Exercises: Measles incidence in the US by state—ggplot2

YSC2210 - DAVis with R

Michael T. Gastner

Introduction

Vaccine sceptics are a vocal minority. According to an opinion poll by the data analytics company Gallup (2019), only 5% of the world's population disagrees with the statement: 'vaccines are effective'. Still, anti-vaccine protests receive much media attention (Stanley-Becker, 2021). How can the public be informed about the positive impact that vaccines have had in the battle against infectious diseases?

The statistical tools needed to assess the effectiveness of vaccines are beyond the scope of this course. However, data visualisation can make the data intuitively understandable. Two journalists at the Wall Street Journal (Tynan and Friedman, 2015) presented a series of heat maps that show how incidence rates of diseases dropped in the US after vaccines became available. Figure 1 is their visualisation of the measles incidence rate.

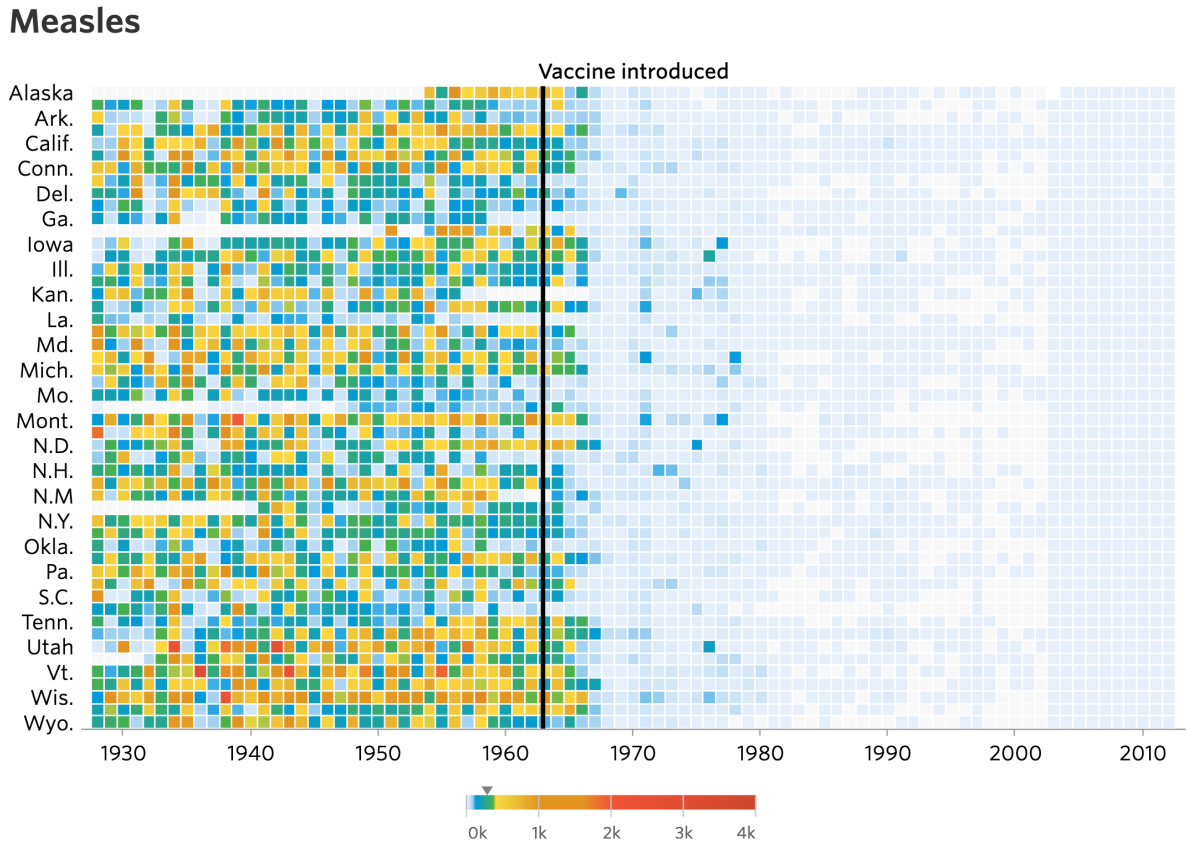


Figure 1: Image from Tynan and Friedman (2015).

Objectives

In these exercises, we practise our **ggplot2** skills by making a plot that improves the visualisation by Tynan and Friedman (2015), shown in figure 1.

Tasks

(1) Discuss with your teammates how figure 1 can be improved. Include your critique as a bulleted list in your R Markdown file.

(2) Download https://michaelgastner.com/DAVisR_data/measles_for_ggplot2.csv. This spreadsheet contains three columns:

- `state`: two-character postal codes of US states.
- `year`.
- `incidence`: numbers of measles cases per 100,000.

The data are extracted from Project Tycho (2019).

Import the CSV file as a tibble.

(3) Make a heat map in which:

- the x-coordinate is the year.
- the y-coordinate is the state.
- the colour of the tiles represents the incidence rate *plus* 10^{-6}. We add a small amount to the incidence rate because we are going to plot the incidence rate on a logarithmic scale later on. Some incidence rates are zero, and the logarithm of zero equals negative infinity. Adding a small positive value to each data point is a common trick to cure this issue.

Do not worry about the scales for the y-axis and the colour bar. We will fix them shortly.

