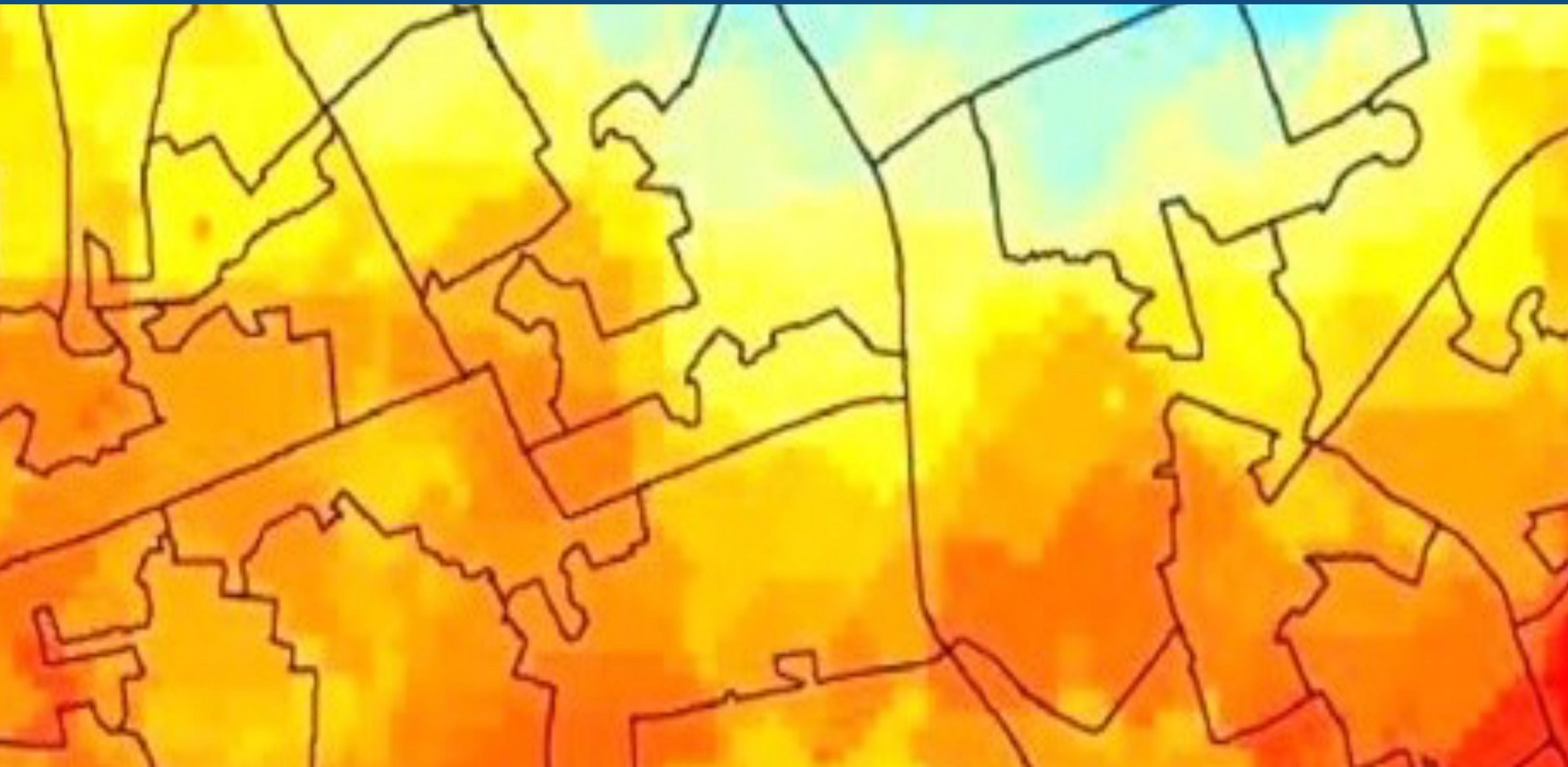
Green River
167 Main Street, Suite 103
Brattleboro, VT 05301
802.257.0641 | www.greenriver.com

Creating HIPAA-compliant Granular Disease Topologies: Better Maps

White Paper

November 2022



Creating HIPAA-compliant Granular Disease Topologies: Better Maps

White Paper
November 2022

Primary authors:

Michael Knapp, PhD, MEM.

Chief Executive Officer, Green River Data Analysis, Inc.
Todd Blackman, MS, BS.

Software Engineer, Green River Data Analysis, Inc.

Contributors:

Matthew Muspratt, JD, MEM, Green River Data Analysis, Inc.

