INTERSTATE RENEWABLE ENERGY COUNCIL

Independent leadership. Trusted clean energy expertise.

Interstate Renewable Energy Council

Grid Disparity Analyses in FirstEnergy Service Territory

March 2025

Authored by Shay Banton

A. Executive Summary

As part of its consumer advocacy efforts, the Ohio Environmental Council (OEC) hired the Interstate Renewable Energy Council (IREC) to research grid infrastructure disparities in communities across Ohio that were victim to historically marginalization through discriminatory housing practices like redlining or that experience significant economic or climate hardship. To investigate, IREC required access to widespread, high-resolution, location-specific grid datasets obtainable only through regulatory discovery. Of the few utility territories with open state regulatory proceedings, only FirstEnergy provided data at a sufficient resolution for this research.

Key findings from this research effort include:

- **Historic redlining is still correlated with racial demographic differences**, but it is a less precise predictor of energy burden or poverty levels compared to modern equity-focused designations, such as those used in the Climate and Economic Justice Screening Tool (CEJST), a tool that identifies disadvantaged census tracts across the United States based on burdens that communities experience.[1]
- **Grid voltage disparities were not clearly correlated with historic redlining but were correlated to CEJST disadvantaged designation.** Communities designated as disadvantaged based on CJEST methodology were twice as likely to have low-voltage circuits, suggesting an ongoing need for equity-informed distribution planning.
- **Grid age disparities were evident in disadvantaged communities** whose grid infrastructure was on average 3.4 to 4.2 years older than that in non-disadvantaged communities, a statistically significant difference that suggests historical patterns of underinvestment in these areas.

[1] Council on Environmental Quality, *Climate and Economic Justice Screening Tool*, White House (2024), *[LINK REMOVED]* – Note: This study was performed prior to the White House's removal of CEJST from all federal websites. An archived version of the data used can be provided upon request.

- **Lower voltage circuits were consistently older and had less capacity than higher voltage circuits**, with differences of 5.4 to 5.7 years in age and significantly reduced normal and overload capacity. Paired with our voltage disparity findings, this furthers the case for prioritizing investments in disadvantaged communities to address inequities in grid infrastructure.
- **Circuits in disadvantaged communities had 10.5% to 23% less normal and overload capacity**, a statistically significant disparity that aligns with prior voltage class and capacity findings. This suggests that historically lower investment levels have led to reduced grid capacity for electrification and clean energy adoption in these areas and reinforces the need for equitable and proactive infrastructure investment.
- **Service quality metrics paradoxically suggested better reliability in disadvantaged and redlined areas**, though this is likely due to the limitations of conventional reliability indices which fail to capture the more granular outage experiences of individual customers.

To address systemic inequities in Ohio's electric grid, regulators and utilities must prioritize transparency, incorporate equity-focused metrics, and proactively invest in historically underserved communities. **Greater access to granular and disaggregated data** is essential for identifying disparities in infrastructure investment, service reliability, and energy access. Without clear and publicly available information, it remains difficult to assess whether marginalized communities are receiving equitable grid improvements or experiencing disproportionate energy insecurity.

Beyond data transparency, Ohio's **grid planning processes must integrate justice-centered metrics** that assess disparities in service quality, infrastructure age, and system capacity. For instance, conventional reliability measures often fail to capture the lived experiences of disadvantaged communities, leading to overlooked inequities in grid performance. By adopting more comprehensive and community-focused assessment methods, utilities can ensure that grid planning decisions account for historic underinvested.

Finally, **ensuring equitable infrastructure investment** is critical for addressing long-standing disparities in grid resilience and energy access. The state must establish mechanisms to prioritize upgrades in communities that have historically faced underinvestment due to systemic barriers such as redlining or economic disinvestment. Embedding justice considerations into decision-making processes will help ensure that the benefits of grid modernization and clean energy expansion are distributed equitably.

By advancing these recommendations, Ohio can move toward a more just energy system where historically marginalized communities are no longer at risk of continued underinvestment, inequitable access to energy decarbonization opportunities, or disproportionate energy-related hardship. These reforms will be essential for ensuring that Ohio's energy transition is both sustainable and equitable.

B. Table of Contents

C. Background

For years, communities historically overlooked within climate change advocacy have voiced concerns about how growing mitigation ambitions have neglected the needs of those most vulnerable to climate impacts. In response to these failings, growing efforts to integrate the concept of equity, or distribution of benefits based on identified need rather than equal treatment, grew into the climate justice movement and the centering of a Just Transition. Several focus areas within climate and energy policy spurred from this larger, overarching justice movement, such as regenerative economics, clean energy workforce development, and fossil fuel community restoration. An area of increased significance of late is ensuring equitable "social and economic participation in the energy system, while also remediating social, economic, and health burdens on those historically harmed by the energy system."[2] This focus is commonly referred to as "energy equity".

Since its inception, advocates interested in applying an energy equity lens have questioned what role the energy system today has played in either creating or exacerbating inequities, and how such systems could be redesigned to instead enable a more just transition. Over the past few years, significant advancements in 'how' to ask such questions have been explored across the energy regulatory landscape. For instance, in Michigan, consumer advocates with Souladarity were some of the first to shed light on how grid infrastructure investment practices resulted in disparities in electric system quality for marginalized communities.[3] In Texas, researchers studying the impacts of Winter Storm Uri found significant disparities in energy resilience that resulted in high minority communities being four times as likely to suffer a blackout than their neighbors.[4] In California, researchers at Berkeley uncovered disparities in the ability of the grid to accommodate new residential solar between black and white communities.[5] And most recently in Minnesota, advocates and researchers collaborated to advocate for new grid metrics and reporting requirements that would ensure better visibility into the potentially disparate experiences of energy customers, thus allowing for more thorough investigations into grid inequities.[6]

These efforts all serve to uncover and address one significant issue: equity is not regulated in grid planning procedures today and can result in hidden yet highly impactful disparities. Today, most distribution planning procedures consider only three main drivers for new investment: equipment failure, prevalent reliability issues, or if significant load growth is expected in an area. Replacing equipment only after it has failed can leave communities vulnerable to prolonged outages, particularly those served by aging and poorly maintained grid equipment. And as evidenced by Minnesota's efforts, most reliability

---

[2] Shalanda Baker et al., *The Energy Justice Workbook. Initiative for Energy Justice,* Initiative for Energy Justice (2019). https://iejusa.org/workbook/

[3] See pages 53-74 of *Direct Testimony of Jackson Koeppel on Behalf of Soulardarity and We Want Green, Too.* Michigan Public Service Commission - Case U-20836-0523. (2022, May 19). https://mi-psc.force.com/sfc/servlet.shepherd/version/download/0688y000002y4SPAAY

[4] Feng Chi Hsu et al., Frozen Out in Texas: Blackouts and Inequity, The Rockefeller Foundation (2021). https://www.rockefellerfoundation.org/insights/grantee-impact-story/frozen-out-in-texas-blackouts-and-inequity/

[5] Anna M. Brockaway et al., *Inequitable access to distributed energy resources due to grid infrastructure limits in California*, University of California, Berkeley (2021), https://www.nature.com/articles/s41560-021-00887-6

[6] Bhavin Pradhan & Gabriel Chan, *Racial and Economic Disparities in Electric Reliability and Service Quality in Xcel Energy's Minnesota Service Area*, University of Minnesota (2024). https://conservancy.umn.edu/items/8121c1ee-b191-4add-ac72-086af690e344

indices deployed today lack the granularity necessary to uncover inequities between groups, thus leaving major gaps in planning and creating pockets of energy insecurity for groups hidden at societies margins. Additionally, because load growth largely stems from greater regional investment, communities experiencing disinvestment in other areas end up being systematically overlooked by their utilities and commissions in planning processes.

Identifying these disparities can be challenging in territories that fail to provide transparency into their grid infrastructure. Ohio is one such case in which there are very limited reporting requirements for investor-owned utilities (IOUs), of which the vast majority is preserved for regulatory discovery and not published publicly. For this reason, the Ohio Environmental Council (OEC)[7] consulted with the Interstate Renewable Energy Council (IREC)[8] to investigate whether there are disparities in grid infrastructure utilizing what limited information there is currently available. Building upon previous research efforts while recognizing constraints to Ohio-specific data availability, the following study objectives were identified:

- **Identify data needs** for investigation to be requested through regulatory discovery.
- Develop a **proxy method of evaluating hosting capacity** disparities.
- **Aggregate and combine data** with demographic and justice-focused datasets.
- **Examine whether correlations exist** between demographics, with particular focus on historically redlined communities.

Though the goal of this research was to perform a widespread analysis of grid disparities across the state, we were ultimately limited in the geography we could examine. This was because data availability was limited to the regulatory discovery process, meaning that only data from IOUs with open proceedings could be obtained. Additionally, as Ohio's utility commission reporting requirements do not currently consider energy equity, the data provided through discovery often lacked sufficient granularity or detail to be useful for such an analysis. Ultimately, of the four IOUs in Ohio, we were only able to obtain sufficient information from one, FirstEnergy Corporation. Fortunately, FirstEnergy's territory contained the most redlined regions to examine (6 of the 11) out of all the IOU territories in the state.

The rest of this paper is organized into seven parts and is focused on providing details into each aspect of this research effort. **Part D - Data Collection & Aggregation** describes what data sources were examined and what post processing was required. **Part E - Relevant Variables for Analysis** defines each of the variables and metrics that were used for this analysis and why they were chosen. **Part F - Methods & Analyses** outlines the statistical analyses performed and the platforms used. **Part G - Research Questions & Results** provides an overview of all findings and the conclusions drawn. **Part H - Summary** provides a shortened version of the results including only those in which important

[7] The OEC is a non-profit organization incorporated in Ohio under Section 501(c)(3) of the U.S. Internal Revenue Code, with thousands of individual and organization members throughout the state of Ohio. The OEC's mission is to protect the environment and health of all Ohio communities through legal and policy advocacy, decision-maker accountability, and civic engagement.

[8] IREC is a 501(c)(3) non-partisan, non-profit organization working nationally to build the foundation for rapid adoption of clean energy and energy efficiency to benefit people, the economy and our planet. IREC's regulatory program responsible for this product works to improve the rules, regulatory policies and technical standards that enable the streamlined, efficient and cost-effective installation of distributed energy resources (DERs).

conclusions could be drawn. **Part I - Conclusion & Recommendations** provides closing remarks and suggested next steps. And **Part J - Appendix** provides additional information that could not be contained within the report narrative.

D. Data Collection & Aggregation

Data collection, processing, and aggregation were important steps in this research project, with ensuring the granularity and quality of the provided utility data being the most critical. Notably, this research sought to correlate demographic and redlining boundaries with grid infrastructure, which do not adhere to most regionally drawn constraints. Thus, there were significant challenges as well as potential errors introduced when aggregating data across the sources without geographically based tools like hosting capacity or grid infrastructure maps as are provided in New York, Minnesota, California, and other states.

Circuit Infrastructure Spreadsheet – FirstEnergy provided a spreadsheet of circuit/feeder characteristics, including circuit name and unique identifier, service area by census tract, supplying substation, voltage classification, operation-based circuit capacity valuations, number of critical customers, grounding configuration, status of bi-directional controls, and circuit in-service date (See **Part E f**or more information on each variable). The following are important notes regarding this data set:

- *Starting Sample Size* **–** There were 2847 circuit data points provided, some with varying levels of information available and most containing all relevant attributes.
- *Out-of-State or Missing* – There were eight circuits that were located in census tracts located in Michigan that were excluded from all analyses. 187 had errors in their census tract designations of which 127 could be addressed with slight modifications to census tract values. A final 60 circuits with errors could not be rectified.
- *Criticality of Census Tract Data Point* – Of the IOUs from which data was requested, only FirstEnergy could provide service area by census tract, an important attribute given the lack of any GIS mapping tools.
- *Limits of Census Tract to Circuit Correlation* – Neither redlined districts nor circuit runs adhere to the drawn boundaries of census tracts, meaning that there is a significant limitation to correlating whether a particular circuit attribute correlates with a particular designation. Additionally, it is still unknown how representative the census tract designations for each circuit were of the types of communities they served, as a later discovery request surmised that many of FirstEnergy's circuit's ran through more than one census tract.

Substation Infrastructure Spreadsheet – FirstEnergy provided a spreadsheet of substation characteristics, including substation name, transformer bank name, transformer nameplate capacity, transmission and distribution side voltage, and transformer in-service date.

- *Starting Sample Size* – There were 1199 substation transformer data points (633 unique substations) provided, some with varying levels of information available and most containing all relevant attributes.

- *Missing or Confusing Census Tracts* – Unlike the circuit spreadsheet, the substation spreadsheet did not include service area census tracts for the substations identified. 39 of the substation data points could not have census tract values determined. Additionally, several of the circuits out of the same substation had different census tracts, meaning it was unknown which census tract actually correlated with the substation's location.

- *More Investigation Needed* – Due to lingering issues that created major uncertainties in the determination of substation transformer census tracts, statistical analyses using transformer nameplate were unable to be performed. These analyses should be conducted in future grid equity assessments.

