Visualizing Measles Incidence in the USA (1928–2001)

Zhang XiangHui, Lin WeiChen, Priscilla Thung, Tang Guan You, Arshad Bin Mazlan, Wong Jing Yong Shawn (Information and Communication Technologies)

SINGAPORE INSTITUTE OF TECHNOLOGY

Introduction

The COVID-19 pandemic has transformed numerous industries, and the food and beverage sector is no exception. One of the most notable developments during this period has been the rapid growth of virtual kitchens, also known as ghost kitchens or cloud kitchens. These innovative culinary ventures have gained significant traction, reshaping the landscape of food service.
Virtual kitchens are commercial cooking facilities that prepare food solely for delivery or takeout, without the traditional dine-in option. Unlike conventional restaurants, virtual kitchens often operate multiple brands or menus from a single location, optimizing efficiency and catering to diverse consumer preferences. This model leverages modern technology, including advanced logistics and digital platforms, to streamline operations and enhance customer experience.

Previous Visualization

Figure 1: Number of virtual restaurants listed on Uber Eats from 2019 to 2023.

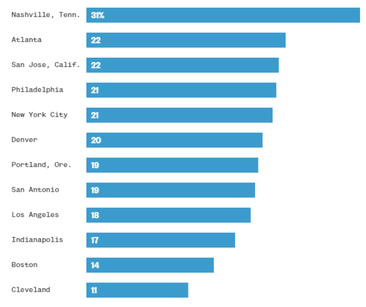

Figure 2: Percentage of restaurant listings in a city.

Strengths

- The heatmap design effectively conveys a high information content without cluttering the plot.
- Pointing with the mouse at a tile opens an infotip, enabling readers to retrieve specific incidence data for a given state and year (**?@fig-infotip_color_change**). The infotip only occludes a small portion of the plot, and the partial transparency of the infotip ensures visibility of the tiles underneath.
- The vertical line indicating the year of vaccine introduction provides valuable contextual information.

Suggested Improvements

1. *Add a plot title and a source note* so that the figure can be understood in isolation (e.g., when shared on social media).
2. *Identify missing data clearly.* Rendering unknown incidence fully transparent will distinguish it from zero incidence,
3. *Include labels for every state.* To avoid overplotting, use two-letter abbreviations instead of full state names and stagger the labels along the y-axis.
4. *Add a title to the color legend.*
5. *Avoid using a rainbow color palette.* It lacks a meaningful progression through color space and is not colorblind-friendly. Consider using a sequential ColorBrewer palette instead.[1]
6. *Use a discrete color palette.* Continuous palettes can make it challenging for humans to detect patterns below just noticeable color differences.
7. *Apply a logarithmic color scale* because most data are below the mean incidence.
8. *The x-axis should end at 2001* as there are no data afterwards.
9. *Add grid lines* in ten-year intervals along the x-axis and for every second state along the y-axis. Grid lines will aid in identifying states and years in the middle of the plot, even without the infotip.
10. Because there are more missing data on the right side of the plot, *shifting y-axis labels to the right* will improve visually matching states with corresponding grid lines.

Implementation

Data

- Weekly counts of measles cases by state were obtained from Project Tycho.[2] The data have missing weeks, which were treated as zero in **?@fig-wsj-on-poster**, potentially underestimating the annual total. Instead, we calculated the weekly mean case count on the basis of non-missing data only.
- Decennial U.S. census data for each state.[3]

Software

We used the Quarto publication framework and the R programming language, along with the following third-party packages:

- *readxl* for data import
- *tidyverse* for data transformation, including *ggplot2* for visualization based on the grammar of graphics
- *knitr* for dynamic document generation
- *zoo* for interpolating annual population data from the decennial U.S. census

Improved Visualization

Further Suggestions for Interactivity

Because our visualization was intended for a poster, we did not implement any interactive features, including the infotip. However, if the data are visualized in an HTML document, interactive features can be achieved using the R packages such as *plotly*. In that case, we recommend that the tile does not change its fill color. In contrast, the original visualization changes the fill color of the activated tile to light blue (see **?@fig-infotip_color_change**), which can be misinterpreted as a change in incidence. Instead, we suggest highlighting the activated tile by thickening its border.

Conclusion

We successfully implemented all suggested improvements for the non-interactive visualization. By labeling every state and choosing a colorblind-friendly palette, the revised plot is more accessible. The logarithmic color scale makes the decrease in incidence after the introduction of the vaccine less striking but enables readers to detect patterns in the low-incidence range more easily.

[1] https://colorbrewer2.org/#type=sequential&scheme=Reds&n=5
[2] https://doi.org/10.25337/T7/ptycho.v2.0/US.14189004
[3] https://www.stats.indiana.edu/population/PopTotals/historic_counts_states.asp

Email IDs (separated by commas)