Katiana Bougouma, MPH, Scriptoria Solutions

Acknowledgments

Much of Green River's work with static and animated heat maps described in this white paper has been conducted for the Delaware Department of Health and Social Services (DHSS), particularly the agency's *My Healthy Community* public health data platform. The authors are grateful to their DHSS colleagues for their collaboration. They especially thank state epidemiologist Dr. Tabatha Offutt-Powell and former Public Health Informatics Bureau chief Marcy Parykaza for the opportunity to apply this mapping methodology and technology in a comprehensive, public-facing environment. Many of the examples herein derive from Green River's work with DHSS. Previously, foundational work in the methodology discussed was conducted and presented in various fora by one author—Michael Knapp—in collaboration with Gary V. Archambault of the Connecticut Department of Public Health. Finally, the authors wish to recognize Professor Emeritus Gerard Rushton of the University of Iowa, whose work in spatial analysis of health was instrumental in developing the authors' methodology and technology.[1]

[1] See, for example, Richards, TB. et al. (1999). Information Technology: Geographic Information Systems and Public Health: Mapping the Future. Public Health Reports. 114: 359-73.; Rushton, G. et al. (1995). A geographic information analysis of urban infant mortality rates. International Journal of Geographical Information Science - GIS. 5.

1. Executive summary

In recent years, the availability of accurate, accessible, and adaptive tools to visualize rapidly evolving health information has become essential to support effective public health response strategies. During the recent COVID-19 pandemic, many state and local governments heavily relied on real-time public health maps for continual disease surveillance. These maps allowed targeted public health decision-making and enabled responses that kept constituents informed of disease dynamics for cohesive action to slow disease transmission. Public health maps have an immense potential to support data-driven and evidence-based decision-making to fight disease spread and improve health systems; they can help health and policy leaders better understand and address the root causes of both infectious disease transmission and chronic disease spread. In the latter, public health maps' potential for targeted action remains mostly untapped. Such maps could illuminate the factors that drive the distribution of non-transmissible, chronic conditions like cancer and help us use readily available ecological data to differentiate random disease occurrences from cases linked to an identifiable cause.

Dynamic and accurate public health maps open up the possibilities for policy and public health action that can significantly positively impact disease control and prevention. Among the ways disease spread is commonly represented in public health communication, *heat maps* are some of the most compelling tools to visualize detailed health information. Heat maps simplify and display complex statistical phenomena occurring over space and time while highlighting the spatial trends of disease distribution. Public health heat maps can display patterns of disease and health in ways that are accessible to audiences at all levels, finely detailed to support decision-making, and fully anonymized to respect the privacy laws to which they must comply.

One major bias that traditional choropleth maps consistently wrestle with and which can skew their interpretations is the Modifiable Area Unit Problem (MAUP). The MAUP stems from using arbitrary, often administrative boundaries to delimit unit areas for disease rate calculation when disease spread does not respect these arbitrary units. This bias does not exclusively affect the public health sector. However, it is an unavoidable consideration given the 'high stakes' nature of disease surveillance in terms of the immediate safety of the population. Two other common biases that affect traditional heat maps are the Large Area Fallacy (LAF) and Disease Rate Over-Sensitivity (DROS). LAF can falsely represent the extent to which large but low-population areas contribute to the overall disease burden. DROS can make heat maps seem unreliable by displaying seemingly drastic disease rate changes that stem from the unstable disease rates often observed in large, low-population rural area units. Beyond these common biases, heat maps also face the additional challenge of complying with the anonymization requirements of the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA).

To address these challenges, Green River has pioneered a methodology for producing static and animated heat maps that can 1) simultaneously circumvent the MAUP, 2) meet the HIPAA privacy requirements that apply to all health data, and 3) render the fluidity of real disease transmission dynamics---despite the use of unit area rates, which bring about the LAF and DROS---while conserving the high accuracy, precision, and detail required of heat maps. This methodology has been successfully deployed since 2018 as part of the State of Delaware's Department of Health and Social Services *My Healthy Community* dashboard. This environmental public health tracking network supports the effective dissemination of state health information.[2]

Supported by the State of Delaware's advanced Information Technology (IT) infrastructure and its funding of the mapping technology described in this paper, Green River's heat map methodology was refined and successfully implemented for COVID-19 surveillance and reporting in 2020. This development marked the creation of one of the most accurate and intuitive COVID-19 data dashboards used to disseminate real-time pandemic information and directly responded to the growing need for technology integration as a crucial health determinant.[3] As the utility and adaptability of this mapping methodology are continually validated, there is an incredible opportunity to scale the mapping technique to more intricate and tenacious issues in public health: chronic disease spread and distribution.

