model Performance: 38.138765651146315
R-squared: 0.9286161659623637
                                                        importance
Total Civilian Noninstitutionalized Population!...      0.329450
Without an Internet subscription                        0.154825
65 years and over                                       0.152115
Percent below poverty level!!Population for who...      0.129262
percent population 25 and over with less than h...      0.101639
No computer                                             0.063635
Under 18 years                                          0.020947
Smartphone with no other type of computing device       0.009893
Civilian veterans                                       0.009631
Households with Cellular data plan with no othe...      0.007591
Less than $20,000:!!Without an Internet subscri...      0.005105
avg_d_mbps_wt_1                                         0.004284
avg_d_mbps_wt                                           0.003875
Estimate!!Percent limited English-speaking hous...      0.003716
avg_u_mbps_wt                                           0.002372
avg_u_mbps_wt_1                                         0.001659
```

Figure 6. Random Forest Model feature importance scores.

With the above interpretation from machine learning model, we plotted the features with top importance against to FL-DDI score to understand the data distribution. Scatter plots in Figure 7 show the comparison between DDI score and two socio economic variables poverty status and Disability Status across various Zip Codes. Each point on the graph represents a different zip code, differentiated by color.

FL-DDI Score vs. Poverty Status (Top Plot): A trend appears to be that as the percentage of the population living below the poverty level increases, the FL-DDI score tends to increase as well. An increased FL-DDI score typically indicates a higher digital divide, suggesting that areas with more poverty might experience less digital inclusion. But there are outliers or exceptions where even though the poverty level is near 100, the FL-DDI score is at an average level between 50-60. This indicates the necessity of considering several factors to be stressed for areas with a high digital divide.

FL-DDI Score Vs Disability Status (Bottom Plot): There seems to be a positive correlation between the disability z-score and the FL-DDI score. This suggests that zip codes with a higher proportion of individuals with disabilities may also have a higher digital divide, facing greater challenges in accessing digital technologies.

ISSN: 2473-4901
v11 n6354

Figure 7. FL-DDI Score vs. Poverty Status & Disability Score.

4.2. DDI Dashboard

A Tableau dashboard, depicted in Figure 8, was developed to present a multi-faceted view of the Florida Digital Divide Index (FL-DDI). By analyzing this dashboard, stakeholders can identify regions at risk, patterns that might inform policy, and the overall landscape of digital inclusion within the state, all of which are critical for designing effective interventions to bridge the digital divide.

DDI > 70 vs. Other Attributes (Bar Graph): This chart compares zip codes with a DDI score greater than 70 against other socio-economic attributes like the percentage of households without a computer, Household with income less than $20000 and without an internet subscription, Household without internet subscription, Percentage of population 25 and above without a high school diploma, percentage below the poverty line, and percentage under 18 years of age. The darker shaded bars represent the zip code with a high DDI score. This visualization suggests that zip codes focus on addressing digital divide issues, which are indicated by higher DDI scores for attributes correlated with providing conditions for digital equality.

