Paper SA03

Detroit, Michigan: A Case Study for DEI via Decentralized Clinical Trials (DCTs)

Mark Rothe, Janssen, New Jersey, United States

ABSTRACT

In April 2022, the FDA issued a draft guidance entitled *Diversity Plans to Improve Enrollment of Participants From Underrepresented Racial and Ethnic Populations in Clinical Trials* (FDA 13-April-2022)—which recommends Sponsors submit enrollment goals to the FDA as part of Race and Ethnicity Diversity Plans. Furthermore, the FDA issued a draft guidance in May 2023 entitled *Decentralized Clinical Trials (DCTs) for Drugs, Biological Products, and Devices: Guidance for Industry, Investigators, and Other Stakeholders* (FDA 01-May-2023). This document advocates for the implementation of DCTs, when appropriate, where all or some study activities occur at patient homes, facilities of local healthcare providers (HCPs), etc.

Taken together these guidance documents are a catalyst for Sponsors to proactively locate diverse and marginalized patient populations while simultaneously connecting them with improved opportunities for study enrollment. Methodologies regarding how to locate patient populations, Sponsor sites, local HCPs, and community organizations will be presented in a case study of the Detroit, Michigan, *immediate* metropolitan area.

INTRODUCTION

Detecting where diverse and marginalized patient populations reside as well as where local HCPs practice medicine can be performed using geographic information systems (GIS). Sponsors have the option to purchase GIS enterprise software or employ opensource software that has GIS capabilities such as R or Python (Wikipedia-GIS). R is a prime candidate, because it contains a rich collection of spatial and spatiotemporal user developed R packages—as illustrated by the CRAN *Spatial* and *Spatiotemporal* Task Views (Bivand and Nowosad; Pebesma and Bivand). Also, many Sponsors are either transitioning from SAS to R for clinical trial reporting or augmenting SAS with R workflows—presenting R as a well-positioned tool for clinical / statistical programmers, data scientists, and statisticians to leverage for a variety of purposes. This paper explores how R can function as a GIS software tool and use open access data to identify areas where diverse and marginalized patients reside as well as nearby potential DCT sites in the immediate Detroit metropolitan area.

METHODS & RESULTS

MACRO Level View of the US State of Michigan: *Economic Inequality*

The City of Detroit is located in the US State of Michigan. *Figure 1* takes a “40,000 foot view” of the United States relative to income inequality as measured by the Ortega $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ parameters—allowing for a bivariate assessment of marginalization (Blesch 25-June-2022). The **Gini** coefficient has historically been a widely used measure of income inequality. However, Blesch, Hauser, and Jachimowicz conclude that the Ortega *two-parameter* model “provides the best overall fit to capture the information contained in the income distributions across U.S. counties. … The first Ortega parameter $\boldsymbol{\alpha}$ captures inequality with a more pronounced focus on concentrations at the *bottom of the income distribution*, while the second Ortega parameter $\boldsymbol{\gamma}$ reflects an emphasis toward inequality concentrated at the very *top of the income distribution*” (Blesch, Hauser, and Jachimowicz 01-November-2022). *Figure 2* illustrates that the State of Michigan has moderate $\boldsymbol{\alpha}$ inequality ($\hat{\boldsymbol{\alpha}}_{\mathbf{MI}} = \mathbf{0.611}$) and minimally high $\boldsymbol{\gamma}$ inequality ($\hat{\boldsymbol{\gamma}}_{\mathbf{MI}} = \mathbf{0.58}$) relative to other US States ($\hat{\boldsymbol{\alpha}}_{\mathbf{US_{median}}} = \mathbf{0.613}$, $\hat{\boldsymbol{\gamma}}_{\mathbf{US_{median}}} = \mathbf{0.55}$).

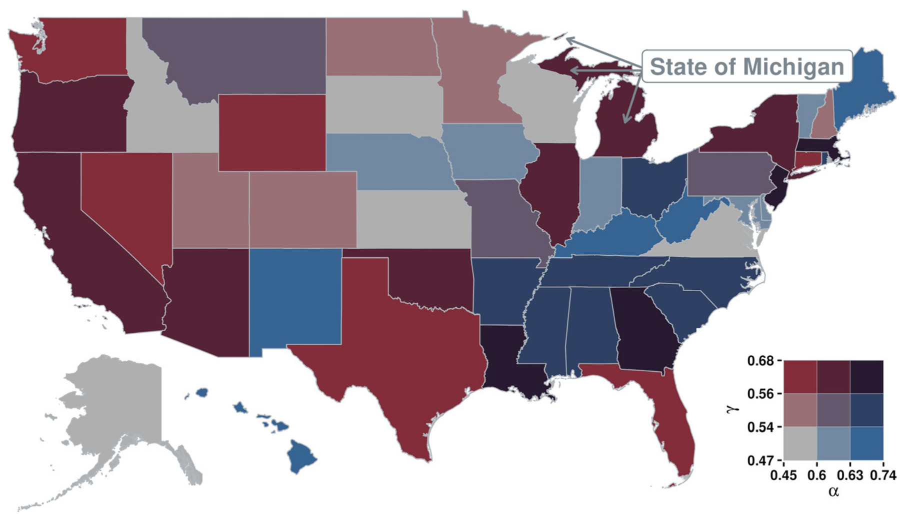


Figure 1: Bivariate map of the Ortega $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ economic inequality parameter estimates by US state. The color of each state corresponds to a combination of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ values—as illustrated in the map legend. [Data sources: Blesch (25-June-2022), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: biscale, ggplot2, ggrepel, patchwork.]

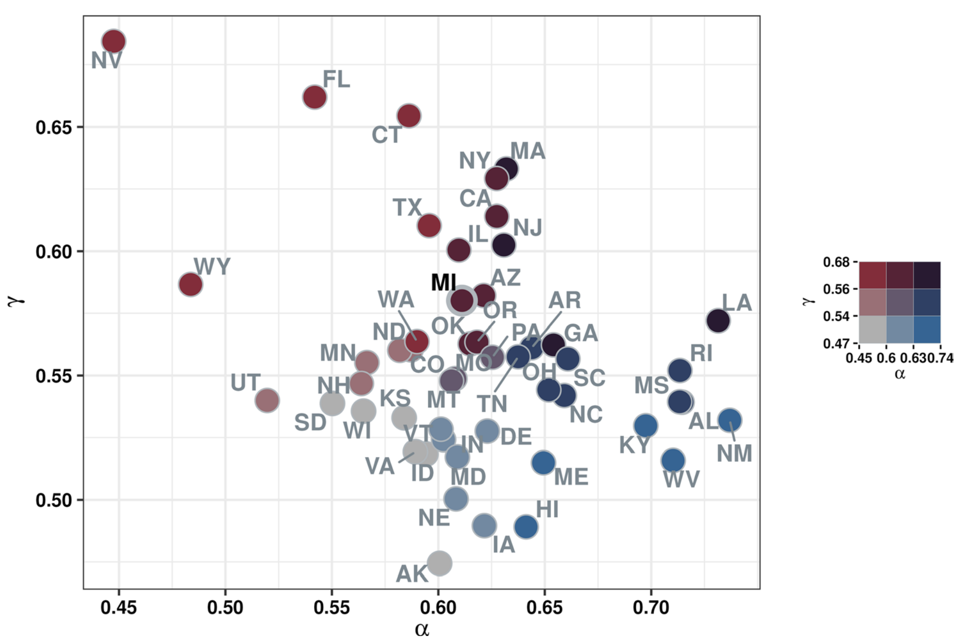


*Figure 2: Bivariate plot of the Ortega $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ economic inequality parameter estimates by US state. The color of each state (data point) corresponds to a combination of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ values—as illustrated in the plot legend. The State of Michigan data label (**MI**) is **black**, and each state is labelled by its two letter abbreviation. [Data source: Blesch (25-June-2022); Salient R plotting / mapping packages: biscale, ggplot2, ggrepel, patchwork.]*

Figure 3 zooms into a "10,000 foot view" of Michigan relative to the Ortega $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ parameter estimates for its **83** counties (Blesch 25-June-2022). The **City of Detroit** is located in **Wayne County**, one of the **5 worst counties** with respect to both $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ measures of income inequality—as can be seen by the 5 counties in the lower peninsula of Michigan which are in both the highest $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ tri-stratification levels. ($\hat{\boldsymbol{\alpha}}_{\mathbf{Wayne}} = \mathbf{0.726}$, $\hat{\boldsymbol{\gamma}}_{\mathbf{Wayne}} = \mathbf{0.621}$; $\hat{\boldsymbol{\alpha}}_{\mathbf{MI_{median}}} = \mathbf{0.58}$, $\hat{\boldsymbol{\gamma}}_{\mathbf{MI_{median}}} = \mathbf{0.537}$). These Ortega estimates derived by Blesch, Hauser, and Jachimowicz correlate well with the 2023 update to the ***County Health Rankings & Roadmaps (CHR&R)*** data created and maintained by the *University of Wisconsin Population Health Institute*—where Wayne County is the **worst** ranked county in Michigan with respect to **Health Outcomes** (Jeden 19-April-2023).

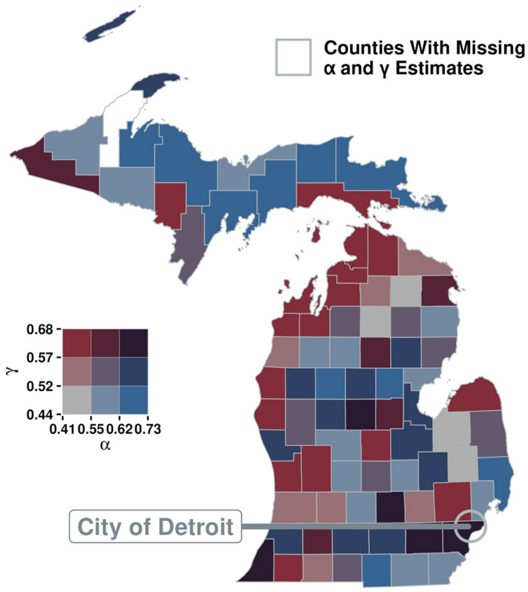


Figure 3: Bivariate map of Ortega **α** *and* **γ** *economic inequality parameter estimates by Michigan county. The color of each county corresponds to a combination of* **α** *and* **γ** *values—as illustrated in the map legend. [Data source: Blesch (25-June-2022), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: biscale, ggnewscale, ggplot2, ggrepel, patchwork.]*

The counties of **Macomb**, **Oakland**, and **Wayne** comprise the *immediate* Detroit metropolitan area. As per *Figure 4*, each county is in a different tri-**α**-stratification level from *"most equal"* (Oakland) to *"least equal"* (Wayne); while Oakland and Wayne are in the *"least equal"* tri-**γ**-stratification level, and Macomb is in the *"most equal"* **γ** grouping. Taken together, these 3 counties represent a wide spectrum of economic inequality within the US State of Michigan.

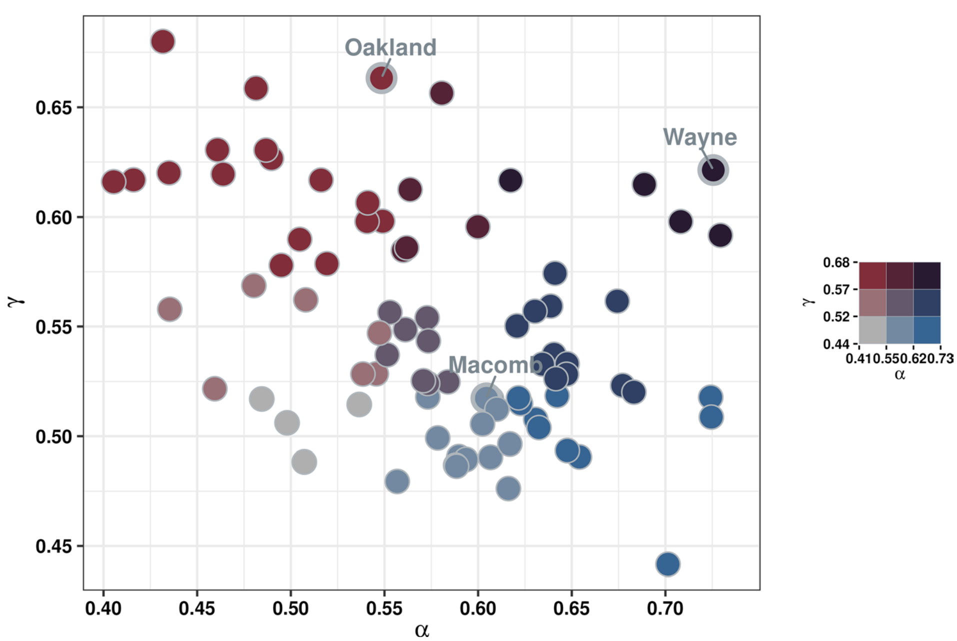


Figure 4: Bivariate plot of the Ortega **α** *and* **γ** *economic inequality parameter estimates by Michigan county. The color of each county (data point) corresponds to a combination of* **α** *and* **γ** *values—as illustrated in the plot legend.* ***Macomb***, ***Oakland***, *and* ***Wayne*** *counties are labelled. [Data source: Blesch (25-June-2022); Salient R plotting / mapping packages: biscale, ggplot2, ggrepel, patchwork.]*

MACRO* Level View of the US State of Michigan: *Health Outcomes & Factors

Figure 5 maps and *Figure 6* respectively plots the 2023 **Health Outcomes** and **Health Factors** rankings for the 83 Michigan counties downloaded from ***County Health Rankings & Roadmaps (CHR&R)*** (CHR&R 2023, Version 2). Both the map and plot illustrate that Oakland, Macomb, and Wayne Counties are each members of the top, middle, and lower performing Michigan counties, respectively.

