

Data Visualization Cluttering

Data visualization is more than pie charts or graphs - the powerful visual art of storytelling with real conflicts, strategies, heroes, and conclusions, hidden in data. It churns actionable insights unveiled by data. Big or small, cluttering Data can transform your visualizations into great decision-makers!

Contents

- What is Data Visualization Cluttering? 2
- What are measures to avoid data cluttering? 3
- Design Recommendations for Data Visualization 3
- Data visualization Case 4
- References 5

What is Data Visualization Cluttering?

Data Visualization is the representation of data using typical graphics such as charts, infographics, animations, and plots. These displays help understand the relationship between different data labels and features available in complex data.

Generally, Data Visualization includes techniques such as tables, pie charts, stacked bars, line charts, area charts, histograms, scattered plots, heat maps & tree maps.

The data cluttering problem occurs when the data dimension is higher.

Data Cluttering is a disordered collection of graphical entities in the formation of data visualization. Data clutter results in misinformation about the data entities.

Decision-making is impossible as it hinders readers' view of observing the patterns in data.

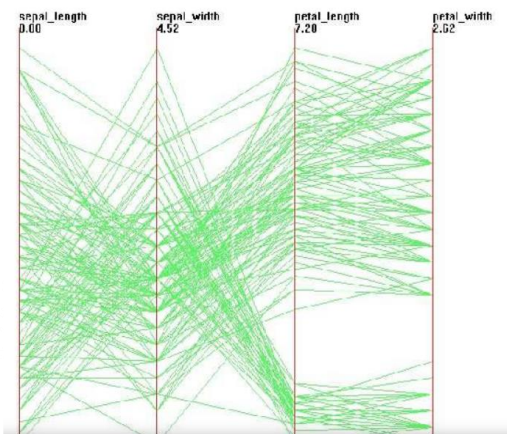


Figure.1.1. Iris dataset (4 dimensions, 150 data items) in Parallel Coordinates.

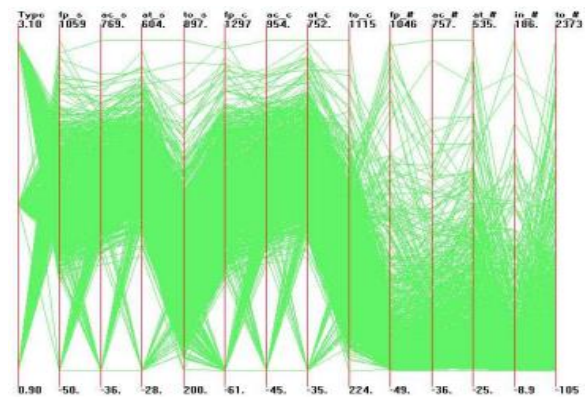


Figure 1.2: AAUP salary dataset (14 dimensions, 1161 data items) in Parallel Coordinates

There is no single type of solution for the data cluttering problem, as every cluttering results due to variation in the visualization techniques and analysis target.

- Does the question remain how to handle the cluttering problem?
- Is there a straightforward solution to solve the cluttering?
- Should I alter the data to address the cluttering problem?

What are measures to avoid data cluttering?

As mentioned above, avoiding data cluttering becomes cumbersome due to variations in the graphing techniques.

Distortion is one of the techniques to solve the problem. But the challenge is the reduction in subset data size into a smaller area, which results in difficulty for the readers to view the data.

A few techniques in the design of the data presented can help reduce the cluttering.

Dimensional Re-ordering :

In dimensional reordering, data coordinates rearrangement helps us position data in a larger area to accommodate the outliers.

But caveats are :

- Top many scattered points in the visualization make it complex
- All the rainbow colors plates usage will look the data similar
- The decision to placement the individual plots will be difficult
- Higher order chances for the audience to get misguided

To ensure a better way of data visualization, avoid the below factors:

- 3D (always)
- Color effects for decoration (always)
- Grid lines (if unnecessary only)
- Axis
- Decimals
- Play with font and text size: some text is more important than others

Design Recommendations for Data Visualization Below are some of the design recommendations for better data visualization

Chart Borders : Remove the chart border. Chart borders often provide no values, representing no clarity and adding no value to the data.