Circuit Reliability Indices Spreadsheet – FirstEnergy provided a spreadsheet of annual reliability indices for each circuit by name and unique identifier over the past six years (2018 to 2023). Reliability indices included are SAIDI (System Average Interruption Duration Index), SAIFI (System Average Interruption Frequency Index), and CAIDI (Customer Average Interruption Duration Index). The following are important notes regarding this data set:

- *Data Errors* – For reliability-focused analyses, some circuit data points were omitted due to data values that appeared to be errors in reporting. If every annual value across all three reliability indices during the six-year period recorded were zero, we concluded they were likely reporting errors.[9]

- *Limitations of Conventional Reliability Indices* – Conventional reliability indices are not considered sufficient in uncovering grid inequities due to their lack of sub-circuit granularity, wide-spread averaging, and poor representation of actual customer outage experiences. See **Part I** for recommendations on how the Ohio Public Utility Commision (OPUC) could rectify these limitations by including other reliability metrics.

Climate and Economic Justice Screening Tool (CEJST)[10] Spreadsheet – We utilized a spreadsheet export of information contained within the CEJST that was developed in 2022 by the White House's Council on Environmental Quality under the Biden Administration and removed by the Trump Administration in 2025. This tool provided a robust list of climate, economic, and demographic identifiers for census tracts across the United States. The data points reported and extracted for this research included census tract state and county, population, racial demographics (e.g., white, black, native-american by percentage), climate or economically disadvantaged census tract designation, and the status of historical housing loan discriminatory practices (i.e., redlining). The following are important notes regarding the CEJST data set:

[9] IREC attempted to cross reference these unlikely zero-value reliability metrics with the age of the circuits in question and found that the average age across error-labeled circuits was 48.5 years, meaning that the low reliability was not a result of the widespread installation of new circuits. This further emphasized the need to remove these data points from the analysis.

[10] Council on Environmental Quality, *Climate and Economic Justice Screening Tool*, White House (2024), *[LINK REMOVED]* – Note: This study was performed prior to the White House's removal of CEJST from all federal websites. An archived version of the data used can be provided upon request.

- *Lacked Home Owner's Loan Corporation (HOLC) Grade* – The housing loan discriminatory practice known as redlining was often conducted using a grading scale (i.e., A, B, C, and D). However, the CEJST does not distinguish the grade for each census tract. Instead, they used the following labels: True - Area did experience housing loan discrimination (likely assigned C or D), False - Area did not experience housing loan discrimination (likely assigned A or B), or "Blank" - No evidence that area experienced systemic housing loan discrimination. Note that some data quality efforts were conducted to cross check assigned census tract labeled with source data[11] from the University of Richmond. Additionally, quality control was performed for data points like population which were cross referenced with data from the U.S. Census Bureau.

Merging and Processing Data Sets

Merging and processing data sets largely occurred within Excel using the VLOOKUP function, however some additional processing for certain analyses were conducted in R using functions like filter() or tidy(). The following is a summary of the variables used to merge and aggregate the datasets. Note that quality control was performed by randomly selecting ten rows after each merger to determine if the values were accurate.

- **Circuit Infrastructure Spreadsheet ↔ CEJST Spreadsheet** – **Census tract identifiers** (e.g., 26091062100) were used to lookup unique demographic values for each tract from CEJST to the Circuit Infrastructure Spreadsheet. This means that CEJST data points would be identical for circuits located within the same census tract.

- **Circuit Infrastructure Spreadsheet ↔ Circuit Reliability Indices Spreadsheet** – Unique circuit identifier numbers (e.g., 500201101) were used to correlate reliability metrics and all other circuit characteristics. Theoretically, there should be no repeat values as all circuits should have their own unique reliability values and characteristics, and our quality control efforts did not identify any significant issues.

- **Circuit Infrastructure Spreadsheet ↔ Substation Infrastructure Spreadsheet** – We attempted to use substation name to determine substation census tracts, however, tracts differed between circuits fed from the same substation and therefore data could not be merged effectively.

In Excel, each of these spreadsheets was added as a tab to a single Excel spreadsheet file then merged (if possible) into a separate tab. Several duplicates were created to test different merging and processing methods until a final version was created. The following post-processing steps were taken:

1. *Remove Out-of-State Circuits* – If census tract was located in a different state (i.e., MI), we filtered out then deleted rows.

2. *Adjust Census Block Group to Census Tracts* – If census tract assigned to a circuit had an error and was not found in CEJST, we replaced the final digit of the tract (also known as the census block group code) with a zero to correlate with the more general census tract.

---

[11] Robert Nelson et al., *Mapping Inequality Redlining Areas*, University of Richmond Digital Scholarship Lab (2023, December 12). https://www.arcgis.com/home/item.html?id=d77c640241d84b6889ab290cd4cb755b

3. *Removed any remaining unknown census tracts* – If census tract unknowns still existed, we removed them from the data set. Note that these missing census tracts often correlated with other missing data points on the circuit's characteristics.

4. *Replaced missing data points with NA* – If any data points remained missing or unidentified, we replaced them with "NA" to make it easier to filter them out in Excel or in R. Note that often some circuits had only partial data sets making them useful for some analyses while needing to be filtered for others. This NA designation allowed us to perform such filtering when needed.

5. *Analysis Specific Processing* – For many statistical analyses, data points had to be converted from strings to numerical values or from booleans to binary values. **See Part E** for more information on exact processing changes for each variable.

As previously noted, it is important to remember that even with all of the quality control and processing steps performed, there is a mismatch in the resolution of each data set which impacts the final results. For instance, a census tract labeled as redlined means it shares some percentage overlap with a redlined region but is not a perfect one-to-one match as census tracts have changed over the past half-century. The same goes for circuits which may start in one census tract and end in a completely different one. As such, all results must be taken with a degree of caution. See **Part I** for recommendations on how to alleviate these resolution issues via more granular and detailed regulatory reporting requirements.

E. Relevant Variables for Analysis

Grid equity evaluations such as this require a thorough understanding of how each grid characteristic correlates to an actual customer experience or outcome. Take a grid characteristic like voltage class for example. If grid voltage is on average lower or higher in redlined areas as compared to non-redlined areas, what exactly does that mean for those in redlined areas? In the table below, we define each of the variables of interest and relate it to an aspect of grid equity. The following labels are assigned to each variable:

Data Type – the type of data point and why it is important

- *Demographic* – Who are the communities being impacted?
- *Infrastructure* – What is the status of grid infrastructure in those communities?
- *Performance* – How reliable is the electricity servicing those communities?

Grid Equity Aspect – the aspect of grid equity this variable helps to uncover

- *Capacity* – Helps assess whether the grid limits access to clean energy or electrification (Note: hosting capacity analysis is the most accurate method, however proxies were used for this research)
- *Disinvestment* – Helps assess whether grid planning procedures have resulted in disparities in infrastructure conditions over time.
- *Reliability* – Helps assess whether communities experience more or less outages
- *Disconnections* – Helps assess whether utilities disconnect certain communities at higher rates (Note: This was not included as part of this project, but is recommended as an improvement to future utility reporting and grid equity assessments)
- *Adoption* – Helps assess whether communities have adopted clean energy (Note: Queue data was provided late in the discovery process and could not be evaluated in time for testimony submittal)

Data Type	**Variable** – *Description*	Grid Equity Aspect
Demographic	**Census Tract** – *Small, somewhat stable geographic areas within counties utilized by the Census Bureau for gathering, organizing, tabulating, and displaying decennial census results.[12] Used to correlate demographic and infrastructure data. (Numeric)*	All
Demographic	**Redlined Status** – *Indicates whether a census tract is located in a previously redlined housing area. Used as an independent variable to evaluate differences in infrastructure or reliability. (Boolean)*	All
Demographic	**HOLC Status** – *Indicates whether a census tract is located within a region (often urban) where there is a record of systemic and discriminatory housing loan policies by banking institutions. Used as a constraint to limit the data set against which redlined communities were evaluated. (Boolean)*	All
Demographic	**Disadvantage Designation** – *Indicates whether a census tract has been identified as a climate, economic, or otherwise disadvantaged area by the federal government. Used as an independent variable for evaluating differences in infrastructure characteristics. (Boolean)*	All
Demographic	**Race (%)** – *Racial makeup of each census tract as a percentage of the total population. Used as an independent variable for infrastructure examinations and dependent variable for determining correlations with other demographic variables. (Numeric)*	All
Demographic	**Poverty (%)** – *Adjusted percentage of individuals living at or below 200% of the federal poverty line. Used as an independent variable for infrastructure examinations and dependent variable for determining correlations with other demographic variables. (Numeric)*	All
Infrastructure	**Circuit Name & Identifier** – *Unique identifiers of circuit infrastructure used to correlate infrastructure and performance data sets. (Character)*	All
Infrastructure	**Age or In-Service Date (Circuit & Substation)** – *Years since major circuit or substation (transformer) equipment were installed or replaced. Used as a dependent variable to discern trends in planning across demographic variables. (Numeric)*	Disinvestment
Infrastructure	**Voltage Class** – *Nominal operating voltage of the distribution circuit used as a dependent proxy variable for evaluating both hosting capacity and disinvestment. This is because lower voltage classes often correlate with lower capacity and higher circuit age. Values were mostly provided in both single/three-phase characters (ex. 7.2/12.5), but were modified to be numeric and represent the numeric three-phase voltage of the circuit (ex. 12.5). (Numeric)*	Disinvestment & Capacity
Infrastructure	**Voltage Class < 5kV** – *Using the nominal three-phase voltage extracted in 'Voltage Class', this variable translates voltage class into a boolean value to examine whether voltage classes are higher or lower than 5kV, a voltage class threshold that has been shown to correlate with lower capacity and grid disinvestment practices. (Boolean)*	Disinvestment & Capacity
Infrastructure	**Capacity** – *The maximum thermal operating capacity of the circuit either "based on summer normal nameplate rating (SNPR) of the substation exit breaker or limiting equipment (i.e. cable or conductor) at exit" (i.e., Overload Capacity) or the "SNPR subtracted from the normal rating of the circuit" (i.e., Normal Capacity).[13] Used as dependent variables to evaluate capacity across demographic independent variables. (Numeric)*	Capacity
Infrastructure	**Transformer Nameplate** – *The maximum thermal operating capacity of the substation bank transformer under normal conditions. Intention was to use this as another dependent variable to further examine differentials in grid capacity for new clean energy projects or electrification between demographic variables. However, we lacked necessary census tract data to correlate with said variables. (Numeric)*	Capacity
Infrastructure	**Bi-Directional Control Status** – *Indicates whether a circuit has bi-directional controls installed which can be used as a dependent proxy variable for capacity. This is because new clean energy projects may require expensive bi-directional upgrades to equipment controls in order to interconnect as circuit penetration levels increase. (Boolean)*	Capacity
Performance	**Reliability Indices (SAIDI, SAIFI, CAIDI)** – *Indices recorded at the circuit level that average the duration, frequency, and restoration time of outages. Provided for the last 6 years. (Numeric)*	Reliability

***Table 1**: Overview of Variables Used in Statistical Analyses. Note: Boolean values were translated into binary values (i.e., "True" = 1, "False" = 0).*

[12] Department of Commerce, *Geographic Areas Reference Manual - Chapter 10: Census Tracts and Block Numbering Areas (pg. 10-1)*, United States Government (1994), https://www2.census.gov/geo/pdfs/reference/GARM/GARMcont.pdf

[13] Definitions were revised based on responses provided in *OEC Set 2 Discovery Requests*

F. Methods & Analyses

To evaluate whether correlations exist between demographic, infrastructure, and performance variables, we deployed a series of statistical analyses to test various hypotheses. The final analysis selected for each test was determined by the structure of the variables themselves. All dependent variables underwent a descriptive statistical analysis to observe whether any trends could be initially observed before undergoing further inferential analyses.

Descriptive Statistics – Establishing a Baseline Understanding
Descriptive statistics were used to summarize key dataset characteristics and identify initial patterns that could inform subsequent inferential analyses.

Mean Comparison – Used to assess the differences in means between justice classifications. Conducted in Excel using the Data Analysis Toolkit and checked manually.

- *Objective:* If substantial differences in means were observed between groups, further statistical testing was pursued to determine whether these differences were statistically significant.
- *Example:* Difference in average circuit infrastructure age between redlined and not-redlined groups.

Visualization of Frequency Distributions or Other Comparative Trends – Used to visually detect patterns between justice classifications and infrastructure characteristics by leveraging bar graphs, histograms, or pie charts. . Conducted in Excel using pivot tables or simple mathematical functions like FREQUENCY() and AVERAGEIFS(), and then visualized using Excel's graphing function.

- *Objective:* Identify visually discernible trends that suggest relationships between variables, warranting further inferential statistical analysis.
- *Example:* Pie chart displaying the number of circuits within each voltage class in redlined versus non-redlined census tracts.

Inferential Statistics – Testing Relationships Between Variables
Inferential analyses were conducted to examine statistical significance, strength of associations, and predictive relationships within the dataset.

T-Test – Used to compare the means of two groups to determine if they are statistically different. Conducted in Excel using the Data Analysis Toolkit.