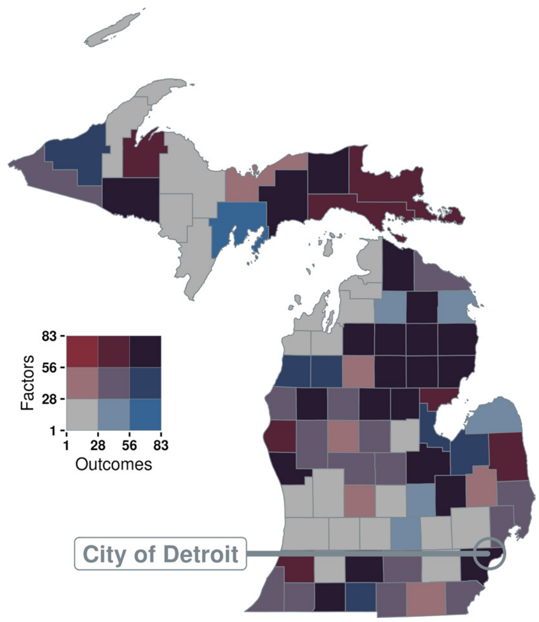


*Figure 5: Bivariate map of the **Health Outcomes** and **Health Factors** rankings by Michigan county. The color of each county corresponds to a combination of **Health Outcomes** and **Health Factors** rankings—as illustrated in the map legend. [Data sources: CHR&R (2023, Version 2), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: biscale, ggplot2, ggrepel, patchwork.]*

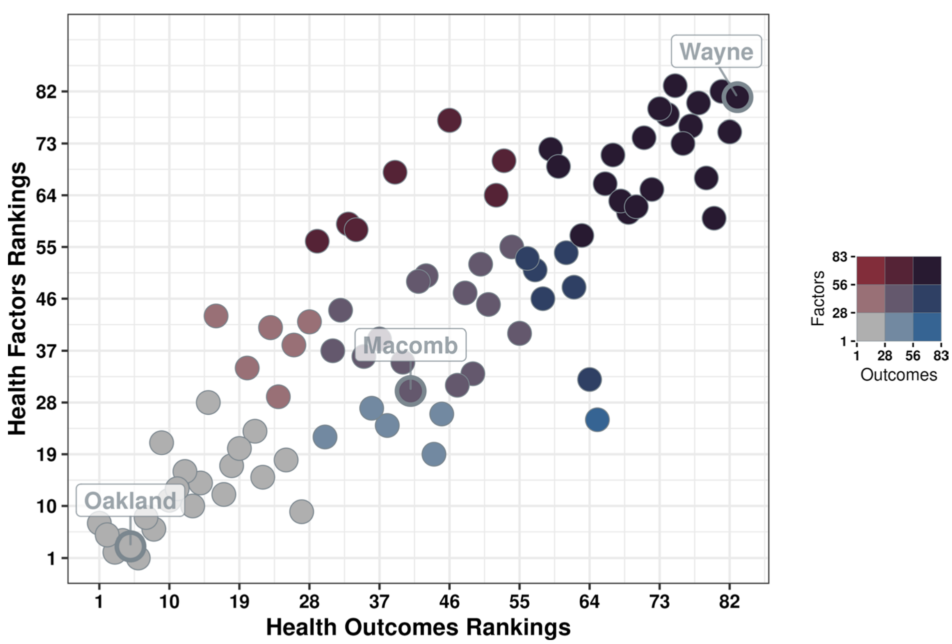


*Figure 6: Bivariate plot of the **Health Outcomes** and **Health Factors** rankings by Michigan county. The color of each county (data point) corresponds to a combination of **Health Outcomes** and **Health Factors** rankings—as illustrated in the plot legend.*

Macomb*, *Oakland*, *and* *Wayne* *counties are labelled. [Data source: CHR&R (2023, Version 2); Salient R plotting / mapping packages: biscale, ggplot2, ggrepel, patchwork.]

MACRO* Level View of the US State of Michigan: *Analyzing Inequality & Health Data Together

Viewing the **α** and **γ** estimates and the 2023 **Health Outcomes** and **Health Factors** rankings side-by-side paints a more complete picture of the economic and health inequities across Michigan's 83 counties. Pooling these different data sources creates the opportunity for similar counties to be more readily grouped together using unsupervised machine learning. The **α** and **γ** estimates were combined with the underlying **subrankings** that formed the basis for the overall outcomes and factors rankings. Below is a breakdown of how the University of Wisconsin Population Health Institute (UWPHI) weighted the subrankings to derive the overall rankings (CHR&R 2023, Version 2):

- *Equally* weighted subranking combination to derive the **health outcomes** ranking:
  - **50%** *Length of Life*
  - **50%** *Quality of Life*
- *Differentially* weighted subranking combination to derive the **health factors** ranking:
  - **30%** *Health Behaviors*
  - **20%** *Clinical Care*
  - **40%** *Social and Economic Factors*
  - **10%** *Physical Environment*

After the data was pooled, a measure of dissimilarity known as the **Gower distance** was derived across all 83 counties—quantifying the dissimilarity between each pair of counties. The same weights employed by UWPHI were used when combining the subrankings during the computation of the Gower distance, and the features **α**, **γ**, combined outcome subrankings, and combined factor subrankings were each equally weighted:

- The Gower distance quantifies the difference between rows of a data matrix (*e.g., counties*) by taking into account the columns of the data matrix (*e.g., Ortega estimates and county subrankings*) when the column variables are of **mixed** data type (in this case, **numeric** and **ordinal** data, respectively) (Maechler et al. v2.1.4).
- Counties with missing **α** or **γ** estimates (*only a **single** county had missing values, as per Figure 3*) were *imputed* using the respective **median** values.
- The **α** and **γ** columns of the pooled data matrix were first **centered** (by subtracting the column's mean from each value) and then **scaled** (by dividing each centered value by the column's standard deviation). The resulting centered and scaled **α** and **γ** columns each had a mean of 0 and standard deviation of 1—preparing them for the future application of machine learning algorithms. One reason for performing these transformations is to ensure that data of different magnitudes are aligned for better comparability.
- The Gower distance was computed using the **daisy** function from the **cluster** R package (Maechler et al. v2.1.4).

The resulting **Gower distance matrix** was then fed into a **self-organizing map (SOM)** using the **SOMbrero** R package (Vialaneix et al. v1.4-1). Self-organizing maps are a form of unsupervised machine learning that help to discern patterns in data by reducing dimensions. Results from the SOM were then clustered using the **partitioning around medoids (PAM)** algorithm via the **pamk** function from the **fpc** R package (Hennig v2.2-10) (to determine the number of clusters) as well as the **pam** function from the **cluster** R package (Maechler et al. v2.1.4). Applying SOM and clustering resulted in 4 different groupings of the 83 counties. To illustrate how each of these 4 clusters is characterized, *Figure 7* was generated as follows and displays 8 heatmaps—1 for each Ortega estimate and CHR&R subranking:

- **X-axis**: The **cluster** each county was assigned via SOM dimensionality reduction and PAM clustering.
- **Y-axis**: The **quartile** corresponding to each county's Ortega estimates or CHR&R subrankings. The subrankings' quartiles assigned by UWPHI during the ranking derivations were used, while the **α** and **γ** estimates were assigned quartiles via the **ntile** function from the **dplyr** R package (Wickham et al. v1.1.2).
- **Color of Heatmap Cell**: The **number of counties** in the corresponding cluster and quartile.
- **Text Label of Heatmap Cell**: The heatmap cell that contains Macomb, Oakland, or Wayne County is **labelled**:
  - Each county never being located in the same heatmap cell is a reflection of the underlying data.
  - The incidence pattern of counties by heatmap cell suggests 3 conclusions:
    - **Cluster No. 4** is the **least** economically equitable (*relative to the **lower** end of income distribution*) and **most** "health challenged" group.
    - **Cluster No. 1** is the **least** economically equitable (*relative to the **higher** end of income distribution*) and **least** "health challenged" group.

- **Cluster Nos. 2 & 3** contain counties in the middle of the continuum as defined by Cluster Nos. 1 and 4.
- The immediate Detroit metropolitan area contains a county in each cluster except Cluster No. 2:
  - *Cluster No. **1***: **Oakland** County
  - *Cluster No. **3***: **Macomb** County
  - *Cluster No. **4***: **Wayne** County

Figure 8 maps cluster membership by Michigan county as well as zooms in to display where the City of Detroit exists in relation to the 3 counties comprising the immediate Detroit metropolitan area.

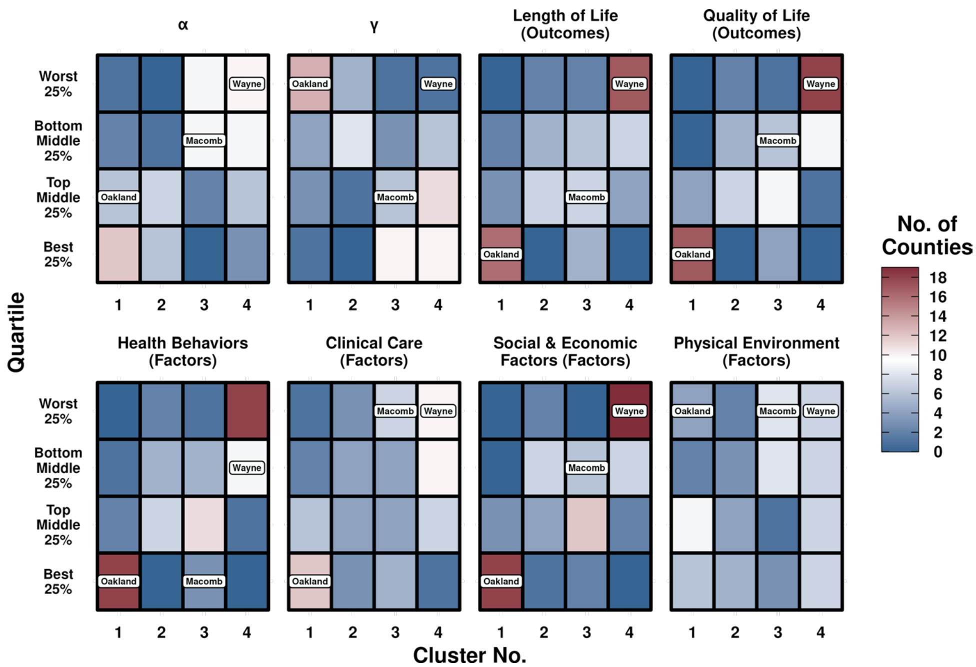


*Figure 7: Heatmaps displaying the number of counties by cluster and quartile for the Ortega **α** and **γ** economic inequality parameter estimates as well as County Health Rankings & Roadmaps subrankings that contribute to **Health Outcomes** and **Health Factors** rankings. The location of **Macomb**, **Oakland**, and **Wayne** text labels indicate the corresponding cluster and quartile for each county. [Data sources: Blesch (25-June-2022), CHR&R (2023, Version 2); Salient R plotting / mapping packages: ggplot2.]*

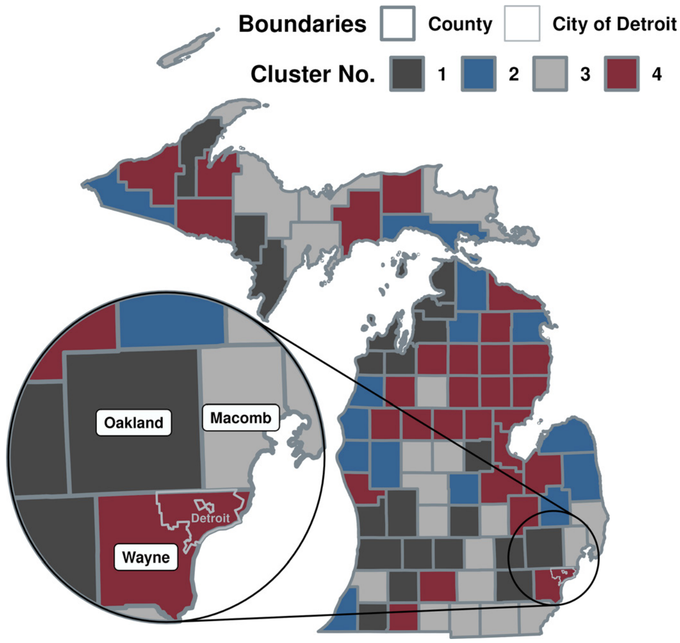


Figure 8: Map of Michigan counties—clustered with respect to their Ortega **α** *and* **γ** *economic inequality parameter estimates as well as County Health Rankings & Roadmaps subrankings that contribute to* ***Health Outcomes*** *and* ***Health Factors*** *rankings. [Data sources: Blesch (25-June-2022), CHR&R (2023, Version 2), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: ggmapinset, ggplot2, ggpubr, patchwork.]*

MICRO **Level View of the** ***Immediate*** **Detroit Metropolitan Area:** ***Analyzing Inequality & Health Data Together***

The prior Gower, SOM, and cluster analyses illustrate how different the 3 counties appear when taking a "10,000 foot view" of Michigan. *Figure 9* depicts a "5,000 foot view" of the immediate Detroit metropolitan area. When zooming in to this height, smaller geographic areas become discernible, allowing for the use of US census tract data.

