**From Data to Viz – leading to the chart you need**

**Blog post for KD nuggets**

**Summary**

[Data-to-viz.com](https://www.data-to-viz.com/) is a classification of chart types based on input data format. It comes in the form of a decision tree leading to a set of potentially appropriate visualizations to represent your dataset.

**Suggested Tags**

[Data Visualization](http://www.kdnuggets.com/tag/data-visualization), [Graphics](http://www.kdnuggets.com/tag/graphics), [Visualization](http://www.kdnuggets.com/tag/visualization), R

**Blog Text**

Data visualization is a key step in the data science process. Choosing the right graphic to explore data or to convey insight efficiently is an everyday task for a data scientist.

[From Data to Viz](https://www.data-to-viz.com/)is a classification of chart types based on input data format. It comes in the form of a decision tree leading to a set of potentially appropriate visualizations to represent the dataset.

## A decision tree for data visualization

The project is built on two underlying philosophies. First, that most data analysis can be summarized in about twenty different dataset formats. Second, that both data and context determine the appropriate chart.

Thus, the suggested method consists in identifying and trying all feasible chart types to find out which suits your data and idea best.

Once this set of graphic identified, data-to-viz.com aims to guide you toward the best decision.

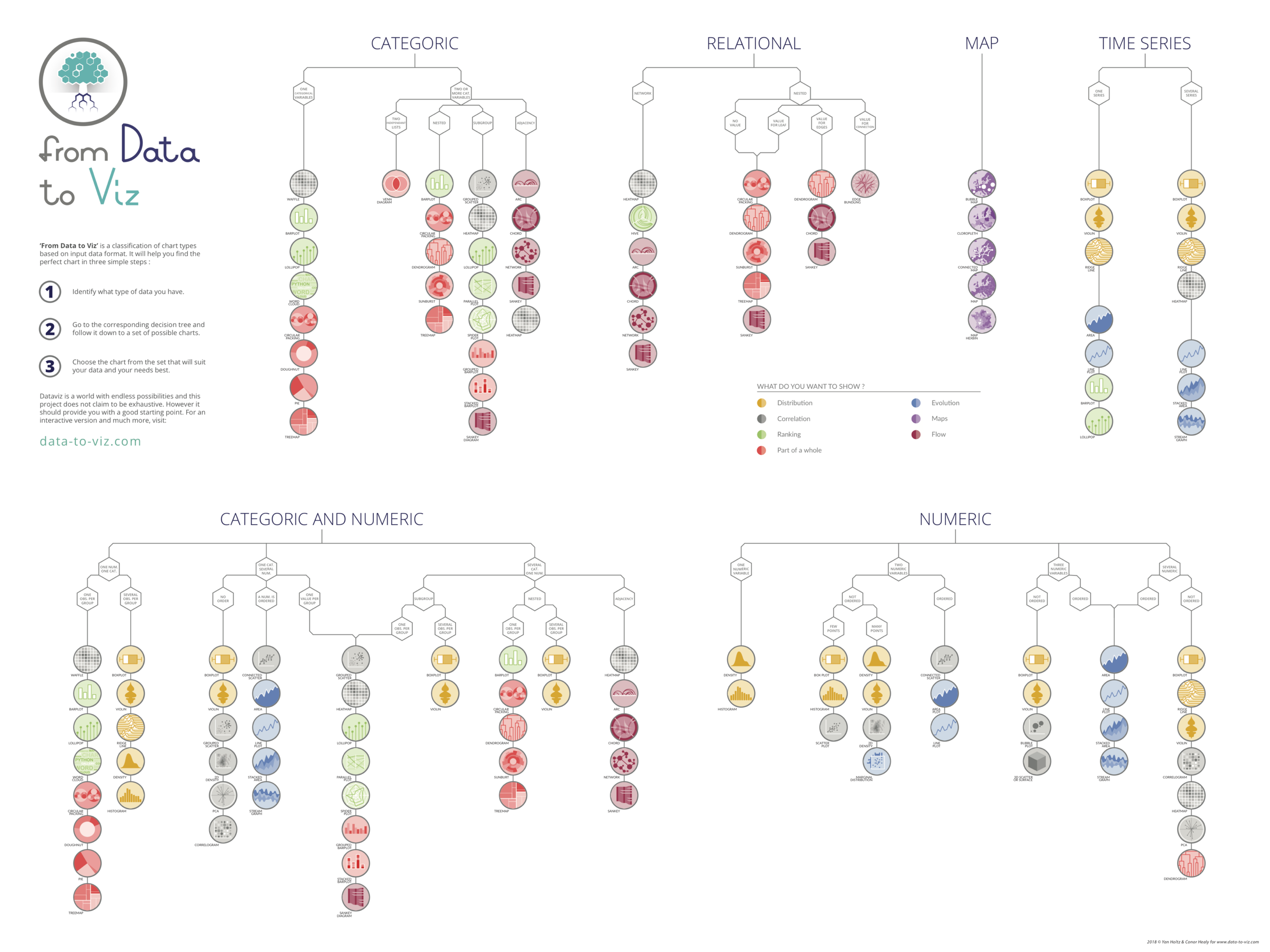


Figure 1: Data to Viz poster displaying the decision trees. See [online](https://www.data-to-viz.com/) for an interactive version. [Prints](https://www.data-to-viz.com/poster.html) are available.

