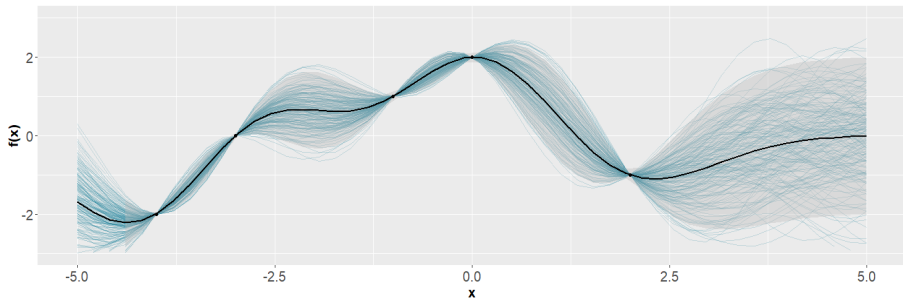


Geovisualization

An introduction to "ggplot2" with R

Joaquin Cavieres

Geoinformatics, Bayreuth University



Outline

- 1 Introduction
 - Data visualization
- 2 The role of visualizations in data analysis
 - Exploratory data analysis
 - Pattern Recognition
 - Detecting errors and anomalies
 - Communication of findings
- 3 Introduction to "ggplot2" with R
- 4 Conclusion

1. Introduction

Data visualization

It is the process of transforming complex datasets into visual formats that are easily understandable, allowing patterns, trends, and relationships within the data to be quickly identified and comprehended.

The goal of data visualization:

- Simplify complex information.
- Reveal hidden patterns.
- Presenting data in a visually appealing and intuitive manner.

Data visualization plays a crucial role in data analysis and enables data-driven decision-making by transforming numbers and statistics into visual narratives that are accessible and impactful.

Benefits of data visualization?

- Comprehension of the data set.
- Identify trend and possible "outliers".
- Effective communication of the analysis.

Good and bad examples of data visualization

Bad data visualization

Refers to visual representations of data that fail to effectively convey the intended information or insights

Bad data visualization fails to convey information clearly and accurately.

Is “China” bigger than “Other”?

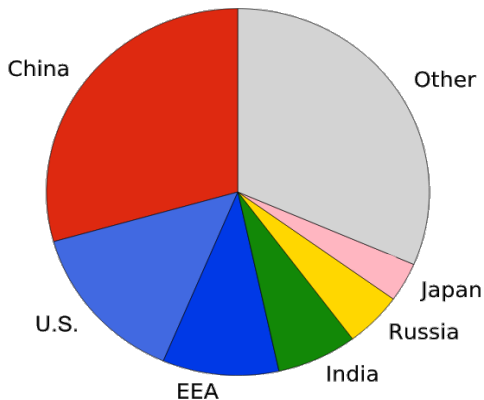


Figure: Source: <https://www.polymersearch.com/>

Good and bad examples of data visualization

Good data visualization

Refers to visual representations of data that effectively communicate information, patterns, and insights to the intended audience.

Good data visualization presents data in a visually appealing, informative, and comprehensible manner, enabling effective analysis and communication.

Notice how China immediately sticks out and we get the point across.

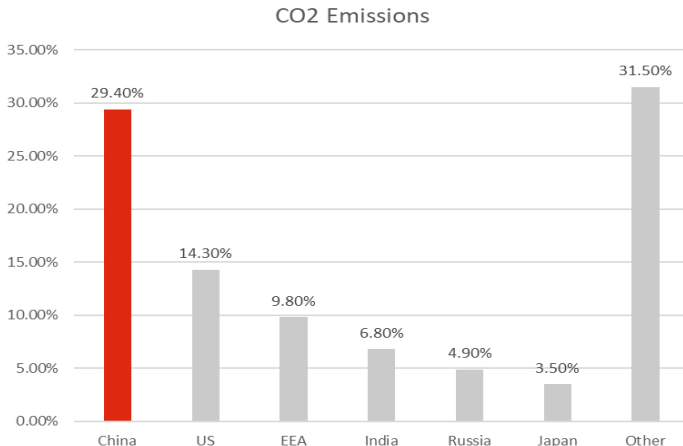


Figure: Source: <https://www.polymersearch.com/>

2. The role of visualizations in data analysis

Visualizations not only aid in understanding the data but also enable more impactful and insightful analysis and interpretation.