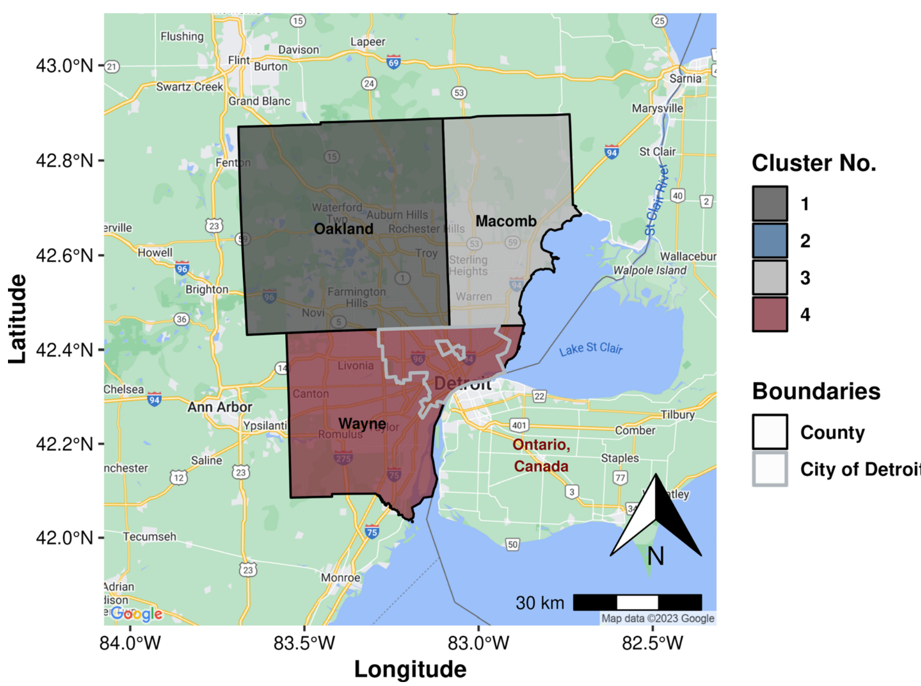


*Figure 9: Google roadmap of the immediate Detroit metropolitan area: The color of the Michigan counties **Macomb**, **Oakland**, and **Wayne** corresponds with the prior Gower, SOM, and cluster analysis results. [Data sources: Blesch (25-June-2022), CHR&R (2023, Version 2), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: ggmap, ggplot2, ggspatial.]*

Unfortunately, Ortega **α** and **γ** economic inequality parameter estimates are unavailable at the census tract level. However, the US Census Bureau's American Community Survey 2011-2015 5-Year Data Release does contain Gini coefficient estimates for these geographic areas. *Figure 10* displays a map and corresponding histogram of Gini estimates for the 1165 census tracts that comprise Macomb, Oakland, and Wayne counties. Each census tract is grouped using the **Jenks natural breaks classification method** via the R package **tmap** (Wikipedia-Jenks; Tennekes v3.3-3). As per *Figure 10*, economic inequality appears to be relatively concentrated within the City of Detroit with some outside the city limits, especially to the northwest in Oakland County. (*Note: As with the Ortega **α** and **γ** parameters, larger **Gini** values indicate greater inequality (Wikipedia-GINI)*.)

More granular data than the county level is not available from County Health Rankings & Roadmaps, but the Centers for Disease Control and Prevention (CDC) maintains the **PLACES** database which "provides model-based, population-level analysis and community estimates of health measures to all counties, places (incorporated and census designated places), census tracts, and ZIP Code Tabulation Areas (ZCTAs) across the United States" (CDC PLACES). The CDC makes this data available through its API powered by Socrata and readily accessible in R via the package **RSocrata** (CDC API 2023; Devlin et al. v1.7.12-4). *Figure 10* displays **2** of the **29** CDC health measures, namely **diabetes** and **asthma** *crude prevalence*, in addition to a map of **life expectancy at birth** estimates from the CDC's U.S. Small-Area Life Expectancy Estimates Project (USALEEP) (CDC LEEP 2023). The City of Detroit appears to be a prominent location for **higher** estimated diabetes and asthma crude prevalence and **lower** life expectancy estimates relative to many of the surrounding census tracts in the immediate Detroit metropolitan area.

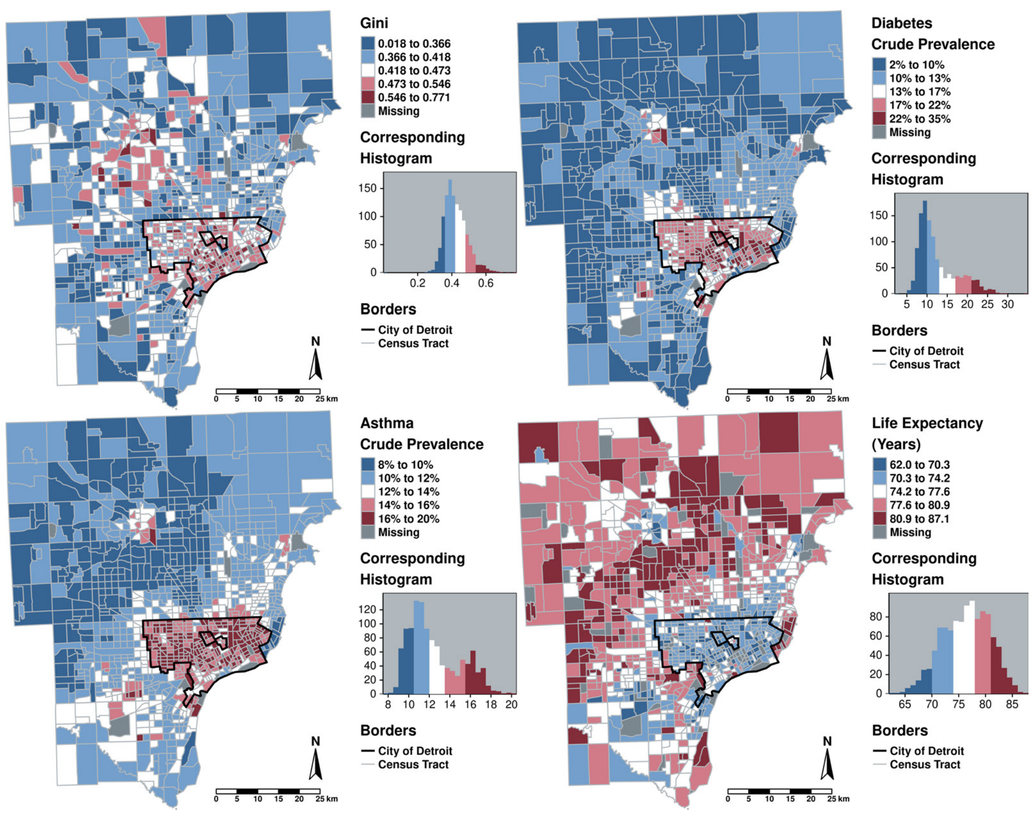


*Figure 10: Maps & corresponding histograms of the 1165 census tracts within the immediate Detroit metropolitan area—colored by **Gini**, **diabetes**, **asthma**, and **life expectancy at birth** estimates via the Jenks natural breaks classification method (Wikipedia-Jenks; Tennekes v3.3-3). [Data sources: USCB via K. Walker and Herman (v1.4.1), CDC API (2023) via Devlin et al. (v1.7.12-4), CDC LEEP (2023), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap.]*

The **29** CDC health measures (**30** in total—as 1 measure is treated separately for females & males), **Gini**, and **life expectancy** data were pooled together for all 1165 census tracts in preparation for grouping similar tracts via unsupervised machine learning—resulting in a $\mathbf{1165} \times \mathbf{32}$ data matrix:

- **14** of the census tracts had **≥ 31** missing values.
- **1** tract had **6** missing values.
- **66** tracts had **1** missing value.
- **1084** census tracts had *no* missing values.

Tracts with **> 6** missing values were dropped from analysis (14 census tracts), and tracts with **≤ 6** missing values underwent imputations via predictive mean matching using the **mice** R package (5 rounds of imputation were performed with the median of the 5 used as the value for imputation) (van Buuren and Groothuis-Oudshoorn v3.16.0). An additional variable was added to the now smaller $\mathbf{1151} \times \mathbf{32}$ data matrix by applying the **Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS)** method—ordering the 1151 census tracts from **most equitable / most healthy** (smaller TOPSIS scores) to **least equitable / least healthy** (larger TOPSIS scores) (Wikipedia-TOPSIS; Mahmoud Mosalman Yazdi v1.0). Addition of this engineered feature resulted in a $\mathbf{1151} \times \mathbf{33}$ data matrix. A map of the TOPSIS scores for each of the 1151 census tracts was generated and is displayed in *Figure 11*: It appears that many of the **least equitable / least healthy** tracts are within the City of Detroit with some smaller concentrations outside the city limits.

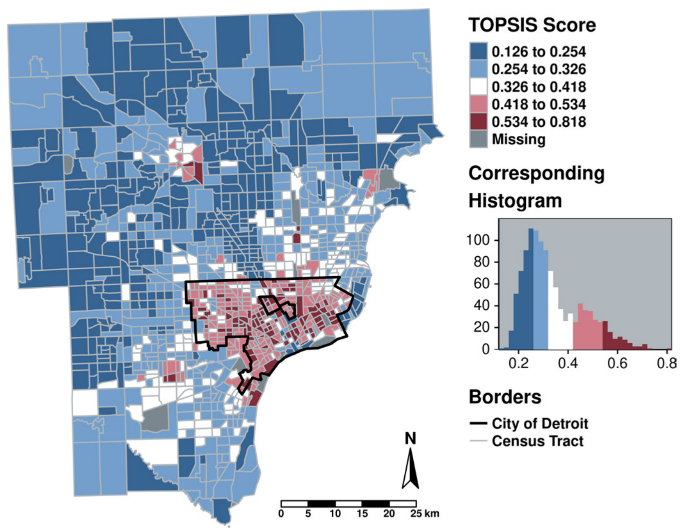


*Figure 11: Map & corresponding histogram of the 1165 census tracts within the immediate Detroit metropolitan area—colored by **TOPSIS** scores via the Jenks natural breaks classification method (Wikipedia-Jenks; Tennekes v3.3-3). [Data sources: USCB via K. Walker and Herman (v1.4.1), CDC API (2023) via Devlin et al. (v1.7.12-4), CDC LEEP (2023), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap.]*

Similar to the previous unsupervised machine learning applied to the county level data, subsequent dimensionality reduction was applied to the $\mathbf{1151} \times \mathbf{33}$ data matrix after the columns were first centered and scaled. Non-linear dimensionality reduction from **33** to **2** dimensions via **Uniform Manifold Approximation and Projection (UMAP)** was performed using the **umap** R package as an attempt to better visualize possible data patterns (Wikipedia-UMAP; Konopka v0.2.10.0). As per *Figure 12*, **3** distinct groups of census tracts can be seen when the resulting two-dimensional census data is plotted. The partitioning around medoids (PAM) algorithm via the **pamk** function from the **fpc** R package confirmed this clustering (Hennig v2.2-10).

