HOCHSCHULE DER MEDIEN

Exploratory data analysis

Effective communication of exploratory results

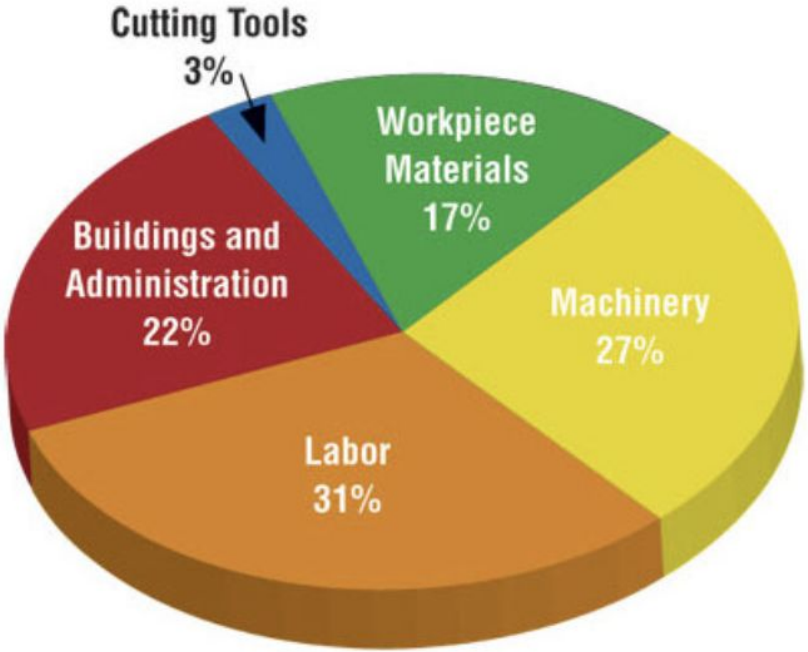
Prof. Dr. Jan Kirenz
HdM Stuttgart

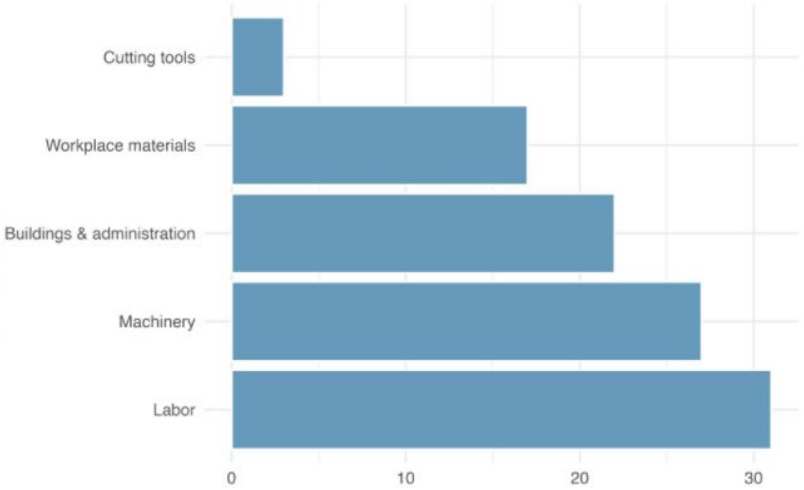
Keep it **simple**

Consider a manufacturing company that has summarized their costs into five different categories.

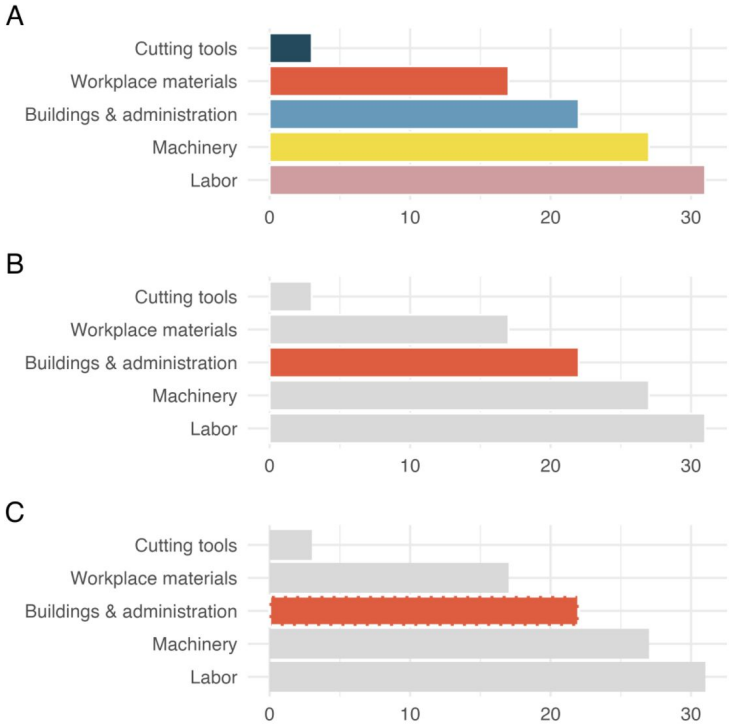
In which of the following plots is it easier to identify the magnitude of the differences across categories?