2.1. Exploratory data analysis

Visualizations allow the analysts to interact with and explore datasets visually, helping them understand the structure, patterns, and relationships within the data.

Titanic survival rate by gender and class

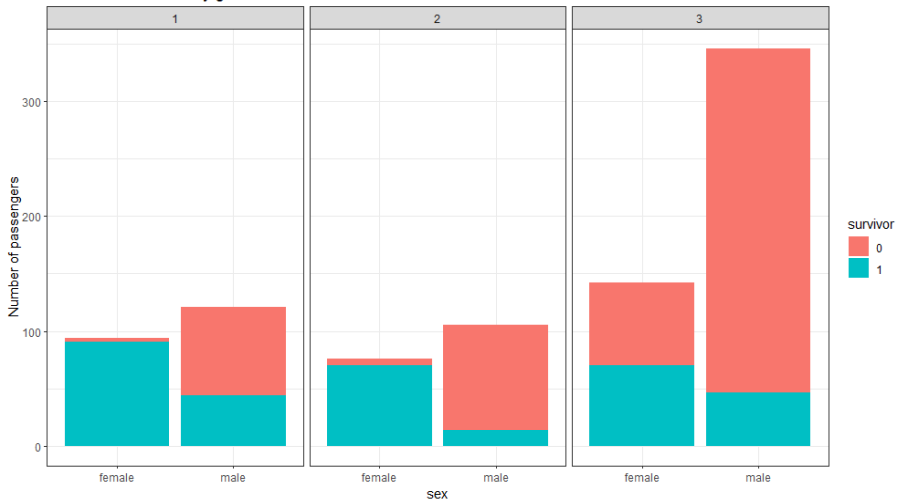


Figure: Titanic survival rate by gender and class

2.2. Pattern Recognition

Through visualizations, analysts can quickly identify clusters, anomalies, trends over time, spatial distributions, and other patterns that aid in understanding the underlying data.

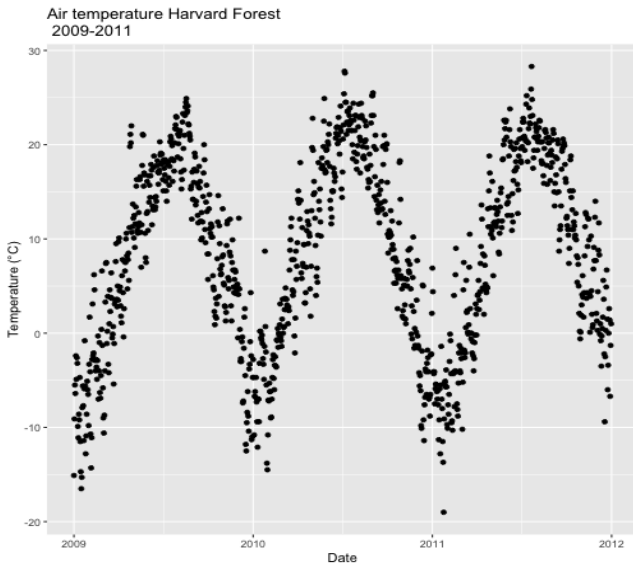


Figure: Air temperature Harvard Forest

2.3. Detecting errors and anomalies

Visualizations can serve as a valuable tool for quality assurance and data validation during the analysis process.