- *Objective:* Assess whether observed differences in means are statistically significant (i.e., p-value < 0.05), providing evidence of disparity between groups
- *Example:* Test whether the differentials between the average age of circuits located in historically redlined or disadvantaged census tracts are statistically significant.

Chi-Square Test – Used to evaluate independence between binary or categorical variables to examine the significance of potential disparities. Conducted in R using chisq.test() and quality checked in Excel using the manual calculations.

- *Objective:* Determine whether observed differences in descriptive analyses are statistically significant (i.e., p-value < 0.05), indicating a meaningful relationship between variables.
- *Example:* Examining the statistical independence between census tracts that experienced redlining and were designated as disadvantaged by CEJST.

Linear Regression Analysis – Used to assess the relationship between two or more continuous variables, determining the statistical significance of differences and the extent to which each independent variable explains variation in the dependent variable. Conducted in R using lm().

- *Objective:* Determine whether demographic variables explain the differentials in grid characteristics by calculating the statistical significance (i.e., p-value < 0.5) and the proportion of the observed variance explained by the independent variables (i.e., $R^2 >> 0$).
- *Example:* Examining whether redlining or disadvantaged status explain observed differences in racial compositions in census tracts.

Logistic Regression Analysis – Used to evaluate the relationship between independent variables and binary or categorical dependent variables to determine significance of differences, the extent to which the independent variables explain the dependent variables, and predicted probabilities of occurrence within categories. Conducted in R using lm() and multinom().

- *Objective:* Determine whether demographic variables explain the differentials in grid characteristics by calculating the statistical significance (i.e., p-value < 0.5) and the proportion of the observed variance explained by the independent variables (i.e., $R^2 >> 0$).
- *Example:* Examining whether certain voltage classes are more likely to be present in redlined or disadvantaged communities than others.

In exploring these relationships using these methodologies, we sought to observe consistent patterns that could help define standardized methodologies for future grid equity investigations. More discussion on these findings will be discussed in later sections.

The following Justice Classifications are used throughout to simplify how the results are displayed in the charts and data tables:

ALL – Analysis conducted across all circuits within FirstEnergy's territory. Automatically excludes errored data points within the variable being evaluated.

HOLC – Subset of "ALL" that includes only circuits located within census tracts identified and then graded by HOLCs.

HOLC-RED – Subset of "HOLC" that includes only circuits that were discriminated against in HOLC grading (i.e., ranked C or D) and thus were victim to redlining.

HOLC-NRED – Subset of "HOLC" that includes only circuits that were not victim to redlining (i.e., ranked A or B).

CEJST-DIS –Subset of "ALL" that includes only circuits within census tracts designated disadvantaged within the CEJST.

CEJST-NDIS – Subset of "ALL" that includes only circuits within census tracts not designated disadvantaged within the CEJST.

G. Research Questions & Results

In this section we provide detail into each of the research questions asked, the analyses selected to evaluate them, their results and their meaning. These are separated into five categories: population dynamics, voltage class composition, infrastructure age, circuit capacity, and service quality.

1. Population Dynamics

For this range of analyses, we conducted statistical analyses to examine if there were any discernible trends in population dynamics that would lend additional insights into later disparity findings. Note that any values labeled BIPOC are essentially the percentage of non-white peoples and are calculated as 100% minus the percentage of white peoples within the census tract.

1.1 – Are there racial demographic differences between redlined and non-redlined communities today?

In recognizing that the primary purpose of redlining was to discriminate against Black Americans, the first question we sought to answer was whether there remained any statistical racial differences between the racial demographics of communities redlined in the mid-century and census data today. **Figure 1.1.1** is a box-and-whiskers plot of Black, White, and BIPOC populations across census tracts in FirstEnergy territory, which shows that redlined tracts have higher percentages of black and BIPOC communities.

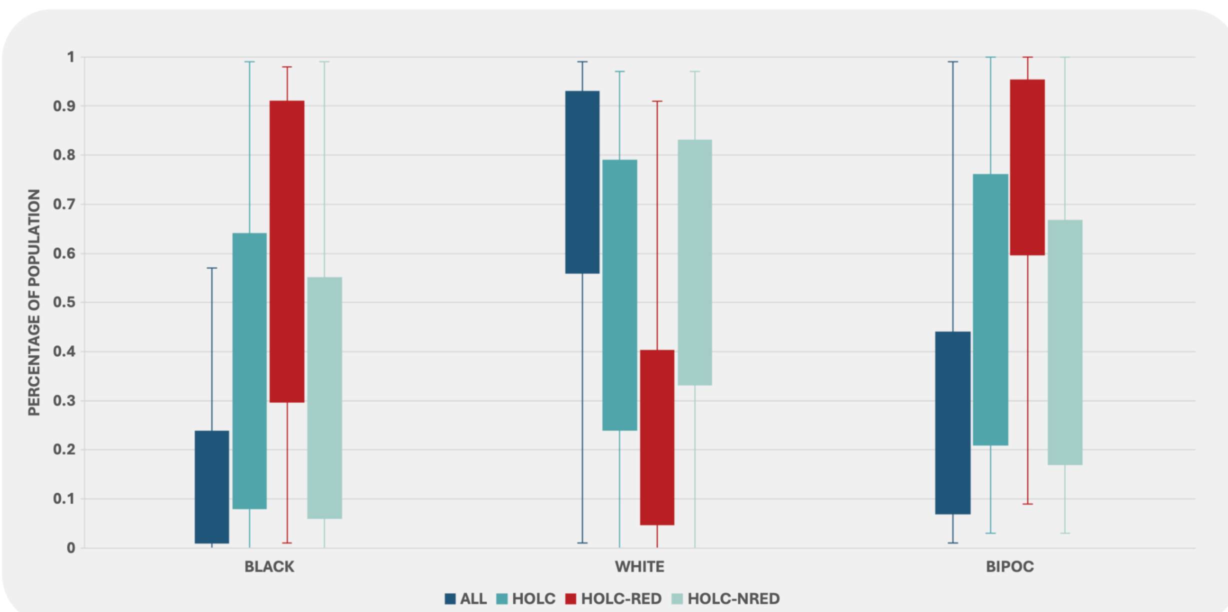

Figure 1.1.1: Racial Composition by Justice Classification *– Box-and-whisker plot showing that redlined tracts have higher black and BIPOC populations as compared to non-redlined and system average.*

While the box-and-whisker plot above showed the magnitude of the differences between labeled datasets, other descriptive measures such as histograms can help push us to explore whether more complex relationships also exist. As is demonstrated by **Figure 1.1.2** below, there is a somewhat linear and contradictory relationship between the frequency of black population percentages and redlining status in HOLC designated areas. Therefore, further investigation through more inferential analysis is necessary.

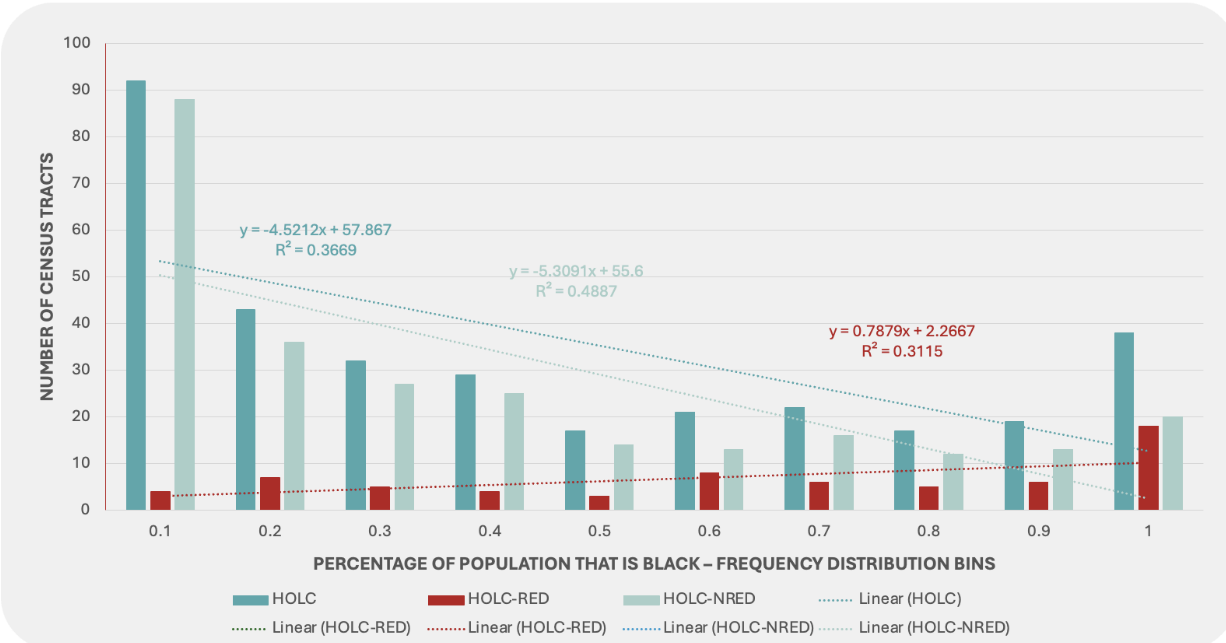

***Figure 1.1.2: Histogram of Black Population Distribution Across HOLC-Designated Census Tracts** – Trend lines show that the number of redlined tracts have a somewhat positive correlation with the composition of black inhabitants as a percentage of total tract population.*

1.2 – Are redlined areas more likely to be identified as climate or economically disadvantaged?

Because redlining occurred over half a century ago, there are questions about whether these regions still experience hardship as a result of such disinvestment. While we cannot perform a complete sociological analysis of these trends, utilizing the correlated data allows us to evaluate trends as a proxy. **Table 1.2.1** below shows a Chi-Square Statistic between redlining and CEJST disadvantaged designations for census tracts located in FirstEnergy's service territory. These results show that in both the all census tracts case (*left*) and HOLC-only tracts (*right*), there is a statistically significant relationship between redlined and disadvantaged tracts (p-value << 0.05). However, this relationship is less pronounced in the HOLC-only scenario (*right*) suggesting that other variables might be better suited than historic redlining in determining modern-day disparities.

REDLINED STATUS vs CEJST DISADVANTAGED DESIGNATION		
CONTINGENGY TABLE		
	NOT CEJST DISADVANTAGED (0)	CEJST DISADVANTAGED (1)
NOT REDLINED (0) *(ALL TRACTS)*	500 61.0%	254 31.0%
REDLINED (1)	3 0.4%	63 7.7%
CHI-SQUARE STATISTIC		
Test Statistic (X^2)	Degrees of Freedom (df)	p-value
95.051	1	2.2E-16

REDLINED STATUS vs CEJST DISADVANTAGED DESIGNATION		
CONTINGENGY TABLE		
	NOT CEJST DISADVANTAGED (0)	CEJST DISADVANTAGED (1)
NOT REDLINED (0) *(ONLY HOLC TRACTS)*	105 31.8%	159 48.2%
REDLINED (1)	3 0.9%	63 19.09%
CHI-SQUARE STATISTIC		
Test Statistic (X^2)	Degrees of Freedom (df)	p-value
28.182	1	1.104E-07

Table 1.2.1: Chi-Square Analysis of Redlined vs. Disadvantaged Status *– Test shows that when all FirstEnergy tracts are considered (left) there is a very highly statistically significant relationship between redlining and disadvantaged status, however, this significance is reduced when considering only HOLC defined tracts (right). See Appendix B for example R-Code.*

1.3 – Can statistical results be improved using both historic and modern indicators?

Learning from the previous test, it appears likely that it would be more advantageous to utilize both historic redlining status and modern CEJST disadvantaged designations to evaluate certain variables. To test this, we performed linear regression models in **Table 1.3.1** below for dependent variables BLACK and BIPOC population percentages that are known to have some correlation to redlining. First, we used a single independent variable (redlining, *left two*) and then multiple independent variables (redlining and CEJST disadvantaged designation, *right two*) to see if the models improved and by how much. While both models showed statistical significance (p-values << 0.01), the R^2 value increased significantly (18.8% and 20.8% variance explained to 39.7% and 47.5%). It should be noted that both redlined status and disadvantaged designations contributed substantially to the model's increased accuracy, although the latter was ultimately more significant. <u>These results indicate that other analyses could benefit from the inclusion of multiple independent variables.</u>

	Dependent variable:			
	BLACK (%)	BIPOC (%)	BLACK (%)	BIPOC (%)
	(1)	(2)	(3)	(4)
Redlined Tracts	0.439***	0.483***	0.269***	0.282***
	(0.032)	(0.033)	(0.029)	(0.029)
Disadvantaged Tracts			0.275***	0.326***
			(0.016)	(0.016)
Constant	0.152***	0.250***	0.060***	0.140***
	(0.009)	(0.009)	(0.010)	(0.009)
Observations	820	820	820	820
R^2	0.188	0.208	0.397	0.475
Adjusted R^2	0.187	0.207	0.395	0.474
Residual Std. Error	0.248 (df = 818)	0.257 (df = 818)	0.214 (df = 817)	0.209 (df = 817)
F Statistic	189.602*** (df = 1; 818)	214.969*** (df = 1; 818)	268.780*** (df = 2; 817)	370.275*** (df = 2; 817)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 1.3.1: Linear Regression Results for Black and BIPOC Population Shares by Redlined and Disadvantaged Tract Designations *– Results show that utilizing both redlined and disadvantaged (right) census tract status*

improves compared to each variable by itself (left) improves model fit for correlating relationships for Black and BIPOC population shares (i.e., increase in R^2).