A **pie chart** as compared to a simple **bar plot**.





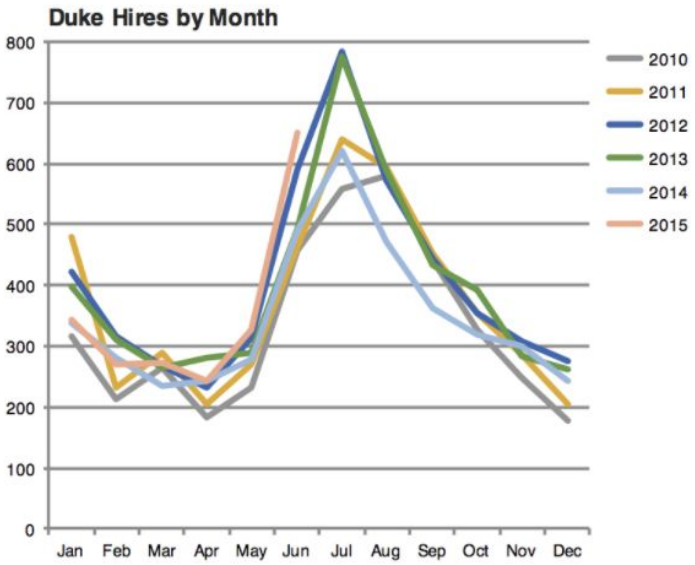
Use **color** to draw attention

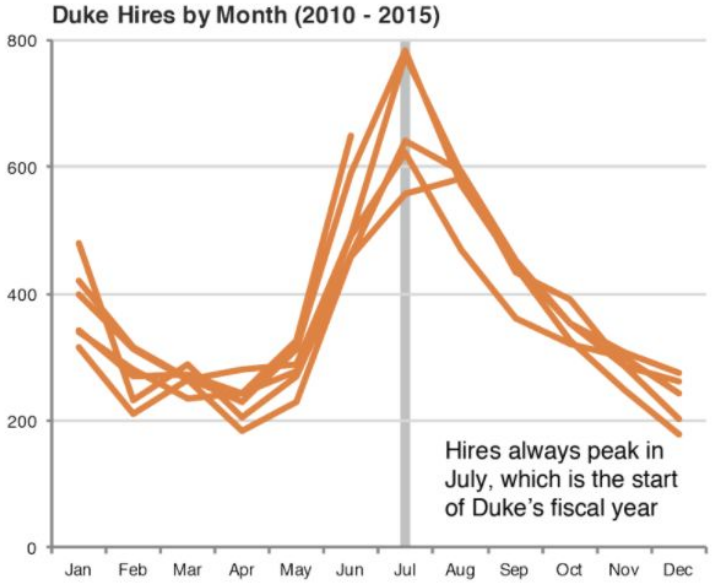


The default coloring in the first bar plot does nothing for the understanding of the data.

In the second and third plot, the color draws attention directly to the bar on Buildings and Administration.