FPN could use this visualization to focus census campaign outreach on high FL-DDI zip codes where households are more likely to lack internet access, live in poverty, have low educational attainment, and include a higher percentage of children—ensuring these digital and socioeconomically disadvantaged communities are accurately counted. This analysis reveals that zip codes with high FL-DDI scores also face compounded socio-economic challenges—such as low internet access, poverty, and limited educational attainment—offering digital divide researchers' critical insights into systemic inequities, while guiding public policy toward targeted investments in broadband access, digital literacy, and educational support to bridge the digital gap.

DDI Score vs. Top Importance Features (Time Series): This graph is represented to track the DDI score concerning top importance features across different zip codes over time. The features include factors identified in the previous section as part of the machine learning model.

ISSN: 2473-4901
v11 n6354

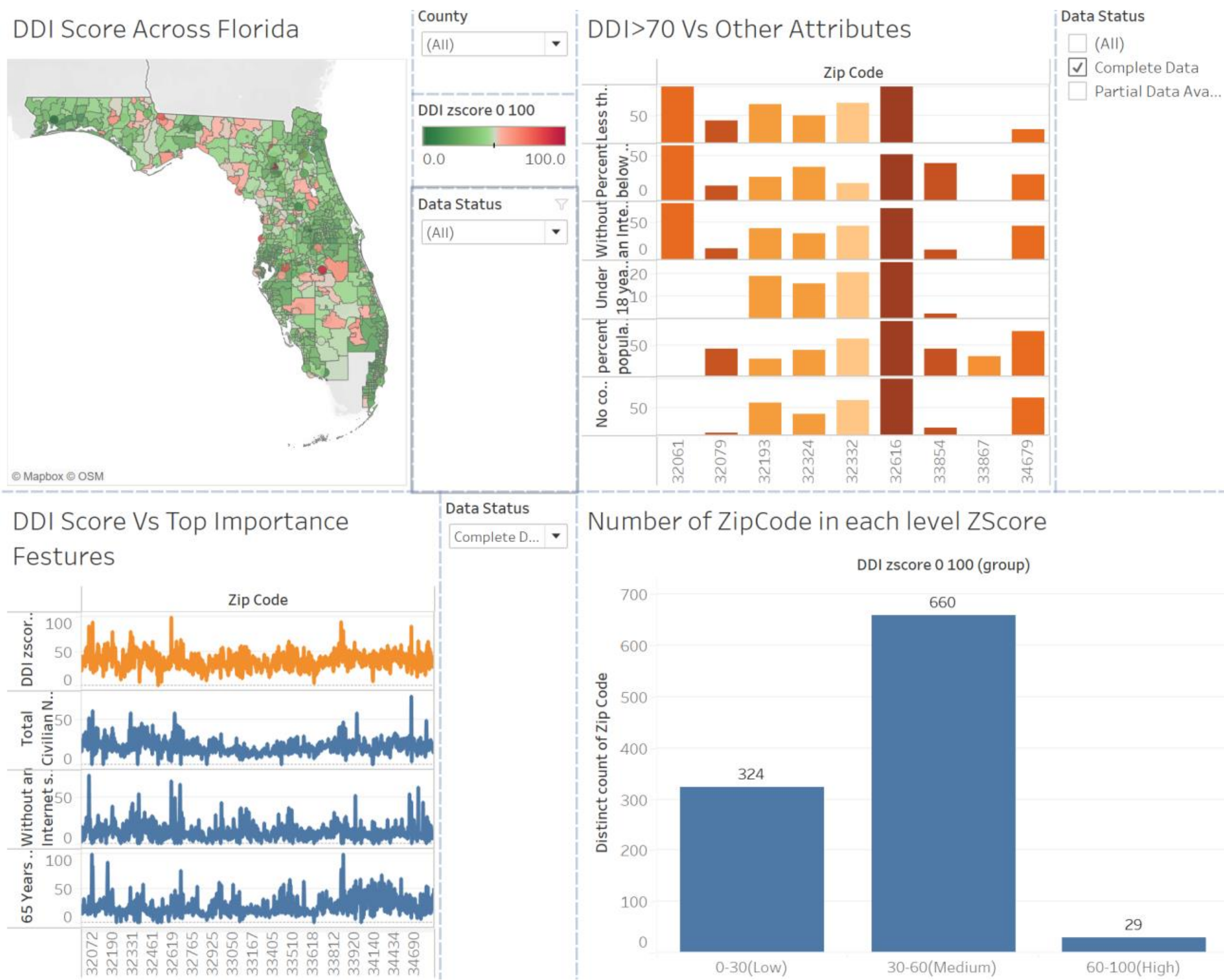

Figure 8. FL-DDI Dashboard.

Specifically attribute "percent of population with disability" shows almost a linear relationship with the DDI score, with both the peaks and lows matching. This graph can also help identify patterns or trends that warrant further investigation.

FPN could use this time series analysis to identify zip codes where rising digital divide risks align with higher percentages of residents with disabilities and seniors, enabling more timely and targeted interventions that address accessibility and connectivity barriers for these vulnerable populations to take part in the Census. Digital divide researchers could use evidence observed from time series visuals to investigate how digital exclusion disproportionately affects vulnerable populations, while policymakers could use this finding to determine zip codes to implement design-inclusive strategies—such as accessible technology programs, senior-focused digital training, and disability-friendly infrastructure—to reduce systemic barriers over time.

Number of Zip Codes in each level Z-Score: This histogram categorizes zip codes into groups based on their DDI z-scores, standardizing DDI values for comparison. It shows the number of zip codes within low, medium, and high DDI score ranges. Although the number of zip codes with a high DDI may be relatively small, it remains crucial to allocate resources effectively to these areas. Ensuring digital inclusivity and providing equitable technological access are essential steps toward integrating all communities into the rapidly evolving tech landscape.

Florida Philanthropic Network can use this histogram to identify and prioritize high-DDI zip codes—despite being fewer in number—as critical areas for targeted investment, ensuring that communities most at risk of digital exclusion receive the necessary support to achieve equitable access to technology and participation in the Census. This distribution of standardized DDI scores highlights a smaller cluster of zip codes facing the highest levels of digital

ISSN: 2473-4901
v11 n6354

exclusion, offering digital divide researchers a clear focus for studying concentrated digital inequities, while guiding public policy toward strategically allocating resources to these high-need areas—ensuring that no community is left behind in the transition to a digitally connected society.

5. CONCLUSIONS