However, it must be noted that such positive relationships are not always the case. As illustrated in the example in **Table 1.3.2**, although all dependent variables were statistically significant with relatively high R^2 values, redlined tracts played a mostly insignificant role in explaining the variance for both rate of poverty ($R^2 = 0.067$) and energy burden ($R^2 = 0.071$) as compared to the disadvantaged designation ($R^2 =$ 0.482 and 0.356, respectively).

	Dependent variable:			
	BLACK (%)	BIPOC (%)	POVERTY (%)	BURDEN (%)
	(1)	(2)	(3)	(4)
Redlined Tracts	0.269***	0.282***	0.067***	0.071***
	(0.029)	(0.029)	(0.021)	(0.026)
Disadvantaged Tracts	0.275***	0.326***	0.482***	0.356***
	(0.016)	(0.016)	(0.012)	(0.015)
Constant	0.060***	0.140***	0.359***	0.463***
	(0.010)	(0.009)	(0.007)	(0.009)
Observations	820	820	820	820
R^2	0.397	0.475	0.713	0.472
Adjusted R^2	0.395	0.474	0.712	0.471
Residual Std. Error (df = 817)	0.214	0.209	0.154	0.192
F Statistic (df = 2; 817)	268.780***	370.275***	1,012.845***	365.506***
Note:				*p<0.1; **p<0.05; ***p<0.01

***Table 1.3.2**: **Linear Regression Results for Population Shares (i.e., Black & BIPOC) and Burden Variables (i.e., Poverty Rate & Energy Burden) by Both Redlined and Disadvantaged Tract Designations** – Results show that while utilizing both redlined and disadvantaged is beneficial to population share models, only disadvantage status has a significant relationship with poverty and [energy] burden.*

Conclusions

Reviewing population dynamics was helpful in concluding the following about census tracts within FirstEnergy's service territory:

- **1C.1** – Historical redlining is still indicative of racial demographic differences; however, other more modern indicators like CEJST disadvantaged designation are slightly more accurate.
- **1C.2** – Historical redlining is not indicative of high poverty levels or energy burden today, both of which are more accurately estimated by CEJST designation.

2. Voltage Class Composition

Next, we moved onto our evaluation of grid infrastructure to determine if there were any disparities in circuit characteristics between redlining status and other independent variables. The first characteristic explored was voltage level and class which can provide insights into grid access inequities related to electrification and the adoption of clean energy. This is because the voltage level of a distribution circuit is directly correlated to a circuit's ability to accommodate higher penetrations of DERs or add new loads through electrification.

2.1 – Are circuits located in historically redlined or disadvantaged census tracts more likely to have lower voltages compared to circuits not in those areas?

We start by employing a range of descriptive analytical visualizations to discern whether any particular trends are visible and should be investigated further. Similar to the analytical process deployed in the Souladarity grid equity investigation in Michigan, our primary concern is the quantity of low-voltage circuits (less than 5kV) within regions of interest.[14] Using pie charts, we reviewed the circuit class composition across the service territory. **Figure 2.1.1 (*top*)** revealed that redlining does not appear to differ significantly from the all circuits baseline aside from the frequency composition of 15kV class circuits (i.e., 11.4 kV to 13.2 kV). However, **Figure 2.1.1 (*bottom*)** showed a much more significant deviation from the baseline as well as major differences in circuit class composition between CEJST designations. This indicates that further investigation is warranted and that CEJST designation will likely be the best independent variable.

[14] *See Note 2*

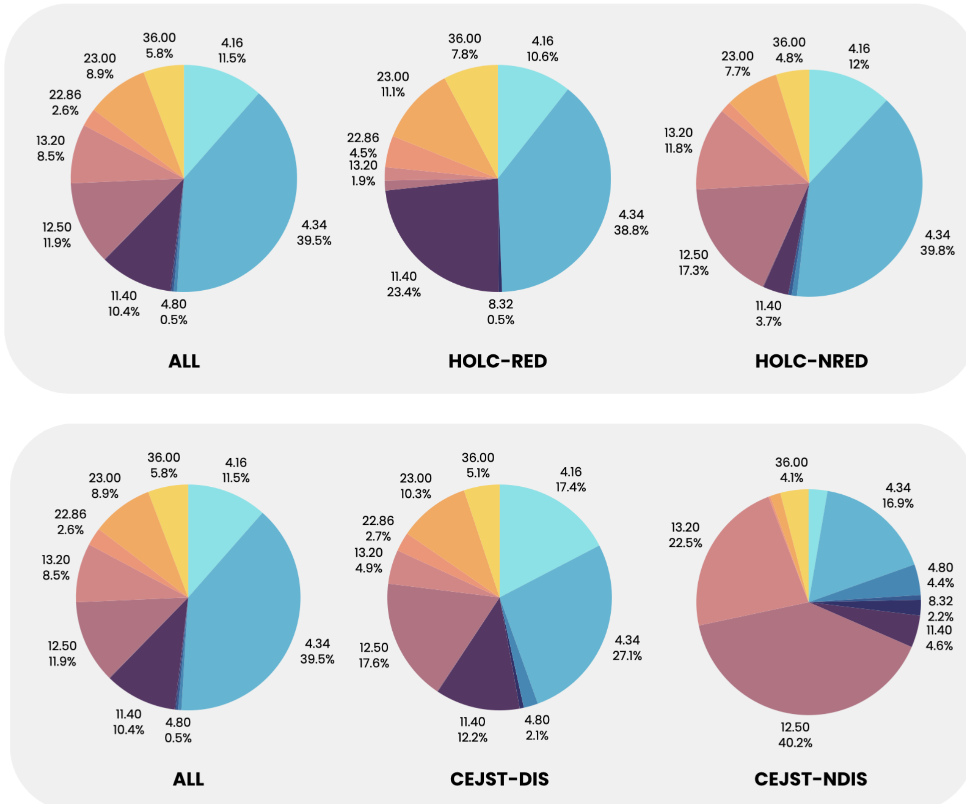

Figure 2.1.1: Pie Chart of Voltage Class Composition by Justice Classification *– Figure shows that class composition is relatively consistent across redlining status (top), but differs across disadvantaged status (bottom).*

Recognizing that there might be some correlation between our independent variables and grid voltage, we then move onto inferential statistical analyses, starting with a series of Chi-Square Tests shown. To do this, we first translate our previous categorical voltage class composition (i.e., 4.8kV, 13.2kV, etc.) into a boolean value where less than 5kV is TRUE (or 1) and greater than 5kV is FALSE (or 0). With our data formatted correctly, we now perform our Chi–Square Tests, the results of which are shown in **Table 2.2.1** below. In the top two tables, we investigate the relationship between redlined status and voltage class < 5kV. These results show that while there is statistical significance when evaluating all circuits (p-value = 2.14e-14), limiting the dataset to HOLC areas only to isolate the impact of redlining yields no statistical significance (p-value = 0.3293). Using disadvantaged status yields statistical significance for all circuits, however, this significance is only marginal for HOLC areas (p-value = 0.06277).

REDLINED STATUS vs VOLTAGE CLASS < 5KV		
	CONTINGENGY TABLE	
	VOLTAGE > 5KV (0)	VOLTAGE < 5KV (1)
NOT REDLINED (0) *(ALL CIRCUITS)*	1597 58.8%	694 25.5%
REDLINED (1)	216 7.9%	211 7.76%
	CHI-SQUARE STATISTIC	
Test Statistic (X^2)	Degrees of Freedom (df)	p-value
58.397	1	2.14E-14

REDLINED STATUS vs VOLTAGE CLASS < 5KV		
	CONTINGENGY TABLE	
	VOLTAGE > 5KV (0)	VOLTAGE < 5KV (1)
NOT REDLINED (0) *(ONLY HOLC CIRCUITS)*	389 31.2%	430 34.5%
REDLINED (1)	216 17.3%	211 16.93%
	CHI-SQUARE STATISTIC	
Test Statistic (X^2)	Degrees of Freedom (df)	p-value
0.95174	1	0.3293

DISADVANTAGED STATUS vs VOLTAGE CLASS < 5KV		
	CONTINGENGY TABLE	
	VOLTAGE > 5KV (0)	VOLTAGE < 5KV (1)
NOT REDLINED (0) *(ALL CIRCUITS)*	1163 42.8%	355 13.1%
REDLINED (1)	650 23.9%	550 20.24%
	CHI-SQUARE STATISTIC	
Test Statistic (X^2)	Degrees of Freedom (df)	p-value
151.04	1	2.20E-16

DISADVANTAGED STATUS vs VOLTAGE CLASS < 5KV		
	CONTINGENGY TABLE	
	VOLTAGE > 5KV (0)	VOLTAGE < 5KV (1)
NOT REDLINED (0) *(ONLY HOLC CIRCUITS)*	158 12.7%	199 16.0%
REDLINED (1)	447 35.9%	442 35.47%
	CHI-SQUARE STATISTIC	
Test Statistic (X^2)	Degrees of Freedom (df)	p-value
3.4626	1	0.06277

Table 2.2.1: Chi-Square Analysis of Justice Classification vs. Voltage Greater or Less than 5kV *– Tests show statistical significance when considering all circuits (left), but little to no significance when only HOLC circuits are considered (right). See Appendix B for example R-Code.*

These results ultimately mean that any perceived relationship between redlining status and voltage class is a result of HOLC-defined circuits having a lower voltage than the larger service territory. **Figure 2.2.1** shows an additional comparative assessment which exemplifies this point. Therefore, deploying the use of disadvantaged designation against circuit voltage to determine disparities is likely a better correlating factor.

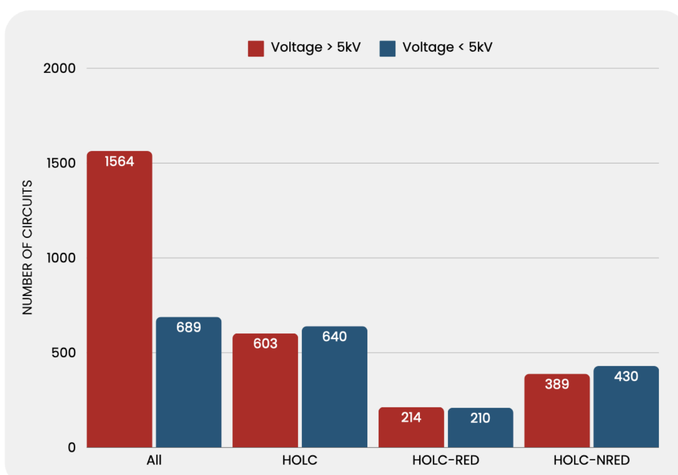

Figure 2.2.1: Histogram of Circuits With Voltages Greater or Less Than 5kV by Redlining Status *– Results show relatively even distribution of circuit voltage classes across redlining status, suggesting little to no relationship.*

We can now perform our final set of inferential regression analyses to discern whether disadvantaged status can help explain the variance in voltage across FirstEnergy's territory. For this series of tests, we will modify the full range of voltage classifications into a more discrete number of categories (i.e., 5kV category includes all 4.16, 4.34, 4.8 classifications; 15kV category includes all 8.32, 11.4, 12.5, 13.2 into 15kV category, etc.). However, in order to perform a regression with a binary independent variable (disadvantaged status) and categorical dependent variable (voltage class – 5kV, 15kV, 24kV, 34kV), we must deploy the use of a multinomial logistic regression (Appendix B for R-Code). **Figure 2.2.2** below provides the results of this analysis.

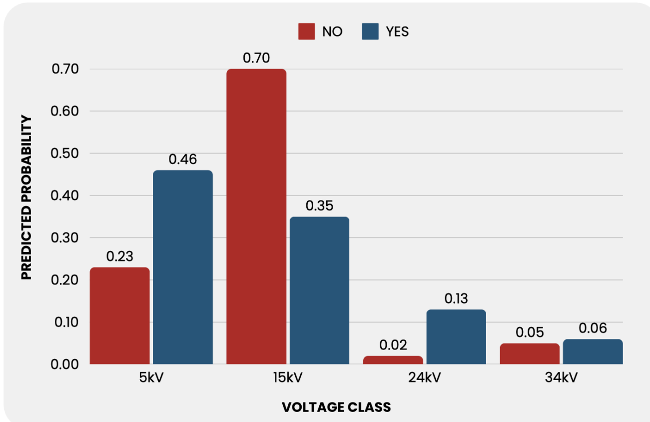

Figure 2.2.2: Predicted Probability Results from Multinomial Regression of Redlining Status to Voltage Less ("YES") or Greater ("NO") Than 5kV *– Results above have the following p-values by voltage class: 5kV (2.2e-16), 15kV (0), 24kV (1.34e-09), 34kV (0.056). These suggest that circuits in disadvantaged tracts are twice as likely to be less than 5kV.*

The p-values by voltage class of these results are: 5kV (2.2e-16), 15kV (0), 24kV (1.34e-09), 34kV (0.056). Thus, all of the predicted probabilities are highly statistically significant aside from 34kV which is only marginally significant. These results show that disadvantaged areas are twice as likely to have circuits with low voltage infrastructure than non-disadvantaged. Additionally, non-disadvantaged communities are twice as likely to have 15kV class infrastructure than disadvantaged areas. <u>These results show that there are clear disparities in the voltage class of distribution infrastructure in communities designated as disadvantaged.</u>

Conclusions

Reviewing voltage class composition was helpful in concluding the following about circuits within FirstEnergy's service territory:

- **2C.1** – Historical redlining is not indicative of disparities in grid voltage, and any differences seen between the larger dataset are a result of all HOLC-designated areas having lower voltages on average.
- **2C.2** – Areas designated as disadvantaged by CEJST are twice as likely to have low voltage circuits, therefore, there is a need to address such disparities through equity-informed distribution planning.

