

Chapter 28

Choosing the right visualization software

Fundamentals of Data Visualization

May 6, 2023

Visualization Software

The best visualization software is the one that allows you to make the figures you need.

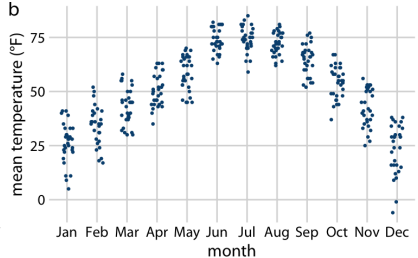
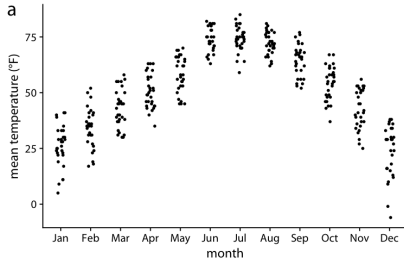
Reproducibility and repeatability

- Scientific work is repeatable if the finding of the work will remain unchanged if a different research group performs the same type of study.
- Work is reproducible if very similar or identical measurements can be obtained by the same person repeating the same procedure on the same equipment.

Reproducibility and repeatability

- A visualization is reproducible if the plotted data are available and any data transformations that may have been applied are exactly specified.
 - For example, if you make a figure and then send me the exact data that you plotted, then I can prepare a figure that looks substantially similar.
- A visualization is repeatable, on the other hand, if it is possible to recreate the exact same visual appearance, down to the last pixel, from the raw data.

Reproducibility and repeatability



- Results on left are reproduced (not repeated) on the right.
- Using interactive software (not R scripts) often produces irreproducible visualizations.

Data exploration versus data presentation

- In the exploration stage, whether the figures you make look appealing is secondary. It's fine if the axis labels are missing, the legend is messed up, or the symbols are too small, as long as you can evaluate the various patterns in the data.
- What is critical, however, is how easy it is for you to change how the data are shown.
- You should be able to rapidly move from a scatter plot to overlapping density distribution plots to boxplots to a heatmap.
- `ggplot2` is designed to facilitate this.

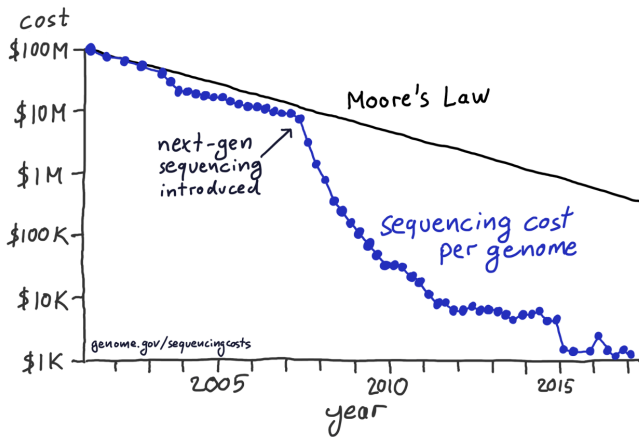
Data exploration versus data presentation

- Once we have determined how we want to visualize our data, what data transformations we want to make, and what type of plot to use, we will want to prepare a high-quality figure for publication.
- We have several different avenues we can pursue:
 - First, we can finalize the figure using same software platform we used for initial exploration.
 - Second, we can switch platform to one that provides us finer control, even if that makes it harder to explore.
 - Third, we can produce a draft figure with a visualization software and then manually post-process with an image manipulation program such as Photoshop.
 - Fourth, we can manually redraw the entire figure from scratch, either with pen and paper or using an illustration program.

Manually touching up visualizations is dangerous

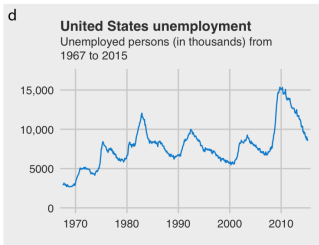
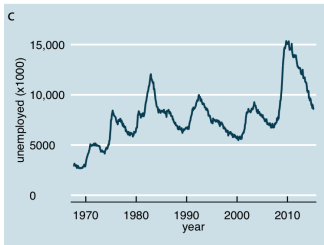
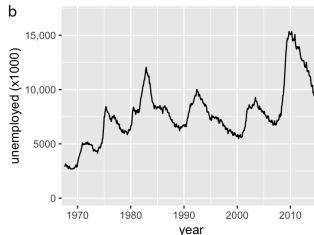
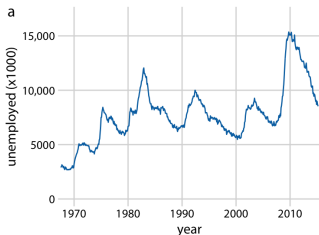
- We rarely make a figure just once.
- Over the course of a study, we may redo experiments, expand the original dataset, or repeat an experiment several times with slightly altered conditions.
- Many times we end up introducing a small modification to how we analyze our data, and consequently all figures have to be redrawn.
- Sometimes a decision is made not to redo the analysis or not to redraw the figures, either due to the effort involved or because the people who had made the original figure have moved on and aren't available anymore.
- An unnecessarily complicated and non-reproducible data visualization pipeline interferes with producing the best possible science.

On the other hand...



- Hand drawn figures are making a comeback.
- There's nothing inherently wrong with sprucing up a visualization.

Separation of content and design



- ggplot2 is designed to facilitate this.