(4) Change the axis labels and give the plot a title. Give credit to Project Tycho as data source in the form of a caption.

(5) Remove any padding between the left (right) edge of the x-axis and the left (right) edge of the panel on which the heat map is drawn. Place tick marks in 10-year intervals: 1930, 1940, . . . , 2000.

(6) Invert the y-axis so that state abbreviations appear alphabetically from top to bottom.

(7) Change the colours in the colour bar so that they are based on a sequential ColorBrewer palette. Represent the 'neutral' point (i.e. nearly zero incidence) by a light colour. Give the colour bar a meaningful legend title. Set the following arguments in `scale_fill_distiller()` so that we get an easily readable logarithmic colour bar:

- `breaks = 10^(-2:3)`
- `limits = c(2e-3, NA)`
- `labels = prettyNum`
- `trans = "log10"`
- `oob = scales::squish`
- `guide = guide_colourbar(frame.colour = "black", ticks.colour = "black")`

You may wish to add more arguments.

(8) Give tiles with missing data a neutral colour (e.g. grey), distinct from any colour in the colour bar. Add a legend for missing data (see: https://stackoverflow.com/questions/29151167/add-na-value-to-ggplot-legend-for-continuous-data-map). Place this legend below the colour bar.

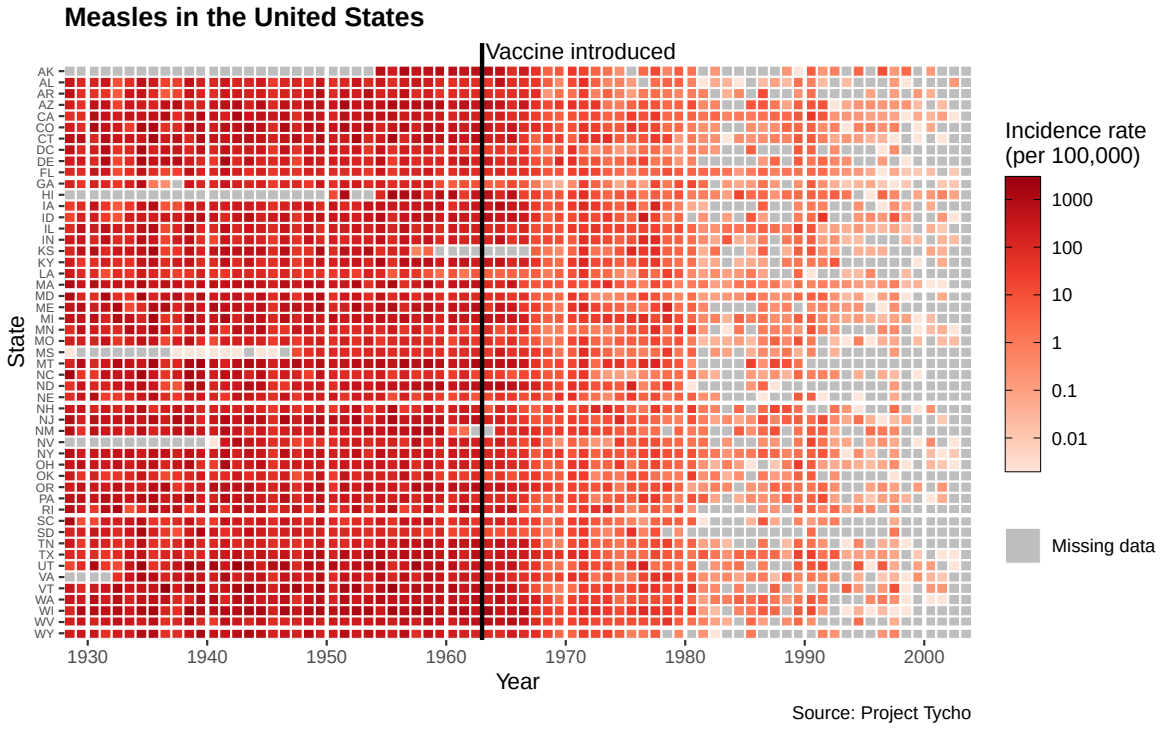
(9) Indicate with a vertical line that the measles vaccine was introduced in 1963. Add a text annotation that explains the meaning of the line.

(10) Change the plot theme with

`theme(panel.background = element_rect(fill = NA))`.

What does this command do?

(11) Feel free to make more adjustments if you think they improve the quality of the plot. Afterwards, adjust the figure dimensions in the knitted file. Labels should be clearly legible without appearing disproportionately large. Here is my attempt.



(12) Write a few sentences about the ways in which you improved the plot compared to figure 1. Were there any ideas you listed in response to task (1) that you were not able to accomplish?

References

Gallup (2019). Wellcome Global Monitor: First Wave Findings. Technical report, Wellcome, London.

Project Tycho (2019). United States of America (measles) dataset. URL: https://www.tycho.pitt.edu/dataset/US.14189004/. Accessed on 2 Feb. 2019.

Stanley-Becker, I. (2021). Anti-vaccine protest at Dodger Stadium was organized on Facebook, including promotion of banned 'Plandemic' video. Washington Post (2 Feb. 2021).

Tynan, D. and Friedman, D. (2015). Battling infectious diseases in the 20th century: the impact of vaccines. Wall Street Journal (11 Feb. 2015).