The rapid development of information and communication technologies (ICT) has exacerbated inequalities between developed and underdeveloped communities. There seems to be limited research focused on the digital divide at the localized or zip code unit level of analysis. In this research paper, we extended existing Digital Divide Index scores that are calculated at the county level; we recalculated them at the zip code level. We calculated the Florida Digital Divide Index (FL-DDI) for Florida zip codes. The FL-DDI scores and related data sets were analyzed using the Random Forest model to identify key important features. Modeling results were utilized as key items for designing and developing visuals for interactive dashboards. We developed a Tableau Dashboard that the Florida Philanthropic Network will utilize as a part of a larger research addressing the Census undercount issue in Florida.

While this study provides valuable insights into Florida's digital divide at the zip code level, future data science and analytics research could explore the impact of targeted interventions in high FL-DDI areas and incorporate Florida-specific data and socio-economic factors such as access to affordable housing, healthcare disparities, and the prevalence of households receiving public assistance. The study's limitations include reliance on available socio-economic data, the inability of the Random Forest model to capture all aspects of digital exclusion, and a lack of data context to inform cultural barriers and local infrastructure variations. Additionally, constraints on using federal data to estimate and calculate the DDI at such a fine-grained level may limit the precision and granularity of certain inputs. The approach presented in this paper can be replicated for other states to identify regional disparities, and comparative studies could help develop best practices for addressing digital inequities nationwide.

6. REFERENCES

America-Counts-Staff. (2022, May 19, 2022). 2020 Census Undercounts in Six States, Overcounts in Eight. Retrieved from https://www.census.gov/library/stories/2022/05/2020-census-undercount-overcount-rates-by-state.html

Atkinson, R. D., & Coduri, R. (2002). *The 2002 State New Economy Index. Benchmarking economic transformation in the states*. Retrieved from Washington, DC:

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi:10.1023/A:1010933404324

Corrocher, N., & Ordanini, A. (2002). Measuring the Digital Divide: A Framework for the Analysis of Cross-Country Differences. *Journal of Information Technology, 17*(1), 9-19. doi:10.1080/02683960210132061

Dasgupta, S., Lall, S., & Wheeler, D. (2005). Policy Reform, Economic Growth and the Digital Divide. *Oxford Development Studies, 33*(2), 229-243. doi:10.1080/13600810500137889

Fuchs, C., & Horak, E. (2008). Africa and the digital divide. *Telematics and Informatics, 25*(2), 99-116. doi:https://doi.org/10.1016/j.tele.2006.06.004

Gallardo, R. (2016). *2014 County-level Digital Divide Index*. Retrieved from https://www.researchgate.net/publication/308514127_2014_County-level_Digital_Divide_Index

Gallardo, R. (2024). Digital Divide Index. Retrieved from http://pcrd.purdue.edu/ddi

Jin, J., & Xiong, C. (2002). *The digital divide in National Informatization Quotient: The perspective of mainland China*. Paper presented at the International Conference on Digital Divides: Technology and Politics in the Information Age, Hong Kong.

Korovkin, V., Park, A., & Kaganer, E. (2023). Towards conceptualization and quantification of the digital divide. *Information, Communication & Society, 26*(11), 2268-2303. doi:10.1080/1369118X.2022.2085612

Mason, J., & Renshaw, J. (2023, June 26). US to spend $42 billion to make internet access universal by 2030. Retrieved from https://www.reuters.com/world/us/biden-

ISSN: 2473-4901
v11 n6354

detail-plans-42-billion-investment-us-internet-access-2023-06-26/

Norris, P. (2001). *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge: Cambridge University Press.