3. Infrastructure Age

Potential grid disinvestment could also result in disparities in the age of grid infrastructure between regions. Our goal here is to examine whether or not grid infrastructure is older in redlined or disadvantaged communities and other communities in FirstEnergy territory.

3.1 – Are circuits located in historically redlined or disadvantaged census tracts more likely to be older than circuits not in those areas?

Starting with a simple series of descriptive statistical analyses to discern any potential trends, **Figure 3.1.1** below shows the box-and-whisker plots for circuit in-service data across our primary independent variables. These indicate that redlined communities might be slightly older on average than the circuit average, circuits identified by HOLCs, and non-red circuits. However, these trends disappear when you include the 239 circuits with zero-value in-service dates (9% of total) which could be due to data errors or wide-spread, recent capital investments.

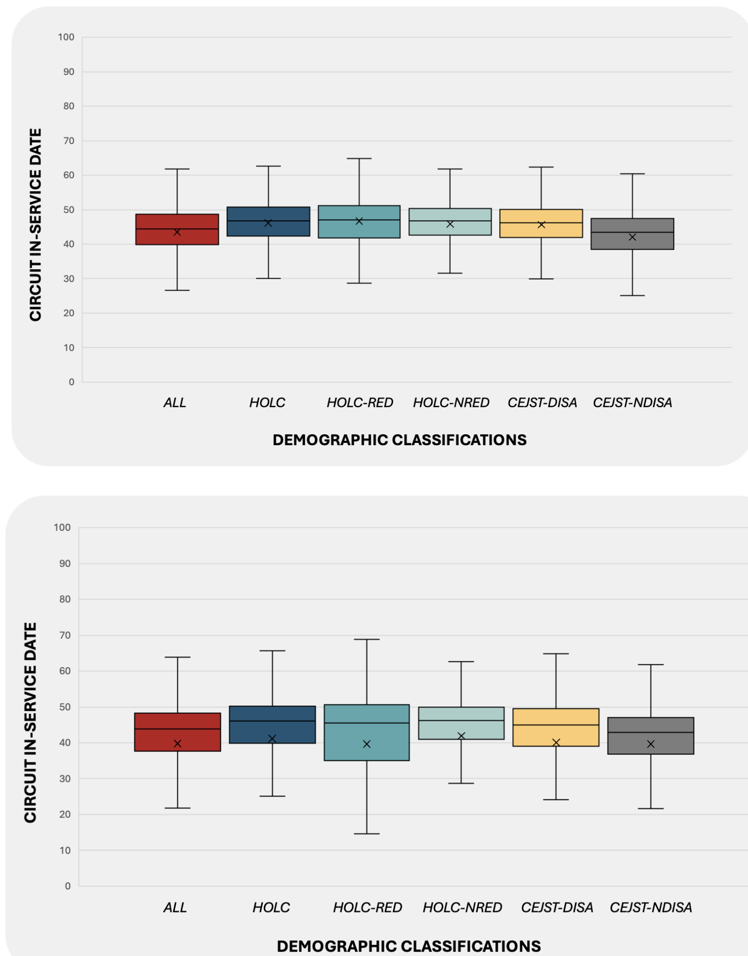

Figure 3.1.1: Infrastructure Age Composition by Justice Classification *– Box-and-whisker plots for infrastructure age by Justice Classification excluding zero in-service values (top) and including zero values (bottom). Results are questionable and show marginal age increase for circuits in redlined tracts, and moderate increase for circuits in disadvantaged tracts when excluding zero values (top). However, these results change drastically when including zero values (bottom).*

Due to the lack of discernible differences in the descriptive statistics, we elected to first perform a simple inferential T-Test analysis to determine which relationships were worth further exploration. **Table 3.1.1** below details these results between HOLC-RED and HOLC-NRED on the left and between CEJST-DIS and CEJST-NDIS on the right. These results show that two relationships are statistically significant. The first is in the zero-inclusive case comparing redlined and non-redlined HOLC-identified regions which shows that the average age of circuits in redlined areas is actually *younger* than that of those located in non-redlined tracts. The second is in the zero-exclusive case comparing circuits in CEJST disadvantaged areas to non-disadvantaged which found that the former is more than 3 years older on average than the later.

F-Test Two-Sample for Variances

	HOLC-RED *	***HOLC-NRED ****	
Mean	39.65693669	41.94100087	
Variance	391.7169096	228.1347157	
Observations	427	819	
df	426	818	
F	1.717042092		F > Fcrit, thus
P(F<=f) one-tail	2.83141E-11		Unequal
F Critical one-tai	1.147225355		Variance T-Test

F-Test Two-Sample for Variances

	HOLC-RED **	***HOLC-NRED *****	
Mean	46.6487933	45.86072058	
Variance	133.9080552	69.48272852	
Observations	363	749	
df	362	748	
F	1.927213541		F > Fcrit, thus
P(F<=f) one-tail	3.87924E-14		Unequal
F Critical one-tai	1.158186193		Variance T-Test

t-Test: Two-Sample Assuming Unequal Variances

	HOLC-RED *	***HOLC-NRED ****	
Mean	39.65693669	41.94100087	
Variance	391.7169096	228.1347157	
Observations	427	819	
Hypothesized Me	0		
df	691		
t Stat	-2.088607201		
P(T<=t) one-tail	0.018554259		p-value < 0.05,
t Critical one-tai	1.647061769		therefore
P(T<=t) two-tail	0.037108517		**statistically**
t Critical two-tai	1.963403002		**significant**

t-Test: Two-Sample Assuming Unequal Variances

	HOLC-RED **	***HOLC-NRED *****	
Mean	46.6487933	45.86072058	
Variance	133.9080552	69.48272852	
Observations	363	749	
Hypothesized Me	0		
df	550		
t Stat	1.159858872		
P(T<=t) one-tail	0.123304759		
t Critical one-tail	1.647628817		p-value > 0.1,
P(T<=t) two-tail	0.246609518		therefore **no**
t Critical two-tail	1.964286551		**significance**

*Including Zero Values
**Excluding Zero Values

F-Test Two-Sample for Variances

	CEJST-DIS *	***CEJST-NDIS ****	
Mean	40.05268783	39.63080104	
Variance	304.2099949	195.4311399	
Observations	1200	1579	
df	1199	1578	
F	1.556609633		F > Fcrit, thus
P(F<=f) one-tail	1.03633E-16		Unequal
F Critical one-tai	1.092900881		Variance T-Test

F-Test Two-Sample for Variances

	CEJST-DIS **	***CEJST-NDIS *****	
Mean	45.64408869	42.08274031	
Variance	91.26142955	104.2767494	
Observations	1053	1487	
df	1052	1486	
F	0.875184833		Close variance,
P(F<=f) one-tail	0.010002321		but selected
F Critical one-tai	0.910099726		Unequal

t-Test: Two-Sample Assuming Unequal Variances

	CEJST-DIS *	***CEJST-NDIS ****	
Mean	40.05268783	39.63080104	
Variance	304.2099949	195.4311399	
Observations	1200	1579	
Hypothesized Me	0		
df	2248		
t Stat	0.686855861		
P(T<=t) one-tail	0.246122205		
t Critical one-tail	1.645531741		p-value > 0.1
P(T<=t) two-tail	0.492244409		therefore **no**
t Critical two-tail	1.961019824		**significance**

t-Test: Two-Sample Assuming Unequal Variances

	CEJST-DIS **	***CEJST-NDIS *****	
Mean	45.64408869	42.08274031	
Variance	91.26142955	104.2767494	
Observations	1053	1487	
Hypothesized Me	0		
df	2353		
t Stat	8.993946008		
P(T<=t) one-tail	2.39307E-19		p-value << 0.05,
t Critical one-tail	1.645501469		therefore
P(T<=t) two-tail	4.78614E-19		**statistically**
t Critical two-tail	1.960972685		**significant**

*Including Zero Values
**Excluding Zero Values

***Table 3.1.1: T-Test Analysis of Justice Classification vs. Average Infrastructure Age** – Tests showing statistical significance of average age between redlining status (left) and disadvantaged status (right) with and without considering zero values. Results show that redlined tracts have younger infrastructure than non-redlined, and that disadvantaged tracts have older infrastructure than non-disadvantaged.*

To investigate the age relationship between CEJST disadvantaged and non-disadvantaged areas, we elected to perform a linear regression analysis using disadvantaged status as our binary independent variable and circuit in-service date as our dependent variable. **Table 3.1.2** below shows the results of this analysis which conclude that while there is a statistically significant difference between the ages of the groups (between 3.4 and 4.2 years older in disadvantaged tracts), the CEJST designation only accounts for between 2.9% and 4.5% of the overall age variance. <u>These results confirma measurable and statistically significant difference in infrastructure age between circuits located in disadvantaged and non-disadvantaged communities. However, the low R^2 value indicates that factors other than disadvantaged designation contribute meaningfully to these age disparities warranting further investigation.</u>

	Dependent variable:	
	CIRCUIT AGE	
	(ALL)	(Excluding 24kV & 34kV)
Disadvantaged Tracts	3.479***	4.254***
	(0.402)	(0.407)
Constant	42.165***	41.828***
	(0.261)	(0.260)
Observations	2,499	2,317
R^2	0.029	0.045
Adjusted R^2	0.029	0.045
Residual Std. Error	9.925 (df = 2497)	9.619 (df = 2315)
F Statistic	74.888*** (df = 1; 2497)	109.394*** (df = 1; 2315)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 3.1.2: Linear Regression Results for Infrastructure Age by Disadvantaged Tract *– Results show high statistical significance for both all circuit voltages (left) and excluding 24kV & 34kV circuits (right), although R^2 values suggests that disadvantaged status has low responsibility for these observed variances.*

3.2 – Are circuits in a lower voltage class (<5kV) older than circuits in a higher voltage class (>5kV)?

Building upon the results in our voltage composition analysis, it is prudent to determine whether the circuit voltage is also indicative of circuit age. Starting with descriptive statistics, box-and-whisker plots were first developed to see if any noticeable trends could be discerned. When reviewing averages in each voltage class shown in **Figure 3.2.1**, there appears to be a slightly higher infrastructure age in the lowest voltage classes. This difference is slightly more pronounced when voltage classes are aggregated into our previously used voltage categories as seen in **Figure 3.2.2**. However, the age of 34kV class infrastructure is a definite outlier but makes up only 6% of the total distribution circuits.

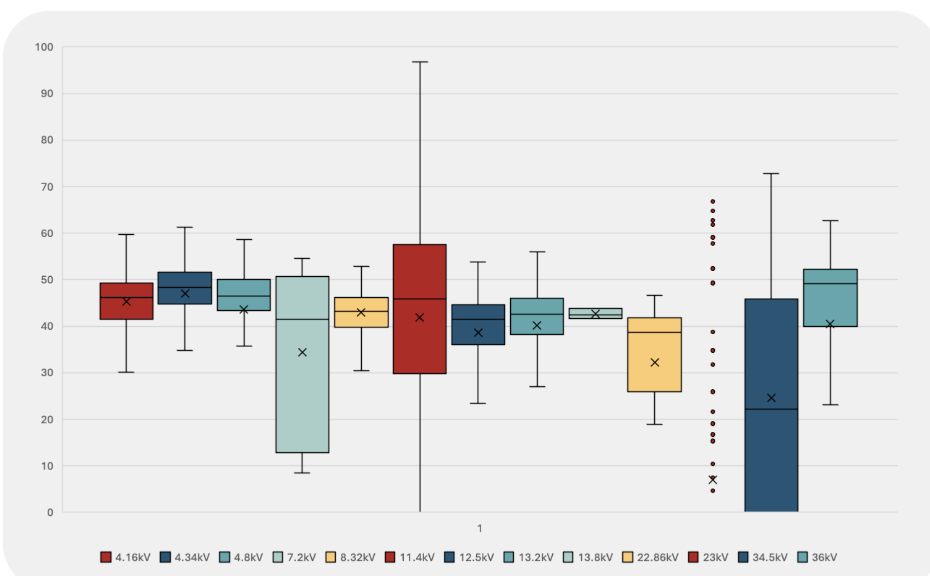

Figure 3.2.1: Infrastructure Age by Circuit Voltage Composition *– Box-and-whisker plot showing marginal increase in infrastructure age for lowest voltage classifications. Note: Voltage classes with less than 10 data points were omitted from this chart.*

