

Visualizing Categorical Data in R

true

2021-10-26

Contents

1	Introduction	1
2	Some Data	2
3	Call The Graphing Library	2
4	Bar Graphs	2
4.1	Stacked Bar Graph	2
4.2	Unstacked Bar Graph	3
5	Pie Chart	3
6	Jittered Points	4
7	Mosaic Plot	4
8	Waffle Plot	5
8.1	Call The Libraries	5
8.2	Make A Data Set Of Counts	5
8.3	Make The Waffle Plot	6
9	Alluvial Diagram	6

1 Introduction

Visualizing categorical data presents unique challenges. A common solution is a bar graph, which may often be the best data visualization solution.

However there are also some alternatives to bar graphs.

Below we present some options for bar graphs, and some possible alternative strategies.

2 Some Data

We create some simulated data with the `tribble` function. The data are created so that the two groups experience the outcomes unequally.

	beneficial outcome	undesirable outcome
Group A	55	40
Group B	50	75

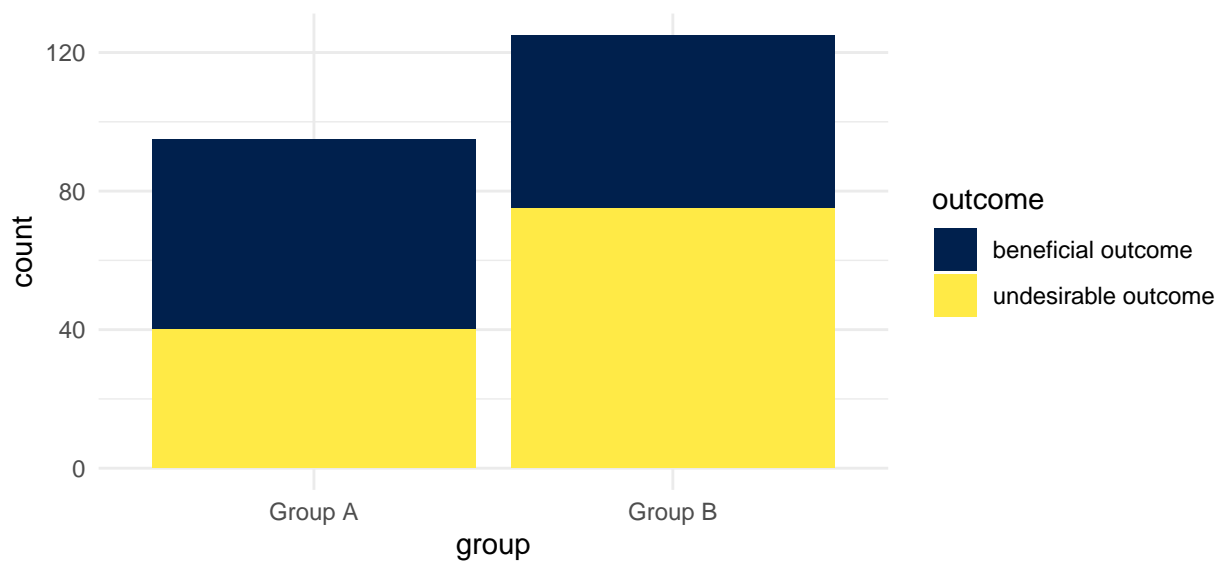
3 Call The Graphing Library

4 Bar Graphs

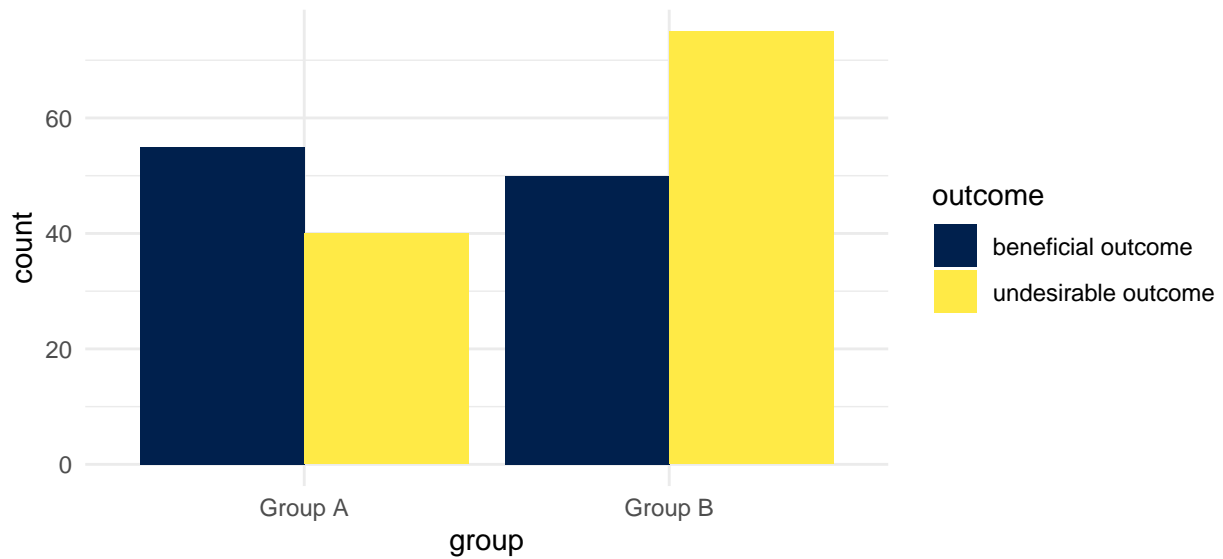
Bar graphs are often the simplest and best option for displaying categorical data. When used with an aesthetically pleasing color scheme, bar graphs can be an effective way of displaying data.

There are several different types of bar graph.