NTIA. (1999). Falling Through the Net: Defining the Digital Divide. Retrieved from https://www.ntia.gov/sites/default/files/data/fttn99/index.html

Ookla. (2024). Speedtest by Ookla Global Fixed and Mobile Network Performance Maps. Retrieved from https://registry.opendata.aws/speedtest-global-performance/

Powell, A., Bryne, A., & Dailey, D. (2010). The Essential Internet: Digital Exclusion in Low-Income American Communities. *Policy & Internet, 2*(2), 161-192. doi:https://doi.org/10.2202/1944-2866.1058

Vassilakopoulou, P., & Hustad, E. (2023). Bridging Digital Divides: a Literature Review and Research Agenda for Information Systems Research. *Information Systems Frontiers, 25*(3), 955-969. doi:10.1007/s10796-020-10096-3

Warschauer, M. (2004). *Technology and Social Inclusion: Rethinking the Digital Divide*. Cambridge, MA: MIT Press.

ISSN: 2473-4901
v11 n6354

APPENDIX A – Data Sources

Table 1. Data Sources Used in FL-DDI Calculation

Attribute Name	Description	Profile Name	Source
Estimate!!Percent!!Total households!!TYPES OF COMPUTER!!No computer	The percentage of households without a computer	ACS – S2801	U.S. Census Bureau
Estimate!!Percent!!Total households!!TYPE OF INTERNET SUBSCRIPTIONS!!Without an Internet subscription	Percentage of households without internet subscription	ACS – S2801	U.S. Census Bureau
Estimate!!Percent!!Total population!!SELECTED AGE CATEGORIES!!65 years and over	Percentage of population above 65 years of age	ACS – S0101	U.S. Census Bureau
Estimate!!Percent!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!Less than 9th grade+Estimate!!Percent!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!9th to 12th grade, no diploma	Percentage of population above 25 years of age with Less Than High school	ACS – S1501	U.S. Census Bureau
Percent!!DISABILITY STATUS OF THE CIVILIAN NONINSTITUTIONALIZED POPULATION!!Total Civilian Noninstitutionalized Population!!With a disability	Percentage of Population with a disability	ACS – DP02	U.S. Census Bureau
Estimate!!Percent below poverty level!!Population for whom poverty status is determined	Percentage of Population below poverty rate	ACS – S1701	U.S. Census Bureau
Estimate!!Total!!Total households!!TYPE OF INTERNET SUBSCRIPTIONS!!With an Internet subscription:!!Dial-up with no other type of Internet subscription	Household with only cellular data plan and no other type of internet subscription	ACS – S2801	U.S. Census Bureau
Estimate!!Total!!Total households!!HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2022 INFLATION-ADJUSTED DOLLARS)!!Less than $20,000:!!Without an Internet subscription	Percentage of household with less than $20000 household income and no internet subscription	ACS – S2801	U.S. Census Bureau
Estimate!!Percent!!Total households!!TYPES OF COMPUTER!!Has one or more types of computing devices:!!Smartphone!!Smartphone with no other type of computing device	Percentage of Households with only smartphone and no other computing device	ACS – S2801	U.S. Census Bureau
Estimate!!Percent!!VETERAN STATUS!!Civilian population 18 years and over!!Civilian veterans	Percentage of Civilian veterans	ACS – DP02	U.S. Census Bureau

ISSN: 2473-4901
v11 n6354

Attribute Name	Description	Profile Name	Source
Estimate!!Percent!!Total population!!SELECTED AGE CATEGORIES!!Under 18 years	Percentage of population below 18 years	ACS – S0101	U.S. Census Bureau
Estimate!!Percent limited English-speaking households!!All households	Limited English-speaking households	ACS – S1602	U.S. Census Bureau
avg_d_mbps_wt	Average download speed	Shapefile & tl_2019_us_zcta510.zip	Ookla Speed Test
avg_u_mbps_wt	Average Upload speed	Shapefile & tl_2019_us_zcta510.zip	Ookla Speed test

APPENDIX B – Python Script for Standardizing Internet Speed Data