Tell a **story** with annotations





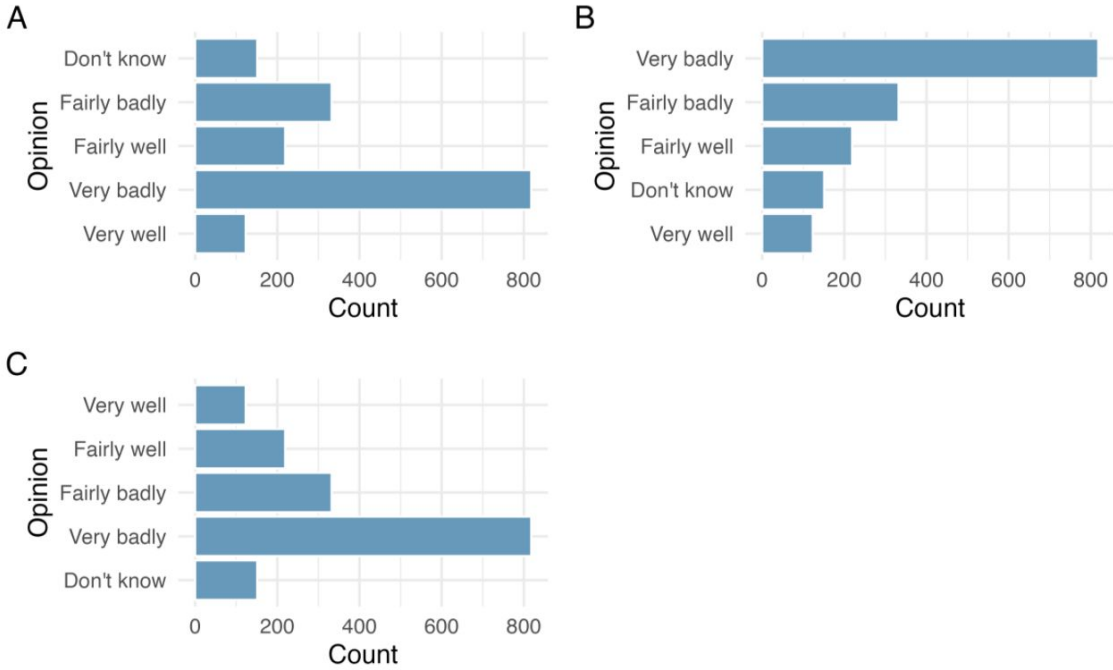
Order matters

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

Order matters: bar plot three different ways.

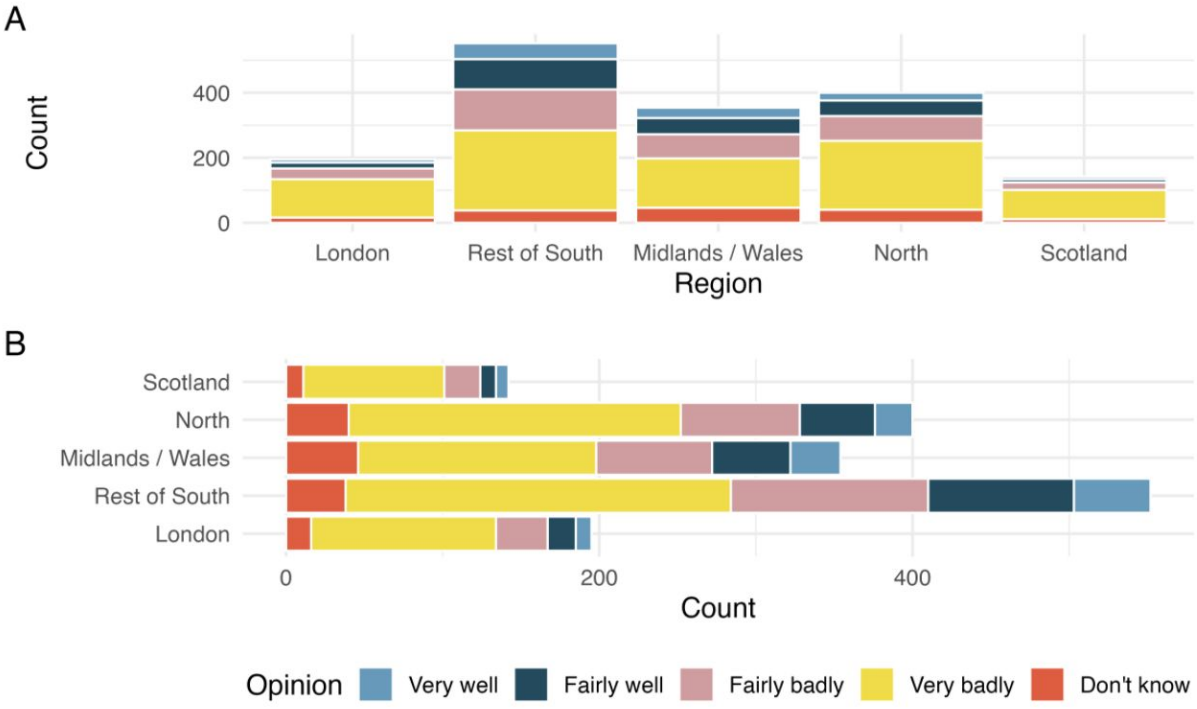


Plot A: Alphabetic ordering of levels,

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.