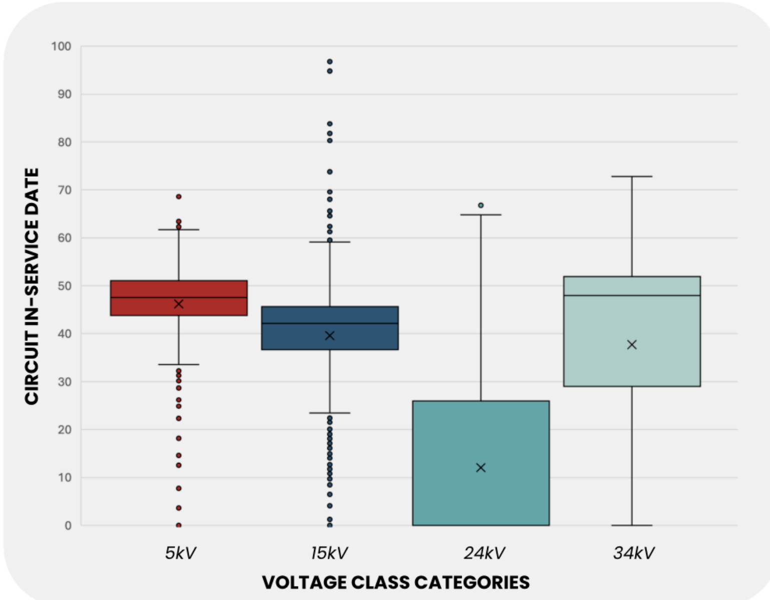

Figure 3.2.2: Infrastructure Age by Circuit Voltage Greater or Less Than 5kV – *Box-and-whisker plot showing that low voltage infrastructure is older on average than those with higher voltages, although low frequency 34kV equipment does not follow this trend.*

To investigate these relationships further, we then employed inferential statistics using linear regression analysis. The results of this regression shown in **Table 3.2.1** show that low voltage circuits (≤ 5kV) are 5.4 years older than higher voltage circuits. However, they also show extremely small R^2 values (R^2 = 0.066 and R^2 = 0.081), indicating that circuit voltage is only responsible for 6.6% of this age variance, and only as much as 8.1% when excluding higher voltage circuits. <u>Therefore, it can be concluded that while there is a difference in circuit age, this is largely due to variables outside of voltage class</u>.

	Dependent variable:	
	CIRCUIT AGE	
	(ALL)	(Exluding 24kV & 34kV)
VOLT $\leq 5KV$	5.426***	5.791***
	(0.405)	(0.402)
Constant	41.645***	41.281***
	(0.241)	(0.248)
Observations	2,540	2,357
R^2	0.066	0.081
Adjusted R^2	0.066	0.081
Residual Std. Error	9.759 (df = 2538)	9.468 (df = 2355)
F Statistic	179.314*** (df = 1; 2538)	207.771*** (df = 1; 2355)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 3.2.1: Linear Regression Results for Infrastructure Age by Voltage Greater or Less Than 5kV *– Results show highly statistically significant relationship between low voltage infrastructure and higher age, with circuits less than 5kV being at least five years older on average both including and excluding highest voltage, lowest frequency circuits. .*

Conclusions

Reviewing infrastructure age against Justice Classifications was helpful in concluding the following about circuits within FirstEnergy's service territory:

- **3C.1** – The circuits in CEJST designated disadvantaged communities are between 3.4 (all voltages) and 4.2 (excluding lower frequency 24kV and 34kV circuits) years older than non-disadvantaged areas. This difference is highly statistically significant, yet other hidden factors may be more responsible for this difference than the designation.
- **3C.2** – Low voltage circuits below 5kV are between 5.4 (all voltages) and 5.7 (excluding lower frequency 24kV and 34kV voltages) years older than higher voltage circuits. This difference is highly statistically significant, although factors other than voltage class may be more responsible for this difference.

---

4. Infrastructure Capacity

The most critical question posed in this research study is whether the grid infrastructure located in redlined or disadvantaged communities has a lower overall capacity, which may prohibit either electrification or clean energy adoption efforts by marginalized communities. For this, we will primarily examine circuit capacity values provided by FirstEnergy labeled as either normal and overload capacity, which were determined using the thermal limits of equipment located at the start of the feeder.[15]

4.1 – Are circuits located in historically redlined or disadvantaged census tracts more likely to have lower capacity than others in FirstEnergy's territory?

To determine whether there are capacity differentials between our independent variables, we start with simple descriptive statistics for both normal and overload circuit capacity, which yields the results in **Table 4.1.1** and visualized for clarity for normal capacity in **Figure 4.1.1**. Notably, these results show that circuit capacity is actually higher in redlined communities as compared to non-redlined for both capacity variables. However, the opposite is true for disadvantaged versus non-disadvantaged communities. Here we see that the circuit capacity is on average below that of non-disadvantaged communities for both capacity variables. Therefore, we will focus on exploring these particular relationships further.

[15] It is important to note that while the evaluation of normal and overload thermal limits provide useful insights, they do not fully capture the system's actual capacity for new loads or renewable energy integration. Capacity can vary significantly depending on the specific circuit equipment near a new DER or load connection. This location-specific capacity, known as hosting capacity, requires detailed power flow analyses using internal utility circuit models.

CAP-NORM	ALL	HOLC	HOLC-RED	HOLC-NRED	CEJST-DIS	CEJST-NDIS
Mean	3.52	3.07	3.70	2.75	3.23	3.60
Standard Error	0.11	0.15	0.30	0.17	0.16	0.15
Median	2.00	1.47	1.89	1.28	1.74	2.18
Mode	0.00	0.00	0.00	0.00	0.00	0.00
Standard Deviation	5.53	5.03	5.80	4.57	5.04	5.58
Sample Variance	30.58	25.28	33.61	20.84	25.35	31.18
Kurtosis	23.52	19.18	13.96	23.33	19.06	26.53
Skewness	4.22	3.90	3.40	4.22	3.88	4.47
Range	62.59	46.78	46.78	42.03	46.78	62.59
Minimum	-5.08	-1.48	-1.48	-1.36	-1.48	-5.08
Maximum	57.51	45.30	45.30	40.67	45.30	57.51
Sum	9042.50	3431.23	1380.92	2050.31	3378.27	5292.95
Count	2568	1119	373	746	1046	1469
CAP-OVER	**ALL**	**HOLC**	**HOLC-RED**	**HOLC-NRED**	**CEJST-DIS**	**CEJST-NDIS**
Mean	8.94	7.35	7.39	7.33	7.44	9.75
Standard Error	0.17	0.26	0.50	0.30	0.32	0.21
Median	7.97	3.51	3.55	3.49	3.77	9.32
Mode	7.97	7.97	7.97	2.74	3.60	11.66
Standard Deviation	8.46	8.75	9.60	8.30	8.82	8.14
Sample Variance	71.50	76.51	92.23	68.82	77.79	66.20
Kurtosis	13.60	14.48	12.26	15.92	16.97	14.18
Skewness	3.24	3.43	3.29	3.50	3.75	3.32
Range	69.46	69.46	69.46	67.54	69.46	66.60
Minimum	1.01	1.01	1.01	1.05	1.01	1.05
Maximum	70.47	70.47	70.47	68.59	70.47	67.65
Sum	22826.68	8172.21	2727.71	5444.50	5529.61	14275.24
Count	2554	1112	369	743	743	1464

***Table 4.1.1: Descriptive Statistics for Circuit Capacity by Justice Classification** – Table shows normal (top) and overload (bottom) circuit capacity. Results show higher to negligible differences in circuit capacities in redlined tracts, and show moderate to severe lower capacity values in disadvantaged tracts. Note that blank values were assumed to be errors and omitted from this analysis.*

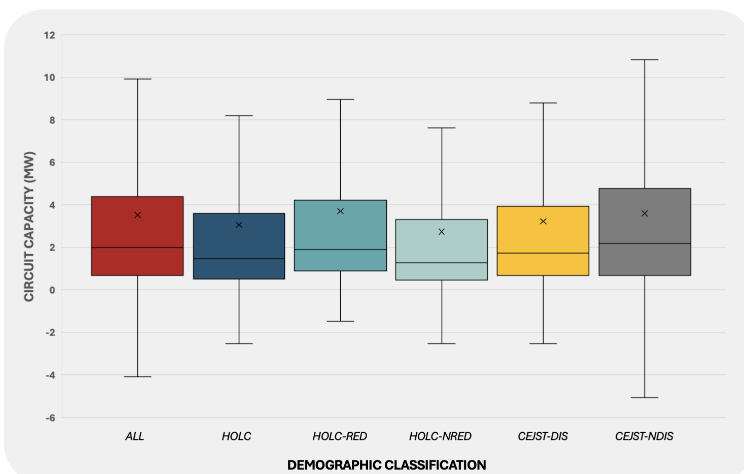

***Figure 4.1.1: Capacity Composition by Justice Classification** – Box-and-whisker plot showing higher circuit capacities in redlined tracts, but lower capacities disadvantaged tracts. Note that blank values were assumed to be errors and omitted from this analysis.*

We began with an inferential T-test to assess whether the differences in mean circuit capacities between disadvantaged and non-disadvantaged tracts are statistically significant and merit further exploration. The results, presented in **Table 4.1.2**, reveal that the capacity differences are marginally statistically significant for normal capacity and highly statistically significant for overload capacity.

T-Test -- Ciircuit Capacity vs. Disadvantaged Designation

CAP-NORM			**CAP-OVER**		
Step 1:			***Step 1:***		
F-Test Two-Sample for Variances			*F-Test Two-Sample for Variances*		
	CEJST-DIS	*CEJST-NDIS*		*CEJST-DIS*	*CEJST-NDIS*
Mean	3.22420248	3.60309735	Mean	7.46341929	9.71765827
Variance	25.3589339	31.1800192	Variance	71.1051753	66.2935005
Observations	1047	1469	Observations	1047	1469
df	1046	1468	df	1046	1468
F	0.81330719		F	1.0725814	
P(F<=f) one-tail	0.00016949		P(F<=f) one-tail	0.10940101	
F Critical one-tail	0.90974413		F Critical one-tail	1.09826226	
Calculated F < F Critical, therefore assume equal variances in t-test			*Calculated F < F Critical, therefore assume equal variances in t-test*		
Step 2:			***Step 2:***		
t-Test: Two-Sample Assuming Equal Variances			*t-Test: Two-Sample Assuming Equal Variances*		
	CEJST-DIS	*CEJST-NDIS*		*CEJST-DIS*	*CEJST-NDIS*
Mean	3.22420248	3.60309735	Mean	7.46341929	9.71765827
Variance	25.3589339	31.1800192	Variance	71.1051753	66.2935005
Observations	1047	1469	Observations	1047	1469
Pooled Variance	28.7580402		Pooled Variance	68.2954941	
Hypothesized Mean D	0		Hypothesized Mean D	0	
df	2514		df	2514	
t Stat	-1.7468998		t Stat	-6.7442374	
P(T<=t) one-tail	0.04038842		P(T<=t) one-tail	9.5096E-12	
t Critical one-tail	1.64545997		t Critical one-tail	1.64545997	
P(T<=t) two-tail	0.08077683		P(T<=t) two-tail	1.9019E-11	
t Critical two-tail	1.96090806		t Critical two-tail	1.96090806	
p-value = 0.04, therefore only marginal statistical signficance.			*p-value << 0.01, therefore statistically significant.*		

***Table 4.1.2: T-Test Analysis of Disadvantaged Status vs. Average Circuit Capacity** – Tests showing statistical significance for normal circuit capacity (left), and very high statistical significance for overload capacity (right). Note that blank values were assumed to be errors and omitted from this analysis.*

Considering the significance identified in these results, we performed a regression analysis to assess the extent to which disadvantaged status influences capacity differentials. These results are presented in **Table 4.1.3** below. On the left, we found a 10.5% reduction in normal capacity (0.366 MW) within disadvantaged tracts that is marginally statistically significant ($p < 0.1$), but that disadvantaged status is not statistically responsible for this difference ($R^2 = 0.001$, or 0.1%). On the right, we found a more notable 23% reduction in overload capacity (2.23 MW) in disadvantaged tracts that is highly statistically significant ($p < 0.01$), yet disadvantaged status is also not statistically responsible for this variance ($R^2 = 0.017$, or 1.7%). <u>Thus, while there is a significant capacity difference (-10.5% to -23%) between circuits in disadvantaged and non-disadvantaged tracts, these differences are not directly attributable to the disadvantaged designation and warrant further investigation to identify the primary contributing factors.</u>

	Dependent variable:	
	CIRCUIT CAPACITY	
	(Normal)	(Overload)
Disadvantaged Tracts	−0.366*	−2.230***
	(0.218)	(0.335)
Constant	3.615***	9.751***
	(0.140)	(0.216)
Observations	2,503	2,503
R^2	0.001	0.017
Adjusted R^2	0.001	0.017
Residual Std. Error (df = 2501)	5.371	8.263
F Statistic (df = 1; 2501)	2.828*	44.258***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 4.1.3: Linear Regression Results for Circuit Capacity by Disadvantaged Status *– Results show that circuits in disadvantaged tracts have lower capacities in both normal (marginally statistically significant) and overload conditions (highly statistically significant). Note that blank values were assumed to be errors and omitted from this analysis.*

4.2 – Do circuits with lower voltage (<5kV) have lower capacity on average than higher voltage classes (>5kV)?

