# Visualizing Measles Incidence in the USA (1928–2001)

*Michael T. Gastner* (Information and Communication Technologies)

SINGAPORE INSTITUTE OF TECHNOLOGY

## Introduction

Measles is a highly contagious viral infection that can lead to severe complications, including pneumonia, encephalitis, and death. The average number of secondary cases resulting from a single infected individual is estimated to be between 10 and 16,[1] but vaccination is reported to be 97% effective at preventing measles[2]. Consequently, the introduction of the vaccine in 1963 was followed by a significant decline in measles incidence in the USA.

Given persistent skepticism in some communities about the efficacy of vaccines, it is important to communicate the impact of vaccination on public health. In this project, we built on a visualization of the measles incidence in the USA published by the Wall Street Journal[3] (Figure 1). This visualization has garnered praise as "One of the more compelling data visualizations produced in recent years"[4]. However, we contend that, despite its effectiveness in summarizing data, several aspects of the plot can be improved.

## Previous Visualization

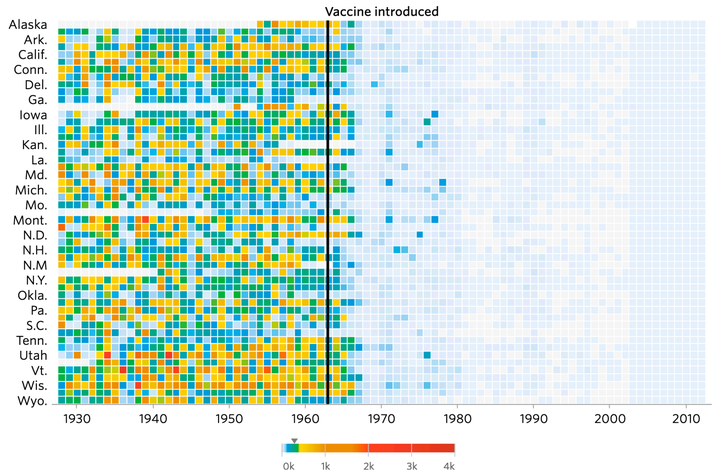


Figure 1: Measles incidence by state of the USA and year, published by the Wall Street Journal.

## 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 (Figure 2). 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.

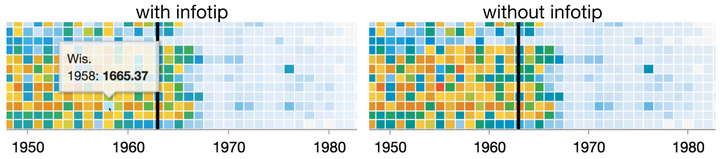


Figure 2: Zoomed-in screenshot of the original plot with and without activated infotip.

## 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.[5]
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. *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.
9. 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.[6] The data have missing weeks, which were treated as zero in Figure 1, 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.[7]

### *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

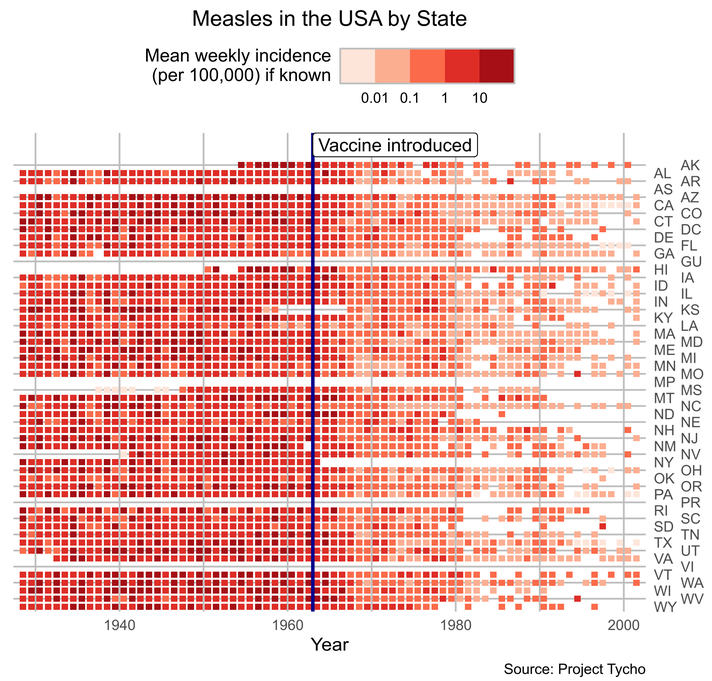


Figure 3: Revised visualization of measles incidence by state of the USA and year.

## 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 Figure 2), 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] F. M. Guerra et al., "The basic reproduction number (R0) of measles: a systematic review," The Lancet Infectious Diseases, vol. 17, no. 12, p. e420-e428, 2017

[2] L. Franconeri et al., "Two-dose measles vaccine effectiveness remains high over time: A French observational study, 2017–2019," Vaccine, vol. 41, no. 39, pp. 5797–5804, 2023]

[3] https://graphics.wsj.com/infectious-diseases-and-vaccines/

[4] https://www.mikelee.co/posts/2017-06-28-wsj-measles-vaccination-chart

[5] https://colorbrewer2.org/#type=sequential&scheme=Reds&n=5

[6] https://doi.org/10.25337/T7/ptycho.v2.0/US.14189004

[7] https://www.stats.indiana.edu/population/PopTotals/historic_counts_states.asp



michael.gastner@singaporetech.edu.sg