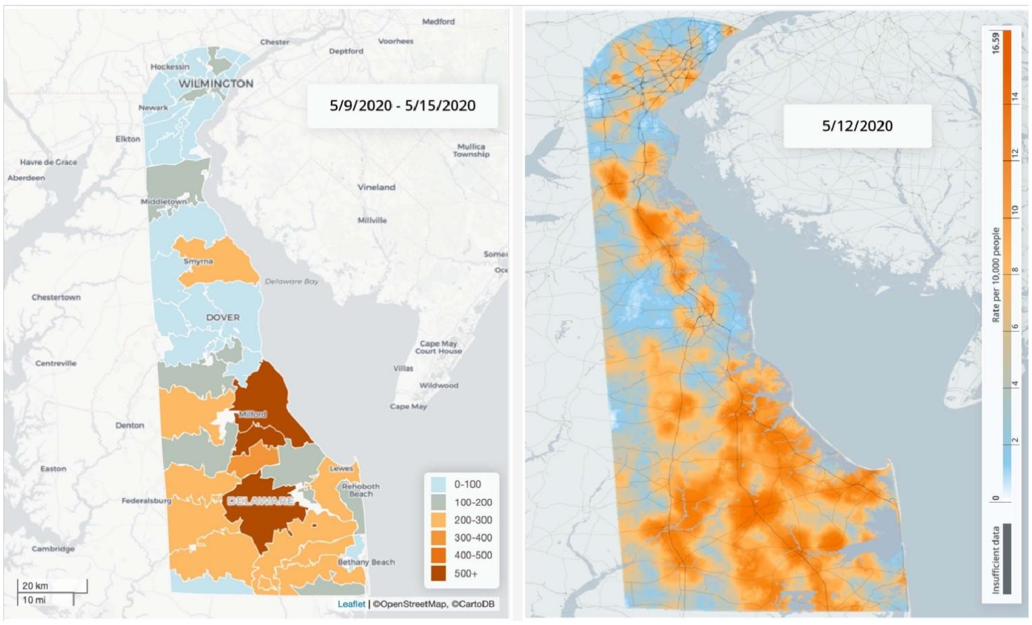
This paper details how Green River's technique deploys novel point-latticing methodologies to overcome the MAUP, LAF, and DROS. The mapping method produces fully HIPAA-compliant heat maps and animations that provide dynamic, accurate, and intuitive health information to communities and decision-makers.

[2] Delaware Department of Health and Social Services, My Healthy Community, https://myhealthycommunity.dhss.delaware.gov/locations/state.

[3] Highberger, JP, and Merriman-Nai. S. (2021). The Value (and Nuances) of Mapping as a Public Health Tool. Delaware Journal of Public Health - Technology and Public Health. (2021). Vol 7, 3.

1. Introduction

Maps A and B in Figure 1 are health information visuals published on the State of Delaware's *My Healthy Community* platform, developed by Green River. Which of the two maps in Figure 1 would give a decision maker more detailed information? Both Map A and Map B employ color shading to indicate data values in a particular geographic area; in this case, the rates of new COVID-19 cases across the U.S. state of Delaware for a specific period. However, Map B offers far more information about where the full spectrum of disease density is located than Map A.



Map A (rate per 100,000) **Map B** (rate per 10,000)

Figure 1: A choropleth map (Map A) vs. heat map (Map B) of new COVID-19 cases in Delaware. For a closer look at animated versions of these maps, please visit greenriver.com/blog.

Map B is a *heat map*, which—aside from the contextual outline of Delaware's border—describes disease case density irrespective of political or geographic boundaries. Map A is a '*choropleth map*', which also describes disease case density but does so according to ZIP code. This administrative boundary is too large to show detailed geographical nuances in disease spread. It is also arbitrary, as it has no bearing on disease and pathogen transmission. There may be appropriate occasions to visualize disease and health by geopolitical region: bird's-eye comparisons by state can be valuable, and county demarcations may govern public budgets and operations. However, heat maps and animations like the above examples offer significantly more comparative advantages than choropleth maps as public health tools. They are a viable way to visually present patterns of disease and health in enough detail to be useful while preserving anonymity.

Moreover, heat maps are a compelling visualization method. They simplify highly complex statistical phenomena occurring over space and time, enabling the public to observe trends and intuitively differentiate between disease clusters---geographically grouped cases of diseases that can be linked to a harmful source or agent---and random disease cases. While essential to engage the right public health response, these differences can be challenging for the public health community to explain to lay audiences when guiding accurate interpretations of health maps. In addition to accuracy, maps must meet minimum privacy standards. Maps A and B (Figure 1) are derived from the same dataset,[4] containing names, addresses, and personal health information that the law and civil liberty sensibilities demand be kept confidential. Both choropleth and heat maps can preserve privacy, but only Map B does so while remaining detailed, informative, and useful for public health decision-makers.