To further support our findings above, we chose to examine what technical circuit attributes might be most responsible for such differences. This began by leveraging our findings in **Question 2.2** where we discovered that circuits located in disadvantaged communities are *twice as likely* to belong to a lower voltage class than those in non-disadvantaged areas. Therefore, there is a need to further investigate the relationship between circuit capacity and voltage class to determine whether voltage differentials may have contributed to the uncovered capacity disparities.

To begin, we reviewed descriptive statistics for both normal and overload circuit capacity categorized by voltage class. The results, shown in **Table 4.2.1**, indicate a clear trend: as voltage class increases, the mean capacity also increases under both capacity values. Additionally, circuits above 5kV exhibit significantly higher capacity than those at or below 5kV. These trends suggest that voltage class may be a key factor underlying the capacity differences observed in **Question 4.1**.

Descriptive Statistics -- Circuit Capacity v. Voltage Classification

CAP-NORM	5kV	15kV	24kV	34kV	< 5kV	> 5kV
Mean	1.19	3.46	4.48	19.72	1.19	4.78
Standard Error	0.04	0.08	0.74	1.16	0.04	0.16
Median	1.03	3.06	2.84	20.92	1.03	3.33
Mode	0.00	0.00	0.00	0.00	0.00	0.00
Standard Deviation	1.26	2.92	4.95	13.33	1.26	6.46
Sample Variance	1.59	8.54	24.47	177.72	1.59	41.72
Kurtosis	9.27	3.29	-0.95	-0.46	9.25	15.96
Skewness	2.14	1.11	0.60	0.31	2.14	3.52
Range	13.60	26.39	17.48	60.87	13.60	61.60
Minimum	-5.08	-4.09	-1.48	-3.36	-5.08	-4.09
Maximum	8.52	22.30	16.00	57.51	8.52	57.51
Sum	1068.03	5148.02	201.59	2622.33	1067.30	7971.94
Count	900	1490	45	133	900	1668
CAP-OVER	**5kV**	**15kV**	**24kV**	**34kV**	**< 5kV**	**> 5kV**
Mean	2.90	9.80	15.97	37.63	2.90	12.20
Standard Error	0.03	0.07	0.71	1.16	0.03	0.22
Median	2.89	9.72	17.82	38.97	2.89	10.20
Mode	2.74	7.97	19.80	43.33	2.74	7.97
Standard Deviation	1.03	2.79	4.77	13.36	1.03	8.90
Sample Variance	1.07	7.80	22.71	178.53	1.07	79.23
Kurtosis	12.01	5.62	0.73	0.10	12.01	12.03
Skewness	2.75	0.79	-0.94	-0.16	2.75	3.31
Range	7.69	26.09	21.80	67.53	7.69	69.46
Minimum	1.05	2.59	1.01	2.94	1.05	1.01
Maximum	8.74	28.68	22.81	70.47	8.74	70.47
Sum	2596.42	14504.77	718.56	5004.86	2598.49	20228.19
Count	895	1480	45	133	896	1658

Table 4.2.1: Descriptive Statistics for Circuit Capacity by Voltage Classification *– Table shows normal capacity (top) and overload capacity (bottom). Results show a consistent trend of circuits with higher voltages having higher circuit capacities.*

To confirm whether these differences are statistically significant, we conducted an inferential t-test comparing capacity values between the low-voltage (≤5kV) and high-voltage (>5kV) categories. As shown in **Table 4.2.2**, the difference in mean capacity is very highly statistically significant ($p << 0.001$) for both normal and overload capacity, reinforcing the idea that voltage level plays a role in capacity variations.

To further quantify this relationship, we performed a regression analysis, with results presented in **Table 4.2.3**. On the left side of the table, circuits under 5kV show a notable reduction in capacity under normal conditions (-3.616 MW), which is highly statistically significant ($p < 0.01$). However, voltage level statistically explains only a small portion of this variance ($R^2 = 0.097$, or 9.7%), suggesting that other factors may be influencing normal capacity values. On the right side, the significant reduction in overload capacity (-9.3 MW) for circuits under 5kV is also highly statistically significant ($p < 0.01$), with voltage class accounting for a much larger share of the variance ($R^2 = 0.276$, or 27.6%).

<u>While voltage level correlates significantly with circuit capacity, statistical analyses suggest that it is a primary driver of capacity differentials only for overload capacity.</u> From a power systems perspective, this relationship should extend to normal capacity as well, yet several data quality issues—including 345 instances of negative or zero normal capacity values—may have weakened the observed relationship. A later response to a discovery request provided some clarity on these negative and zero normal capacity values, ultimately indicating that overload capacity is the most accurate attribute for use in this comparative analysis. However, addressing these data inconsistencies through standardization and

transparency will be critical for obtaining a more accurate understanding of capacity constraints across voltage classes.

T-Test -- Ciircuit Capacity vs. Voltage Level

CAP-NORM			CAP-OVER		
Step 1: *F-Test Two-Sample for Variances*			***Step 1:*** *F-Test Two-Sample for Variances*		
	< 5kV	*> 5kV*		*< 5kV*	*> 5kV*
Mean	1.18538291	4.77934053	Mean	2.90010045	12.2003559
Variance	1.58888777	41.7192629	Variance	1.06659384	79.2274372
Observations	901	1668	Observations	896	1658
df	900	1667	df	895	1657
F	0.03808523		F	0.01346243	
P(F<=f) one-tail	0		P(F<=f) one-tail	0	
F Critical one-tail	0.9074974		F Critical one-tail	0.90724301	
Calculated F < F Critical, therefore assume equal variances in t-test			*Calculated F < F Critical, therefore assume equal variances in t-test*		
Step 2: *t-Test: Two-Sample Assuming Equal Variances*			***Step 2:*** *t-Test: Two-Sample Assuming Equal Variances*		
	< 5kV	*> 5kV*		*< 5kV*	*> 5kV*
Mean	1.18538291	4.77934053	Mean	2.90010045	12.2003559
Variance	1.58888777	41.7192629	Variance	1.06659384	79.2274372
Observations	901	1668	Observations	896	1658
Pooled Variance	27.6494002		Pooled Variance	51.8160129	
Hypothesized Mean D	0		Hypothesized Mean D	0	
df	2567		df	2552	
t Stat	-16.531355		t Stat	-31.16009	
P(T<=t) one-tail	1.0254E-58		P(T<=t) one-tail	3.189E-181	
t Critical one-tail	1.64544744		t Critical one-tail	1.64545093	
P(T<=t) two-tail	2.0508E-58		P(T<=t) two-tail	6.379E-181	
t Critical two-tail	1.96088855		t Critical two-tail	1.96089399	
p-value << 0.001, therefore very highly statistically significant			*p-value << 0.001, therefore very highly statistically significant.*		

***Table 4.2.2: T-Test Analysis of Circuit Capacity vs. Voltage Greater or Less Than 5kV** – Tests show very high statistical significance across both normal (left) and overload (right) capacity values, and reinforce higher voltage higher capacity trend.*

	Dependent variable:	
	CIRCUIT CAPACITY	
	(Normal)	(Overload)
Voltage $\leq 5kV$	-3.616^{***}	-9.300^{***}
	(0.218)	(0.298)
Constant	4.808^{***}	12.200^{***}
	(0.129)	(0.177)
Observations	2,554	2,554
R^2	0.097	0.276
Adjusted R^2	0.097	0.275
Residual Std. Error (df = 2552)	5.265	7.198
F Statistic (df = 1; 2552)	274.404^{***}	970.951^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

***Table 4.2.3: Linear Regression Results for Circuit Capacity by Voltage Greater or Less Than 5kV** – Results reinforce trend of higher voltage, higher capacity with very high statistical significance across both normal (left) and overload (right) capacity values.*

Conclusions

Reviewing circuit capacity against justice classifications was helpful in concluding the following about circuits within FirstEnergy's service territory:

- **4C.1** – The circuits in CEJST designated disadvantaged tracts have less capacity than non-disadvantaged areas ranging between 10.5% to 23% for normal and overload capacity respectively. The results for normal capacity are marginally statistically significant while the results for overload capacity are highly statistically significant. However, other hidden variables may be more responsible for these differences than disadvantaged designation.

- **4C.2** – Lower voltage circuits (less than 5kV) have significantly lower capacities than higher voltage circuits, ranging between 3.6 MW less normal capacity and 9.3 MW less overload capacity. These results are highly statistically significant, however, variables other than low or high voltage may be more responsible for the differences for normal capacity.

---

5. Service Quality

Service quality, or the rates of electricity disruption in terms of frequency or duration, are highly important in determining whether there are disparities in customer experience between populations of interest. Our goal is therefore to examine whether such disparities exist between communities with differing redlined or disadvantaged status.

5.1 – Do communities in historically redlined or disadvantaged census tracts have lower service quality than others in FirstEnergy's service territory?

Starting with descriptive statistics that evaluate the average of each reliability metric across census tracts, we found that the differences were most effectively assessed using visualizations, as shown in **Figure 5.1.1**. Notably, these charts clearly illustrate that both SAIDI and SAIFI values for redlined and disadvantaged communities are significantly lower than their counterparts and well below the system average.[16] Only CAIDI shows a slight indication of reduced service quality in these communities; however, this metric is generally considered less representative of customer experience than SAIDI and SAIFI. Given the extent of these differences, which suggest counter disparities, further inferential analyses are not necessary at this time.

Additionally, we examined other demographic factors, including reliability differences based on the percentage of BIPOC and Black populations within census tracts. However, due to the very low frequency of high-percentage Black and BIPOC communities in the dataset, no definitive conclusions could be drawn.

Conclusions

[16] Note that a lower SAIDI or SAIFI value indicates better service quality. A reduction in SAIDI means the total duration of service interruptions per average customer is shorter, while a reduction in SAIFI indicates that customers experience fewer sustained service interruptions.

Reviewing service quality against justice classifications was helpful in concluding the following about circuits within FirstEnergy's service territory:

- **5C.1** – Service quality in redlined or disadvantaged communities is higher than in non-designated communities using conventional reliability metrics.

Figure 5.1.1: Trends in Reliability Indices by Justice Classification *– Results show higher service quality in both redlined and disadvantaged tracts as compared to variable alternatives and to the system average.*

H. Summary

Below is a summary table of the results of each research question.

Sorting	Area	Research Question	*Descriptive Statistics*			*Inferential Statistics*				Results
			Simple	B-and-W	Histogram	Chi-Sq.	T-Test	Linear	Logistic	
1.1	Demographics	Are there racial demographic differences between redlined and non-redlined communities today?	✓	✓	✓	-	-	-	-	Redlined tracts have larger black and BIPOC populations than non-redlined.
1.2	Demographics	Are redlined areas more likely to be identified as climate or economically disadvantaged?	-	-	-	✓	-	-	-	Redlined tracts and disadvantaged designation have a statistically significant relationship.
1.3	Demographics	Can statistical results be improved using both historic and modern indicators?	-	-	-	-	-	✓	-	Utilizing both Redlining and Disadvantaged status provide analytial benefits for increasing model accuracy.
2.1	Voltage	Are circuits located in historically redlined or disadvantaged census tracts more likely to have lower voltages compared to circuits not in those areas?	?	-	-	?	-	-	✓	**RED**: Voltage class does not differ amongst redlined status. **DIS**: Circuits within disdvantaged communities are twice as likely to be lower voltage.
3.1	Age	Are circuits located in historically redlined or disadvantaged census tracts more likely to be older than circuits not in those areas?	-	✓	-	-	?	✓	-	**RED**: Infrastructure is 2 years younger on average in redlined areas, **DIS**: Infrastructure is 3.4 to 4.2 years older on average in disadvantaged areas.
3.2	Age	Are circuits in a lower voltage class (< 5kV) older than circuits in higher voltage classes (> 5kV)?	-	?	-	-	-	✓	-	Lower voltage circuits are 5 years older on average than higher voltage circuits.
4.1	Capacity	Are circuits located in historically redlined or disadvantaged census tracts more likely to have lower capacity than others in FirstEnergy's territory?	✓	?	-	-	✓	✓	-	**RED**: Circuits *may* have higher capacities on average in redlined tracts. **DIS**: Circuits have lower capacities on average in disadvantaged tracts.
4.2	Capacity	Do circuits with lower voltage (<5kV) have lower capacity on average than higher voltage classes (>5kV)?	✓	-	-	-	✓	✓	-	Lower voltage circuits have significantly lower capacities (-3.6 to -9.3 MW) as compared to higher voltage circuits.
5.1	Service Quality	Do communities in historically redlined or disadvantaged census tracts have lower service quality than others in First Energy's service territory?	✓	-	-	-	-	-	-	Service quality in redlined or disadvantaged communities is higher on average when using conventional reliability metrics.