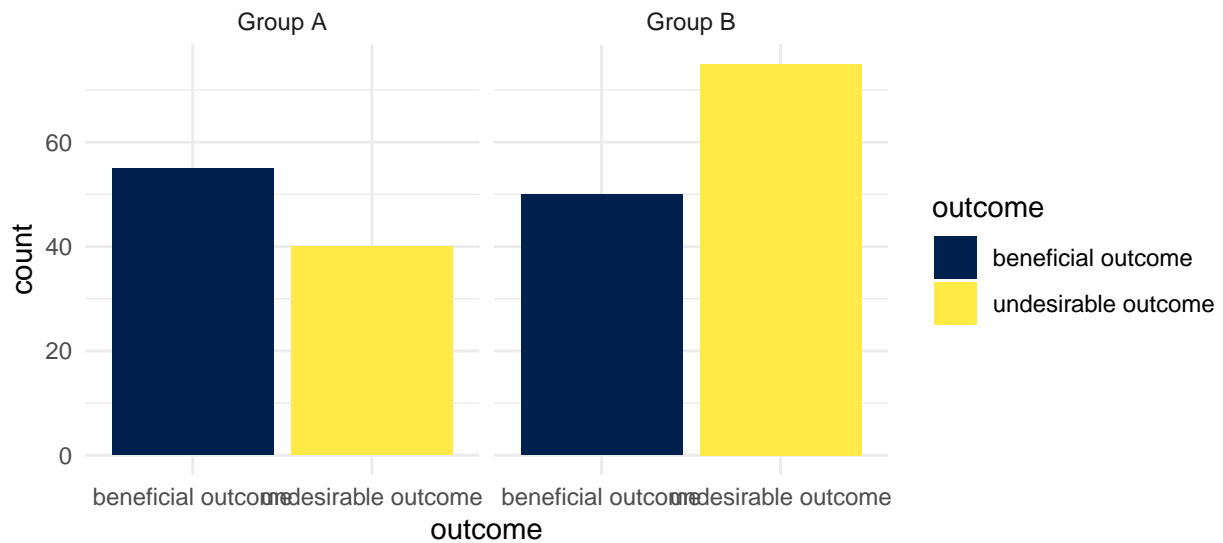
4.1 Stacked Bar Graph



4.2 Unstacked Bar Graph

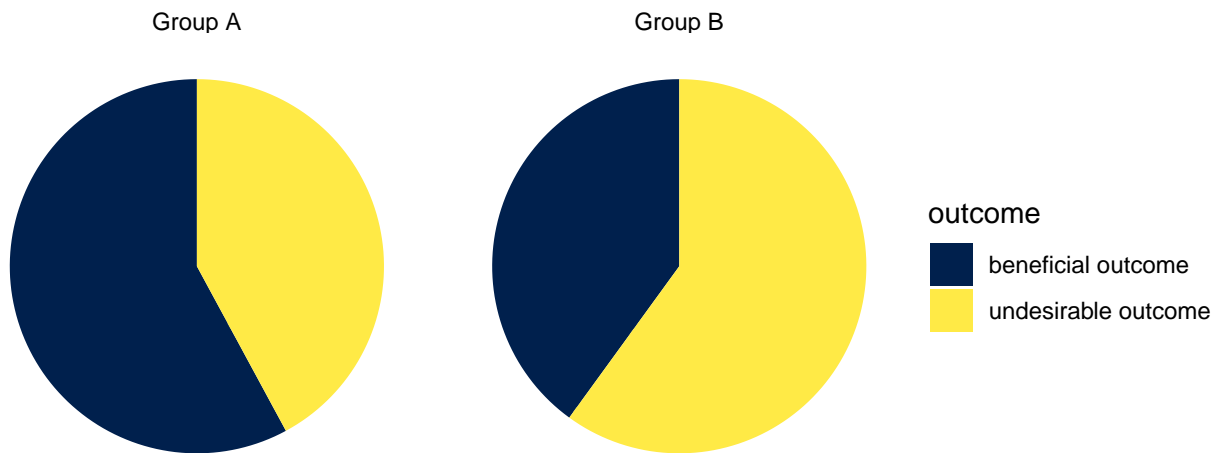


Faceted Bar Graph



5 Pie Chart

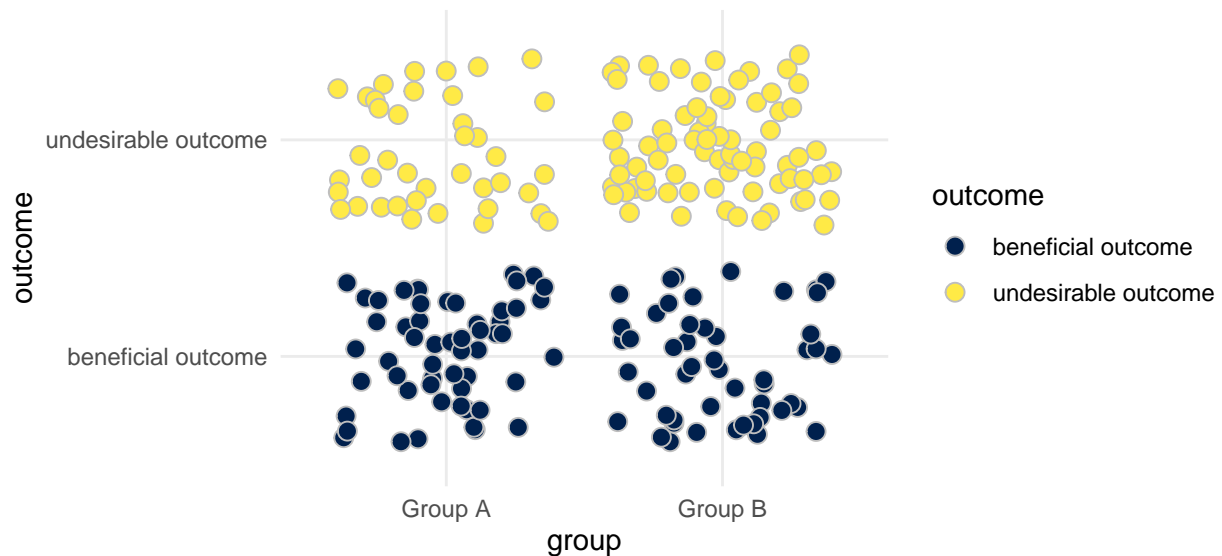
In `ggplot` terms, pie charts are bar graphs displayed with *polar coordinates*.