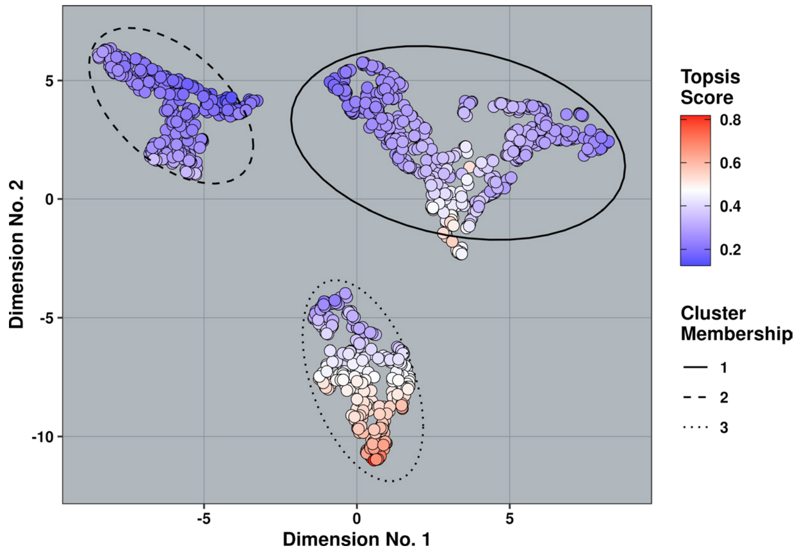


Figure 12: Two-dimensional plot of the 1151 of the 1165 census tracts within the immediate Detroit metropolitan area—colored by ***TOPSIS*** *score, clustered via* ***PAM****, and reduced from the original 33 dimensions via* ***UMAP****. [Data sources: USCB via K. Walker and Herman (v1.4.1), CDC API (2023) via Devlin et al. (v1.7.12-4), CDC LEEP (2023); Salient R plotting / mapping packages: ggplot2.]*

Figure 13 displays a map of the 1165 census tracts colored by cluster membership—including the 14 tracts dropped from the analysis (**Cluster Membership** set to **Missing**):

- The previous Gower, SOM, and PAM county-level analysis resulted in Macomb, Oakland, and Wayne counties being assigned to 3 different clusters.
- The UMAP and PAM tract-level analysis resulted in the majority of Oakland County being assigned to one cluster (Cluster No. 2), almost all of Macomb County and most non-Detroit Wayne tracts being assigned to a second cluster (Cluster No. 1), and almost the entirety of the City of Detroit categorized as a third cluster (Cluster No. 3).
- *Figure 14* illustrates that Cluster No. 3 is predominantly, non-White, while Cluster Nos. 1 and 2 are largely White. Furthermore, Cluster No. 3 appears to be **less equitable** (*visibly* **higher** Gini) and **less healthy** (*visibly* **higher** asthma, coronary heart disease, stroke, diabetes, blood pressure, smoking, and obesity crude prevalence & *visibly* **lower** life expectancy estimates) relative to the other 2 clusters. Formal hypothesis tests could be conducted for more definitive and statistically defensible conclusions. However, **visible** comparisons of the 3 clusters in *Figure 14* should be sufficient to deduce that Cluster No. 3 is the most marginalized of the 3 and would likely benefit the greatest from increased awareness of and access to DCTs.
- These *same* **visible** conclusions can be drawn from *Figure 15*—where the color of each census tract corresponds to the majority race / ethnicity in each tract. Additionally, it is interesting to observe that sometimes the smaller-in-number census tracts in clusters where the larger-in-number census tracts have a different majority race / ethnicity often visibly mirror their counterparts in other clusters (e.g., The obesity crude prevalence of the few majority Black census tracts in Cluster No. 1 appears to be in line with the obesity crude prevalence of the majority Black census tracts in Cluster No. 3.)

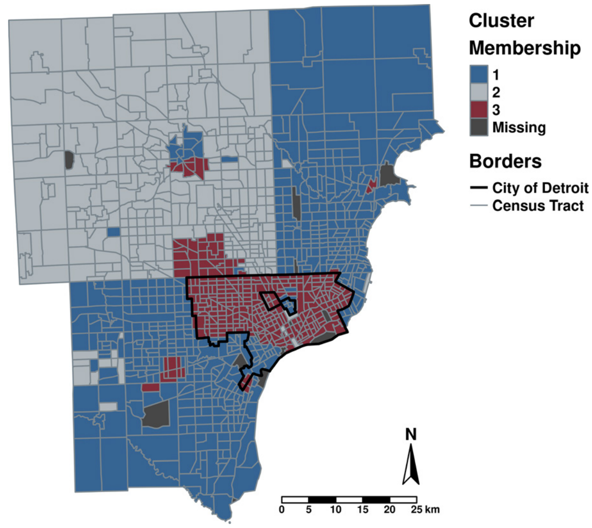


*Figure 13: Map of the 1165 census tracts within the immediate Detroit metropolitan area—colored by **Cluster Membership**. [Data sources: USCB via K. Walker and Herman (v1.4.1), CDC API (2023) via Devlin et al. (v1.7.12-4), CDC LEEP (2023), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap.]*

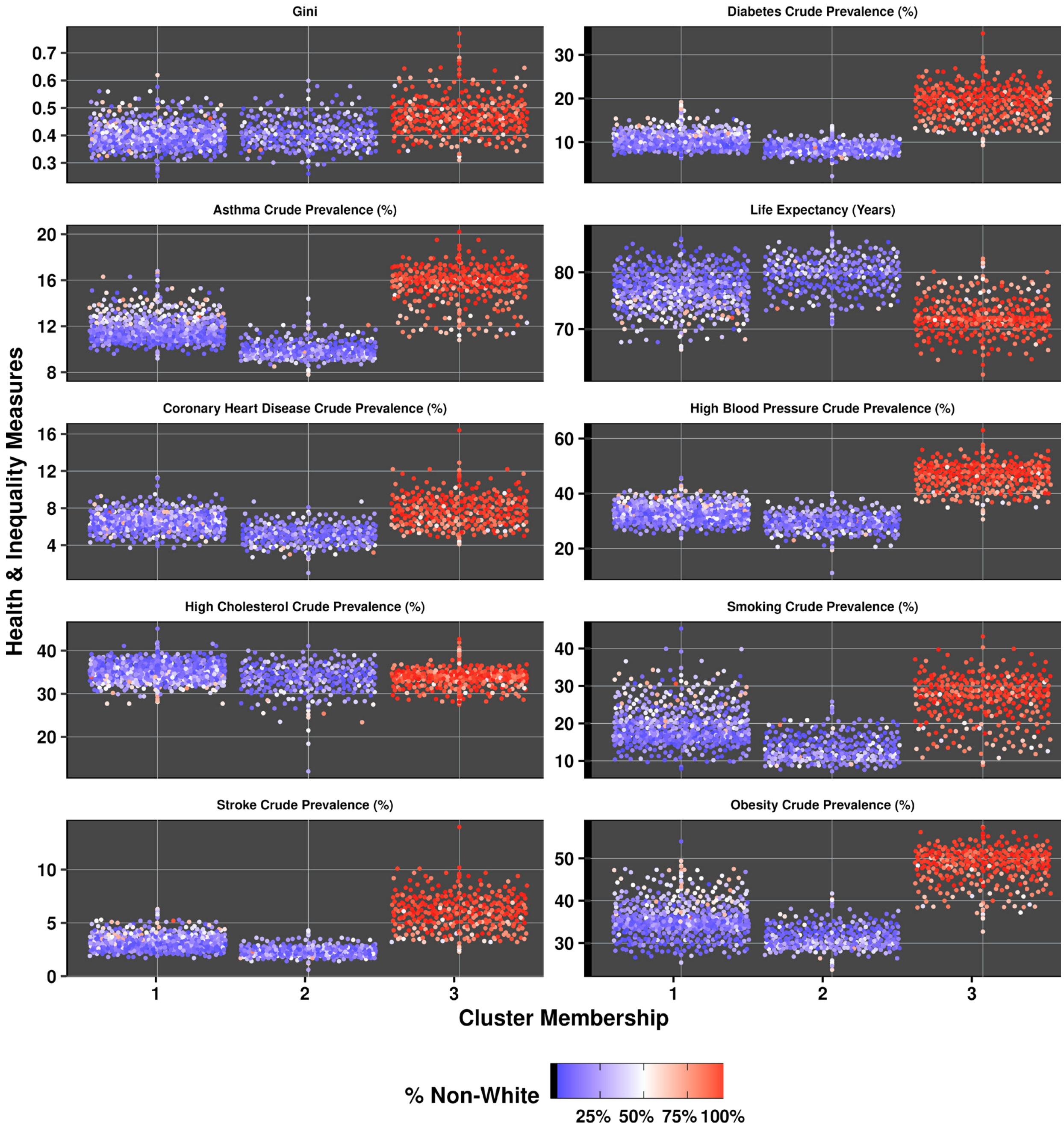


*Figure 14: Plot of 1151 of the 1165 census tracts within the immediate Detroit metropolitan area—colored by the % of residents who are not White with respect to **race**. [Data sources: USCB via K. Walker and Herman (v1.4.1), CDC API (2023) via Devlin et al. (v1.7.12-4), CDC LEEP (2023), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: ggbeeswarm, ggplot2.]*

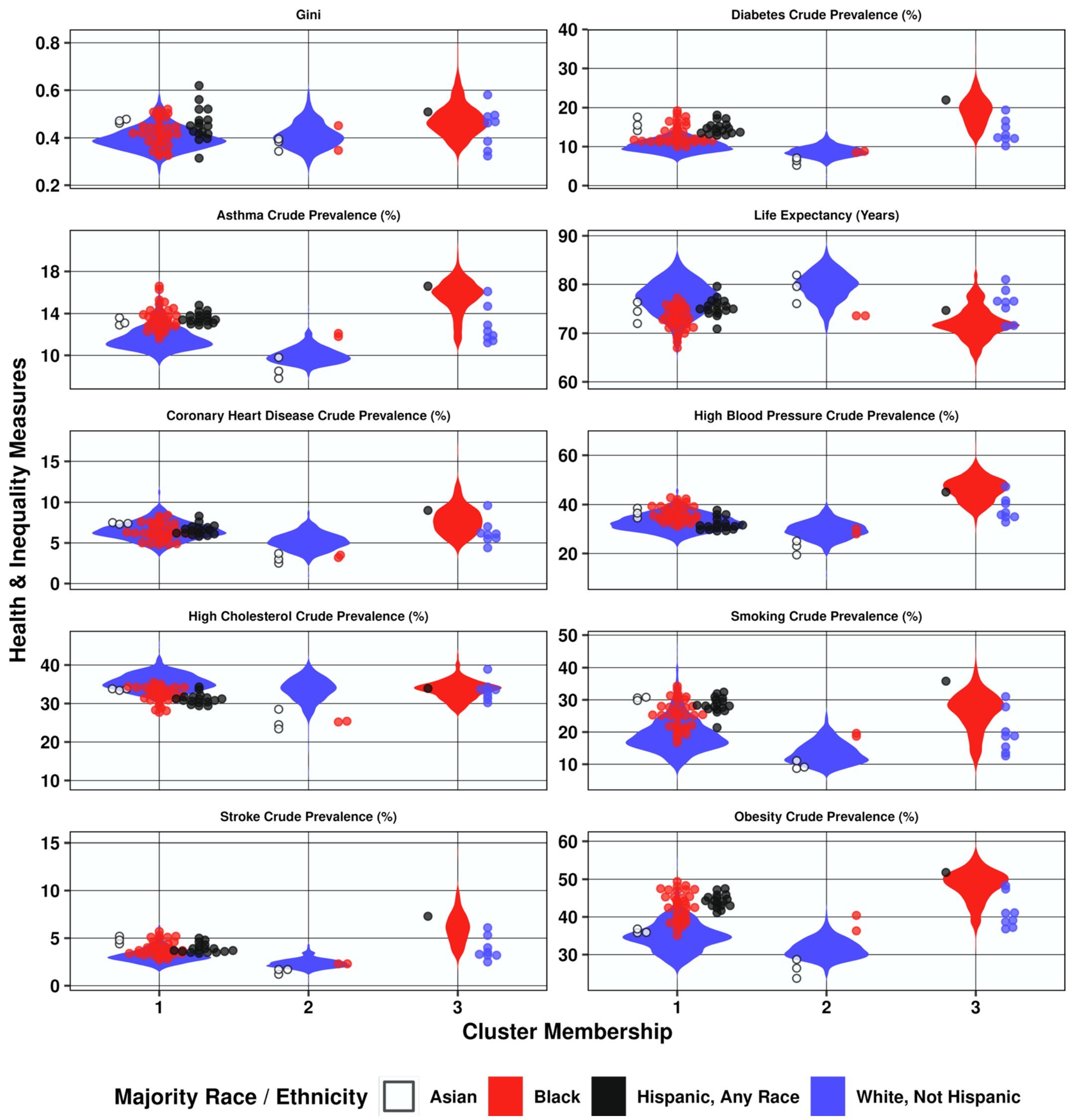


Figure 15: Plot of 1151 of the 1165 census tracts within the immediate Detroit metropolitan area—colored by the majority race / ethnicity in each census tract. Violin plots are used in lieu of beeswarm plots when there are substantively many census tracts belonging to a majority race / ethnicity. [Data sources: USCB via K. Walker and Herman (v1.4.1), CDC API (2023) via Devlin et al. (v1.7.12-4), CDC LEEP (2023), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: ggbeeswarm, ggplot2.]