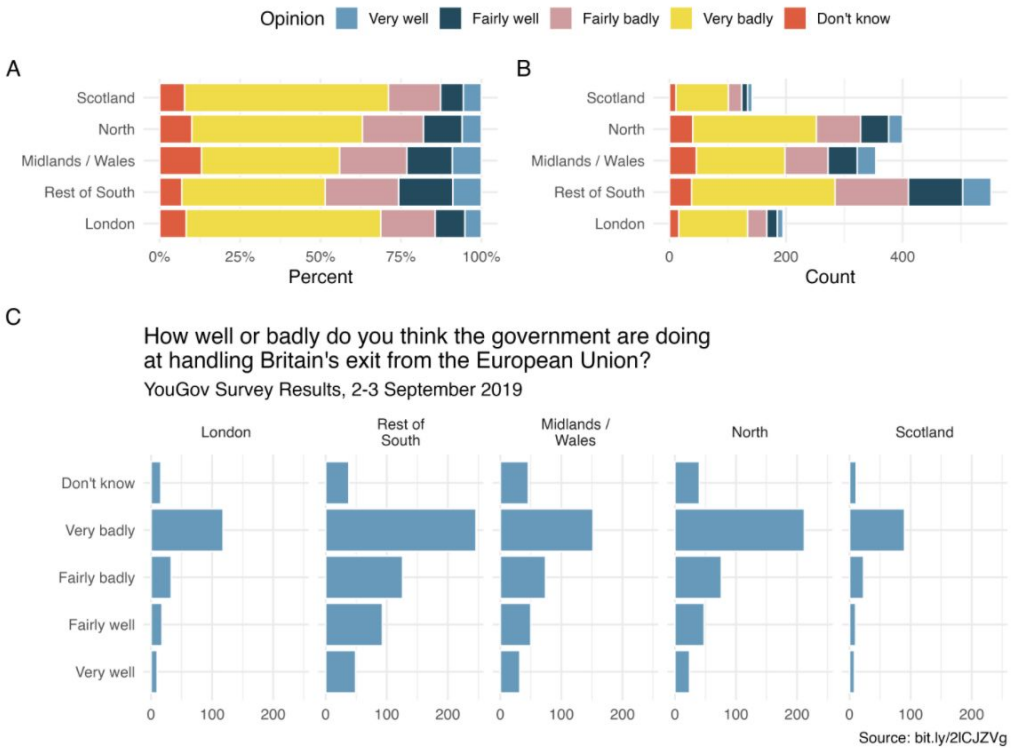
Make the labels as **easy to read** as possible



Stacked bar plots vertically and horizontally.

The horizontal orientation makes the region labels easier to read

Pick a **purpose**



Three different representations of the two variables including survey opinion and region.

A: across regions in proportion

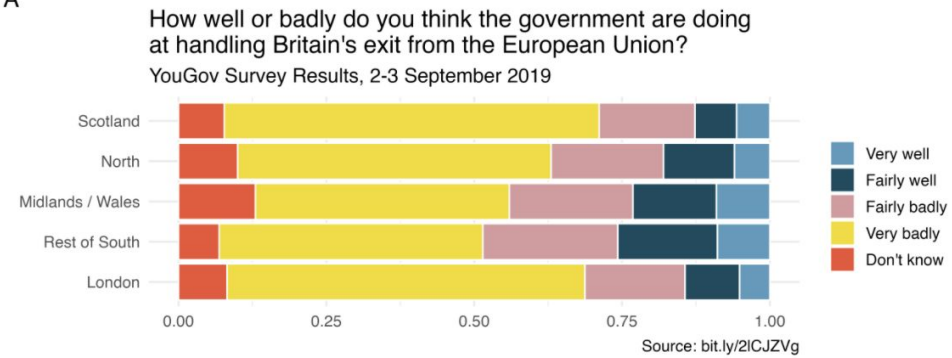
B: across region with consideration of total numbers of individuals

C: separate bar plot for each region

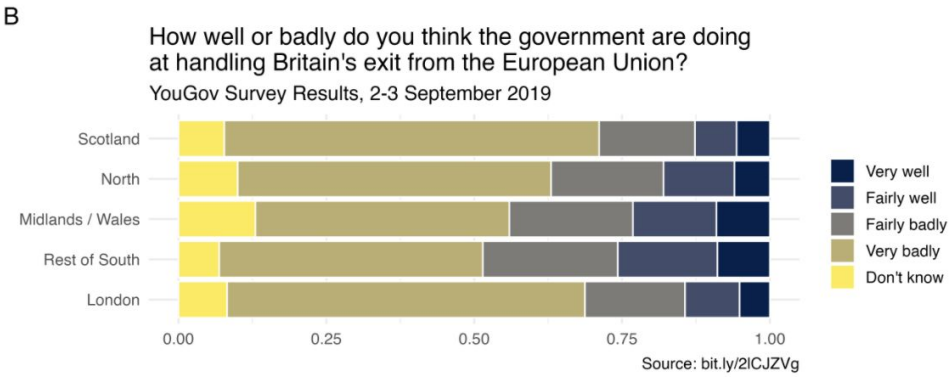
Use the graphic that best conveys the data narrative at hand.

Select meaningful colors

A



B



Identical bar plots with two different coloring options.

Plot A uses a default color scale,

Plot B uses colors from the cividis scale (works well with ordinal data)