Legend: '-' - Not Performed, '?' Perfomed but Uncertain Relationship, '✓' Performed and Relationship Detected, 'B-and-W' Box and Whisker Plot

Figure 6*: Summary table of analyses conducted and results concluded.*

I. Conclusions and Recommendations

This study sought to assess the extent to which electric grid infrastructure disparities exist in FirstEnergy's service territory by leveraging available utility data and integrating it with demographic and justice-focused datasets. Specifically, our objectives were to:

- **Identify data needs** for investigation to be requested through regulatory discovery.
- Develop a **proxy method of evaluating hosting capacity disparities**.
- **Aggregate and combine data** with demographic and justice-focused datasets.
- **Examine whether correlations exist** between demographics, with particular focus on historically redlined communities.

Despite significant data limitations, this study represents a critical step toward assessing energy equity in Ohio's electric grid planning processes. By applying an innovative data integration approach and leveraging existing regulatory records, we were able to illuminate structural inequities in grid conditions within historically marginalized communities. Our analysis revealed several critical insights:

- **Historic redlining is still correlated with racial demographic differences**, but it is a less precise predictor of energy burden or poverty levels compared to modern equity-focused designations, such as those used in the Climate and Economic Justice Screening Tool (CEJST), a tool that identifies disadvantaged census tracts across the United States based on burdens that communities experience.[17]
- **Grid voltage disparities were not clearly correlated with historic redlining but were correlated to CEJST disadvantaged designation.** Communities designated as disadvantaged based on CJEST methodology were twice as likely to have low-voltage circuits, suggesting an ongoing need for equity-informed distribution planning.
- **Grid age disparities were evident in disadvantaged communities** whose grid infrastructure was on average 3.4 to 4.2 years older than that in non-disadvantaged communities, a statistically significant difference that suggests historical patterns of underinvestment in these areas.
- **Lower voltage circuits were consistently older and had less capacity than higher voltage circuits**, with differences of 5.4 to 5.7 years in age and significantly reduced normal and overload capacity. Paired with our voltage disparity findings, this furthers the case for prioritizing investments in disadvantaged communities to address inequities in grid infrastructure.
- **Circuits in disadvantaged communities had 10.5% to 23% less normal and overload capacity**, a statistically significant disparity that aligns with prior voltage class and capacity findings. This suggests that historically lower investment levels have led to reduced grid capacity

[17] Council on Environmental Quality, *Climate and Economic Justice Screening Tool*, White House (2024), *[LINK REMOVED]* – Note: This study was performed prior to the White House's removal of CEJST from all federal websites. An archived version of the data used can be provided upon request.

for electrification and clean energy adoption in these areas and reinforces the need for equitable and proactive infrastructure investment.

- **Service quality metrics paradoxically suggested better reliability in disadvantaged and redlined areas**, though this is likely due to the limitations of conventional reliability indices which fail to capture the more granular outage experiences of individual customers.

These findings underscore the urgent need for greater transparency and equity considerations in Ohio's grid planning and investment processes. The lack of publicly available data remains a significant barrier to fully understanding and addressing systemic disparities in the energy system. Without robust reporting requirements, marginalized communities remain at risk of continued underinvestment and potentially disproportionate exposure to energy insecurity. To advance energy equity, we recommend that Ohio regulators and utilities take the following actions:

Expand Data Transparency and Reporting Requirements – Ensuring public access to granular, disaggregated, equity-relevant grid data is critical for assessing disparities and enabling informed decision-making. Ohio utilities and regulators should:

- ***Enhance geospatial identification of grid infrastructure*** by requiring utilities to publicly provide:
    - **Reporting on the primary census tract(s) served by each circuit**, including the percentage of the circuit that runs through each tract.
    - **Reporting on the substation transformers' primary service areas by census tract** to allow for a more precise analysis of infrastructure disparities.
- ***Adopt and publicize more customer-centric reliability metrics*** to better capture disparities in service quality. In addition to standard reliability indices, utilities should:
    - **Report Customers Experiencing Long Interruption Durations (CELID)**, as has been implemented and publicized in Minnesota, to provide a clearer picture of how outages impact disadvantaged communities.
    - **Report Customers Experiencing Multiple Interruptions (CEMI)**, as has been implemented and publicized in Minnesota, to provide a clearer picture of how outages impact disadvantaged communities.
- ***Increase transparency in grid access and interconnection data*** by requiring utilities to:
    - **Publish sub-circuit hosting capacity maps** to help identify disparities in grid access for clean energy deployment.
    - **Publish interconnection queue data with geographic identifiers** to assess whether marginalized communities face additional barriers to clean energy access.

    - **Publish assigned grid upgrade costs with geographic identifiers** to highlight any inequities in how infrastructure upgrade costs are distributed across different communities.

- ***Report and publicize energy insecurity and poverty data with geographic identifiers*** to illuminate disparities in energy affordability by requiring utilities to:
    - **Disaggregate and publicize disconnection rates** by circuit, substation, and census tract to allow for more detailed assessments of energy burden and shutoff disparities.
    - **Perform comprehensive energy burden assessments** that identify both visible and hidden burdened communities.

Integrate Justice-Focused Metrics Into Grid Planning – To ensure infrastructure investments actively address historical inequities, Ohio utilities and regulators must integrate justice-focused metrics into grid planning. Specifically, they should:

- ***Prioritize investments in historically underserved communities*** by incorporating Climate and Economic Justice Screening Tool (CEJST) disadvantaged status (or similar state-wide demographic resources) and historical redlining data into:
    - **Grid infrastructure investment decision-making**, ensuring that historically marginalized communities are prioritized for upgrades.
    - **Hosting capacity maps**, so that communities with documented inequities in grid conditions can be better positioned for proactive investments.
- ***Mandate independent third-party equity assessments*** as part of all Ohio utilities' grid planning processes to:
    - **Conduct comprehensive statistical analysis** of infrastructure conditions, reliability, and investment patterns to uncover systemic inequities.

While data limitations posed challenges for this study, our findings establish a critical framework for assessing energy equity in Ohio's electric grid. These recommendations offer a path forward for enhancing transparency, embedding equity into grid planning, and addressing long-standing disparities in infrastructure investment and service quality.

Moving forward, regulatory reforms, stronger data access requirements, and proactive investment strategies will be essential to achieving a more just, resilient, and equitable energy system in Ohio. Continued advocacy, stakeholder engagement, and interdisciplinary research will be necessary to ensure that the needs of historically marginalized communities are not only considered but actively prioritized in the energy transition.

J. Appendix

Appendix A – Population Dynamics: Descriptive Statistics Summary Table

Descriptive Statistics -- Demographics

DEMO-BLACK	ALL	HOLC	HOLC-RED	HOLC-NRED	CEJST-DIS	CEJST-NDIS
Mean	0.19	0.38	0.59	0.33	0.39	0.06
Standard Error	0.01	0.02	0.04	0.02	0.02	0.01
Median	0.05	0.29	0.61	0.21	0.29	0.01
Mode	0.00	0.06	0.93	0.06	0.00	0.00
Standard Deviation	0.28	0.32	0.31	0.30	0.32	0.13
Sample Variance	0.08	0.10	0.10	0.09	0.10	0.02
Kurtosis	1.45	-1.14	-1.30	-0.69	-1.13	14.10
Skewness	1.63	0.54	-0.32	0.80	0.55	3.60
Range	0.99	0.99	0.97	0.99	0.99	0.89
Minimum	0.00	0.00	0.01	0.00	0.00	0.00
Maximum	0.99	0.99	0.98	0.99	0.99	0.89
Sum	153.97	125.16	39.02	86.14	123.08	30.89
Count	820	330	66	264	317	503
DEMO-WHITE	ALL	HOLC	HOLC-RED	HOLC-NRED	CEJST-DIS	CEJST-NDIS
Mean	0.66	0.49	0.27	0.55	0.48	0.86
Standard Error	0.01	0.02	0.03	0.02	0.02	0.01
Median	0.81	0.50	0.24	0.58	0.49	0.90
Mode	0.95	0.00	0.00	0.79	0.00	0.95
Standard Deviation	0.31	0.30	0.24	0.29	0.30	0.15
Sample Variance	0.10	0.09	0.06	0.09	0.09	0.02
Kurtosis	-0.59	-1.27	-0.16	-1.07	-1.21	8.33
Skewness	-0.89	-0.14	0.76	-0.39	-0.07	-2.65
Range	0.99	0.97	0.91	0.97	0.98	0.95
Minimum	0.00	0.00	0.00	0.00	0.00	0.04
Maximum	0.99	0.97	0.91	0.97	0.98	0.99
Sum	334.12	162.51	17.63	144.88	151.56	431.76
Count	503	330	66	264	317	503
DEMO-BIPOC	ALL	HOLC	HOLC-RED	HOLC-NRED	CEJST-DIS	CEJST-NDIS
Mean	0.29	0.51	0.73	0.45	0.52	0.14
Standard Error	0.01	0.02	0.03	0.02	0.02	0.01
Median	0.16	0.51	0.76	0.42	0.51	0.10
Mode	0.05	1.00	1.00	0.21	1.00	0.05
Standard Deviation	0.29	0.30	0.24	0.29	0.30	0.15
Sample Variance	0.08	0.09	0.06	0.09	0.09	0.02
Kurtosis	0.13	-1.27	-0.16	-1.07	-1.21	8.33
Skewness	1.17	0.14	-0.76	0.39	0.07	2.65
Range	0.99	0.97	0.91	0.97	0.98	0.95
Minimum	0.01	0.03	0.09	0.03	0.02	0.01
Maximum	1.00	1.00	1.00	1.00	1.00	0.96
Sum	236.68	167.49	48.37	119.12	165.44	71.24
Count	820	330	66	264	317	503

Appendix B – R-Code Examples

Data Import & Cleanup – Importing, labeling, and cleansing data

```
# Load necessary libraries
library(readxl)
# Import data from an Excel file
data_all <- read_excel("path_to_file.xlsx")
# Subset the data for specific analysis criteria
data_subset <- data_all[data_all$Category_Column == TRUE | data_all$Category_Column == 1, ]
# View the dataset
View(data_all)
# Check for missing or problematic values in key variables
sum(is.na(data_all$Variable1))  # Count NAs
sum(is.nan(data_all$Variable1)) # Count NaNs
sum(is.infinite(data_all$Variable1))  # Count Inf values
# Remove rows with missing values for specific analysis
data_clean <- data_all[!(data_all$Variable1 == "NA") & !is.na(data_all$Variable1) & !(data_all$Variable2 == "NA") &
!is.na(data_all$Variable2), ]
# Ensure categorical variables are properly formatted
data_all$Categorical_Variable <- as.factor(data_all$Categorical_Variable)
```

Chi-Square Tests – Performing an inferential chi-square statistical test

```
# Create a contingency table for two categorical variables
contingency_table <- table(data$Category1, data$Category2)
# Print the contingency table
print("Contingency Table:")
print(contingency_table)
# Perform Chi-Square test
chi_square_result <- chisq.test(contingency_table)
# Print test results
print("Chi-Square Test Result:")
print(chi_square_result)
```

Regression – Performing an inferential linear regression statistical analysis

```
# Fit a linear regression model with a continuous dependent variable and a categorical independent variable
linear_model <- lm(Continuous_Variable ~ Categorical_Variable, data = data_clean)
# View summary of the regression model
summary(linear_model)
# Perform a second regression on another dependent variable
linear_model_2 <- lm(Continuous_Variable2 ~ Categorical_Variable, data = data_clean)
# Display regression results using stargazer (for LaTeX output)
library(stargazer)
stargazer(linear_model, linear_model_2, type = "text")
```

Logistic Regression – Performing an inferential multinomial logistic regression statistical analysis

```
# Load necessary libraries
library(nnet)
library(reshape2)
library(ggplot2)
# Fit a multinomial logistic regression model
multi_logit_model <- multinom(Categorical_Variable ~ Independent_Variable, data = data_all)
# View summary of the model
summary(multi_logit_model)
# Calculate p-values for coefficients
```

```r
z_values <- summary(multi_logit_model)$coefficients / summary(multi_logit_model)$standard.errors
p_values <- (1 - pnorm(abs(z_values), 0, 1)) * 2
print("P-values for model coefficients:")
print(p_values)
# Predict probabilities for each category
predicted_probabilities <- predict(multi_logit_model, type = "probs")
print("Predicted Probabilities (first few rows):")
head(predicted_probabilities)
# Create a data frame for predicted probabilities based on a specific independent variable
new_data <- data.frame(Independent_Variable = c(0, 1))  # Example: Binary categories
predicted <- predict(multi_logit_model, newdata = new_data, type = "probs")
predicted <- data.frame(predicted)
predicted$Category_Label <- c("No", "Yes")
# Reshape data for visualization
predicted_long <- melt(predicted, id.vars = "Category_Label", variable.name = "Category", value.name = "Probability")
# Plot predicted probabilities
ggplot(predicted_long, aes(x = Category_Label, y = Probability, fill = Category)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Category Label", y = "Predicted Probability", fill = "Category") +
  theme_minimal()
# Check model fit with AIC
aic_value <- AIC(multi_logit_model)
print(paste("AIC of the model:", aic_value))
# Perform likelihood ratio test
library(lmtest)
lr_test <- lrtest(multi_logit_model)
print("Likelihood Ratio Test:")
print(lr_test)
```