Heat maps' comparative advantage to choropleth maps

Heat maps and animations are dynamic visuals that can allow public health professionals to communicate health information in more detail than traditional choropleth maps. Heat maps' color gradations finely and continuously illustrate the slightest variations in disease rate. Their boundary-independence can also reflect the apolitical nature of pathogen transmission, non-infectious condition occurrence, and environmental disease spread, which afflict constituents without respecting arbitary administrative boundaries like ZIP codes.

This white paper describes how Green River—an impact-focused software and analytics firm based in Brattleboro, Vermont—uses software to convert raw data containing individual personal and health details into anonymized public health heat maps. Since presenting data in informative-yet-anonymized forms is a high priority in public health research, we propose this methodology as one solution to the ongoing conundrum of preserving anonymity while providing sufficient detail to make information useful in public health and its communication.

2. Compliance with the HIPAA Privacy Regulations

Community members need and want to know about health risks in their immediate vicinity, including COVID-19 infection, unsafe drinking water, carcinogens, opioid use disorder, or any other health threats, events, and outcomes that may impact them. Maps are superior tools for communicating such information to a broad audience, as visual representations are far more accessible than endless rows and columns of data. However, other easily understood graphics, such as tables, bar charts, and trend lines, lack the geographic dimension that enables a lay observer to understand where, when, and in what pattern health risk events occur.

Law and ethics preclude dropping map pins that would correlate health conditions with personal addresses. Such a map would be unacceptably invasive and illegal for most handlers of health data, per HIPAA rules. However, even anonymizing a dataset's key identifying variables is often insufficient to preserve privacy. Laranya Sweeney of Harvard University has shown that nearly 90 percent of Americans can be uniquely identified by just their ZIP code, gender, and birthday.[5]

---

[4] Both sets of visualizations display rates of new cases across the same period, and both derive from the same dataset. However, neither display rates from a single day and, relatedly, these particular choropleth and heat maps employ different multi-day averaging conventions. The result is two sets of visualizations that are not perfectly comparable in terms of the rates displayed for a particular day, but which nonetheless suitably demonstrate the contrasting granularity and patterns of these distinct plotting tools.

[5] Sweeney, L. (2000). Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3, Pittsburgh. Available at: https://dataprivacylab.org/projects/identifiability/paper1.pdf

In the United States, most use and disclosure of health information must employ de-identification methods that meet a standard under the Health Insurance Portability and Accountability Act of 1996, or HIPAA. The act's Privacy Rule, expressed in 45 C.F.R. § 164.514(b) Implementation specifications: Requirements for de-identification of protected health information, deems "*health information*"—a comprehensive term spanning an individual's condition, care, and health expenses—to be "*…not individually identifiable…*" unless *"…the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information"*.[6]

Green River's anonymized heat maps have been HIPAA certified since 2020,[7,8], thanks to our statistical methodology and mapmaking process that (1) does not employ *direct identifiers*, e.g., names or telephone numbers, to produce data and reports, and (2) requires *indirect identifiers* or non-unique attributes or any combination thereof, to meet an acceptable anonymity rule. Such a rule establishes minimum counts and threshold proportions for population groups under investigation. The general principles in the mapmaking process are described below, and a separate white paper detailing Green River's application of this anonymity technique to public health data is forthcoming.[9]

3. Addressing MAUP bias

There are, of course, many ways to ensure maps comply with HIPAA. Choropleth maps can readily satisfy HIPAA requirements by selecting convenient, i.e., large enough, geographic unit sizes or aggregating and suppressing counts. Yet, as suggested above, such maps are chunky, unrefined, and lack the details necessary for actionable insight. Choropleth maps also suffer a problematic yet unavoidable geographic partitioning bias.

The modifiable areal unit problem, or MAUP, is a bias to which choropleth maps are particularly prone. The MAUP, in short, stems from how the patterns of choropleth maps can be misleading, as they are determined by arbitrary geographic grouping that in no way serve as boundaries for public health risks. Choropleth maps calculate disease events per population total to calculate the disease rate. As such, the population groups —as determined by the chosen geographic grouping—dictate which disease events are associated with which populations, and the corresponding generate choropleth visualizations.

As a blunt example, imagine Figure 2 depicts a territory of 144 residents, 24 of whom are afflicted with a health condition X, for a total rate of 24/144, or 16.7%.

