Visualizing Mode of Transport Choices by Employed Residents (2000–2020)

NavyBlue (Computing Science)

SINGAPORE INSTITUTE OF TECHNOLOGY

Introduction

This visualization shows how transportation habits of employed residents have changed over the decade. This shows how the residents' choices of transportation has changed as the public transportation infrastructure evolves over the years.

Previous Visualization

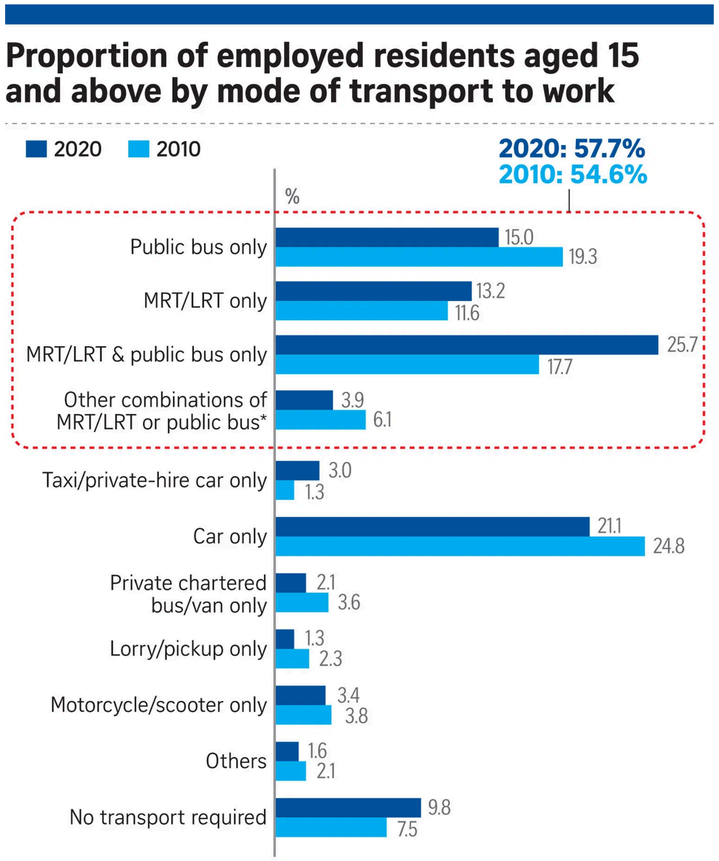

Figure 1

Strengths and weakness

- Strength: The visualization uses the bar chart format, which is clear and easy to read.
- Weakness: While different shades of blue are used to differentiate the years, it may be difficult for the color blind to distinguish between the datasets. Some categories are too specific while others are more vague. This makes it difficult to interpret the data

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. *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

- Data was collected from 2000 to 2020 at 5-Year intervals from the Department of Statistics Singapore.[2] Some data were obmitted and combined during the cleaning process.The data includes rows that are not present in the entire dataset for certain years. Therefore, rows that are not in common throughout the dataset were removed.
- 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.[3]

Software

The Quarto publication framework and the R programming language were used, 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

[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

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.