

6 Applications: Explore

6.1 Case study: Effective communication of exploratory results

Graphs can powerfully communicate ideas directly and quickly. We all know, after all, that “a picture is worth 1000 words.” Unfortunately, however, there are times when an image conveys a message which is inaccurate or misleading.

This chapter focuses on how graphs can best be utilized to present data accurately and effectively. Along with data modeling, creative visualization is somewhat of an art. However, even with an art, there are recommended guiding principles. We provide a few best practices for creating data visualizations.

6.1.1 Keep it simple

When creating a graphic, keep in mind what it is that you’d like your reader to see. Colors should be used to group items or differentiate levels in meaningful ways. Colors can be distracting when they are only used to brighten up the plot.

Consider a manufacturing company that has summarized their costs into five different categories. In the two graphics provided in Figure 6.1, notice that the magnitudes in the pie chart are difficult for the eye to compare. That is, can your eye tell how different “Buildings and administration” is from “Workplace materials” when looking at the slices of pie? Additionally, the colors in the pie chart do not mean anything and are therefore distracting. Lastly, the three-dimensional aspect of the image does not improve the reader’s ability to understand the data presented.

As an alternative, a bar plot has been provided. Notice how much easier it is to identify the magnitude of the differences across categories while not being distracted by other aspects of the image. Typically, a bar plot will be easier for the reader to digest than a pie chart, especially if the categorical data being plotted has more than just a few levels.

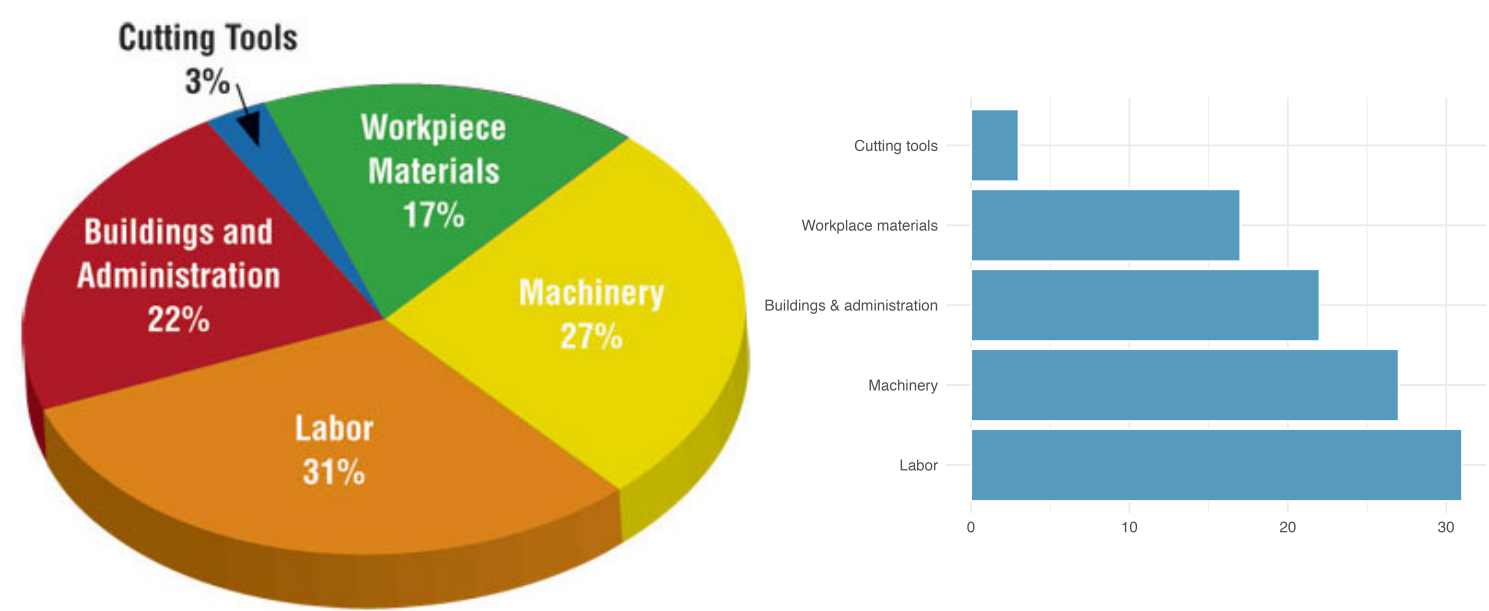


Figure 6.1: A pie chart (with added irrelevant features) as compared to a simple bar plot.

6.1.2 Use color to draw attention

There are many reasons why you might choose to add **color** to your plots. An important principle to keep in mind is to use color to draw attention. Of course, you should still think about how visually pleasing your visualization is, and if you’re adding color for making it visually pleasing without drawing attention to a particular feature, that might be fine. However, you should be critical of default coloring and explicitly decide whether to include color and how. Notice that in Plot B in Figure 6.2 the coloring is done in such a way to draw the reader’s attention to one particular piece of information. The default

On this page

- 6 Applications: Explore
- 6.1 Case study: Effective communication of exploratory results
 - 6.1.1 Keep it simple
 - 6.1.2 Use color to draw attention
 - 6.1.3 Tell a story
 - 6.1.4 Order matters
 - 6.1.5 Make the labels as easy to read as possible
 - 6.1.6 Pick a purpose
 - 6.1.7 Select meaningful colors
- 6.2 Interactive R tutorials
- 6.3 R labs

[View source](#)

[Edit this page](#)

coloring in Plot A can be distracting and makes the reader question, for example, is there something similar about the red and purple bars? Also note that not everyone sees color the same way, often it's useful to add color and one more feature (e.g., pattern) so that you can refer to the features you're drawing attention to in multiple ways.

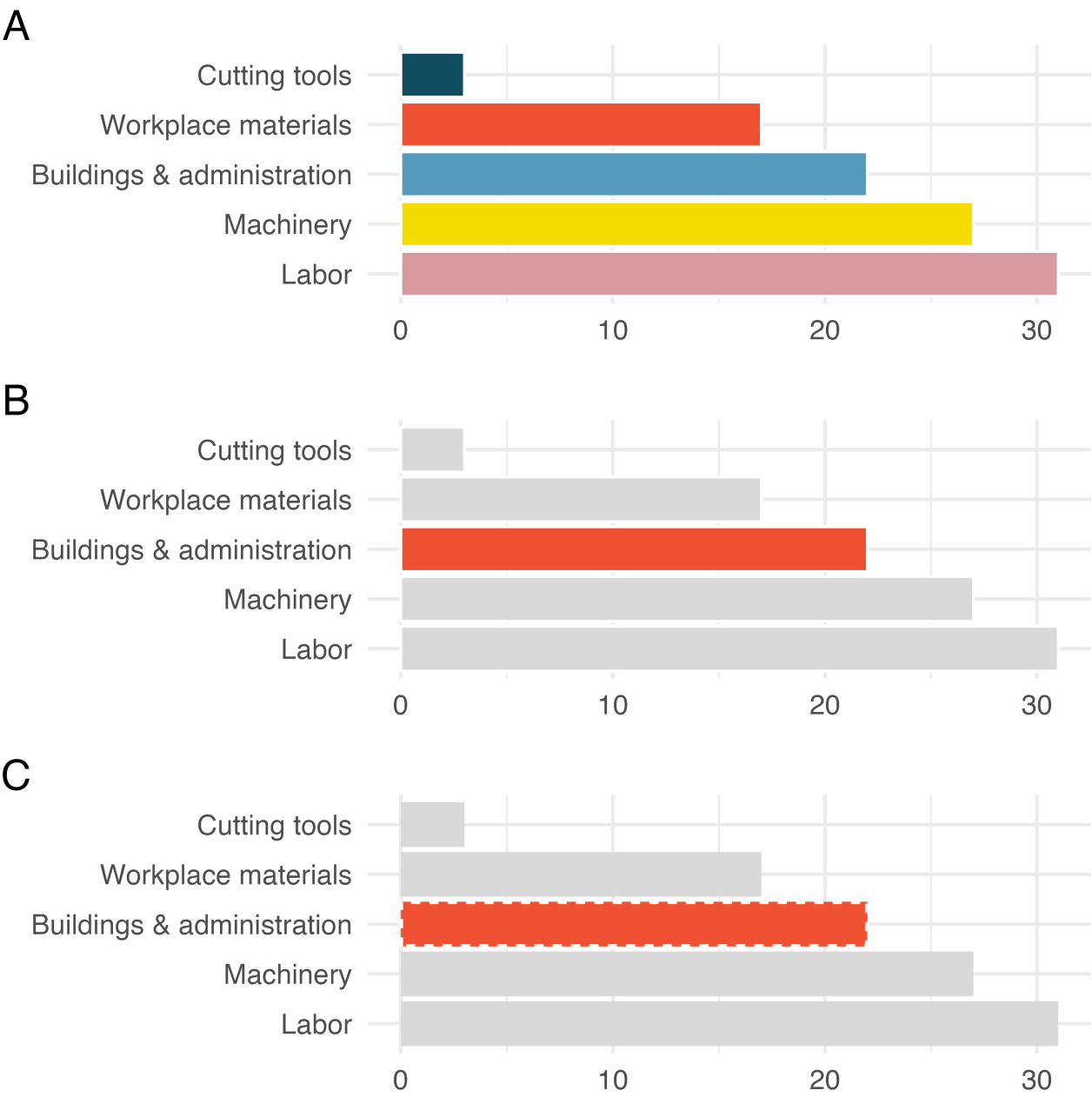


Figure 6.2: The default coloring in the first bar plot does nothing for the understanding of the data. In the second plot, the color draws attention directly to the bar on Buildings and Administration.

6.1.3 Tell a story

For many graphs, an important aspect is the inclusion of information which is not provided in the dataset that is being plotted. The external information serves to contextualize the data and helps communicate the narrative of the research. In Figure 6.3, the graph on the right is **annotated** with information about the start of the university's fiscal year which contextualizes the information provided by the data. Sometimes the additional information may be a diagonal line given by $y = x$, points above the line quickly show the reader which values have a y coordinate larger than the x coordinate; points below the line show the opposite.

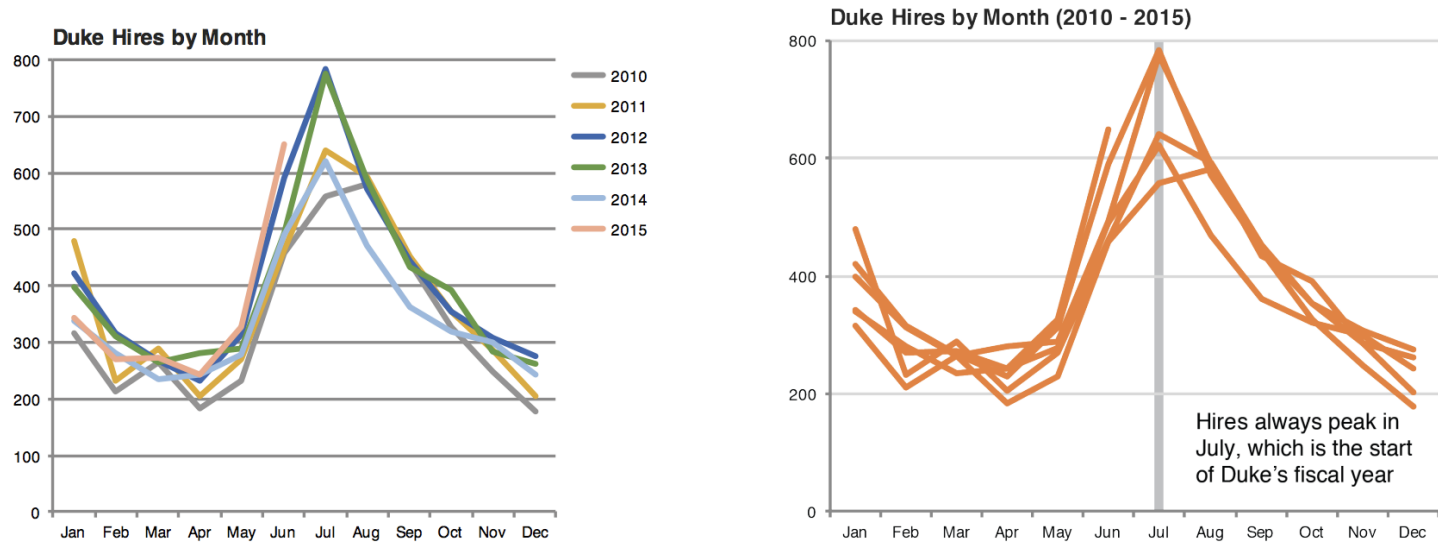


Figure 6.3: Credit: Angela Zoss and Eric Monson, Duke Data Visualization Services

6.1.4 Order matters

Most software programs have built in methods for some of the plot details. For example, the default option for the software program used in this text, R, is to order the bars in a bar plot alphabetically. As seen in Figure 6.4, the alphabetical ordering isn't particularly meaningful for describing the data. Sometimes it makes sense to **order** the bars from tallest to shortest (or vice versa). But in this case, the best ordering is probably the one in which the questions were asked. An ordering which does not make sense in the context of the problem (e.g., alphabetically here), can mislead the reader who might take a quick glance at the axes and not read the bar labels carefully.

In September 2019, YouGov survey asked 1,639 Great Britain adults the following question⁷¹:

- How well or badly do you think the government are doing at handling Britain's exit from the European Union?
- Very well
 - Fairly well
 - Fairly badly
 - Very badly
 - Don't know

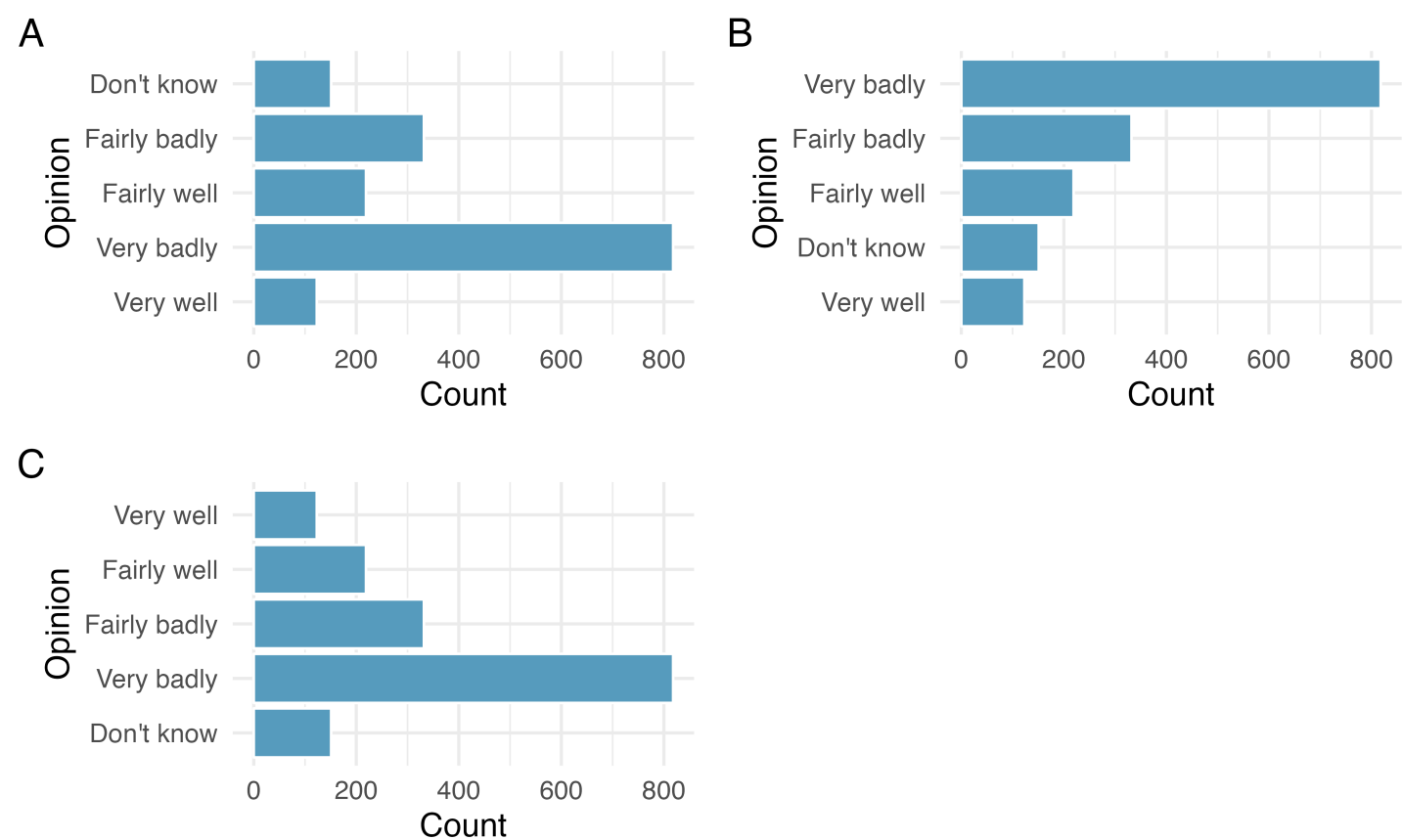


Figure 6.4: Bar plot three different ways. Plot A: Alphanumeric ordering of levels, Plot B: Bars ordered in descending order of frequency, Plot C: Bars ordered in the same order as they were presented in the survey question.

6.1.5 Make the labels as easy to read as possible

The Brexit survey results were additionally broken down by region in Great Britain. The stacked bar plot allows for comparison of Brexit opinion across the five regions. In Figure 6.5 the bars are vertical in Plot A and horizontal in Plot B. While the quantitative information in the two graphics is identical, flipping the graph and creating horizontal bars provides more space for the **axis labels**. The easier the categories are to read, the more the reader will learn from the visualization. Remember, the goal is to convey as much information as possible in a succinct and clear manner.

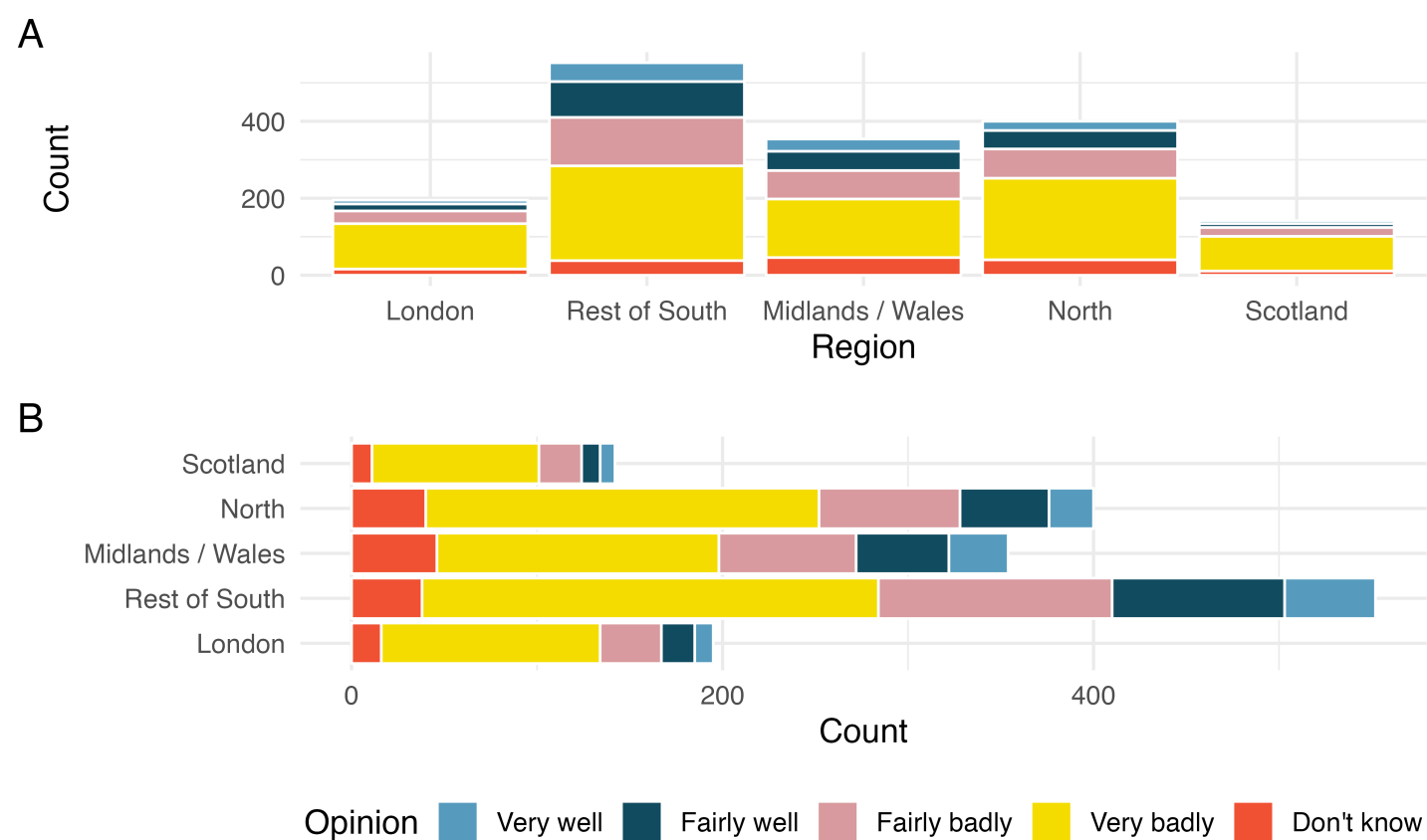


Figure 6.5: Stacked bar plots vertically and horizontally. The horizontal orientation makes the region labels easier to read.

6.1.6 Pick a purpose

Every graphical decision should be made with a **purpose**. As previously mentioned, sticking with default options is not always best for conveying the narrative of your data story. Stacked bar plots tell one part of a story. Depending on your research question, they may not tell the part of the story most important to the research. Figure 6.6 provides three different ways of representing the same information. If the most important comparison across regions is proportion, you might prefer Plot A. If the most important comparison across regions also considers the total number of individuals in the region, you might prefer Plot B. If a separate bar plot for each region makes the point you'd like, use Plot C, which has been **faceted** by region.

Plot C in Figure 6.6 also provides full titles and a succinct URL with the data source. Other deliberate decisions to consider include using informative labels and avoiding redundancy.

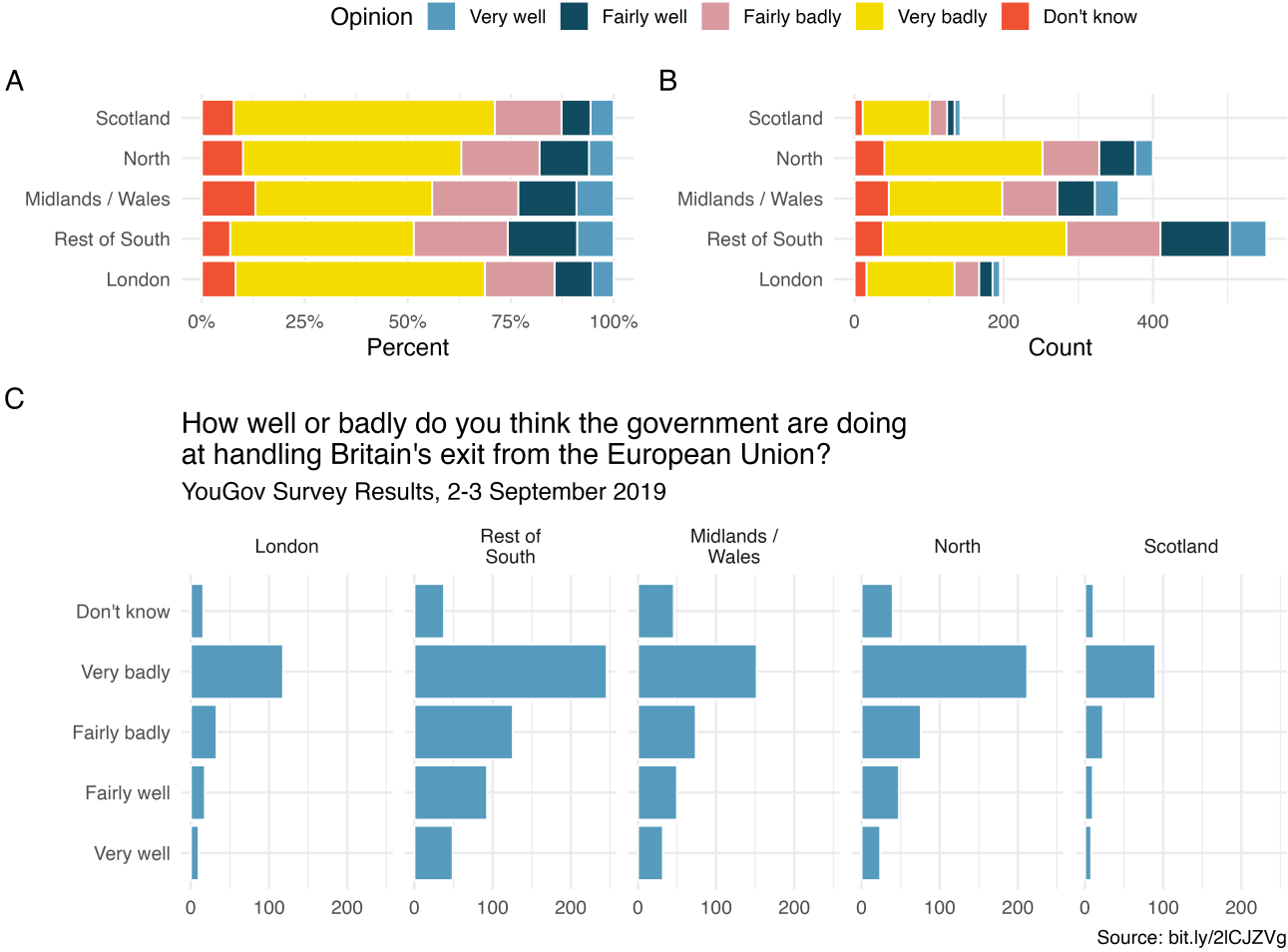


Figure 6.6: Three different representations of the two variables including survey opinion and region. Use the graphic that best conveys the data narrative at hand.

6.1.7 Select meaningful colors

One last consideration for building graphs is to consider color choices. Default or rainbow colors are not always the choice which will best distinguish the level of your variables. Much research has been done to find color combinations which are distinct and which are clear for differently sighted individuals. The *cividis* scale works well with ordinal data. (Nuñez, Anderton, and Renslow 2018) Figure 6.7 shows the same plot with two different colorings.

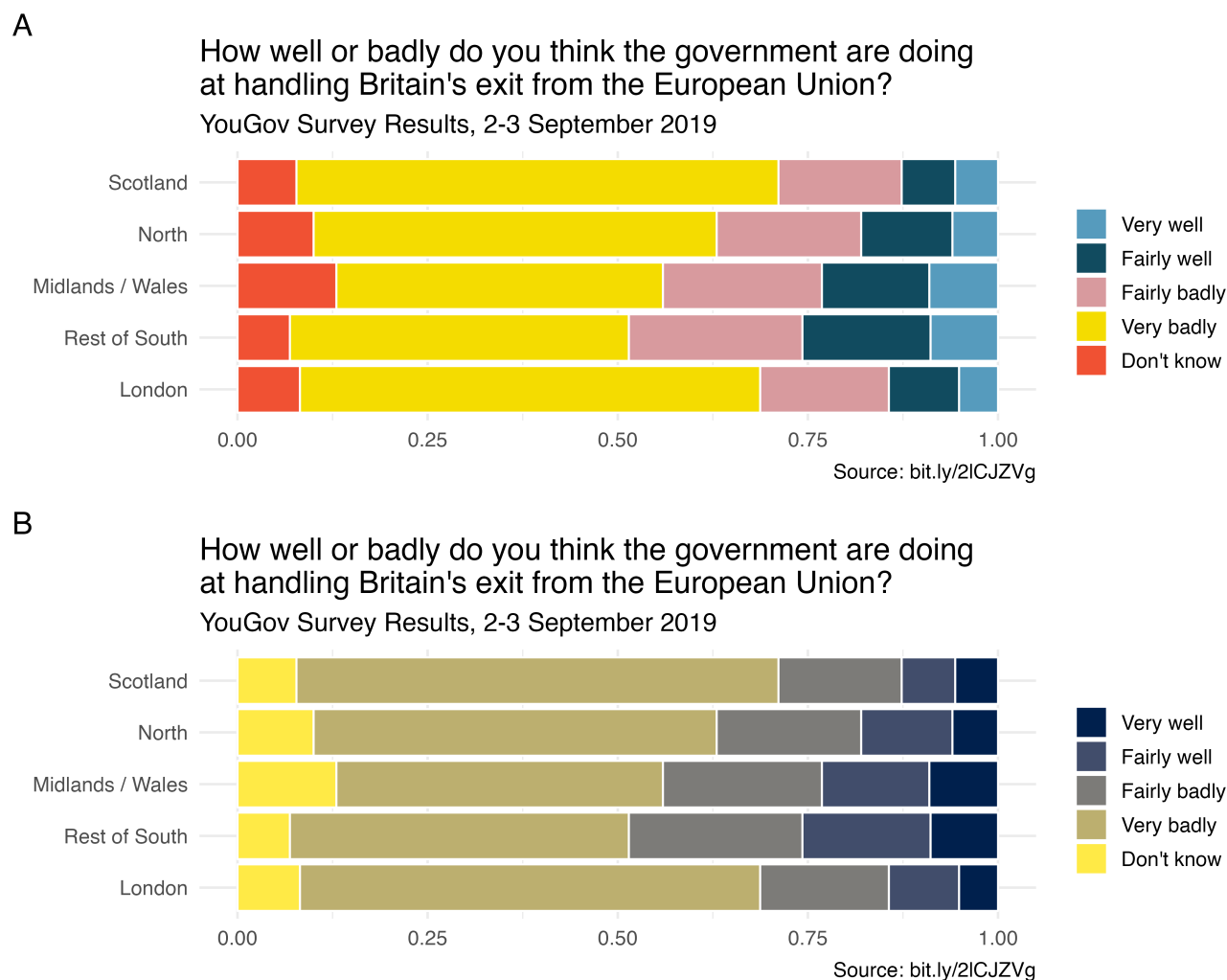


Figure 6.7: Identical bar plots with two different coloring options. Plot A uses a default color scale, Plot B uses colors from the cividis scale.

In this chapter different representations are contrasted to demonstrate best practices in creating graphs. The fundamental principle is that your graph should provide maximal information succinctly and clearly. Labels should be clear and oriented horizontally for the reader. Don't forget titles and, if possible, include the source of the data.

6.2 Interactive R tutorials

Navigate the concepts you've learned in this chapter in R using the following self-paced tutorials. All you need is your browser to get started!

 [Tutorial 2: Exploratory data analysis](#)

 [Tutorial 2 - Lesson 1: Visualizing categorical data](#)

 [Tutorial 2 - Lesson 2: Visualizing numerical data](#)

 [Tutorial 2 - Lesson 3: Summarizing with statistics](#)

 [Tutorial 2 - Lesson 4: Case study](#)

You can also access the full list of tutorials supporting this book [here](#).

6.3 R labs

Further apply the concepts you've learned in this part in R with computational labs that walk you through a data analysis case study.



[Intro to data - Flight delays](#)

You can also access the full list of labs supporting this book [here](#).

[« 5 Exploring numerical data](#)

[7 Linear regression with a single predictor »](#)

"Introduction to Modern Statistics" was written by Mine Çetinkaya-Rundel and Johanna Hardin.

This book was built by the [bookdown](#) R package.