[6] 45 C.F.R. § 164.514(b) Implementation specifications: Requirements for de-identification of protected health information. https://www.law.cornell.edu/cfr/text/45/164.514. The law applies to "covered entities," which include healthcare providers, health plans, healthcare clearinghouses, and "business associates," a term that spans a broad range of data handlers and their subcontractors.

[7] 45 C.F.R. § 164.514(b)(1).

[8] Scheuren, F. and Baier, P. (2020). HIPAA Certification for Green River Data Analysis, LLC. On file with the authors.

[9] Please visit https://www.greenriver.com/blog for the latest white papers from Green River.

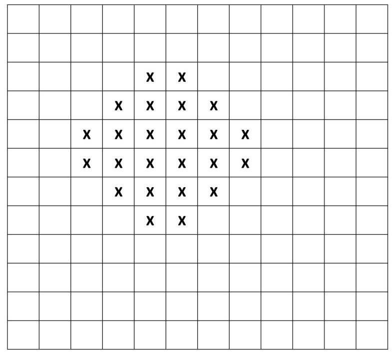
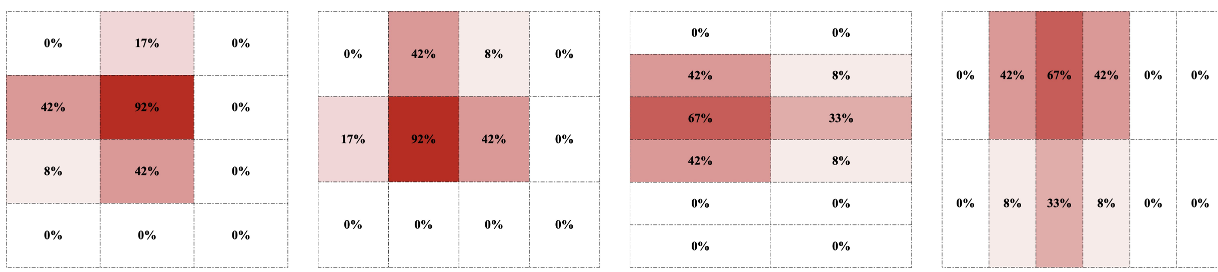


Figure 2: Grid representation of 144 residents (individual squares), including 24 afflicted by condition X (squares containing crosses)

In a choropleth map, map coloration depends entirely on which geographic grouping is chosen. Figure 3 demonstrates four possible choropleth maps that could be produced from the same data based on four different methods of geographic grouping. All four methods aggregate the 144 residents into 12 equal areas and populations that could represent, for example, (a) 12 ZIP codes, (b) census tracts, (c) congressional districts, or (d) school districts.



(a) ZIP code (b) Census tract (c) Congressional district (d) School district

Figure 3: Possible choropleth maps that can be generated from Condition X disease status data using (a) zip code, (b) census tract, (c) congressional district, or (d) school district geographical boundaries

Observers could easily make conflicting conjectures about the pattern depending on which map they were viewing. The *ZIP code* and *census tract* maps imply a significant central hotspot, with up to 92% of the population affected; the *congressional district* map suggests the condition is more pronounced in the territory's western reaches, while the *school district* map places the trend in the north. These four possible mapping outcomes perfectly illustrate the MAUP bias, where different conclusions can be drawn from the same data based on arbitrary geographic grouping. Similarly, maps of COVID-19 vaccination rates in Delaware by ZIP code and census tract impact our understanding of precisely where rates are high and low, particularly in the state's southeast corner (Figure 4). The true pattern is unclear because the geographic grouping biases the observable patterns. Gerrymandering, for instance, is yet another form of MAUP that is notorious for making its way into American politics.[10]