```
# %%
%matplotlib inline

from datetime import datetime

import geopandas as gp
import matplotlib
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

from shapely.geometry import Point
from adjustText import adjust_text

# %%
def quarter_start(year: int, q: int) -> datetime:
    if not 1 <= q <= 4:
        raise ValueError("Quarter must be within [1, 2, 3, 4]")

    month = [1, 4, 7, 10]
    return datetime(year, month[q - 1], 1)

def get_tile_url(service_type: str, year: int, q: int) -> str:
    dt = quarter_start(year, q)

    base_url = "https://ookla-open-data.s3-us-west-2.amazonaws.com/shapefiles/performance"
    url = f"{base_url}/type%3D{service_type}/year%3D{dt:%Y}/quarter%3D{q}/{dt:%Y-%m-
%d}_performance_{service_type}_tiles.zip"
    return url

# %%
tile_url = get_tile_url("fixed", 2020, 2)
tile_url

# %%
tiles = gp.read_file(tile_url)

# %%
print(tiles.head())

# %%
# zipfile of U.S. county boundaries
```

ISSN: 2473-4901
v11 n6354

```
state_url = "https://www2.census.gov/geo/tiger/TIGER2019/STATE/tl_2019_us_state.zip"
States = gp.read_file(state_url)

# %%
print(States.head())

# %%
# zipfile of U.S. county boundaries
ZipCode_url = "https://www2.census.gov/geo/tiger/TIGER2019/ZCTA5/tl_2019_us_zcta510.zip"
ZipCodes = gp.read_file(ZipCode_url)

# %%
print(ZipCodes.head())

# %%
ky_States = States.loc[States['STATEFP'] == '12'].to_crs(4326)

# %%
ky_States.head()

# %%
ky_States = ky_States.to_crs(ZipCodes.crs)

# %%
florida_zipcodes = gp.sjoin(ZipCodes, ky_States, how="inner", op='intersects')

# %%
florida_zipcodes.head()

# %%
florida_zipcodes['ZCTA5CE10'].nunique()

# %%
florida_zipcodes.info()

# %%
if 'index_left' in tiles.columns:
    tiles = tiles.rename(columns={'index_left': 'index_left_old'})

if 'index_right' in tiles.columns:
    tiles = tiles.rename(columns={'index_right': 'index_right_old'})

# Similarly, check and rename for 'florida_zipcodes' if needed
if 'index_left' in florida_zipcodes.columns:
    florida_zipcodes = florida_zipcodes.rename(columns={'index_left': 'index_left_old'})

if 'index_right' in florida_zipcodes.columns:
    florida_zipcodes = florida_zipcodes.rename(columns={'index_right': 'index_right_old'})

# %%
tiles_in_florida_zipcodes = gp.sjoin(tiles, florida_zipcodes, how="inner", op='intersects')

# %%
# convert to Mbps for easier reading
tiles_in_florida_zipcodes['avg_d_mbps'] = tiles_in_florida_zipcodes['avg_d_kbps'] / 1000
tiles_in_florida_zipcodes['avg_u_mbps'] = tiles_in_florida_zipcodes['avg_u_kbps'] / 1000

# %%
tiles_in_florida_zipcodes.head()
```

 ISSN: 2473-4901
v11 n6354

```
# %%
tiles_in_florida_zipcodes.info()

# %%
zipcode_stats = (
    tiles_in_florida_zipcodes.groupby(["GEOID10", "ZCTA5CE10"])
    .apply(
        lambda x: pd.Series({
            "avg_d_mbps_wt": np.average(x["avg_d_mbps"], weights=x["tests"]),
            "avg_u_mbps_wt": np.average(x["avg_u_mbps"], weights=x["tests"])
        })
    )
    .reset_index()
    .merge(
        # Aggregate total tests for each group
        tiles_in_florida_zipcodes.groupby(["GEOID10", "ZCTA5CE10"])
        .agg(tests=("tests", "sum"))
        .reset_index(),
        on=["GEOID10", "ZCTA5CE10"],
    )
)

# %%
zipcode_stats.head()

# %%
zipcode_stats.to_csv(r'zipcode_stats_output.csv', index=False)

# %%
table_stats = (
    zipcode_stats.loc[zipcode_stats["tests"] >= 50]
    .nlargest(20, "avg_d_mbps_wt")
    .append(
        zipcode_stats.loc[zipcode_stats["tests"] >= 50].nsmallest(20, "avg_d_mbps_wt")
    )
    .sort_values("avg_d_mbps_wt", ascending=False)
    .round(2) # round to 2 decimal places for easier reading
)

# %%
header = ["GEOID10", "ZCTA5CE10", "Avg download speed (Mbps)", "Tests"]

table_stats.rename(columns=dict(zip(table_stats.columns, header)))

# %%
zipcode_data_map = tiles_in_florida_zipcodes[['GEOID10', 'ZCTA5CE10']].merge(zipcode_stats,
on='GEOID10')

# %%
labels = ["0 to 25 Mbps", "25 to 50 Mbps", "50 to 100 Mbps", "100 to 150 Mbps", "150 to 200 Mbps"]

zipcode_data_map['group'] = pd.cut(
    zipcode_data_map.avg_d_mbps_wt,
    (0, 25, 50, 100, 150, 200),
    right=False,
    labels = labels
)
```

ISSN: 2473-4901
v11 n6354

```
# %%
zipcode_data_map.head()

# %%
ky_places = gp.read_file("ftp://ftp2.census.gov/geo/tiger/TIGER2019/PLACE/tl_2019_12_place.zip")

# %%
ky_places = ky_places.loc[ky_places['PCICBSA'] >= "Y"].sample(15, random_state=1).to_crs(26916)
ky_places["centroid"] = ky_places["geometry"].centroid
ky_places.set_geometry("centroid", inplace = True)

# %%
ky_places.head()

# %%
zipcode_data=tiles_in_florida_zipcodes[['ZCTA5CE10','REGION','DIVISION','STATEFP','NAME','avg_d_
mbps','avg_u_mbps']]
print(zipcode_data.head())

# %%
zipcode=tiles_in_florida_zipcodes['ZCTA5CE10'].nunique()
print(zipcode)
```