6 Jittered Points

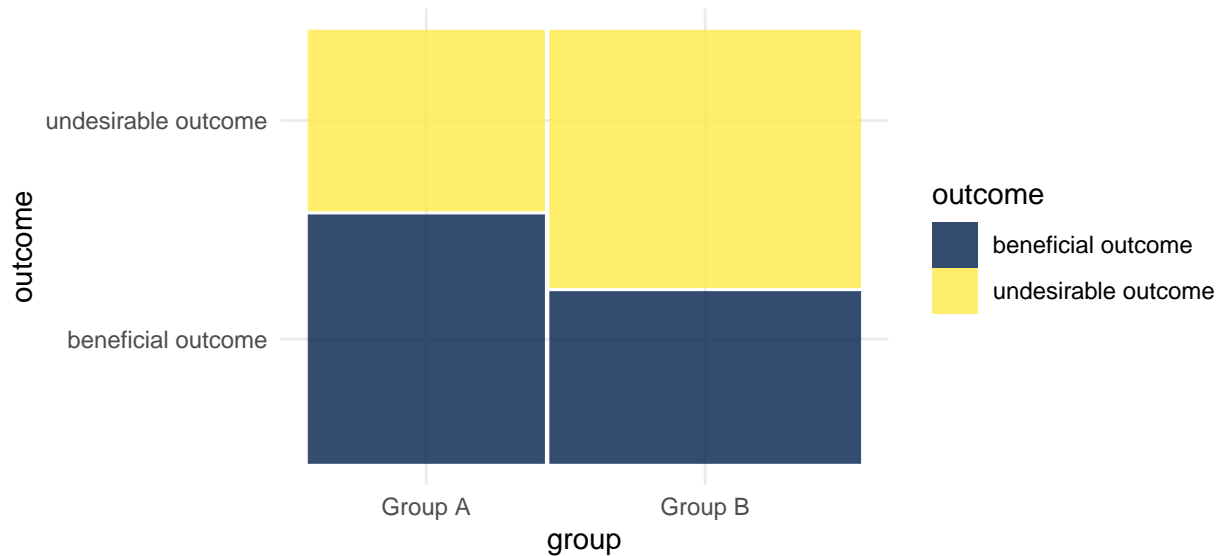
Jittered points may be a good choice because every point represents an individual in the data set. However, it may be difficult to draw exact conclusions from jittered points.

Jittered points *may* (or may not) benefit from having an outline in a different color to make them more distinct.



7 Mosaic Plot

Mosaic plots are another way to display data. They are especially effective for being clear about the relative membership in different groups, and about the proportion of each group experiencing each outcome.



8 Waffle Plot

Lastly, waffle plots may be a useful way to display information. Waffle plots are aesthetically appealing. The aesthetic appeal of a waffle plot may, however, obscure the fact that they may not provide the clearest presentation of quantitative information. Waffle plots work best when the sample size is several hundred or fewer.

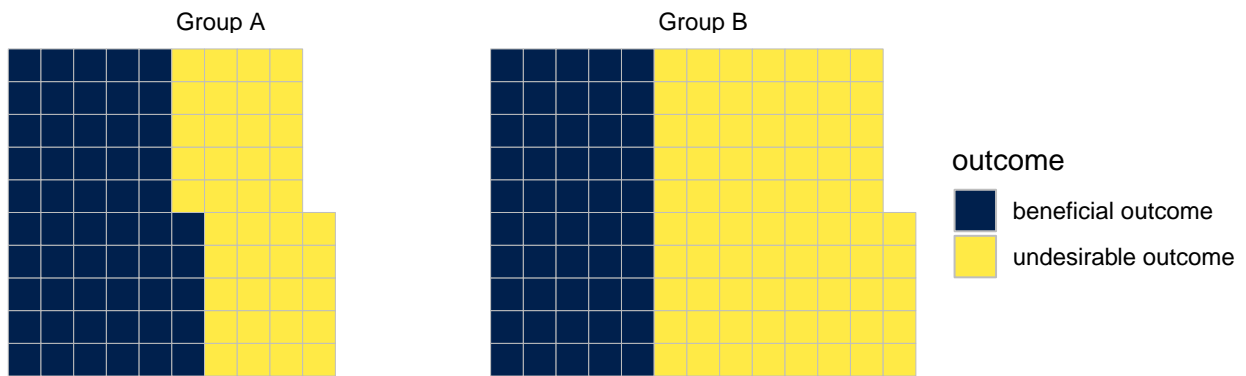
Waffle plots require some data wrangling.

8.1 Call The Libraries

8.2 Make A Data Set Of Counts

group	outcome	n
Group A	beneficial outcome	55
Group A	undesirable outcome	40
Group B	beneficial outcome	50
Group B	undesirable outcome	75

8.3 Make The Waffle Plot



9 Alluvial Diagram

Lastly, an alluvial diagram may be useful to illustrate a *flow* from one status to another.

We will use the data set of `mycounts` that we generated above.