MICRO* Level View of the *Immediate* Detroit Metropolitan Area: *Analyzing Race & Ethnicity Data

The same data that is the basis for *Figure 14* and *Figure 15* is displayed in the top 2 maps of *Figure 16*, respectively. The maps in *Figure 16*, *Figure 17*, and *Figure 18* display the race / ethnicity data from the US Census Bureau's American Community Survey 2015-2019 5-Year Data Release made available via the R package **tidycensus** (USCB 10-December-2020; K. Walker and Herman v1.4.1):

- Sponsors could use maps in each of these 3 figures to identify census tracts whose residents could benefit from increased awareness of and access to DCTs. For example,
  - *Figure 16*: Census tracts in the southern portion of the City of Detroit and within the central part of Oakland County could be potential sources of prospective patients who happen to be Hispanic.
  - *Figure 17*: A small number of census tracts in dark red exist for prospective patients who happen to identify as either American Indian / Alaska Native or Native Hawaiian / Pacific Islander. Patient recruitment efforts could be focused in these areas.

- *Figure 18*: Higher concentrations of prospective patients who identify as being of another race or 2 or more races reside in the southern portion of the City of Detroit. DCT advertising for these patients could coincide with Hispanic outreach efforts.
- A number of maps in *Figure 16*, *Figure 17*, and *Figure 18* could be used for **geo targeting** for DCTs: Focused advertising to census tracts of interest or larger corresponding ZIP codes to reach prospective patients who might not otherwise be aware of enrolling clinical trials or possibly may not be able to travel to the larger clinics / hospitals traditionally used for studies (Google).
- The map of the **Gini-Simpson Diversity Index** in *Figure 16*—when viewed in the context of the other maps in *Figure 16*, *Figure 17*, and *Figure 18*, is visual evidence that supports the ***Roots of Structural Racism Project*** finding that the **Detroit-Warren-Livonia Metropolitan Statistical Area** is the **4th most segregated** in the US (Mendendian, Gambhir, and Gailes 30-June-2021). In the context of US census data, the Gini-Simpson Diversity Index (*as derived for this PHUSE paper*) is an estimate of the probability of encountering 2 people of 2 different races (specifically, American Indian / Alaska Native, Single Race; Asian, Single Race; Black, Single Race; Native Hawaiian / Pacific Islander, Single Race; Other, Single Race; Two or More Races; White, Single Race) within the *same census tract* (Wikipedia-Diversity; Guiasu and Guiasu 16-February-2012).

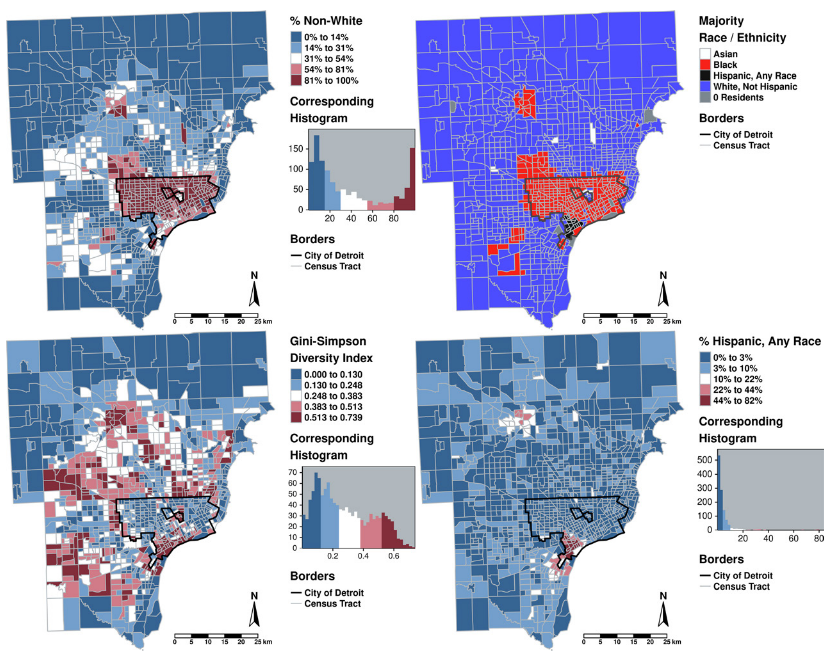


*Figure 16: Maps & corresponding histograms of the 1165 census tracts within the immediate Detroit metropolitan area—colored by **% Non-White**, **Majority Race / Ethnicity**, **Gini-Simpson Diversity Index**, and **% Hispanic, Any Race** via the Jenks natural breaks classification method (except for the **Majority Race / Ethnicity** map) (Wikipedia-Jenks; Tennekes v3.3-3). [Data sources: USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap.]*

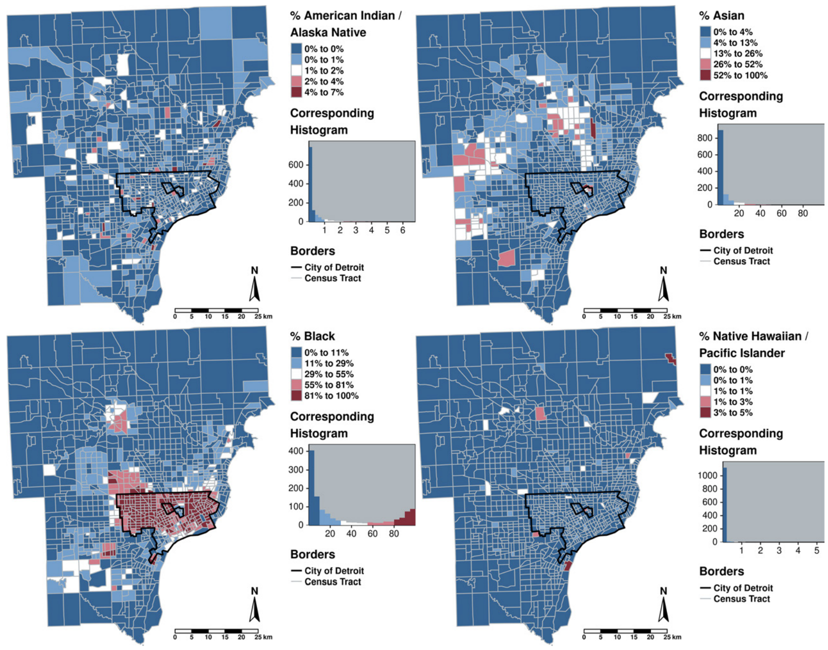


*Figure 17: Maps & corresponding histograms of the 1165 census tracts within the immediate Detroit metropolitan area—colored by **% American Indian / Alaska Native**, **% Asian**, **% Black**, and **% Native Hawaiian / Pacific Islander** via the Jenks natural breaks classification method (Wikipedia-Jenks; Tennekes v3.3-3). [Data sources: USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap.]*

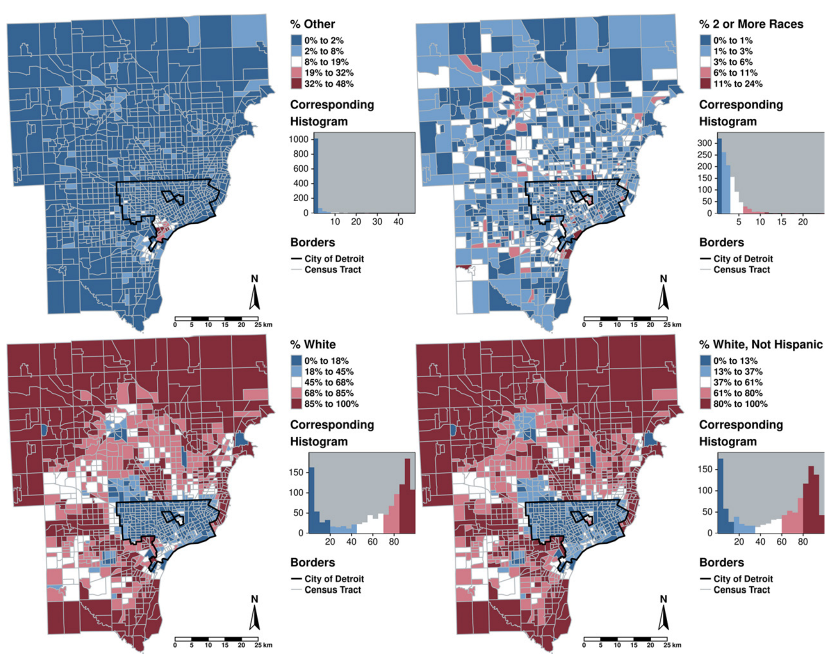


*Figure 18: Maps & corresponding histograms of the 1165 census tracts within the immediate Detroit metropolitan area—colored by **% Other**, **% 2 or More Races**, **% White**, and **% White, Not Hispanic** via the Jenks natural breaks classification method (Wikipedia-Jenks; Tennekes v3.3-3). [Data sources: USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap.]*

MICRO* Level View of the *Immediate* Detroit Metropolitan Area: *Applying Geospatial Statistics

The same underlying data used for *Figure 16*, *Figure 17*, and *Figure 18* has been used to create *Figure 19* and *Figure 20*. These latter 2 figures display maps of **Local Getis-Ord G** *hot spots* and their *neighboring* census tracts (Li and Anselin v0.0.10-2; Anselin 10-October-2020). The hot spot census tracts **and** their neighbors need to be analyzed together as local areas of interest, because the local Getis-Ord G statistic of spatial autocorrelation accounts for data in the **combined set of tracts** (Geographic Information Systems Stack Exchange 16-September-2014). For example, the map of the % of Native Hawaiian / Pacific Islander residents in *Figure 17* indicates there is a census tract in the northeast part of Macomb County which contains **3% to 5%** Native Hawaiian / Pacific Islander residents. This same tract is a neighbor (**not** a hot spot) of a Getis-Ord G hot spot for Native Hawaiian / Pacific Islander residents. Thus, local areas of interest identified by the Local Getis-Ord G statistic need to be further explored for specific areas to focus on for DCT efforts.

The local areas of interest identified in *Figure 19* and *Figure 20* have been combined together and termed "consolidated hot spots". Maps of these consolidated hot spots and Cluster No. 3–identified as the least equitable / least healthy cluster by the UMAP and PAM tract-level analysis–are displayed in *Figure 21*. The consolidated hot spots and Cluster No. 3 have been combined together in the map on the right in *Figure 21* with Cluster No. 3 renamed as a "health disparities focus area". The orange census tracts in this map taken together have the potential to serve as areas where Sponsors can reach out to diverse and / or marginalized prospective patients for DCTs. The health disparities focus area could be viewed as the DCT priority area for diverse and marginalized patients. However, within this priority area (Cluster No. 3), some residents may live closer than others to local HCPs for DCTs—further compounding the challenges of enrolling prospective patients.

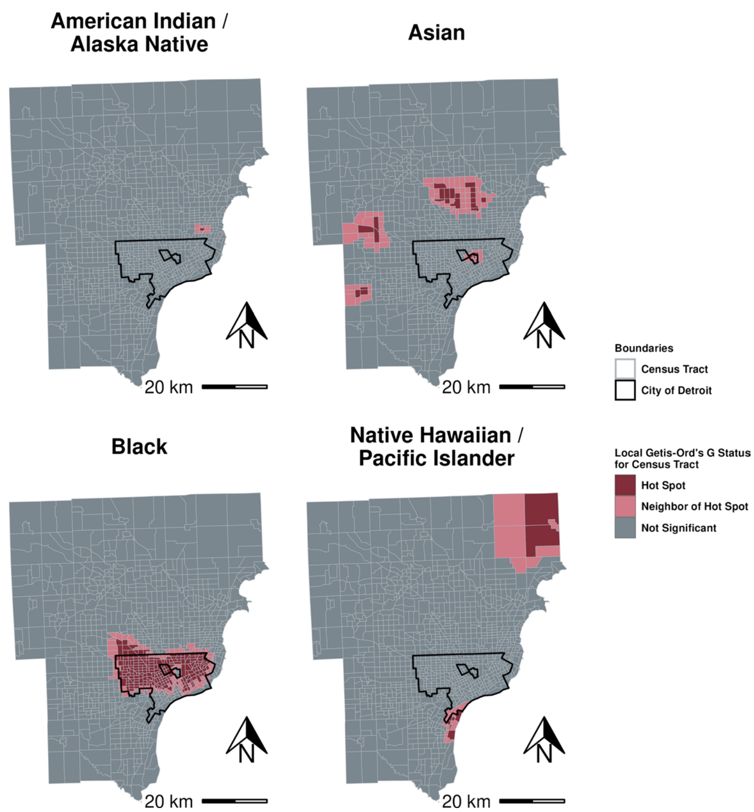


*Figure 19: Maps of Local Getis-Ord G hot spots and their neighbors for different populations within the immediate Detroit metropolitan area—specifically, **American Indian / Alaska Native**, **Asian**, **Black**, and the **Native Hawaiian / Pacific Islander** (Li and Anselin v0.0.10-2; Anselin 10-October-2020). [Data sources: USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: ggplot2, ggspatial; Analysis Notes: Queen method applied when identifying geographic neighbors; a pseudo p-value significance cutoff of 0.01 was used in combination with a multiple comparison False Discovery Rate (FDR) cutoff of 0.01 and 999,999 complete permutations.]*

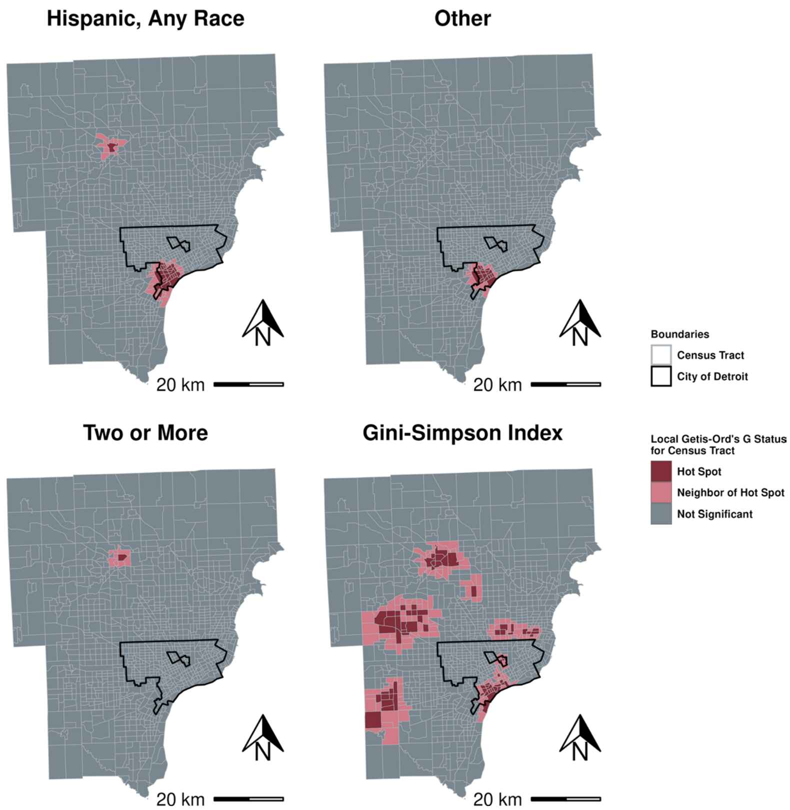


*Figure 20: Maps of Local Getis-Ord G hot spots and their neighbors for different populations within the immediate Detroit metropolitan area—specifically, **Hispanic, Any Race**, **Other Race**, **Two or More Races**, and the **Gini-Simpson Diversity Index** (Li and Anselin v0.0.10-2; Anselin 10-October-2020). [Data sources: USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: ggplot2, ggspatial; Analysis Notes: Queen method applied when identifying geographic neighbors; a pseudo p-value significance cutoff of 0.01 was used in combination with a multiple comparison False Discovery Rate (FDR) cutoff of 0.01 and 999,999 complete permutations.]*

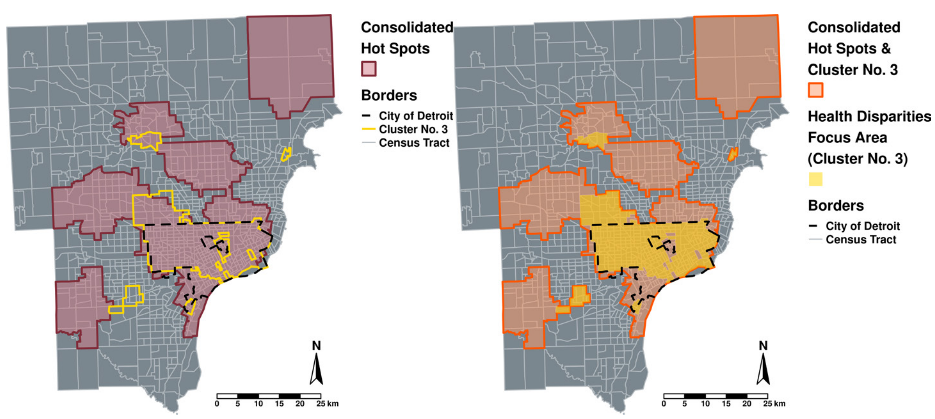


Figure 21: Maps of consolidated hot spots—containing all local areas of interest as identified in Figure 19 and Figure 20; map on the right combines the consolidated hot spots with Cluster No. 3. (Li and Anselin v0.0.10-2; Anselin 10-October-2020). [Data sources: USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap; Analysis Notes: Queen method applied when identifying geographic neighbors; a pseudo p-value significance cutoff of 0.01 was used in combination with a multiple comparison False Discovery Rate (FDR) cutoff of 0.01 and 999,999 complete permutations.]

MICRO* Level View of the *Immediate* Detroit Metropolitan Area: *Locating Local HCPs & Non-Traditional DCT Sites

The proximity of local HCPs to residents within the DCT priority area was explored by using a number of R packages. The key R packages are listed below:

- **npi** (Farach v0.2.0; US Centers for Medicare & Medicaid Services 02-August-2023; American Medical Association 01-July-2023): To obtain local HCP U.S. **National Provider Identifier (NPI)** numbers and street address information. (*Note: Local HCP data was downloaded from **National Plan and Provider Enumeration System (NPPES) Registry** to speed data processing time.*)
- **zipcodeR** (Rozzi v0.3.5): To select only local HCPs who have street address ZIP codes within Macomb, Oakland, and Wayne Counties.
- **tidygeocoder** (Cambon et al. v1.0.5): To obtain the longitude and latitude for the local HCP street addresses.
- **sf** (Pebesma v1.0-13): To manipulate spatial data objects.
- **tmap** (Tennekes v3.3-3) and **ggplot2** (Wickham v3.4.2): To plot spatial data objects.
- **eks** (Duong v1.0.4): To determine the density of local HCPs in the immediate Detroit metropolitan area.

Individual local HCPs who practice medicine within Macomb, Oakland, and Wayne Counties were identified using data downloaded from the ***National Plan and Provider Enumeration System (NPPES) Registry***, geographically subset using the **zipcodeR** R package, and geocoded via the **tidygeocoder** R package (US Centers for Medicare & Medicaid Services 02-August-2023; American Medical Association 01-July-2023; Rozzi v0.3.5; Cambon et al. v1.0.5). These providers were then further subset to obtain all who were identified as either **Allopathic & Osteopathic Physicians** or **Physician Assistants & Advanced Practice Nursing Providers**. This data was mapped along with geographic density estimates of their practice locations in *Figure 22*. Density estimates were mapped in addition to the number of local HCPs per census tract as a direct result of the **Modifiable Areal Unit Problem (MAUP)**:

> "One common technique for analyzing many point phenomena such as crime (***or** the number of local HCPs for this PHUSE paper*), is to aggregate (count the number of) events that occur within **predefined areas** like neighborhoods, zip codes, or **census tracts**.
>
> However, the choice of what area to aggregate in can result in dramatically different analysis. If you modify the boundaries of aggregation areas, that affects which areas points are aggregated into and, therefore, the counts of points for each area. Boundaries of areas used for aggregation like neighborhoods or census tracts are often historical and political, and *may not be perfectly suited to the aggregation analysis you are performing*. **This is the modifiable areal unit problem.**"
>
> *—Micheal Minn (Minn)*

As per Figure 22, there are many low density areas of local HCPs within the City of Detroit.

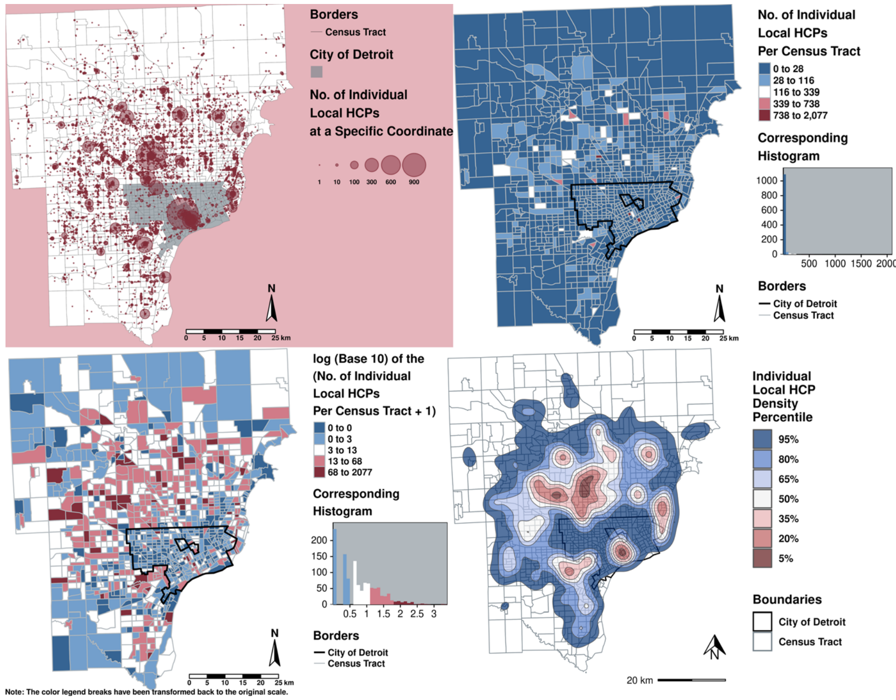


*Figure 22: Maps of **individual** local HCPs (**Allopathic & Osteopathic Physicians** or **Physician Assistants & Advanced Practice Nursing Providers**) within the immediate Detroit metropolitan area by number per longitude and latitude, number per census tract, $\log_{10}$**(number per census tract + 1)**, and by density. The 2 maps displaying numbers per census tract color each tract via the Jenks natural breaks classification method (Wikipedia-Jenks; Tennekes v3.3-3). The **color legend breaks** for the $\log_{10} + 1$ map have been transformed back to the **original** scale. [Data sources: US Centers for Medicare & Medicaid Services (02-August-2023), American Medical Association (01-July-2023), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: colorspace, eks, ggnewscale, ggplot2, ggspatial, tmap.]*

The low density areas of individual local HCPs were selected—as displayed in the left-hand map in *Figure 23*, and these were spatially intersected with the health disparities focus area (DCT priority area) to yield the regions in **green**—as per the right-hand map in *Figure 23*. The resulting health disparities / low individual HCP density focus area covers a sizable portion of the City of Detroit and provides added visual impetus for Sponsors to begin DCT recruitment campaigns in these green regions (termed the "DCT focus area"). This importance is underscored by a result of the Lown Institute's analysis concluding that Detroit hospitals were the most racially segregated in the US market (Liss 21-March-2022; Walsh 17-March-2022).

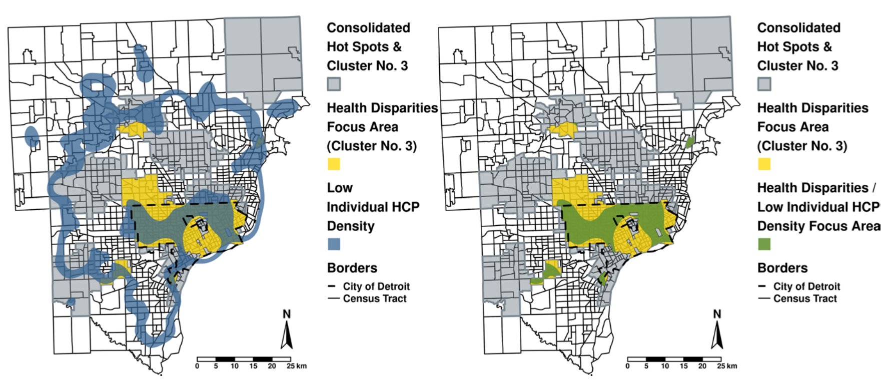


*Figure 23: Maps of low density areas of individual local HCPs and their spatial intersection with the health disparities focus area (intersection in **green**, the "DCT focus area"). [Data sources: US Centers for Medicare & Medicaid Services (02-August-2023), American Medical Association (01-July-2023), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap.]*

After geolocating local HCPs who practice within the green DCT focus area, Sponsors would need to adjudicate their qualifications and contact prioritized providers. Data available from the National Plan and Provider Enumeration System (NPPES) Registry contains individual provider as well as provider organization **names**, **credentials**, **street addresses**, **phone numbers**, **fax numbers**, **license numbers**, etc. *Figure 24* displays both geocoded locations of **303 *individual*** local HCPs as well as **143** local HCP ***organizations*** who identify as either **Allopathic & Osteopathic Physicians** or **Physician Assistants & Advanced Practice Nursing Providers** and practice within the green DCT focus area.

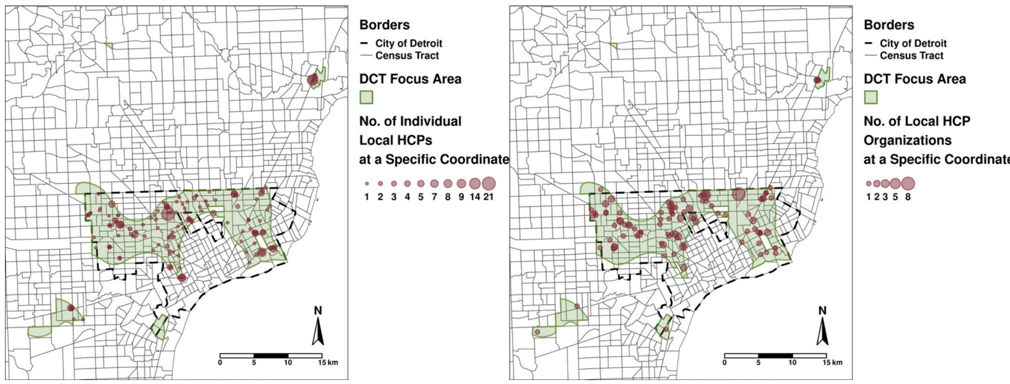


*Figure 24: Maps of geocoded locations of **303 individual** local HCPs as well as **143** local HCP **organizations** who identify as either **Allopathic & Osteopathic Physicians** or **Physician Assistants & Advanced Practice Nursing Providers** and practice within the green DCT focus area. [Data sources: US Centers for Medicare & Medicaid Services (02-August-2023), American Medical Association (01-July-2023), USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap.]*

Even if Sponsors identify qualified and strategically located local HCPs for DCTs, Sponsors could face the added challenge posed by the sizable mistrust that exists between many diverse / marginalized communities and the medical establishment—as discussed by an article published in *Fortune* and written by **Sanofi CEO Paul Hudson** and **Michelle A. Williams, Dean of the Harvard T. H. Chan School of Public Health**:

> “This stark disparity surfaced in new polling commissioned by Sanofi. The global survey of 11,500 people from across the U.S., France, the U.K., Japan, and Brazil provides a first-of-its-kind look at how people from a wide variety of backgrounds experience the healthcare system.
>
> The results are worrying. In the U.S., 77% of people with disabilities, 69% of people from ethnic minority groups, and 70% of people from the LGBTQ+ community say they have had experiences that damaged their trust in the healthcare system.
>
> The survey also showed the compounding effect of marginalization: People who fell into more than one of these groups were significantly more likely to have had a bad experience while seeking care. For example, in the U.S., 80% of people who have disabilities and identify as LGBTQ+ reported losing trust in the healthcare system due to their personal experience, compared with 56% of people without those backgrounds.”
>
> *—Paul Hudson and Michelle A. Williams (Hudson and Williams 31-January-2023)*

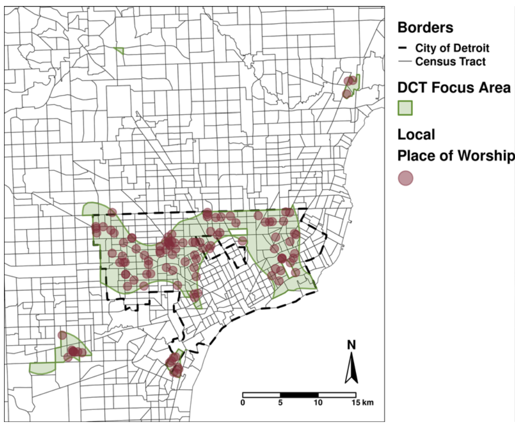
In an attempt to overcome the challenge that many prospective patients in the green DCT focus area may mistrust local HCPs, Sponsors should consider pairing local HCPs with non-traditional healthcare facilities such as places of worship and community centers. These facilities may be more proximal to some prospective patients in the green DCT focus area, and their leaders may have a greater trust of the local communities. A similar partnership was developed in some New York City Black Churches during the COVID-19 pandemic:

> “To address the gap, health officials and some Black churches have sought to use the power of the pulpit to vouch for the safety of vaccines and to push back against misinformation. They have also hosted vaccination events in church halls or from mobile vans parked outside of churches after Sunday services.
>
> ‘These cultural institutions are a safe space to have discussions—you go to your faith leader and they’ll answer questions’ said Dr. Torian Easterling, the first deputy commissioner and chief equity officer at the New York City Department of Health and Mental Hygiene.

Pastor Butts said vaccine advocacy had to overcome the legacy of those conspiracies, including the Tuskegee syphilis study, during which government researchers withheld syphilis treatment from hundreds of Black men for 40 years so they could observe the course of the disease."

—Liam Stack (Stack 13-October-2021)

Using the R packages **sf** (Pebesma v1.0-13), **stats** (R Core Team 2023), **dismo** (Hijmans et al. v1.3-14), **terra** (Hijmans v1.7-39), and **googleway** (Cooley v2.7.7), **113 *places of worship*** and **41 *community centers*** located within the green DCT focus area have been mapped in *Figure 25*. In addition to finding and geocoding these places, the package **googleway** also provides detailed contact information such as **name**, **street address**, **phone number**, and **website**, if available via its access to the **Google Maps API** (Google).



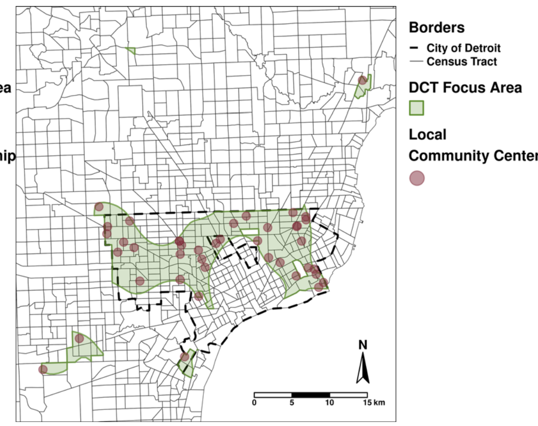


*Figure 25: Maps of geocoded locations of **113 places of worship** and **41 community centers** located with the green DCT focus area. [Data sources: Google, USCB (10-December-2020) via K. Walker and Herman (v1.4.1); Salient R plotting / mapping packages: tmap.]*

CONCLUSION

The draft guidance documents *Diversity Plans to Improve Enrollment of Participants From Underrepresented Racial and Ethnic Populations in Clinical Trials* (FDA 13-April-2022) and *Decentralized Clinical Trials (DCTs) for Drugs, Biological Products, and Devices: Guidance for Industry, Investigators, and Other Stakeholders* (FDA 01-May-2023) are a call to action for Sponsors to bridge the gap between diverse / marginalized prospective patients and clinical trial access via DCTs. Many Sponsors are already using R for data analysis and reporting. It has been demonstrated that this same open source language can be used as a GIS software tool to assist with matching patient populations with potential DCT sites such as local HCP offices, places of worship, community centers, etc.

Some R Packages and Data Resources to Aid Finding Patient Populations and Potential DCT Locations *Outside the U.S.*

- **R Packages**
  - **Brazil** (K. E. Walker 17-February-2023)
    - ***geobr**: Download Official Spatial Data Sets of Brazil*
  - **Canada**
    - ***cancensus**: Access, Retrieve, and Work with Canadian Census Data and Geography*
    - ***cansim**: Accessing Statistics Canada Data Table and Vectors*
    - ***cmhc**: Access, Retrieve, and Work with CMHC Data*
    - ***tongfen**: Make Data Based on Different Geographies Comparable*
  - **Europe** (K. E. Walker 17-February-2023)
    - ***eurostat**: Tools for Eurostat Open Data*
  - **France** (K. E. Walker 17-February-2023)
    - ***insee**: Tools to Easily Download Data from INSEE BDM Database*
  - **International** (K. E. Walker 17-February-2023)
    - ***idbr**: R Interface to the US Census Bureau International Data Base API*
    - ***ipumsr**: An R Interface for Downloading, Reading, and Handling IPUMS Data*
  - **Kenya** (K. E. Walker 17-February-2023)

    - ***rKenyaCensus**: The goal of rKenyaCensus is to provide tidy datasets obtained from the Kenya Population and Housing Census results.*
  - **Mexico** (K. E. Walker 17-February-2023)
    - ***mxmaps**: Mexico Choropleths*
    - ***inegiR**: Integrate INEGI's (Mexican Stats Office) API with R*
  - **United Kingdom** (K. E. Walker 17-February-2023)
    - ***nomisr**: Access 'Nomis' UK Labour Market Data*
- **Data Resources**
  - **Canada**
    - *The Open Database of Healthcare Facilities*
  - **Ireland**
    - *Central Statistics Office*
    - *eHealth Ireland*
  - **United Kingdom**
    - *HDRUK: Health Data Research Innovation Gateway*
    - *NHS Digital: Data Access Request Service (DARS)*
    - *NHS Digital: The NHS England Secure Data Environment*
    - *NHS Digital: NHS DigiTrials*
    - *NHS Digital: Data Access Environment (DAE)*
    - *Office of National Statistics*

SALIENT R PLOTTING / MAPPING PACKAGES

- **biscale** (Prener, Grossenbacher, and Zehr v1.0.0)
- **colorspace** (Zeileis et al. v2.1-0)
- **eks** (Duong v1.0.4)
- **ggbeeswarm** (Clarke, Sherrill-Mix, and Dawson v0.7.2)
- **ggmap** (Kahle and Wickham v3.0.2)
- **ggmapinset** (Suster v0.3.0)
- **ggnewscale** (Campitelli v0.4.9)
- **ggplot2** (Wickham v3.4.2)
- **ggpubr** (Kassambara v0.6.0)
- **ggrepel** (Slowikowski v0.9.3)
- **ggspatial** (Dunnington v1.1.8)
- **patchwork** (Pedersen v1.1.2)
- **tmap** (Tennekes v3.3-3)

REFERENCES

Allaire, JJ. v1.2. *quarto: R Interface to 'Quarto' Markdown Publishing System*. *https://CRAN.R-project.org/package=quarto*.

American Medical Association. 01-July-2023. "National Uniform Claim Committee: Health Care Provider Taxonomy Code Set CSV (Version 23.1)." *https://www.nucc.org/index.php/code-sets-mainmenu-41/provider-taxonomy-mainmenu-40/csv-mainmenu-57*.

Anselin, Luc. 10-October-2020. "Local Spatial Autocorrelation (2): Other Local Spatial Autocorrelation Statistics." GeoDa: An Introduction to Spatial Data Science Website: *https://geodacenter.github.io/workbook/6b_local_adv/lab6b.html*.

Bivand, Roger, and Jakub Nowosad. "CRAN Task View: Analysis of Spatial Data." *https://cran.r-project.org/web/views/Spatial.html*.

Blesch, K. 25-June-2022. "Repository Providing Code to Reproduce Results for "Measuring Inequality Beyond the Gini Coefficient May Clarify Conflicting Findings" by Blesch Et Al. (2022)." GitHub Repository: *https://github.com/kristinblesch/inequality*.

Blesch, K., O. P. Hauser, and J. M. Jachimowicz. 01-November-2022. "Measuring Inequality Beyond the Gini Coefficient May Clarify Conflicting Findings." *Nature Human Behavior* 6 (11): 1525–36. *https://doi.org/10.1038/s41562-022-01430-7*.

Cambon, Jesse, Diego Hernangómez, Christopher Belanger, and Daniel Possenriede. v1.0.5. *tidygeocoder: Geocoding Made Easy*. *https://CRAN.R-project.org/package=tidygeocoder*.

Campitelli, Elio. v0.4.9. *ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'*. *https://CRAN.R-project.org/package=ggnewscale*.

CDC API. 2023. "PLACES: Local Data for Better Health, Census Tract Data 2023 Release." *https://dev.socrata.com/foundry/chronicdata.cdc.gov/cwsq-ngmh*.

CDC LEEP. 2023. "U.S. Small-Area Life Expectancy Estimates Project - USALEEP." *https://www.cdc.gov/nchs/nvss/usaleep/usaleep.html#data*.