Gridlines: Delete the gridlines. Gridlines are the lower level usage where the audience never uses it. Primarily the gridlines indicate the values spread across. None will drag across the gridlines to know individual values

Data Labels:

Ensure the minimal usage of data labels is present. Highlight the essential data labels which will make the audience informative

Axis Titles:

Remember, the most critical feature of visualization is the axis titles.

The Y-axis should be descriptive

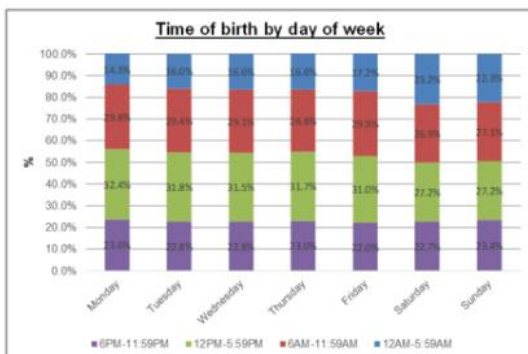
The X-axis should be clean

Data visualization Case

The dataset comprises data for the percentage of babies born within 24 hours, broken down by day of the week.

Related note: this dataset didn't include the absolute number of babies born daily. Ideally, we'd want that information for context. Still, for this illustrative example, we'll assume the numbers are large enough to compare across days of the week accurately.

BEFORE

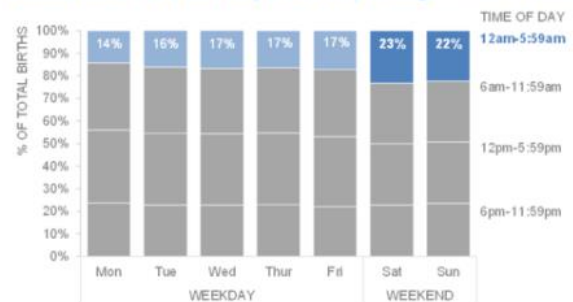


Data source: CDC (National Vital Statistics Reports, Vol. 67, No. 1, January 31, 2016)

AFTER

When babies are born

Weekend deliveries are more likely to be in early morning, compared to weekdays



Data source: CDC (National Vital Statistics Reports, Vol. 67, No. 1, January 31, 2016)

By reducing clutter, the audience can use their precious brainpower to decide what potential actions might be warranted rather than trying to figure out how to read the graph. Taking time to modify the default settings means we can focus on the data and the message.

References

1. Wei Peng, M. O. Ward, and E. A. Rundensteiner, "Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering," IEEE Symposium on Information Visualization, 2004, pp. 89-96, DOI: 10.1109/INFVIS.2004.15.

http://www.sci.utah.edu/~beiwang/publications/HighDim_Survey_TVCG_BeiWang_2017.pdf

3. Visual thinking lab. VTL Projects, Declutter, and Focus. (n.d.). Retrieved August 16, 2022, from <https://visualthinking.psych.northwestern.edu/projects/DeclutterFocus/>

4. <https://www.storytellingwithdata.com/blog/2019/5/13/declutter-and-question-defaults>

imagine.

Imagine Tech is the world's leading product engineering, innovation and talent solution provider. We're here to help you transform and elevate your business through innovation and technology-led disruption, so that you can build better products, faster. We're here to fuel the 2.0 of who you are. See how at <https://imaginetech.io/>

India
TVH Belicia Towers, 1st St, MRC Nagar, Raja Annamalai Puram,
Chennai, Tamil Nadu 600024

USA
1560 East Southlake Blvd, Southlake,
Tx 76092

© Copyright 2022, imagine Tech. All rights reserved. No part of this document may be reproduced, stored in a retrieval system, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the express written permission of Imagine Tech. The information contained herein is subject to change without notice. All other trademarks mentioned herein are the property of their respective owners.