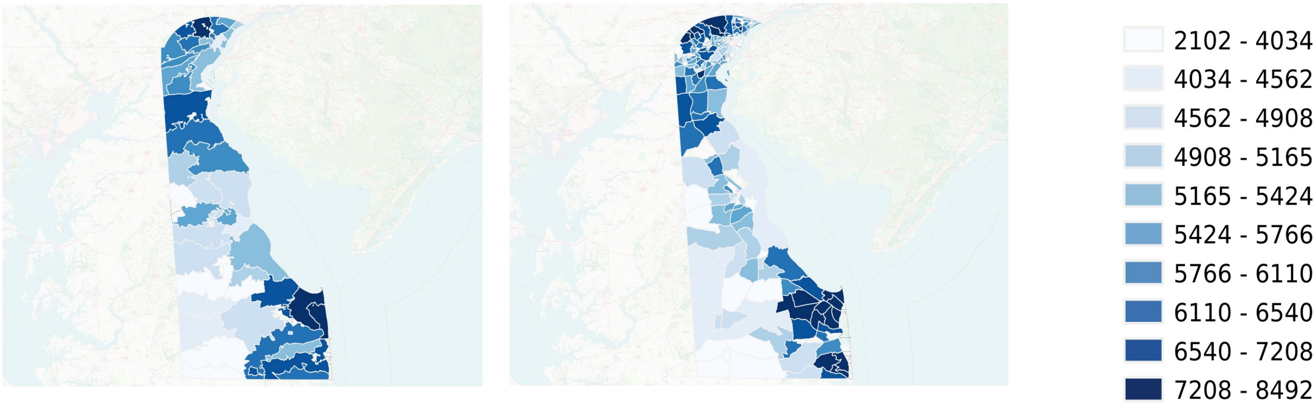


Figure 4: Full COVID-19 vaccination rate per 10,000 people by ZIP code (left) and census tract (right) in Delaware, November 2021

Green River's heat map generation methodology avoids the MAUP. As detailed below, the methodology defines health events and population groups independently from geographic groupings. Instead, groupings are governed by proximity to defined lattice points uniformly distributed across the map area. This method renders a depiction of the true underlying disease pattern.

4. Addressing rural visual misperception biases

Another challenge in visualizing health data is that, ideally, data should be conveyed per unit population, not per unit geographic grouping, as not all areas are equally populated. The challenge manifests in at least two related ways: 1) Large Area Fallacy (LAF), whereby maps can offer a misleading 'footprint' across geographies of varying size and population density, and 2) Disease Rate Over-Sensitivity (DROS), whereby lower total population numbers in rural areas make the diseased-to-total population ratio highly sensitive to small changes in the number of disease cases, causing disease rates to appear to fluctuate dramatically.

Overcoming rural visual misperception biases in heat maps

Green River accounts for visual misperception biases of rural areas depicted on heat maps by minimizing the appearance of heat map spikes or hotspots in low-density regions. By "borrowing" populations and events from nearby areas, this method allows local rates in sparsely populated rural areas to approach a stability resembling that of more densely populated areas. This approach ensures that changes in heat map coloration are linked to significant rate changes, regardless of where they occur.

4.1 The Large Area Fallacy bias

[10] See, *for example,* Ingraham, C. (2015), "This is the best explanation of gerrymandering you will ever see," *The Washington Post* Wonk Blog, March 1, 2015; Buzzell, M. (2020). Modifiable Areal Unit Problem. International Encyclopedia of Human Geography, 169–173. https://doi.org/10.1016/B978-0-08-102295-5.10406-8.

The LAF stems from rural areas typically occupying larger physical spaces than urban population centers. As a result, the eye can overestimate the impact of large, scantly populated rural areas on the overall public health situation. For example, dark red across the 8,000 square miles of the lightly populated Carbon County, Wyoming, could be interpreted as contributing much more to the national disease burden than the dark red across 300 square miles around densely populated New York City; in reality, New York City may house many more cases. Similarly, a 1% increase in Carbon County's rate —an additional 150 cases —would have a much smaller impact on national public health than a 1% increase in New York City's rate, which would add 84,000 extra cases.

Map-stretching techniques, which involve collapsing less dense areas and expanding urban centers, are one way to reduce LAF, where large areas do not represent large populations. Figure 5 shows one example of a *density-equalizing* method obtained by map stretching, where the denser areas are visually expanded on the maps. As Figure 5 demonstrates, however, one clear shortcoming of these visualizations is that the maps are deformed and look unfamiliar, potentially confusing readers.

Figure 5: Lung cancer cases among males in the State of New York, 1993–1997[11]

4.2 The Disease Rate Over-Sensitivity bias

Disease rate expresses the percentage of the total population affected by a given disease. This percentage fluctuates as the number of disease cases in a geographic unit changes. In rural areas with small populations, even one additional disease case can appear to increase the disease rate drastically, resulting in DROS. On a map, DROS may appear to be a rapid shift from light to dark red, for example, when the number of cases in the area may have changed only slightly. Accounting for DROS allows us to differentiate between changes in case numbers involving a few random occurrences from those that suggest significant local shifts in disease transmission. In our example, a fluctuation of 2–3 cases in Carbon County would greatly affect its disease rate as its population is very small. However, 2–3 additional cases could realistically also occur by chance, so mobilization of public health resources would be an inappropriate response. Meanwhile, a proportionally similar fluctuation in New York City would correspond to an additional 1,100

[11] Gastner, MT, and Newman, MEJ. (2004). Diffusion-based method for producing density-equalizing maps. Proceedings of the National Academy of Sciences May 2004, 101 (20) 7499-7504; DOI: 10.1073/pnas.0400280101

cases. That figure is far beyond what could simply be attributed to chance and would likely require public health action.