CDC PLACES. "PLACES: Local Data for Better Health." *https://www.cdc.gov/places/*.

CHR&R. 2023, Version 2. "County Health Rankings & Roadmaps." CHR&R Rankings Data & Documentation Website: *https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation*, University of Wisconsin Population Health Institute.

Clarke, Erik, Scott Sherrill-Mix, and Charlotte Dawson. v0.7.2. *ggbeeswarm: Categorical Scatter (Violin Point) Plots*. *https://CRAN.R-project.org/package=ggbeeswarm*.

Cooley, David. v2.7.7. *googleway: Accesses Google Maps APIs to Retrieve Data and Plot Maps*. *https://CRAN.R-project.org/package=googleway*.

Devlin, Hugh, Ph. D., Tom Schenk, Jr., Gene Leynes, Nick Lucius, John Malc, Mark Silverberg, and Peter Schmeideskamp. v1.7.12-4. *RSocrata: Download or Upload 'Socrata' Data Sets*. *https://CRAN.R-project.org/package=RSocrata*.

Dunnington, Dewey. v1.1.8. *ggspatial: Spatial Data Framework for ggplot2*. *https://CRAN.R-project.org/package=ggspatial*.

Duong, Tarn. v1.0.4. *eks: Tidy and Geospatial Kernel Smoothing*. *https://CRAN.R-project.org/package=eks*.

Farach, Frank. v0.2.0. *npi: Access the U.S. National Provider Identifier Registry API*. *https://CRAN.R-project.org/package=npi*.

FDA. 01-May-2023. "Decentralized Clinical Trials for Drugs, Biological Products, and Devices." FDA Website: *https://www.fda.gov/regulatory-information/search-fda-guidance-documents/decentralized-clinical-trials-drugs-biological-products-and-devices*.

———. 13-April-2022. "Diversity Plans to Improve Enrollment of Participants from Underrepresented Racial and Ethnic Populations in Clinical Trials." FDA Website: *https://www.fda.gov/regulatory-information/search-fda-guidance-documents/diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations*.

Geographic Information Systems Stack Exchange. 16-September-2014. "Hot Spot Analysis Indicated the Areas with 0 Value as Hot Spot?" *https://gis.stackexchange.com/questions/114073/hot-spot-analysis-indicated-the-areas-with-0-value-as-hot-spot*.

Google. "Google Maps Platform." *https://developers.google.com/maps*.

———. "Target Ads to Geographic Locations." Google Ads Help: *https://support.google.com/google-ads/answer/1722043?hl=en*.

Guiasu, R., and S. Guiasu. 16-February-2012. "The Weighted Gini-Simpson Index: Revitalizing an Old Index of Biodiversity." *International Journal of Ecology* 2012 (Article ID 478728): 1–10. *https://doi.org/10.1155/2012/478728*.

Hennig, Christian. v2.2-10. *fpc: Flexible Procedures for Clustering*. *https://CRAN.R-project.org/package=fpc*.

Hijmans, Robert J. v1.7-39. *terra: Spatial Data Analysis*. *https://CRAN.R-project.org/package=terra*.

Hijmans, Robert J., Steven Phillips, John Leathwick, and Jane Elith. v1.3-14. *dismo: Species Distribution Modeling*. *https://CRAN.R-project.org/package=dismo*.

Hudson, P., and M. A. Williams. 31-January-2023. "People Are Much Less Likely to Trust the Medical System If They Are from an Ethnic Minority, Have Disabilities, or Identify as LGBTQ+, According to a First-of-Its-Kind Study by Sanofi." Fortune Health Commentary by Paul Hudson and Michelle A. Williams: *https://fortune.com/2023/01/31/people-trust-health-medical-system-ethnic-minority-disabilities-identify-lgbtq-study-sanofi-hudson-williams/*.

Jeden, P. 19-April-2023. “A Renewed Focus on Health Equity in Wayne County, Michigan.” Network for Public Health Law Website: *https://www.networkforphl.org/news-insights/a-renewed-focus-on-health-equity-in-wayne-county-michigan/*.

Kahle, David, and Hadley Wickham. v3.0.2. *ggmap: Spatial Visualization with ggplot2*. *https://CRAN.R-project.org/package=ggmap*.

Kassambara, Alboukadel. v0.6.0. *ggpubr: ’ggplot2’ Based Publication Ready Plots*. *https://CRAN.R-project.org/package=ggpubr*.

Konopka, Tomasz. v0.2.10.0. *umap: Uniform Manifold Approximation and Projection*. *https://CRAN.R-project.org/package=umap*.

Li, Xun, and Luc Anselin. v0.0.10-2. *rgeoda: R Library for Spatial Data Analysis*. *https://CRAN.R-project.org/package=rgeoda*.

Liss, Samantha. 21-March-2022. “Detroit Rated as Most Racially Segregated Hospital Market in US, Report Shows.” *https://www.healthcaredive.com/news/detroit-rated-as-most-racially-segregated-hospital-market-in-us-reports/620708/*.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. v2.1.4. *cluster: Cluster Analysis Basics and Extensions*. *https://CRAN.R-project.org/package=cluster*.

Mahmoud Mosalman Yazdi. v1.0. *topsis: TOPSIS Method for Multiple-Criteria Decision Making (MCDM)*. *https://CRAN.R-project.org/package=topsis*.

Mendendian, S., S. Gambhir, and A. Gailes. 30-June-2021. “The Roots of Structural Racism Project: Twenty-First Century Racial Residential Segregation in the United States.” Othering and Belonging Institute, University of California, Berkeley: *https://belonging.berkeley.edu/roots-structural-racism*.

Minn, Michael. “Basic Spatial Point Pattern Analysis in R.” *https://michaelminn.net/tutorials/r-point-analysis/*.

Pebesma, Edzer. v1.0-13. *sf: Simple Features for R*. *https://CRAN.R-project.org/package=sf*.

Pebesma, Edzer, and Roger Bivand. “CRAN Task View: Handling and Analyzing Spatio-Temporal Data.” *https://cran.r-project.org/web/views/SpatioTemporal.html*.

Pedersen, Thomas Lin. v1.1.2. *patchwork: The Composer of Plots*. *https://CRAN.R-project.org/package=patchwork*.

Prener, Christopher, Timo Grossenbacher, and Angelo Zehr. v1.0.0. *biscale: Tools and Palettes for Bivariate Thematic Mapping*. *https://CRAN.R-project.org/package=biscale*.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. *https://www.R-project.org/*.

Rozzi, Gavin C. v0.3.5. *zipcodeR: Data & Functions for Working with US ZIP Codes*. *https://CRAN.R-project.org/package=zipcodeR*.

Slowikowski, Kamil. v0.9.3. *ggrepel: Automatically Position Non-Overlapping Text Labels with ’ggplot2’*. *https://CRAN.R-project.org/package=ggrepel*.

Stack, L. 13-October-2021. “"A Safe Space": Black Pastors Promote Vaccinations from the Pulpit—More Than 80 Percent of Adults in New York City Have Received at Least One Dose of the Coronavirus Vaccine, but There Are Significant Racial Disparities in the Vaccination Rate.” New York Times: *https://www.nytimes.com/2021/10/09/nyregion/covid-vaccinations-black-churches.html*.

Suster, Carl. v0.3.0. *ggmapinset: Add Inset Panels to Maps*. *https://CRAN.R-project.org/package=ggmapinset*.

Tennekes, Martijn. v3.3-3. *tmap: Thematic Maps*. *https://CRAN.R-project.org/package=tmap*.

US Centers for Medicare & Medicaid Services. 02-August-2023. “National Plan and Provider Enumeration System (NPPES) Registry.” NPPES Registry Downloadable Data: *https://www.cms.gov/medicare/regulations-guidance/administrative-simplification/data-dissemination*; NPPES Registry API: *https://npiregistry.cms.hhs.gov/api/*.

USCB. “American Community Survey 2011-2015 5-Year Data Release.” U.S. Census Bureau Website: *https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2015/5-year.html*.

———. 10-December-2020. “American Community Survey 2015-2019 5-Year Data Release.” U.S. Census Bureau Website: *https://www.census.gov/newsroom/press-kits/2020/acs-5-year.html*.

van Buuren, Stef, and Karin Groothuis-Oudshoorn. v3.16.0. *mice: Multivariate Imputation by Chained Equations*. *https://CRAN.R-project.org/package=mice*.

Vialaneix, Nathalie, Elise Maigne, Jerome Mariette, Madalina Olteanu, Fabrice Rossi, Laura Bendhaiba, and Julien Bolaert. v1.4-1. *SOMbrero: SOM Bound to Realize Euclidean and Relational Outputs*. *https://CRAN.R-project.org/package=SOMbrero*.

Walker, K. E. 17-February-2023. “Working with Census Data Outside the United States.” *https://walker-data.com/census-r/working-with-census-data-outside-the-united-states.html*.

Walker, Kyle, and Matt Herman. v1.4.1. *tidycensus: Load US Census Boundary and Attribute Data as tidyverse and sf-Ready Data Frames*. *https://CRAN.R-project.org/package=tidycensus*.

Walsh, Dustin. 17-March-2022. “Metro Detroit Hospitals the Most Racially Segregated in the U.S., Study Says.” *https://www.crainsdetroit.com/health-care/metro-detroit-hospitals-most-racially-segregated-us-study-says*.

Wickham, Hadley. v3.4.2. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. *https://CRAN.R-project.org/package=ggplot2*.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. v1.1.2. *dplyr: A Grammar of Data Manipulation*. *https://CRAN.R-project.org/package=dplyr*.

Wikipedia-Diversity. “Diversity Index.” *https://en.wikipedia.org/wiki/Diversity_index*.

Wikipedia-GINI. “Gini Coefficient.” *https://en.wikipedia.org/wiki/Gini_coefficient*.

Wikipedia-GIS. “Geographic Information System.” *https://en.wikipedia.org/wiki/Geographic_information_system*.

Wikipedia-Jenks. “Jenks Natural Breaks Optimization.” *https://en.wikipedia.org/wiki/Jenks_natural_breaks_optimization*.

Wikipedia-TOPSIS. “TOPSIS.” *https://en.wikipedia.org/wiki/TOPSIS*.

Wikipedia-UMAP. “Nonlinear Dimensionality Reduction: UMAP.” *https://en.wikipedia.org/wiki/Nonlinear_dimensionality_reduction#Uniform_manifold_approximation_and_projection*.

Zeileis, Achim, Jason C. Fisher, Kurt Hornik, Ross Ihaka, Claire D. McWhite, Paul Murrell, Reto Stauffer, Claus O. Wilke, Georg J. Mayr, and Markus Dabernig. v2.1-0. *colorspace: A Toolbox for Manipulating and Assessing Colors and Palettes*. *https://CRAN.R-project.org/package=colorspace*.

ACKNOWLEDGEMENTS

*A **special thanks** to **Amy Cramer**, **Annemie Chiau**, **Axel Muehlig**, and **Tad Lewandowski** for their tremendous support and expert guidance.*

This paper was produced using **Quarto** (Allaire v1.2).

RECOMMENDED READING

- **Online**
  - *Analyzing US Census Data: Methods, Maps, and Models in R*
  - *Basic Spatial Point Pattern Analysis in R*
  - *Decentralized Clinical Trials for Drugs, Biological Products, and Devices*
  - *Diversity Plans to Improve Enrollment of Participants From Underrepresented Racial and Ethnic Populations in Clinical Trials*
  - *Geocomputation with R*
  - ***GeoDa**—An Introduction to Spatial Data Science*

- *Geographic Data Science with R: Visualizing and Analyzing Environmental Change*
    - *Repository providing code to reproduce results for "Measuring Inequality Beyond the Gini Coefficient May Clarify Conflicting Findings" by Blesch et al. (2022)*
    - *Spatial Data Science With Applications in R*
- **Print**
    - *Analyzing US Census Data: Methods, Maps, and Models in R*
    - *Geocomputation with R*
    - *Geographic Data Science with R: Visualizing and Analyzing Environmental Change*
    - *Spatial Analysis with R Statistics, Visualization, and Computational Methods*
    - *Spatial Data Science With Applications in R*
- **R Packages**
    - ***googleway**: Accesses Google Maps APIs to Retrieve Data and Plot Maps*
    - ***npi**: Access the U.S. National Provider Identifier Registry API*
    - ***rgeoda**: R Library for Spatial Data Analysis*
    - ***sf**: Simple Features for R*
    - ***tidycensus**: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*
    - ***tmap**: Thematic Maps*

CONTACT INFO

Your comments and questions are valued and encouraged. Contact the author at:

Mark Rothe
Janssen
1000 U.S. Route 202 South
Raritan, New Jersey, United States 08869
Email: mrothe@its.jnj.com