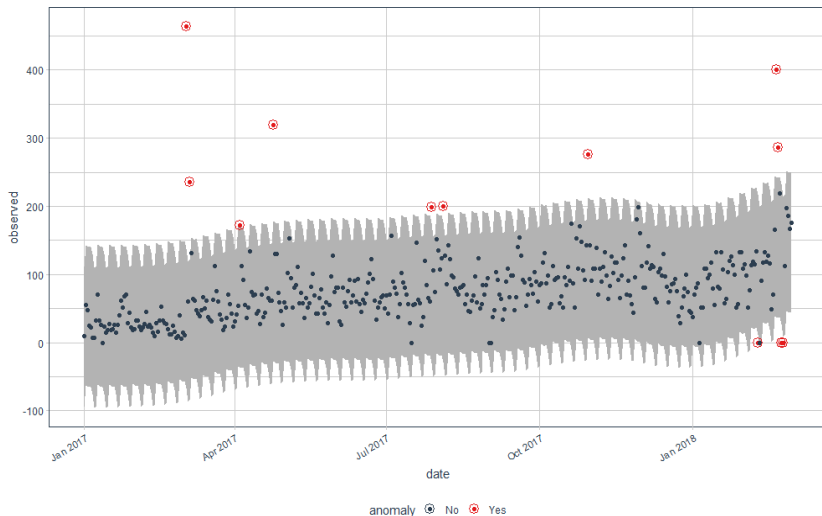


Figure: Detecting anomalies in a timeseries data using the "anomalies" package

2.4. Communication of findings

Good visualizations can communicate complex relationships and trends in a concise and compelling manner, enabling make informed decisions based on the presented data.

2. Introduction to "ggplot2" with R

The "ggplot2" package is the most elegant and aesthetically pleasing graphics framework available in R. It has a nicely planned structure to it. It allows users to define the visual properties of a plot, such as:

- Colors
- Shape
- Sizes.

besides, the user can choose from various geometric objects (geoms), such as points, lines, bars, or areas, to represent the data.

By following the principles of grammar of graphics, "ggplot2" promotes a coherent and structured approach to data visualization. This makes it easier to understand, reproduce, and modify visualizations, fostering efficient data exploration, analysis, and communication.

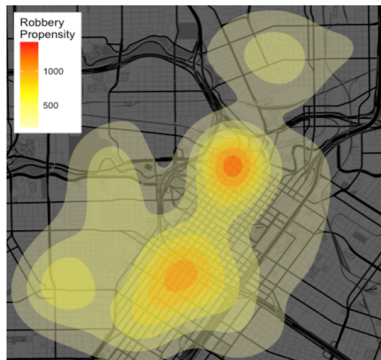
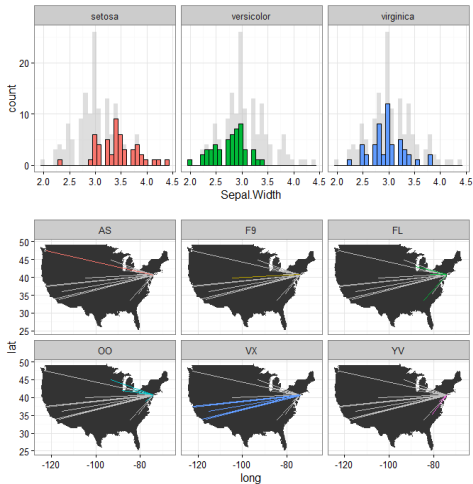


Figure: Source: <https://mode.com/>

Some extensions of the "ggplot2" package.

- "gganimate": The result is an animation built from various frames of the same plot.
- "ggtheme": with this package you can replicate data viz like The Economist, The Wall Street Journal, Highcharts JS, and more.
- "patchwork": The user can arrange multiple R plots in the same window.
- "ggribbles": It is especially useful for visualizing changes in distributions over time, space, or another continuous variable.

4. Conclusion

Advantages of "ggplot2"

- It follows a systematic approach to create visualizations, providing code clarity and reusability.
- Also it offers diverse plot types for different data types and analysis needs.
- Provides extensive customization options for plot aesthetics, labels, titles, etc.
- It supports creating subplots based on different subsets of data.

Disadvantages of "ggplot2"

- It may require some initial effort to understand and master its syntax.
- It initially provides static visualizations, requiring additional packages for interactivity.

 ggplot2: Elegant Graphics for Data Analysis.
<https://ggplot2-book.org/>

 Ozancan, O, 2022. An Introduction to ggplot2. <https://bookdown.org/ozancanozdemir/introduction-to-ggplot2/>

 r-statistics.co: <http://r-statistics.co/>