4.3 Addressing rural visual misperception biases

Green River accounts for the LAF and DROS in rural areas by minimizing the appearance of heat map spikes or hotspots in low-density regions. This approach better aligns the visual effect of the change with its impact on broader public health by minimizing the visual effect of changes in events that do not have a broad public health impact or may be attributable to chance. This is achieved by considering a larger total population, including population totals from nearby areas, to calculate disease rates in specific locations and employing a smoothing technique in animated heat maps. By 'borrowing' populations and events from nearby areas, this method allows local rates in sparsely populated rural areas to approach a stability resembling that of more densely populated areas. This approach ensures that changes in heat map coloration are linked to significant rate changes, regardless of where they occur. Further details of this approach, including the rate grouping rules that inform the color gradient levels, are described below.

5. Methodology: Green River's static and animated maps

Green River's method for processing raw health data and rendering it as static or animated heat maps is complex and computationally intensive. The method employs iterative software-based calculations that apply spatial data and geographic information system tools to census data. As described above, the approach leverages rates rather than counts, avoids MAUP and rural visual misperception biases, and preserves the HIPAA-compliant anonymity of health data while depicting rate changes over space and time. Here are the key steps of this process:

1. Our heat map comprises a large set of points, each assigned x, y, and z values. x and y represent the point's longitude and latitude coordinates, while z represents the disease rate under investigation, e.g., the rate of positive COVID-19 cases, calculated for that point. z will correspond to a particular coloration on a color gradient in the final heat map.

   The first step is the generation of a lattice or net of these points, which are equally spaced and arranged to cover the entire map area. There are several options for the precise spacing of the points, each with advantages and disadvantages. A dense lattice may produce more precise visual details in highly populated regions but requires higher computer processing. A less dense lattice requires less processing power but, in extreme cases of very few points, might render a more pixelated product with a level of detail resembling a choropleth map. For reference, in Figure 5 above, the point spacing is 0.004 degrees.

2. At each lattice point, a circle is initialized with a radius equal to the point spacing. This ensures that the circumference of a given lattice point's circle touches the four closest points to its north, south, east, and west, and that the entire map surface is covered by at least once circle, as shown in Figure 6.

Figure 6: Sample map with the entire surface covered by at least one circle

3. The lattice is then populated with event locations, e.g., the address of an individual with a positive COVID-19 case, which are '*snapped'* from their true location to the nearest lattice point, to be combined and tallied for that point.

   Snapping is a component of our HIPAA-compliant anonymization process. It is not required for calculating rates or generating heat maps and could be foregone in mapping contexts where privacy is not of concern.

4. Next, an anonymity test is performed based on event rates within circles. For each circle's associated lattice point, a rate is assigned, determined by the following factors:
   - The number of events within the point's associated circle as the rate's numerator
   - The total population, an estimate based on the circle's area and census data, as the rate's denominator.

   Generally, a population estimate of ≥ 500 people and an event count of *either* 0 *or* ≥ 5 is sufficient for a circle to be considered anonymous as long as the resulting rate calculated from that circle is less than 90 percent. The identification of individuals is more feasible at rate extremes. For example, a 450/500 (90 percent) rate discloses that nearly everyone in a location associates with the investigated condition. In contrast, a 1/500 rate exposes an individual to the re-identifying techniques and probabilities that Sweeney discusses. If a circle is not sufficiently anonymous, its size can be increased to cover the next closest lattice points, and the rate re-calculated. Expansion to the next closest lattice points is, again, a concession to HIPAA compliance and the *snapping* of events to lattice points in Step 3. In contexts with less stringent privacy concerns, where events remain at their original coordinates, circle expansion can progress more conservatively by expanding only to encompass the next closest event before re-calculating. At this point, there are two possible outcomes:
   - The rate for the lattice point satisfies the test for sufficient anonymity, or
   - A value of zero (0) is assigned for the point's rate after *n* iterations of circle expansion, with the number *n* determined experimentally.

   An important consequence of this process is that the disease events *snapped* to any given lattice point may count towards the rate calculation in more than one circle and therefore be represented in the calculated rates of more than one lattice point. Double-counting certain individuals is not problematic in this case since the health data is not tallied across more than one lattice point at once. Rather, the map displays the spatial spread of disease rates within each point's respective radius. This means that every pixel in the final heat map presents information *snapped* from the surrounding area and accurately represents location-specific rates for the entire area it covers.

For map areas with varying population densities, the circle expansion process also applies the smoothing technique described previously; circle expansion increases rate denominators by including surrounding populations in their total population count from nearby *snapped* points. Smoothing makes rates more resistant to small changes in disease events and lends rural rates a stability rate similar to that of denser regions. This, in turn, minimizes misleading or distracting coloration changes in heat map animations.

Depending on the event and population being investigated, large circles may return rate information of little use; a map generated from only a few large circles will lack the granularity these visualizations are meant to achieve.

At this stage, the process has yielded a collection of circles, each with a lattice point that has been assigned a 0 or a number between 0 and 1, representing a rate.

5. The above steps are repeated for data across each available time period. To produce animated maps, the static map for each time period merges into a single frame in a final animated heat map.

6. The full set of calculated rates across all time periods is ordered, separated into ten equal groups (deciles). A color scale is chosen to represent the entire range of rates, essentially assigning a color to the *z* value for each point. The set is modified—and the standard definition of decile is deviated from it—in a few ways:

   - Zero rates are counted only once in a set to maximize the color range. Otherwise, zeros would consume most of the color gradient if they were common across the map.
   - The lowest and highest 2 percent of values are excluded from the set to reduce the impact of outliers on the color gradient.
   - The lowest limit of the lowest rate decile is pre-defined as zero, regardless of the lowest value in the full set of rates.
   - An off-gradient color, e.g., gray, is assigned to rates with a value of 0, representing the lattice points and associated circles that did not reach sufficient anonymity.

7. Because the lattice points themselves are not sufficiently numerous and dense to constitute a high-resolution heat map, a smoothing technique called inverse distance weighted (IDW) interpolation is applied for each time period to generate additional points (pixels) between all the initial lattice points, each with their own *z* value and associated color based on neighboring points' rate values.

   In this step, upon IDW interpolation, both the original lattice points and the interpolated points can adopt different colorations based on their *z* values, which change along with new disease event counts. Point data has then effectively been converted to a true, static heat map image colored according to the scale determined from the deciles described above for each time period.

The smoothing process in this step also provides an additional layer of anonymity. For example, in the case of a single high-rate lattice point surrounded by zero-rate points, smoothing *blurs* the distinctive rate by generating points of intermediary rates nearby. Where a map contains areas of high population density with relatively few events, this smoothing also ensures the visualization does not suggest that the few disease events are more likely to occur in one neighborhood than another.

8. Time intervals in public health data are typically quite long since data might be collected monthly, yearly, or even every five years. Due to this, the above process typically uses data from distant points in time and results in only a few static heat maps. The result is a choppy animation switching rapidly from one coloration to another. Another interpolation technique can create a fluid animation by generating additional intermediate static maps. This can be done by estimating rates at each lattice point, i.e., z values, based on the known z values at the known time intervals. The number of additional frames generated is determined by balancing processing power limitations and the frame rate required to produce a sufficiently fluid animation experience.

6. Caveats

Heat maps are not without shortcomings and limitations. The potential for rural visual misinterpretation remains despite our mitigation techniques. Depending on the topic of investigation, heat maps may obscure or blur disease sources or causes. For example, a heat map of lung cancer rates would be overwhelmingly influenced by cases of smoking and secondhand smoke exposure, making the tool largely unhelpful as a means to investigate lung cancer rates stemming from local industry. Conversely, other cancer types might render as persistent, crystal-clear clusters in the industrial area. Similarly, heat maps are best applied to tracking the dynamic disease rate and spread, i.e., surveillance methodologies, and performing ecological analyses. They are a far less relevant tool for some longitudinal studies whose participants' health data is periodically gathered and updated collectively, e.g., in cohort studies, or case-control studies that compare afflicted and healthy individuals to identify possible disease sources.

7. Conclusion

The utility of heat maps stems from their robust depiction of disease rates across space and time in an accessible format for all. Raw data, or even considerable statistical results, cannot match the efficiency of a heat map in intuitively revealing disease patterns to the human eye. Standard mapping techniques suffer from MAUP and rural visualization misperception biases, i.e., LAF and DROS. In contrast, Green River's innovative methodology overcomes these potential biases, offers more refined detail, and preserves HIPAA-compliant confidentiality.

Public health is intrinsically concerned with trends in disease across space and time but often finds it challenging to effectively communicate data to experts and lay audiences while preserving personal privacy. The challenge is amplified when it comes to visual communication products like graphs, trend lines, and maps, which are more easily digestible but more prone to misinterpretation. Thus, public health has a unique opportunity to leverage the power of maps that convey detailed information while respecting privacy and overcoming common map biases. Green River's heat maps equip communities and public health professionals with an accurate intuitive tool to better understand their disease risk, sustain their well-being, and make informed decisions for themselves and others.