## Example

## Let’s consider a dataset composed by one numeric and one categoric variable. For instance, the [quantity of weapons](https://www.data-to-viz.com/story/OneNumOneCat.html) exported (numeric) per country (categoric). What kind of visualization can we apply to it?

This data format is represented by a branch of the decision tree. We have a dataset composed by both numeric and categoric data, we have one variable of each, and we have only one observation per group. The decision tree suggests many appropriate chart types:



Of course, the most common solution is probably to build a [barplot](https://www.data-to-viz.com/graph/barplot.html). The [lollipop](https://www.data-to-viz.com/#lollipop) plot is a good alternative with so many groups, resulting in a less cluttered figure. Note that both can be done in a [circular version](https://www.data-to-viz.com/#circularbarplot), giving a more eye-catching but less accurate output. Last but not least, [treemap](https://www.data-to-viz.com/#treemap) and [circular packing](https://www.data-to-viz.com/graph/circularpacking.html) are good options if you’re interested in how the whole is divided.

Here is an overview of the chart you get when applied to the weapon dataset:

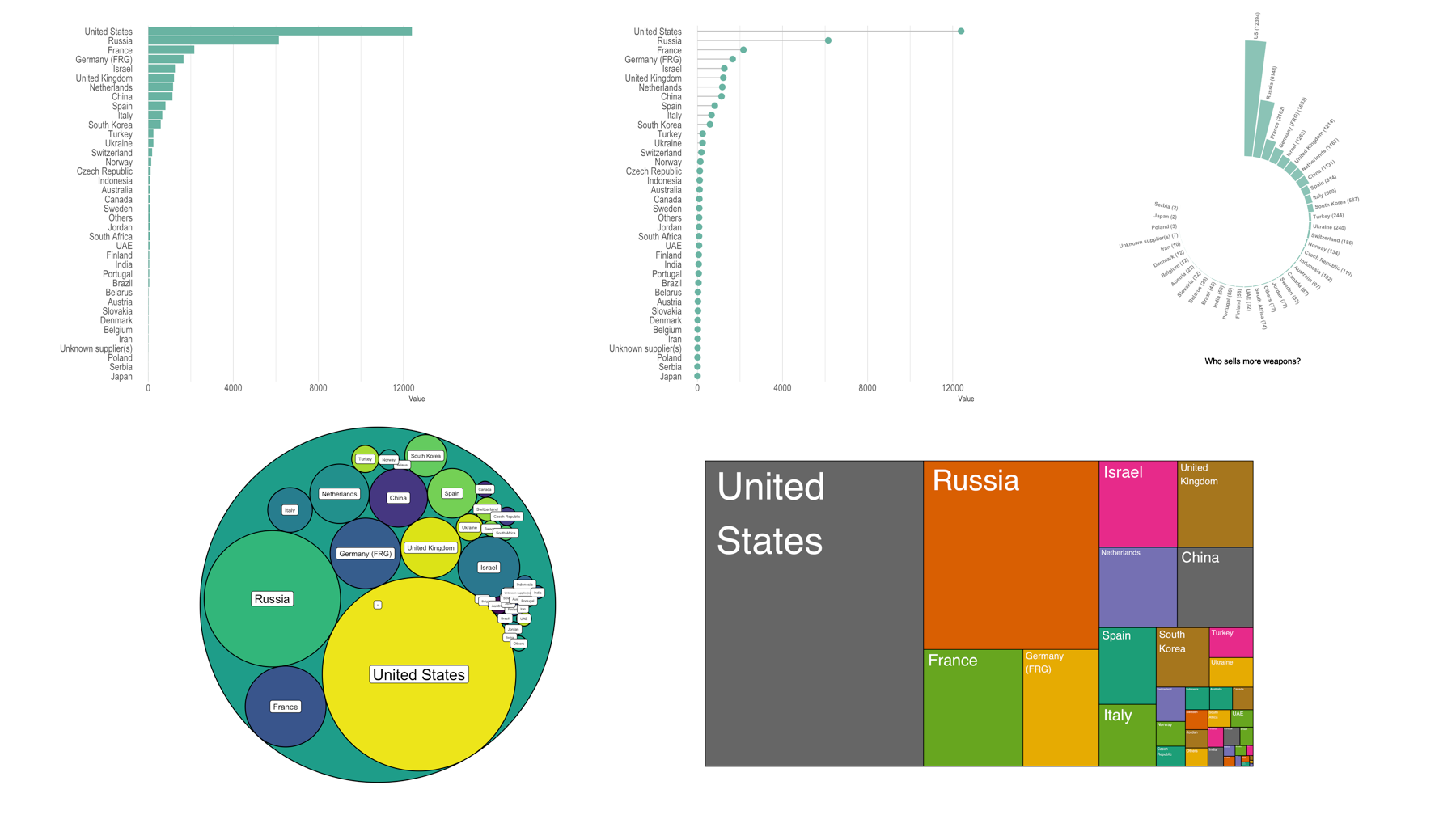


Figure 2: five representations of the same dataset: from left to right, top to bottom: barplot, lollipop plot, circular barplot, circular packing, treemap.

## Content

The website does not only lead you to a set of potential visualization. It also aims to help you picking the right one. Several sections help in this task:

– [portfolio](https://www.data-to-viz.com/index.html#portfolio) – an overview of all chart possibilities. For each, an extensive description is given, showing variations, pros and cons, common pitfalls and more.  
– [*stories*](https://www.data-to-viz.com/index.html#story) – for each input data format, a real-life example is analyzed to illustrate the different chart types applicable to it. It also explains which graphic is good to answer which question.  
– [*caveat gallery*](https://www.data-to-viz.com/caveats.html) – a list of common dataviz pitfalls, with suggested workarounds.

## About the caveat gallery

The best way to visualize data efficiently is probably to avoid the most common pitfalls. The [caveat gallery](https://www.data-to-viz.com/caveats.html) lists about 40 common caveats, and the list is still growing.

For instance, it points out that a barplot is much more insightful when [ordered](https://www.data-to-viz.com/caveat/order_data.html). Moreover, it is a good practice to make it horizontal when you have [long labels](https://www.data-to-viz.com/caveat/hard_label.html). If you have many bars, the lollipop plot is probably a good alternative to declutter the graphic and avoid a [Moire effect](http://gravyanecdote.com/uncategorized/wrappedbarsthoughts/). Of course, the many downsides of [pie](https://www.data-to-viz.com/caveat/pie.html) and donut charts are described, even if they are still suggested in the decision tree.

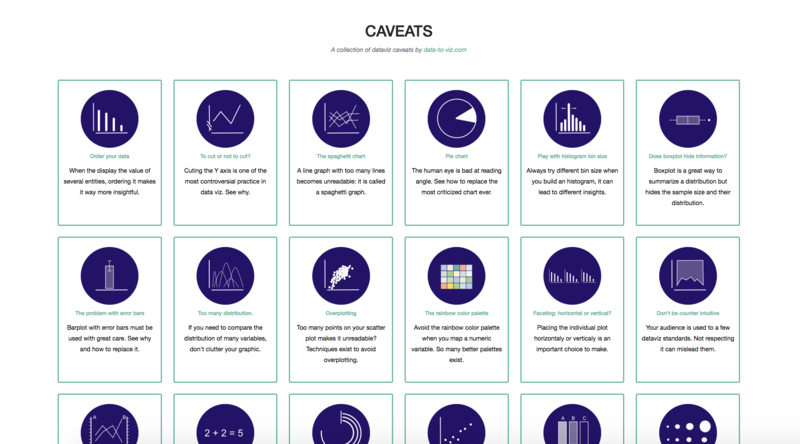


Figure 3: overview of the gallery of dataviz caveats

## Building your chart.

From [*Data to Viz*](http://www.data-to-viz.com/) aims to give general advices for data visualization in general and is not targeting coders especially.

However, 100% of the charts are made using [*R*](http://www.r-graph-gallery.com/), mostly using [*ggplot2*](https://www.r-graph-gallery.com/portfolio/ggplot2-package/)and the tidyverse. The reproducible code snippets are always available. The biggest part of the website is built using R Markdown, using a good amount of[hacks described *here*](https://holtzy.github.io/Pimp-my-rmd/).

The website is tightly linked with the [*R graph gallery*](https://www.r-graph-gallery.com/)and the [Python graph gallery](https://www.python-graph-gallery.com/). Once you’ve identified the graphic that suits your needs, you will be redirected to the appropriate section of the gallery to get the code in your favourite language.

## Conclusion

Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However, it should provide the user with a good starting point. Moreover, it is a valuable tool for students and people willing to learn more about data visualization best practices.

The project is hosted on [Github](https://github.com/holtzy/data_to_viz). Any comment, issue or pull request is very welcome. You can also reach me on twitter ([@R\_Graph\_Gallery](https://twitter.com/R_Graph_Gallery?lang=en)) or drop me an email at [yan.holtz.data@gmail.com](mailto:yan.holtz.data@gmail.com).

**Author Bio**

[Yan Holtz](https://holtzyan.wordpress.com/) is a passionate data analyst and bio-informatician currently working for the Queensland Brain Institute of Brisbane. He has a special attraction for data visualization which led him to build the [R](http://www.r-graph-gallery.com/) and the [Python](https://www.python-graph-gallery.com/) graph galleries. He can be reached at: yan.holtz.data@gmail.com.

Homepage: https://www.linkedin.com/in/yan-holtz-2477534a/

Picture:

