

Tips for effective data visualization

Data Science in a Box

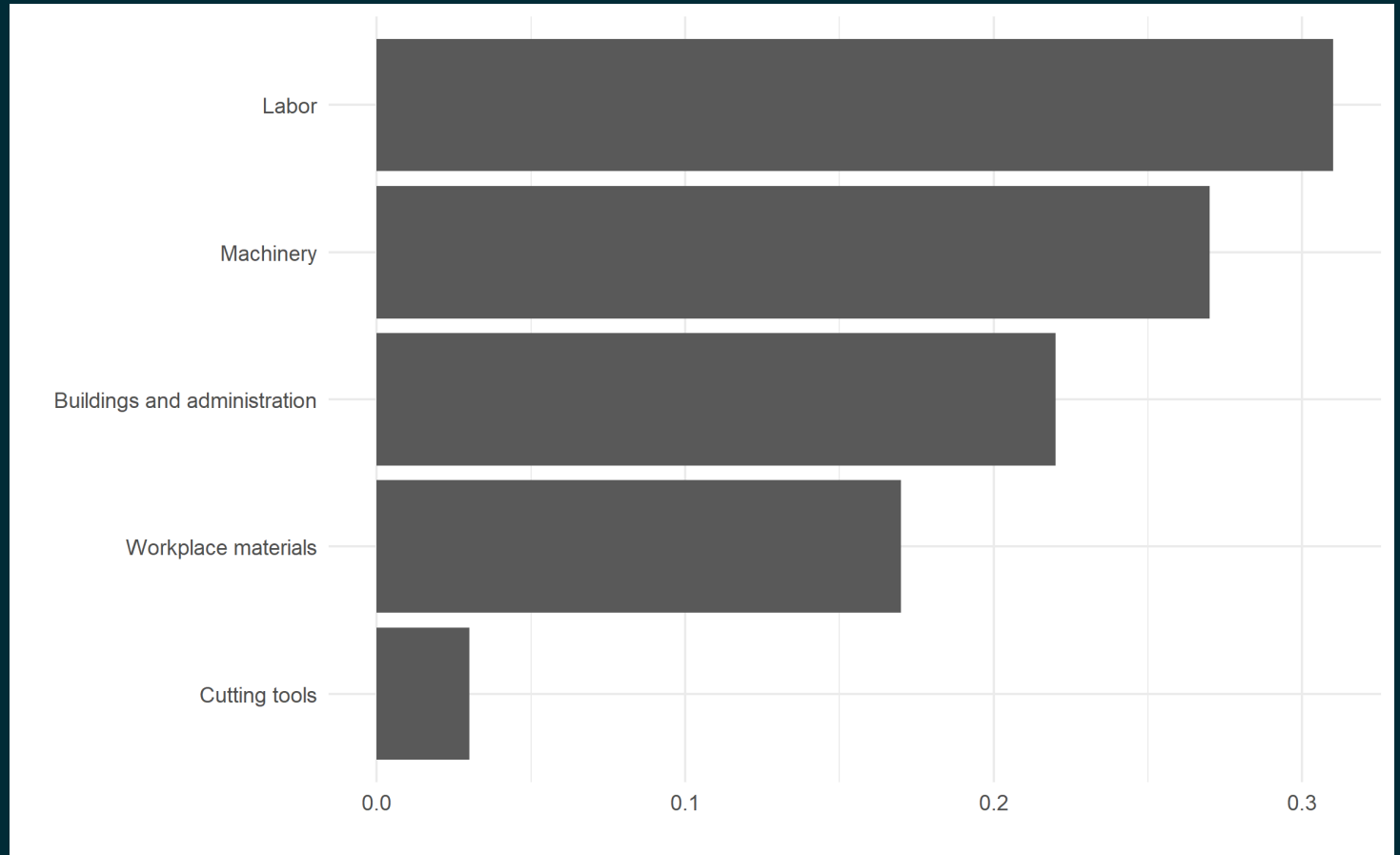
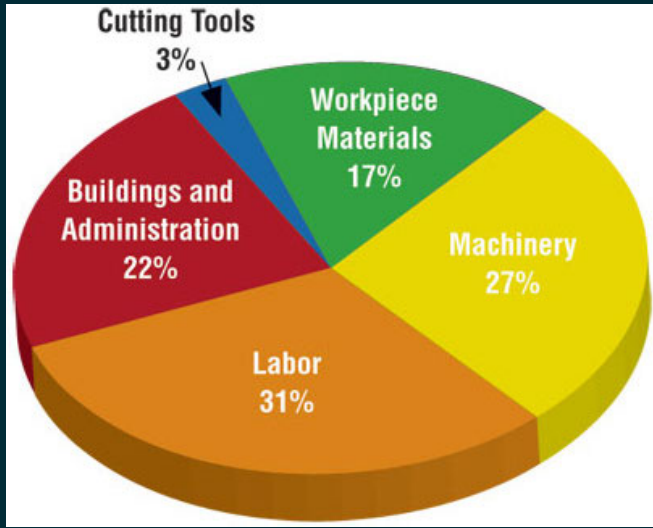
datasciencebox.org



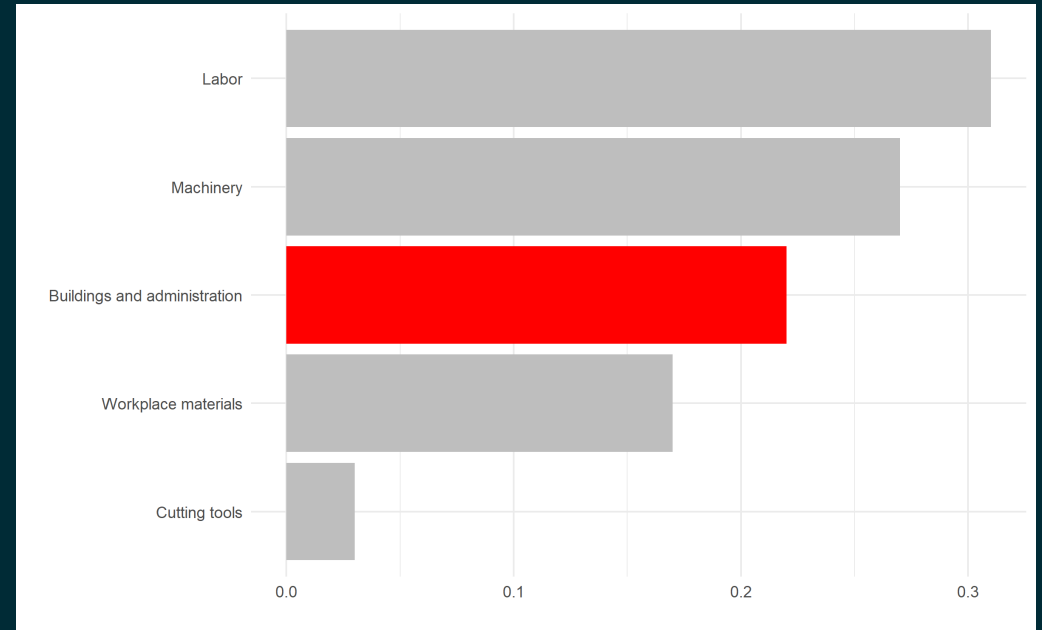
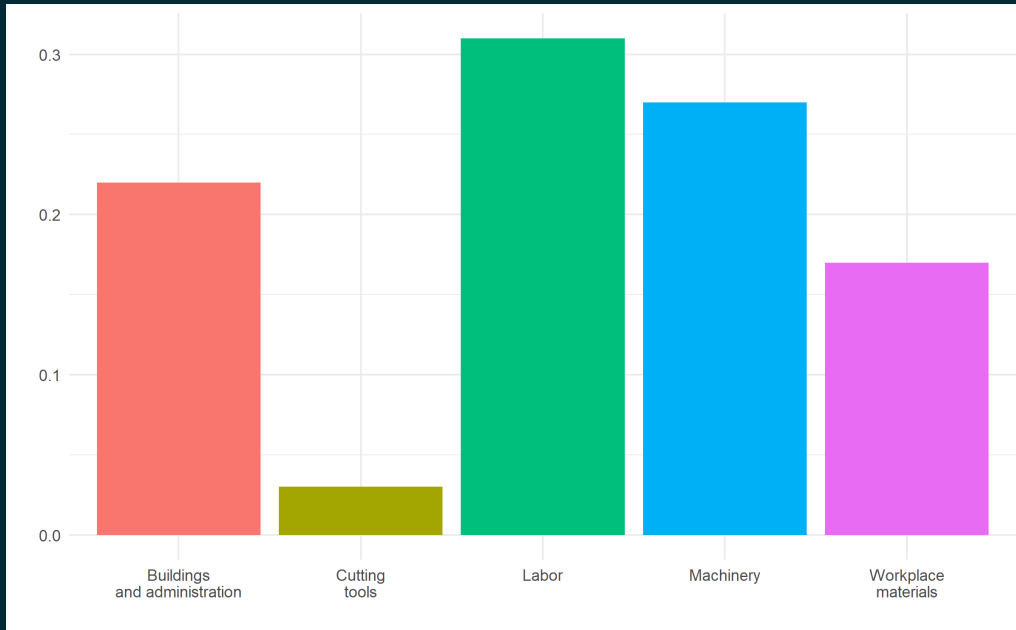
Designing effective visualizations



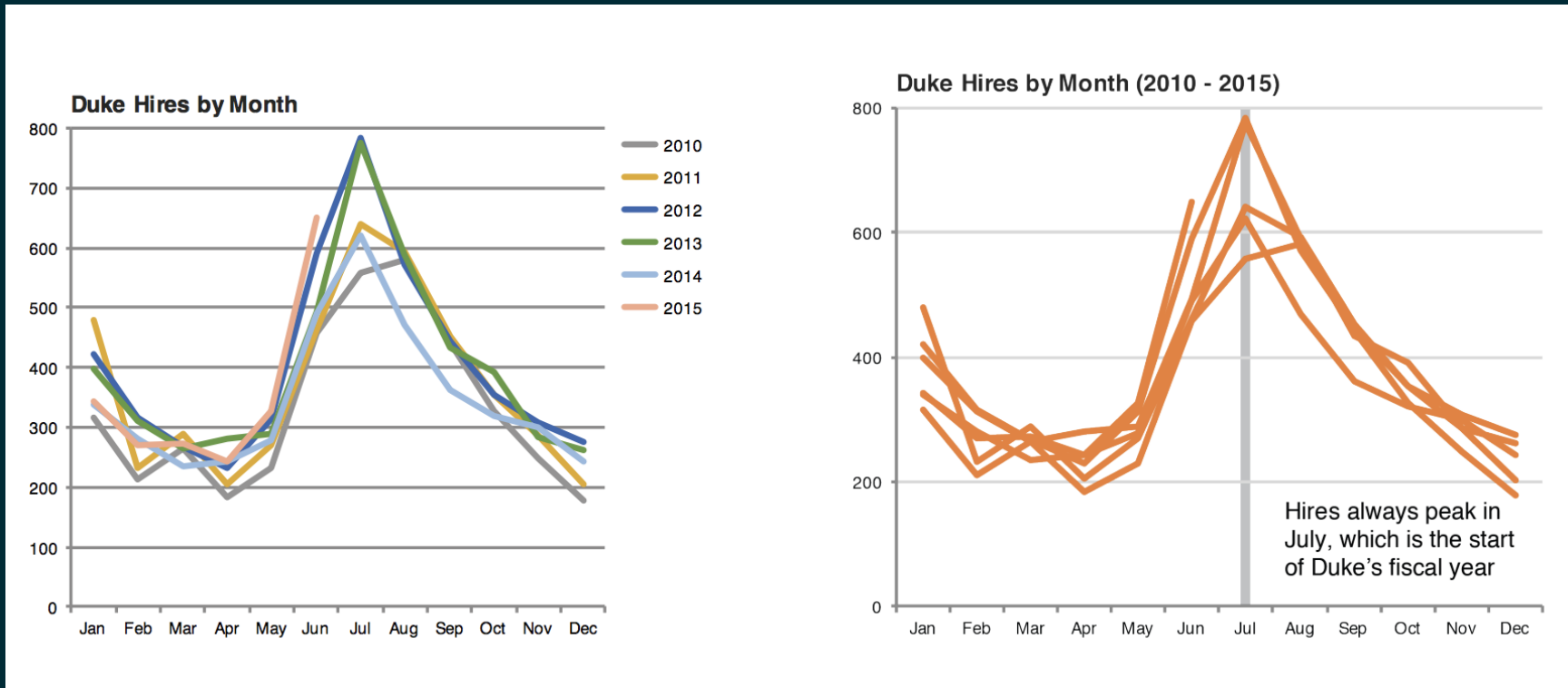
Keep it simple



Use color to draw attention



Tell a story



Credit: Angela Zoss and Eric Monson, Duke DVS



Principles for effective visualizations



Principles for effective visualizations

- Order matters
- Put long categories on the y-axis
- Keep scales consistent
- Select meaningful colors
- Use meaningful and nonredundant labels

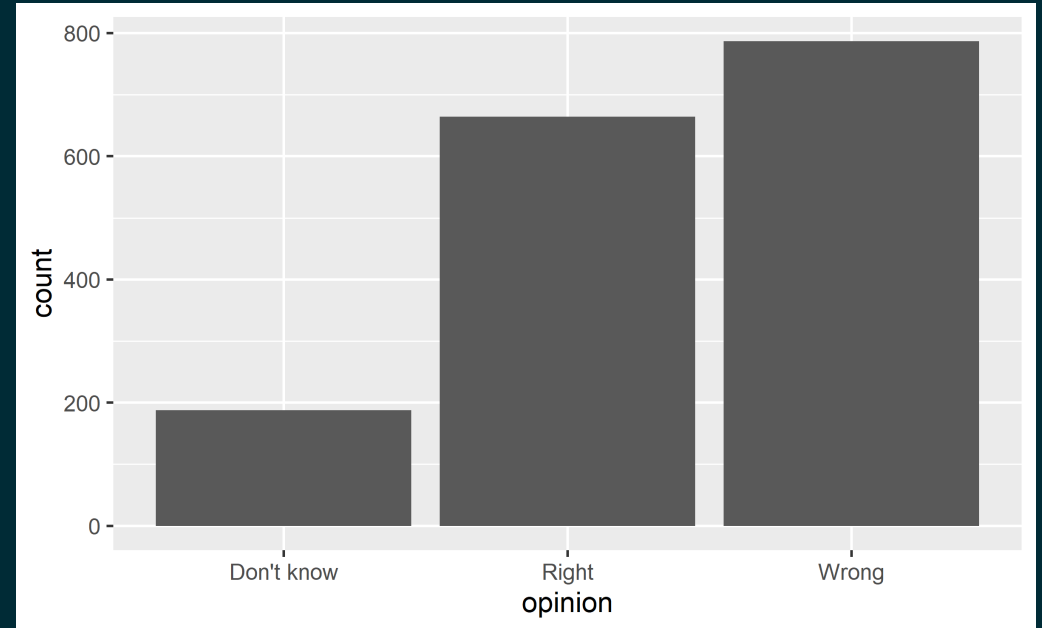


Data

In September 2019, YouGov survey asked 1,639 GB adults the following question:

In hindsight, do you think Britain was right/wrong to vote to leave EU?

- Right to leave
- Wrong to leave
- Don't know



Source: YouGov Survey Results, retrieved Oct 7, 2019



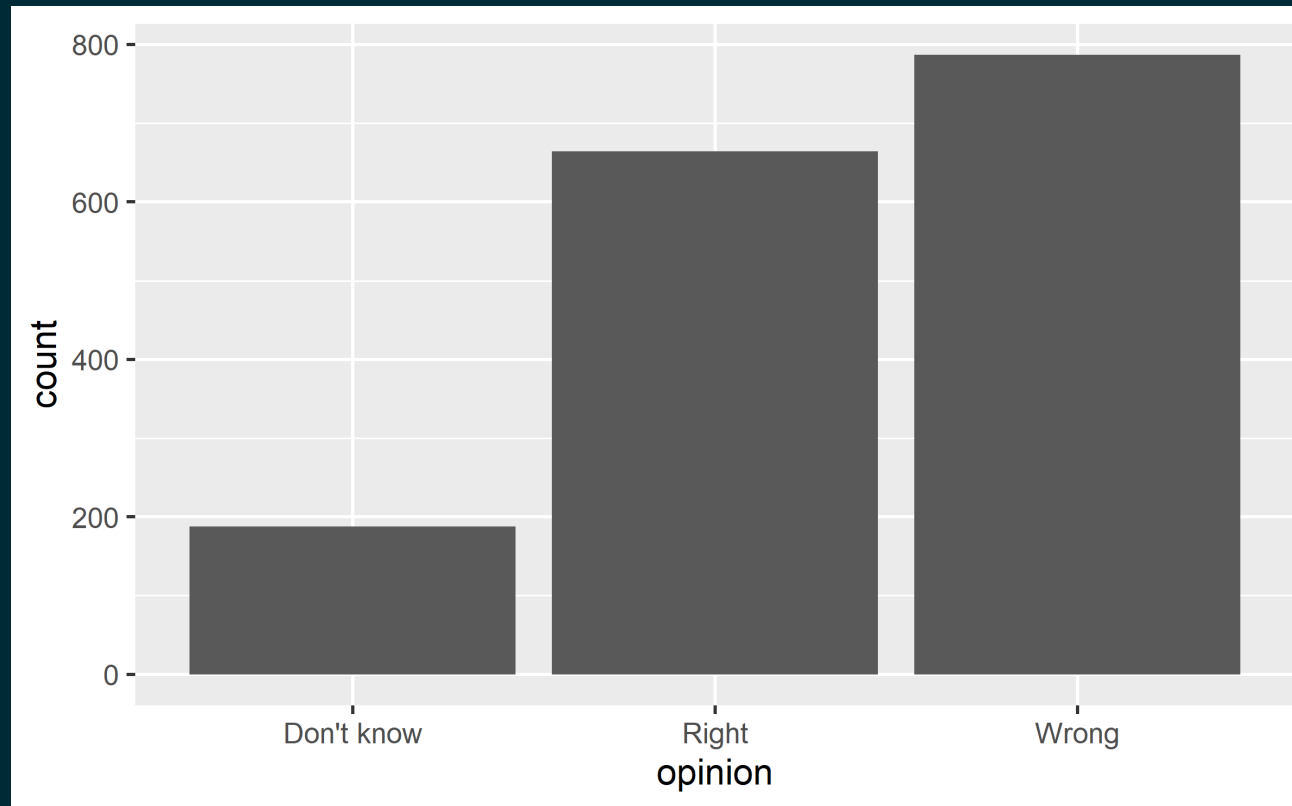
Order matters



Alphabetical order is rarely ideal

Plot

Code



Alphabetical order is rarely ideal

Plot

Code

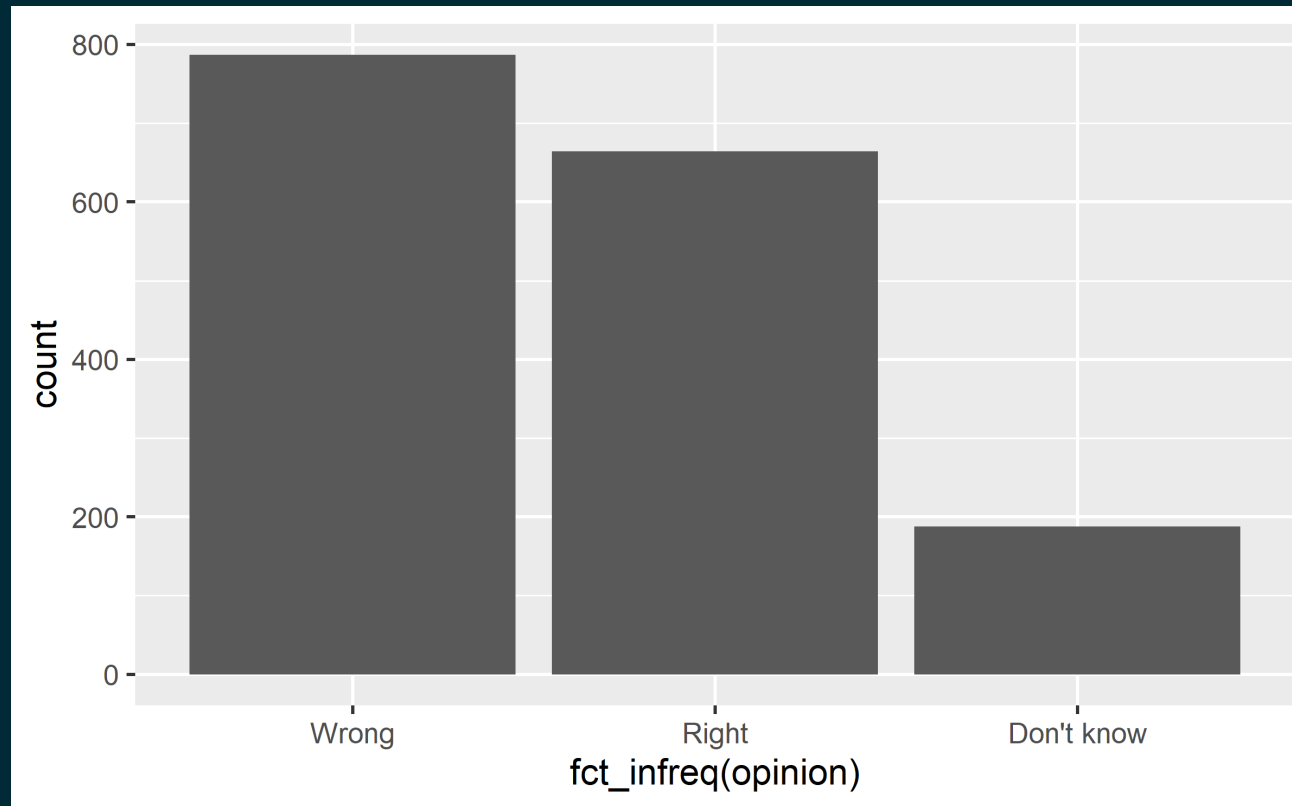
```
ggplot(brexit, aes(x = opinion)) +  
  geom_bar()
```



Order by frequency

Plot

Code



Order by frequency

Plot

Code

`fct_infreq`: Reorder factors' levels by frequency

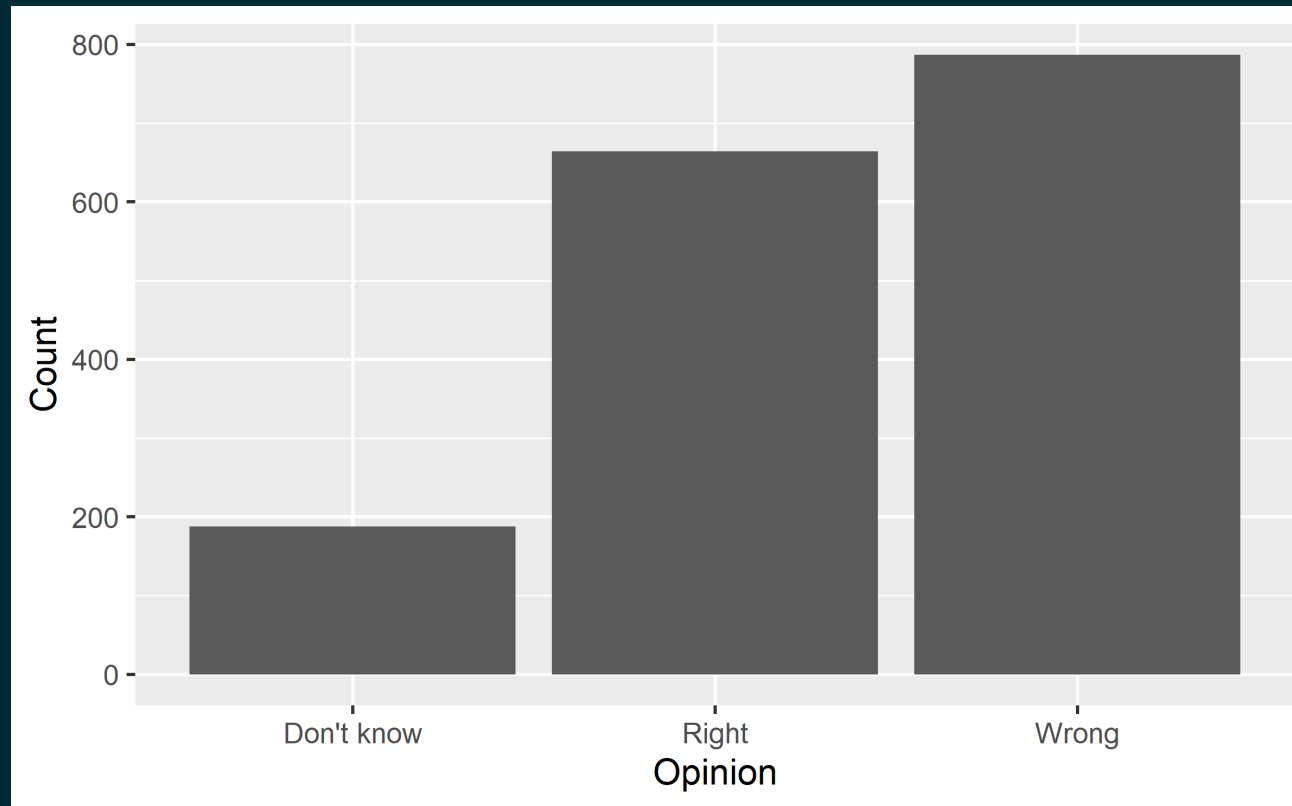
```
ggplot(brexit, aes(x = fct_infreq(opinion))) +  
  geom_bar()
```



Clean up labels

Plot

Code



Clean up labels

Plot

Code

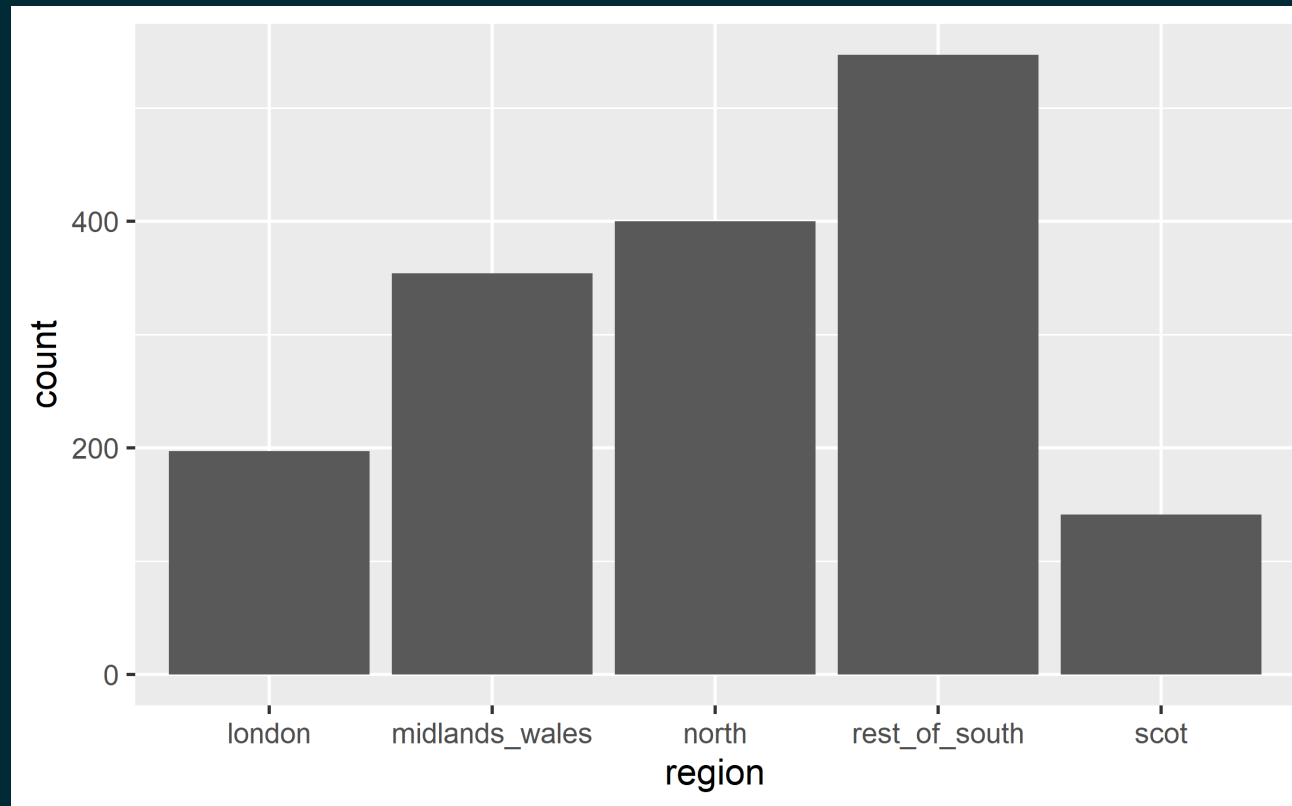
```
ggplot(brexit, aes(x = opinion)) +  
  geom_bar() +  
  labs(  
    x = "Opinion",  
    y = "Count"  
  )
```



Alphabetical order is rarely ideal

Plot

Code



Alphabetical order is rarely ideal

Plot

Code

```
ggplot(brexit, aes(x = region)) +  
  geom_bar()
```



Use inherent level order

Relevel

Plot

`fct_relevel`: Reorder factor levels using a custom order

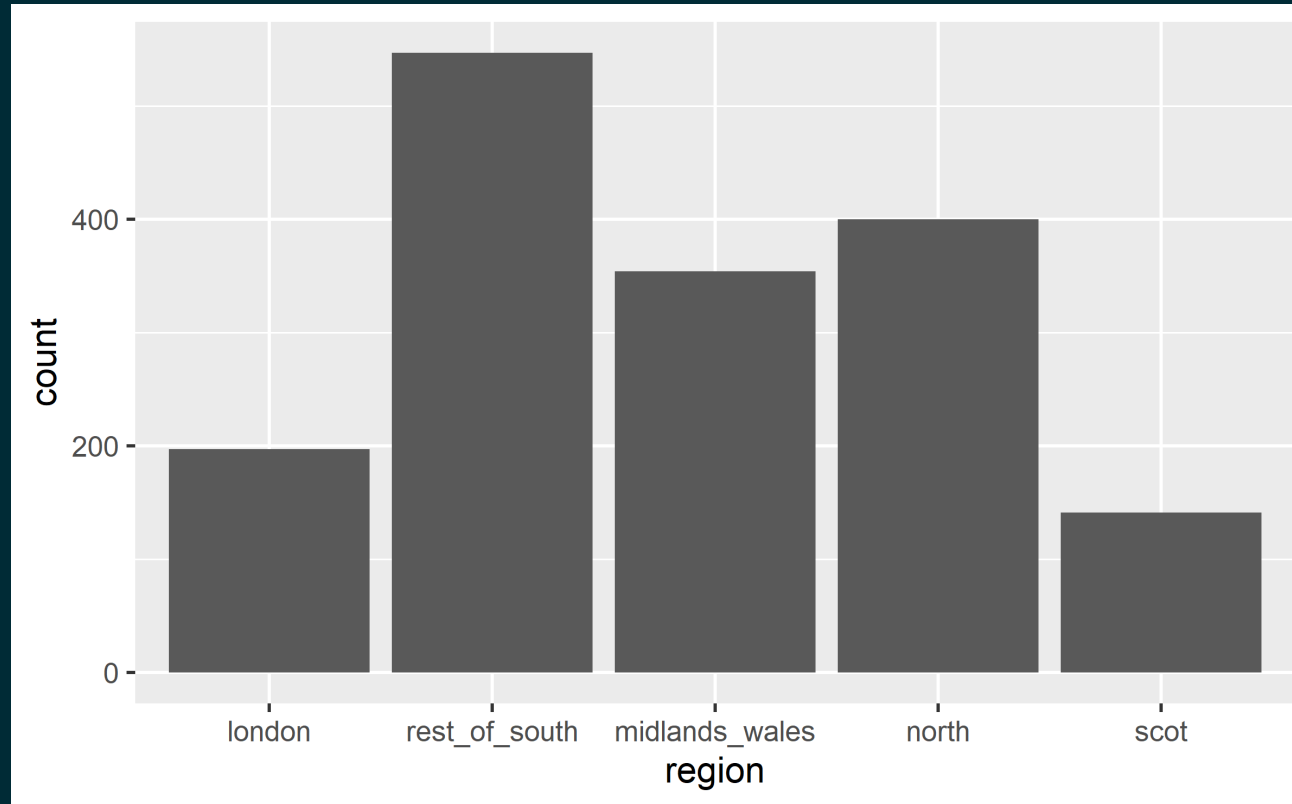
```
brexit <- brexit %>%  
  mutate(  
    region = fct_relevel(  
      region,  
      "london", "rest_of_south", "midlands_wales", "north", "scot"  
    )  
  )
```



Use inherent level order

Relevel

Plot



Clean up labels

Recode

Plot

fct_recode: Change factor levels by hand

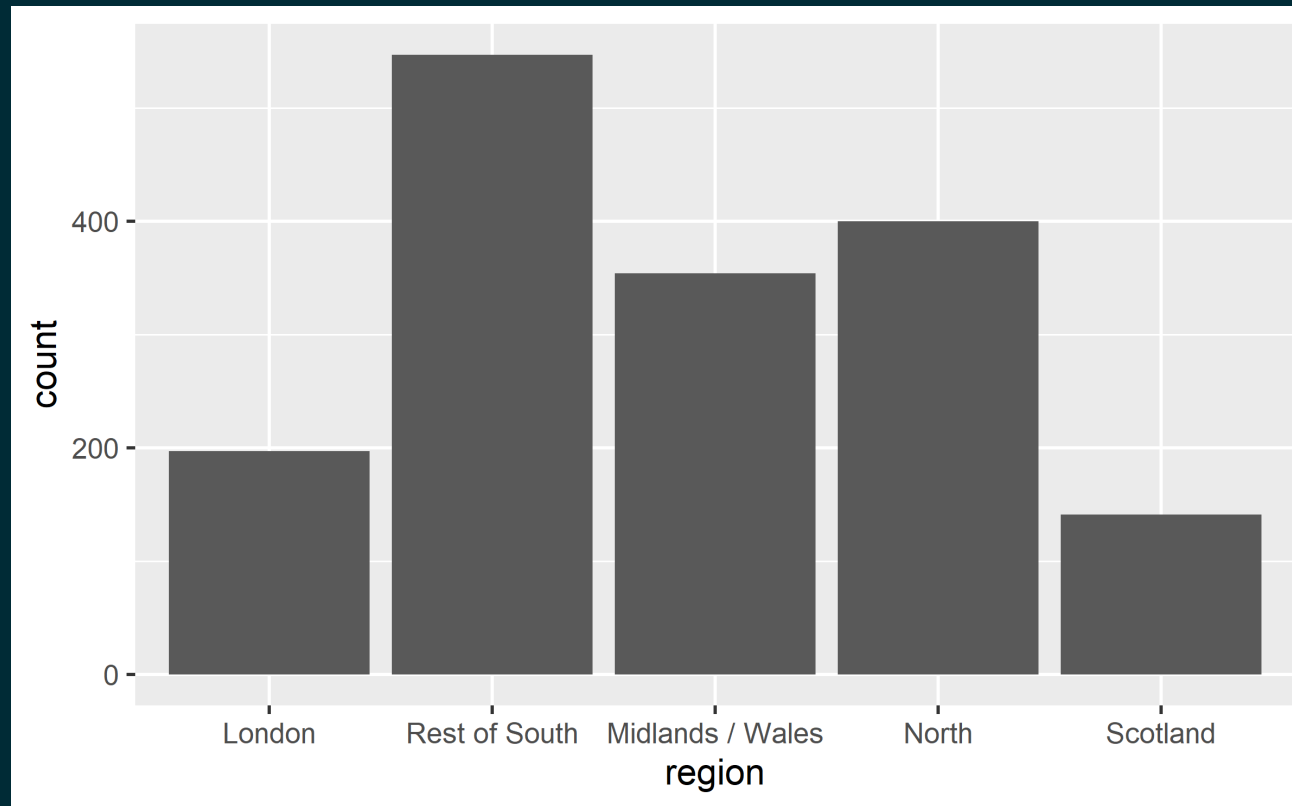
```
brexit <- brexit %>%
  mutate(
    region = fct_recode(
      region,
      London = "london",
      `Rest of South` = "rest_of_south",
      `Midlands / Wales` = "midlands_wales",
      North = "north",
      Scotland = "scot"
    )
  )
```



Clean up labels

Recode

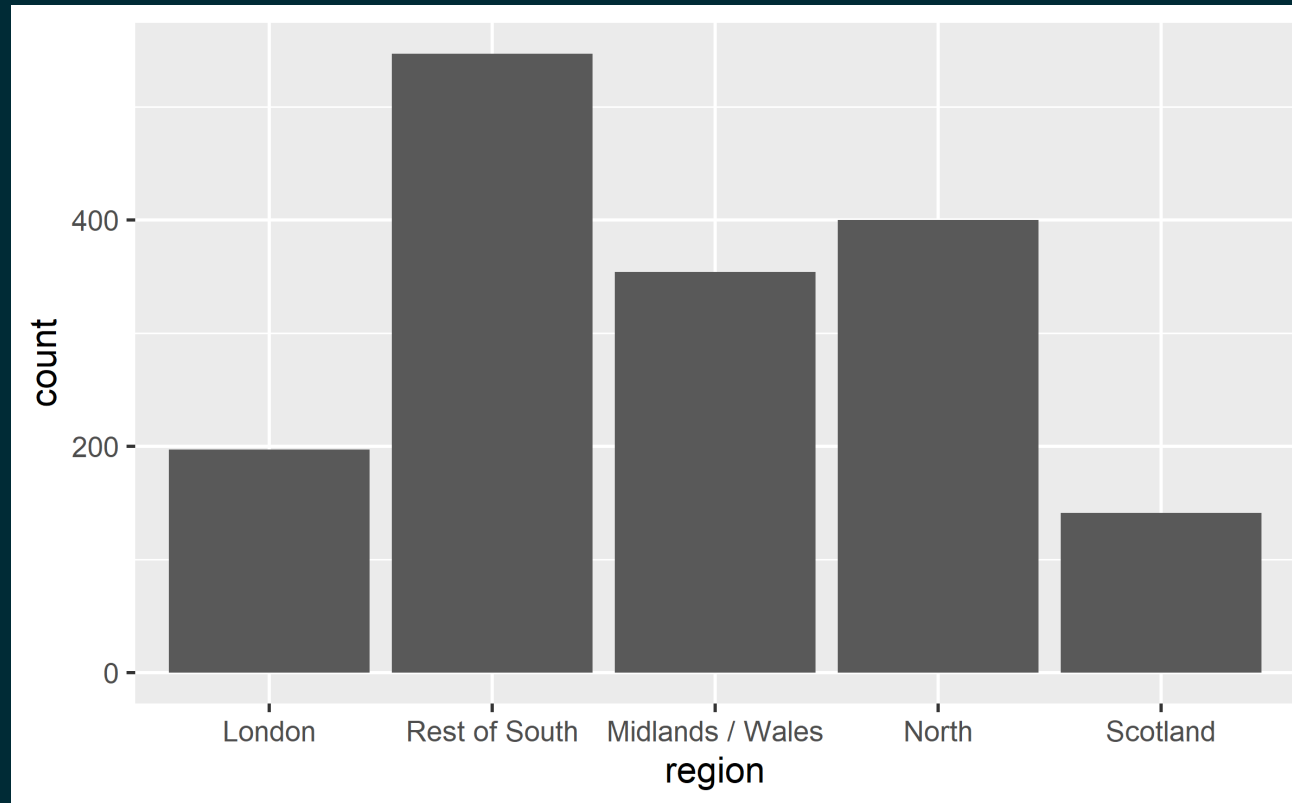
Plot



Put long categories on the y-axis



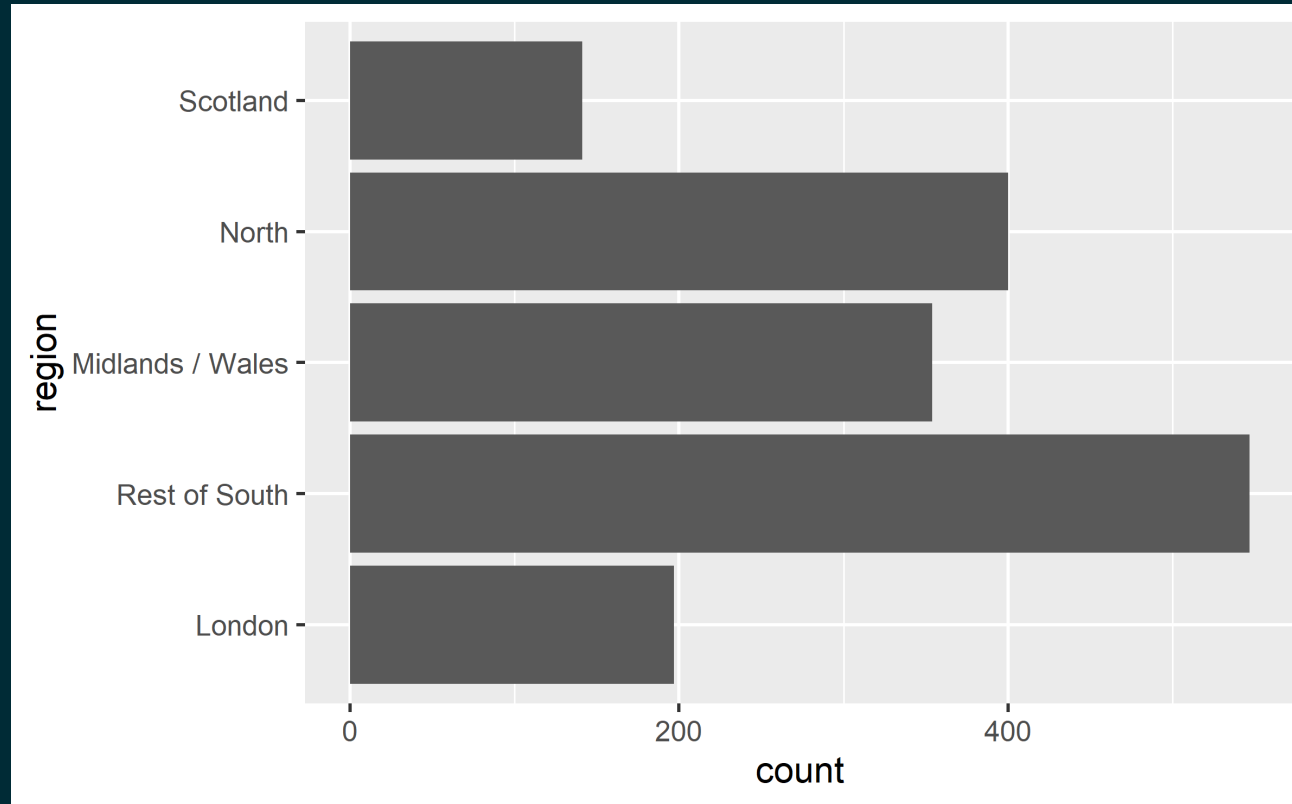
Long categories can be hard to read



Move them to the y-axis

Plot

Code



Move them to the y-axis

Plot

Code

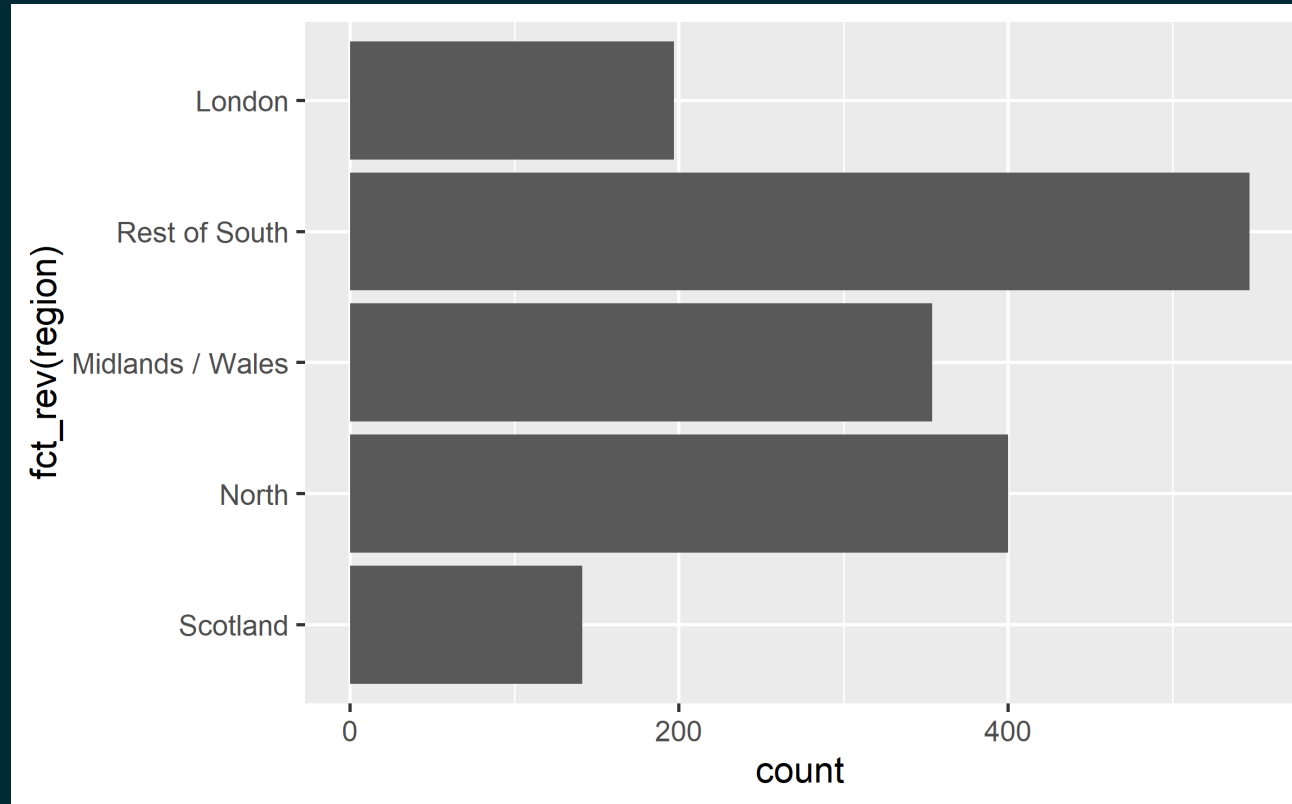
```
ggplot(brexit, aes(y = region)) +  
  geom_bar()
```



And reverse the order of levels

Plot

Code



And reverse the order of levels

Plot

Code

fct_rev: Reverse order of factor levels

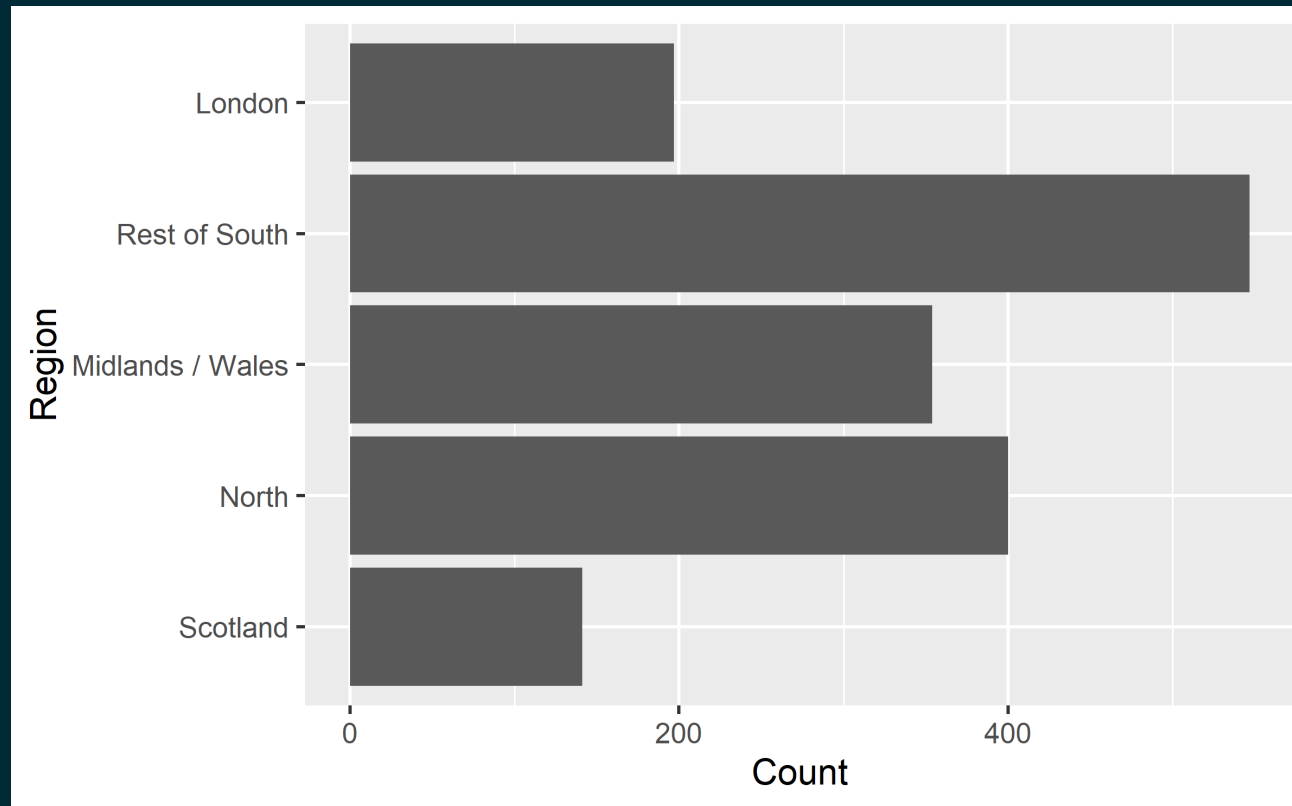
```
ggplot(brexit, aes(y = fct_rev(region))) +  
  geom_bar()
```



Clean up labels

Plot

Code



Clean up labels

Plot

Code

```
ggplot(brexit, aes(y = fct_rev(region))) +  
  geom_bar() +  
  labs(  
    x = "Count",  
    y = "Region"  
  )
```



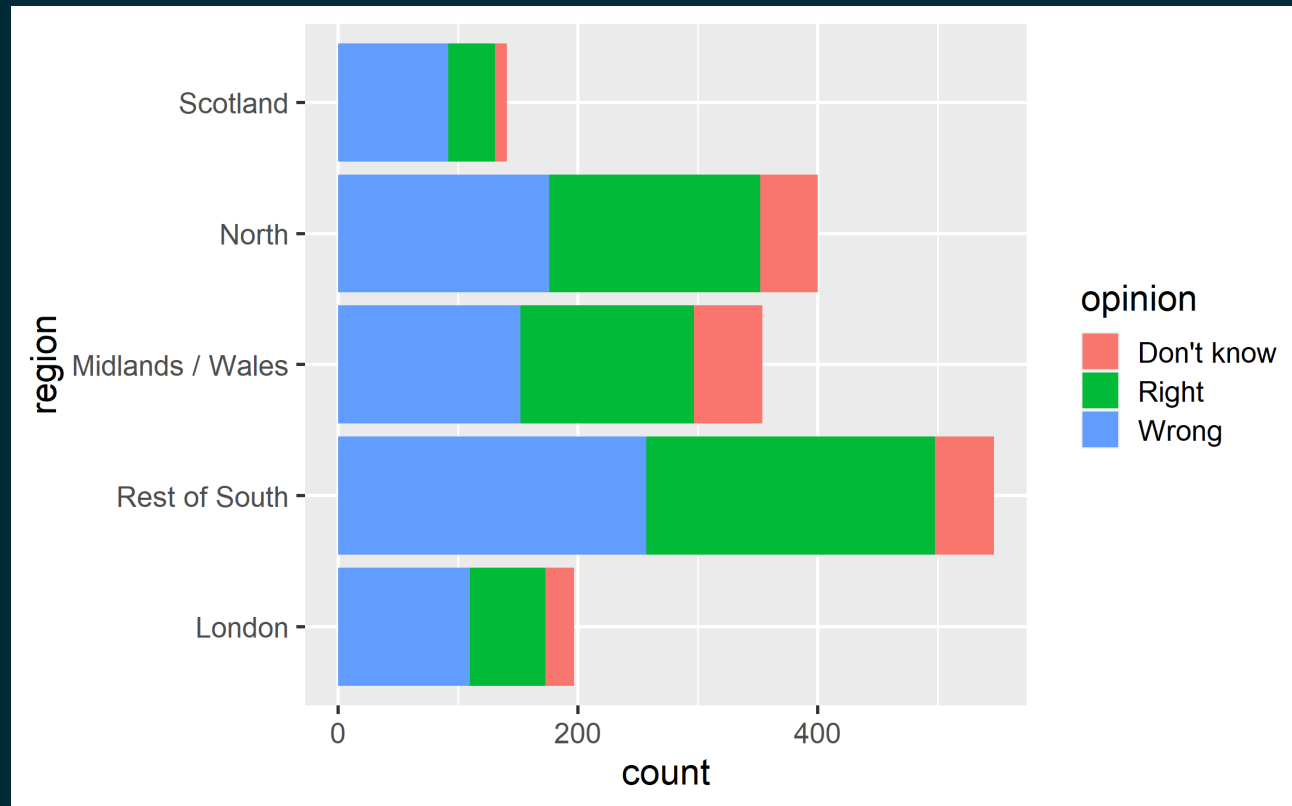
Pick a purpose



Segmented bar plots can be hard to read

Plot

Code



Segmented bar plots can be hard to read

Plot

Code

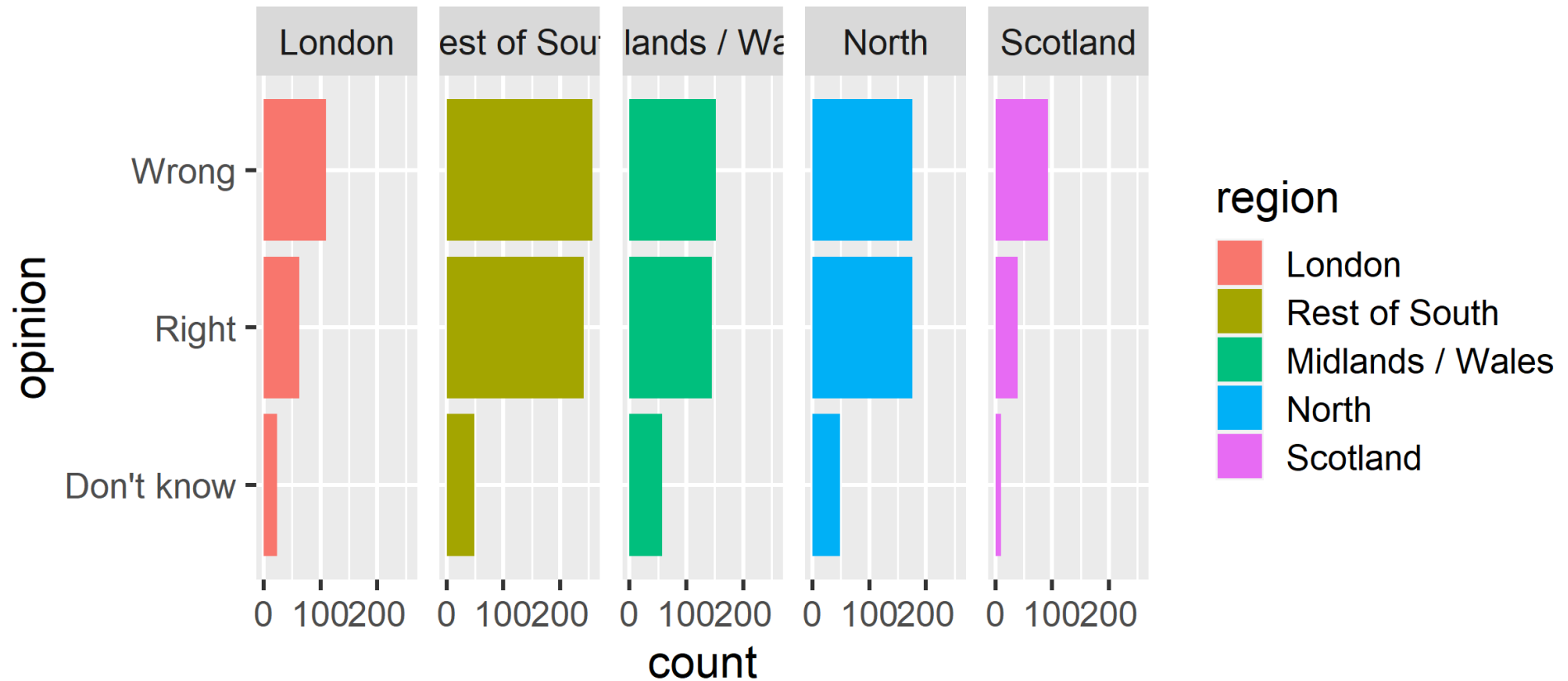
```
ggplot(brexit, aes(y = region, fill = opinion)) +  
  geom_bar()
```



Use facets

Plot

Code



Use facets

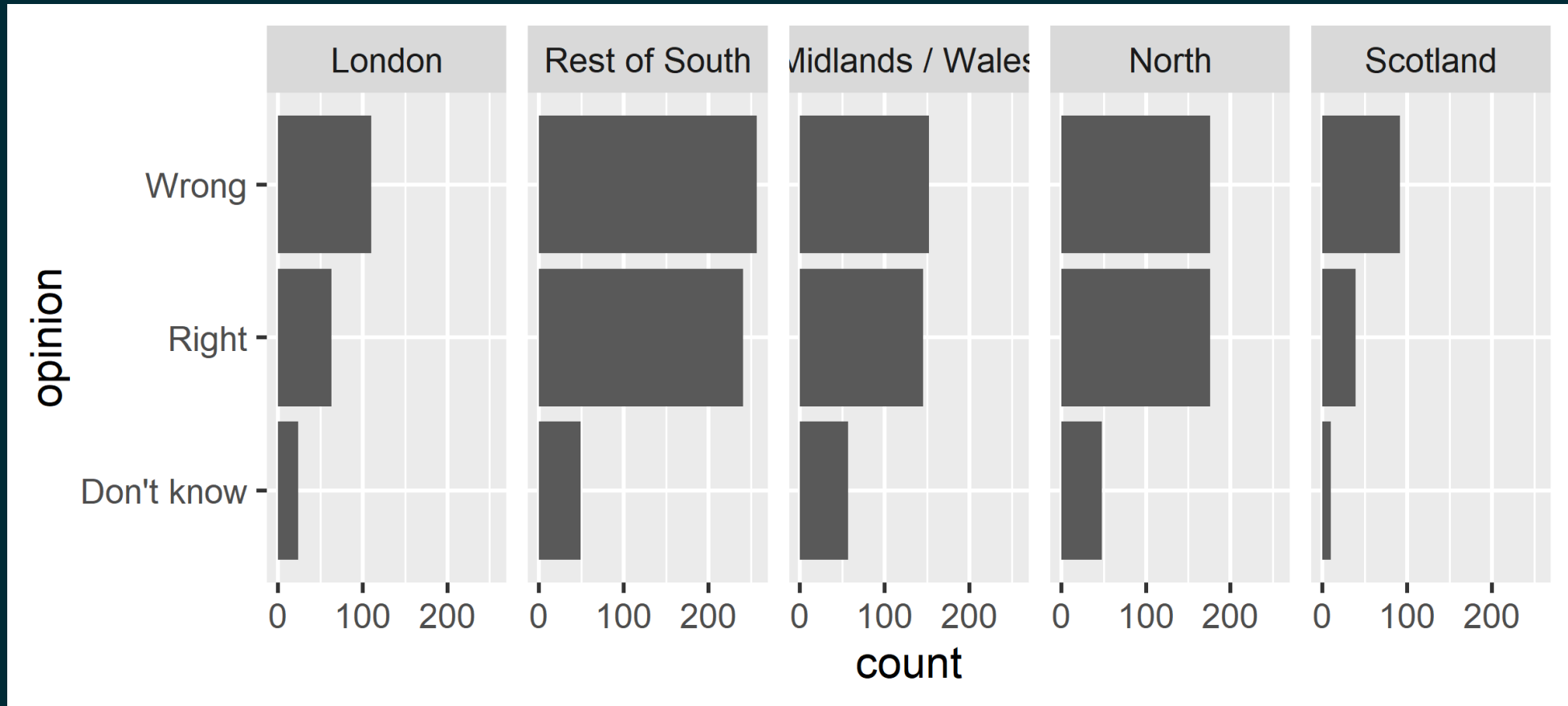
Plot

Code

```
ggplot(brexit, aes(y = opinion, fill = region)) +  
  geom_bar() +  
  facet_wrap(~region, nrow = 1)
```



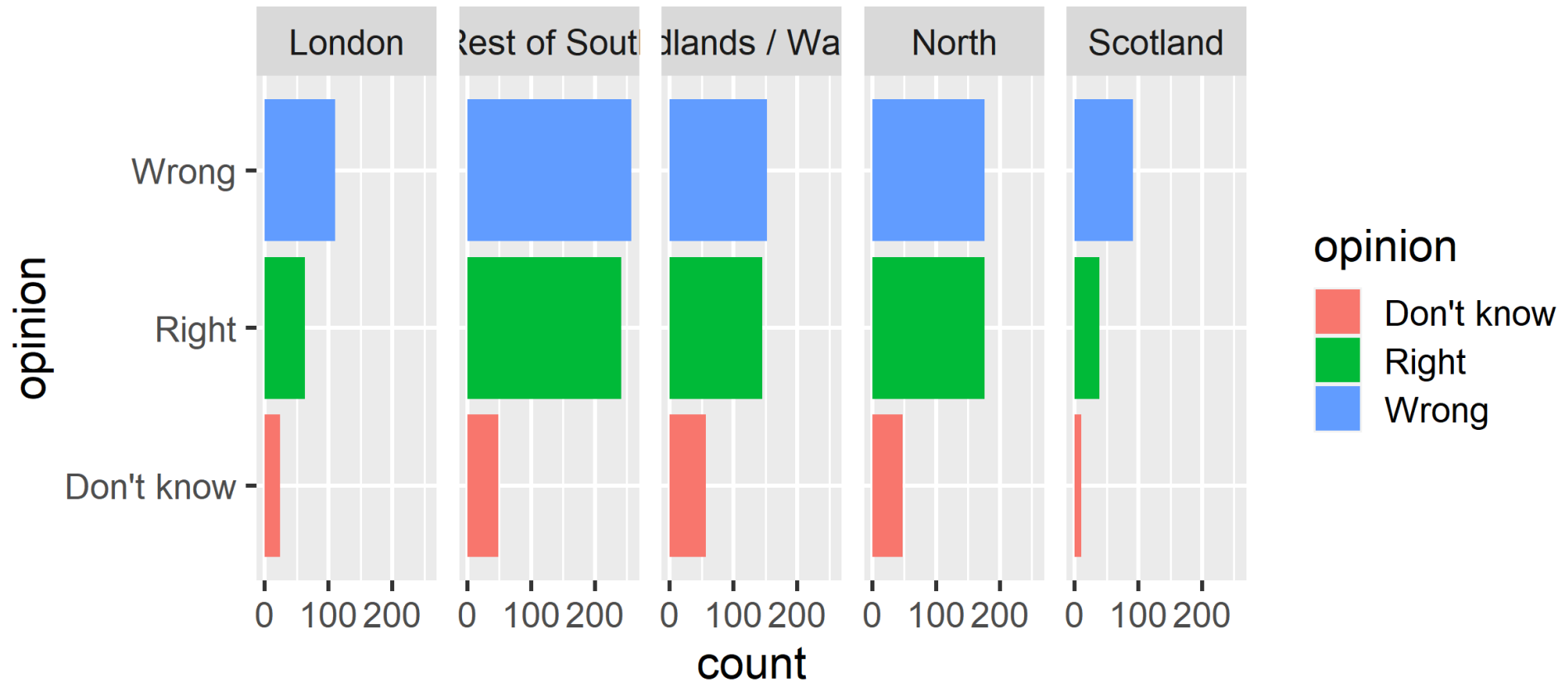
Avoid redundancy?



Redundancy can help tell a story

Plot

Code



Redundancy can help tell a story

Plot

Code

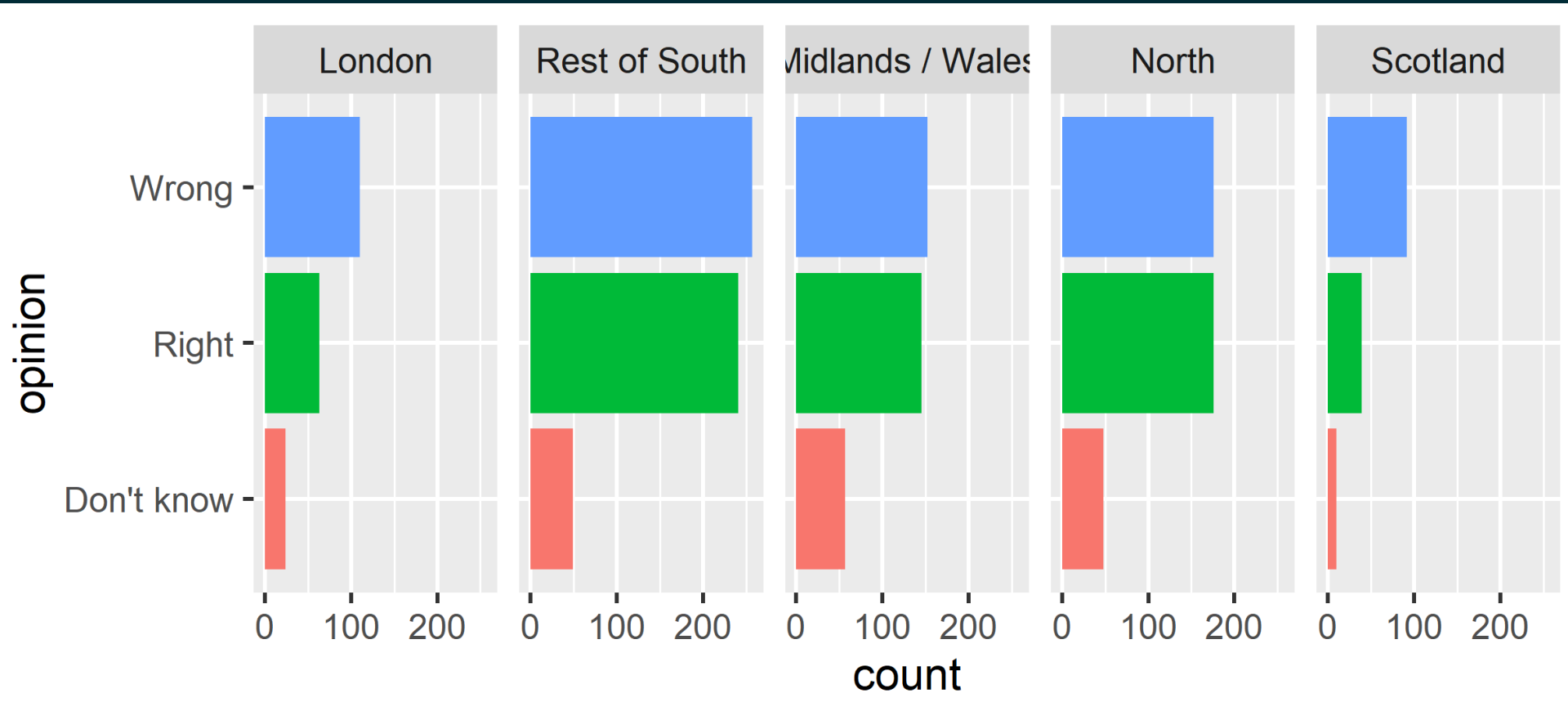
```
ggplot(brexit, aes(y = opinion, fill = opinion)) +  
  geom_bar() +  
  facet_wrap(~region, nrow = 1)
```



Be selective with redundancy

Plot

Code



Be selective with redundancy

Plot

Code

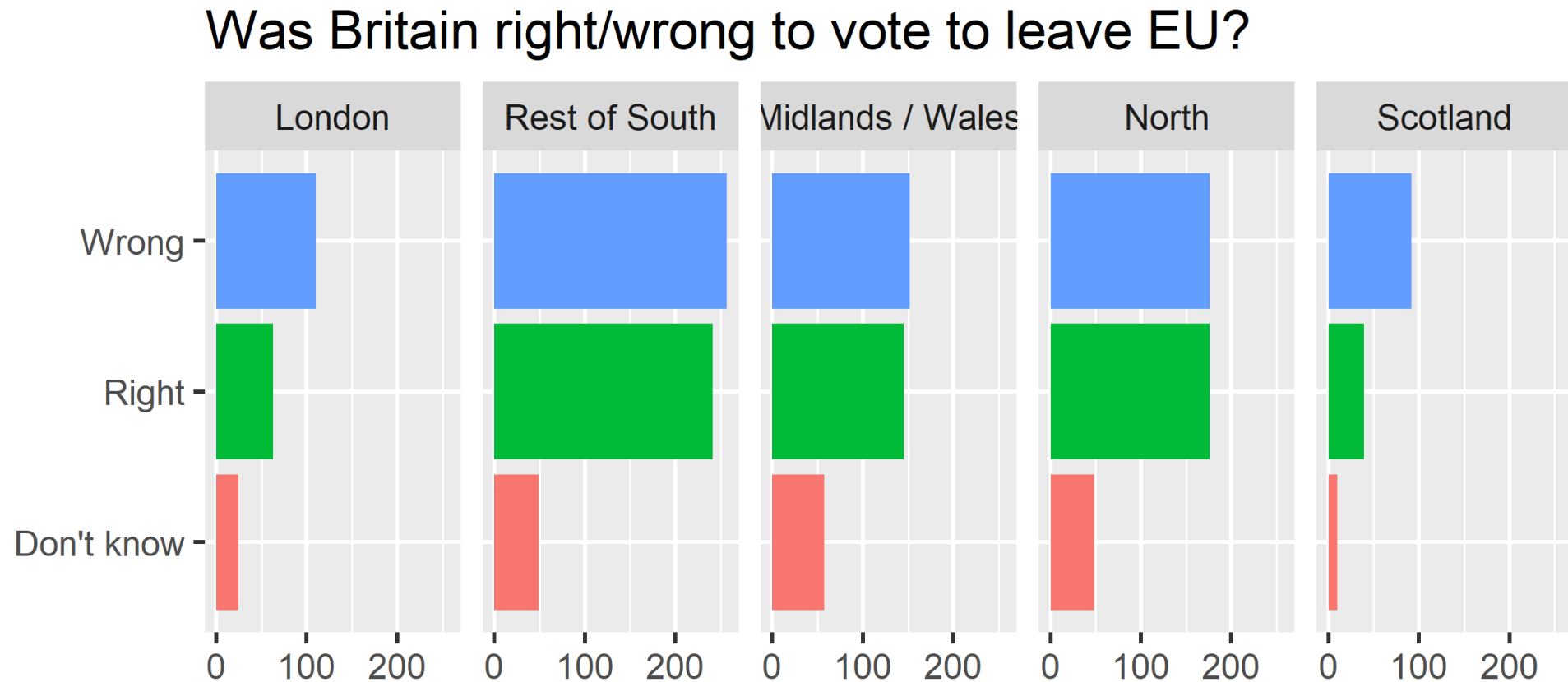
```
ggplot(brexit, aes(y = opinion, fill = opinion)) +  
  geom_bar() +  
  facet_wrap(~region, nrow = 1) +  
  guides(fill = "none")
```



Use informative labels

Plot

Code



Use informative labels

Plot

Code

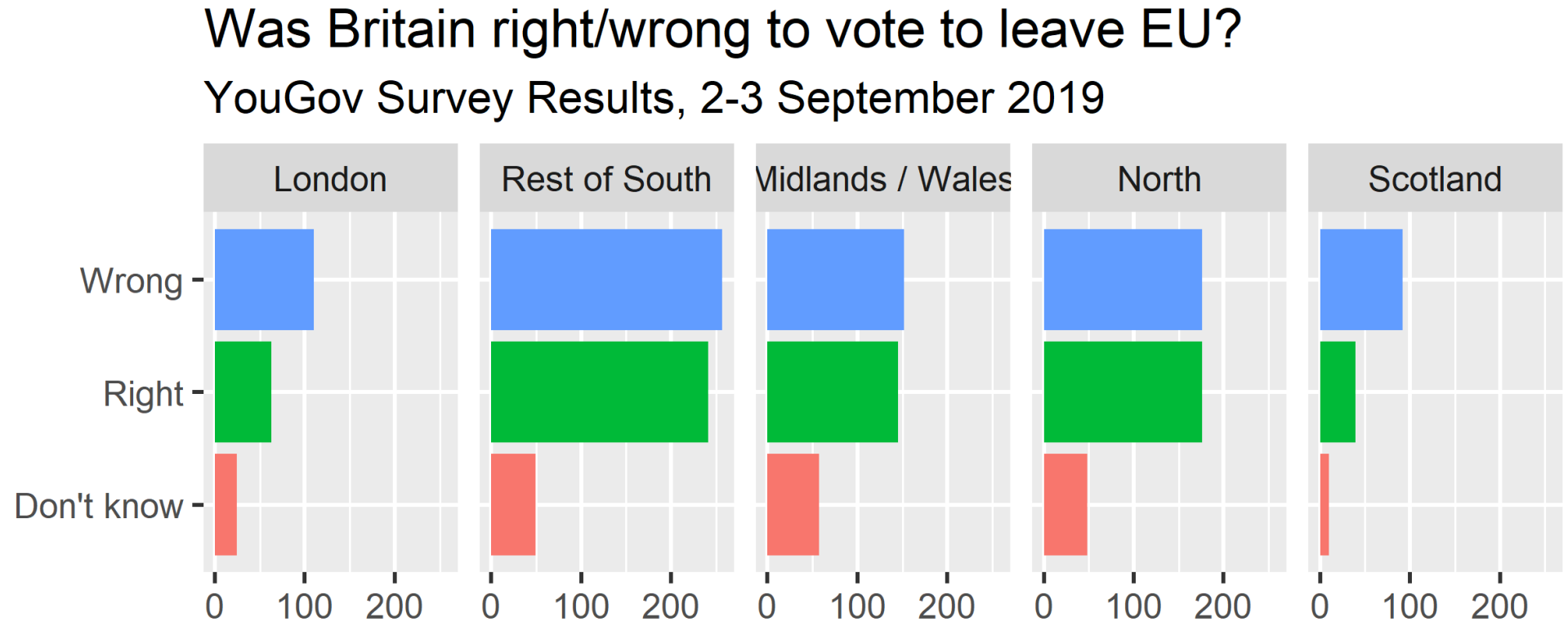
```
ggplot(brexit, aes(y = opinion, fill = opinion)) +  
  geom_bar() +  
  facet_wrap(~region, nrow = 1) +  
  guides(fill = "none") +  
  labs(  
    title = "Was Britain right/wrong to vote to leave EU?",  
    x = NULL, y = NULL  
  )
```



A bit more info

Plot

Code



https://www.yougov.com/cumulus_uploads/document/x0msmgx08/YouGov%20-%20Brexit%20and%202019%20election.pdf



A bit more info

Plot

Code

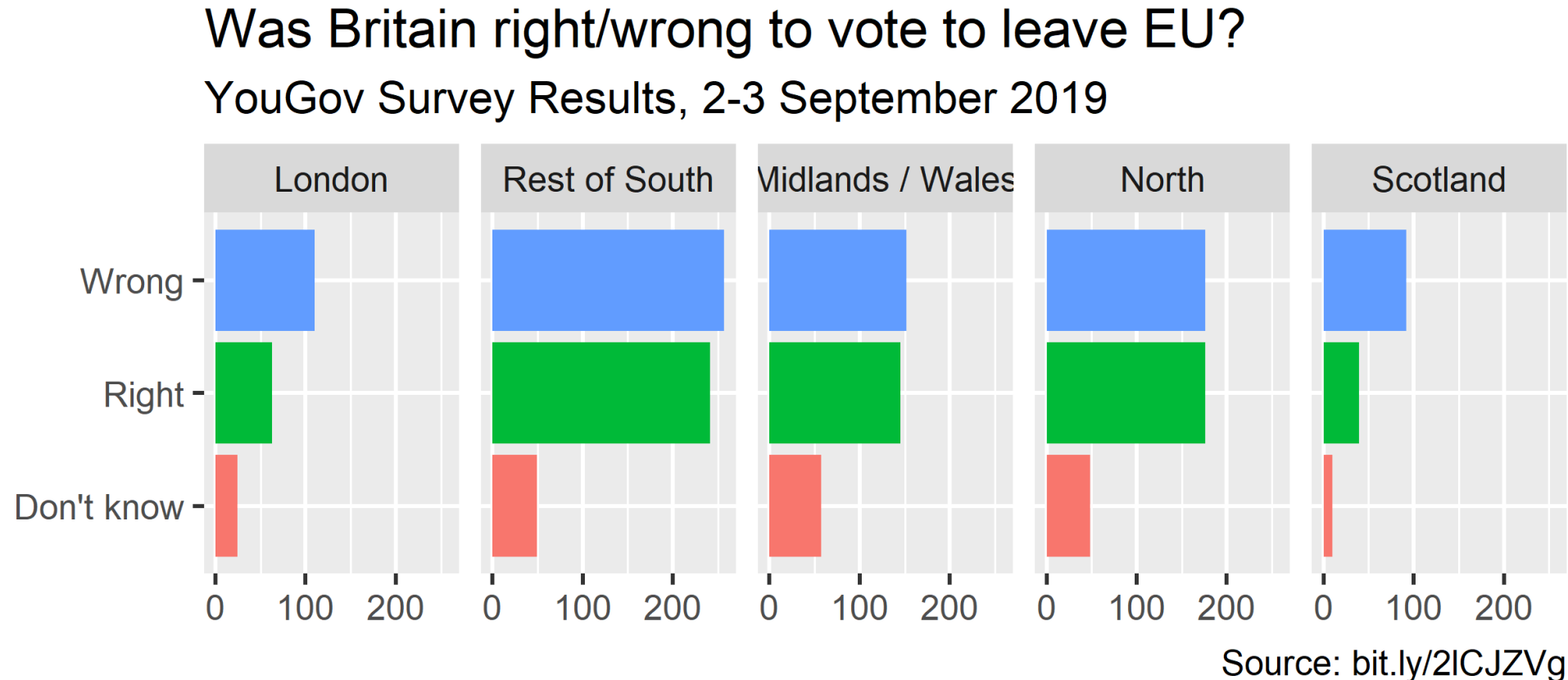
```
ggplot(brexit, aes(y = opinion, fill = opinion)) +  
  geom_bar() +  
  facet_wrap(~region, nrow = 1) +  
  guides(fill = "none") +  
  labs(  
    title = "Was Britain right/wrong to vote to leave EU?",  
    subtitle = "YouGov Survey Results, 2-3 September 2019",  
    caption = "Source: https://d25d2506sfb94s.cloudfront.net/cumulus\_uploads/document/x0msmggx08/Y  
    x = NULL, y = NULL  
  )
```



Let's do better

Plot

Code



Let's do better

Plot

Code

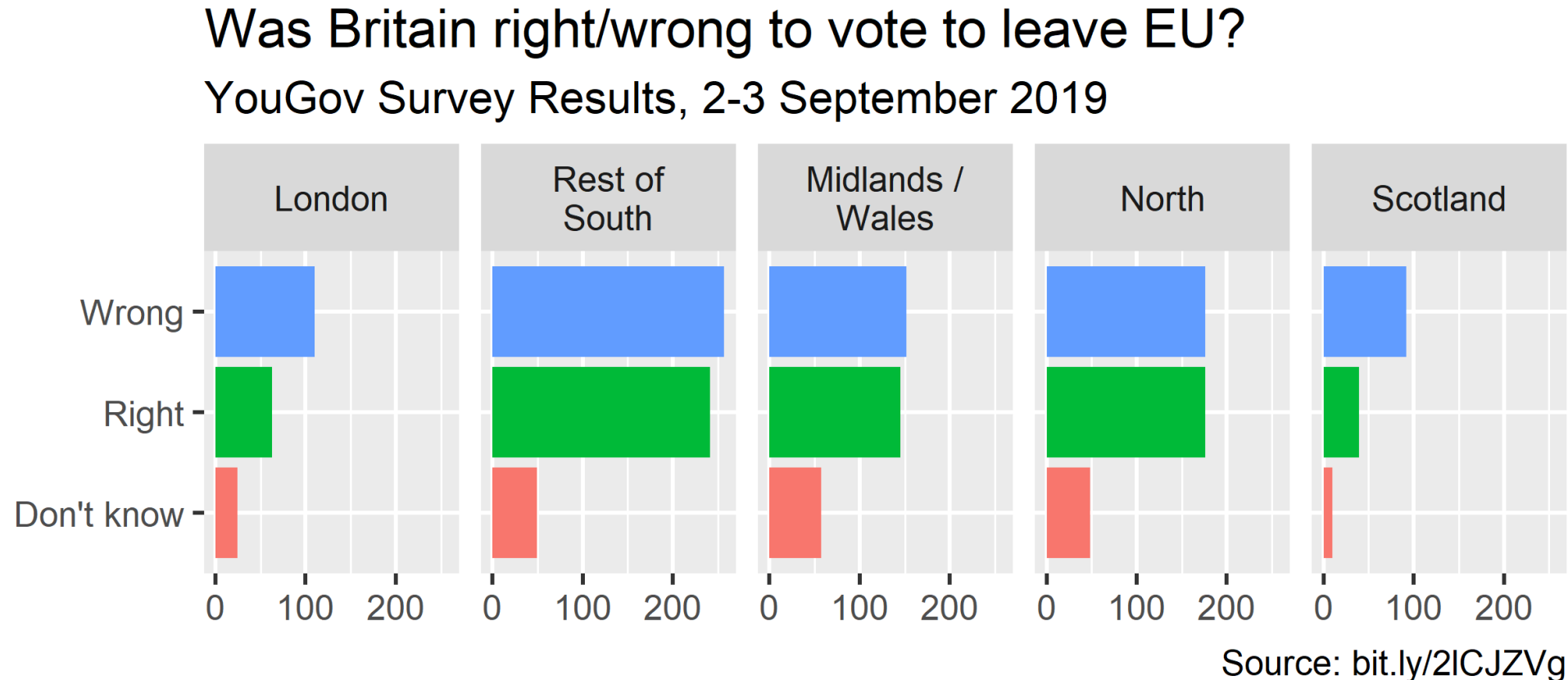
```
ggplot(brexit, aes(y = opinion, fill = opinion)) +  
  geom_bar() +  
  facet_wrap(~region, nrow = 1) +  
  guides(fill = "none") +  
  labs(  
    title = "Was Britain right/wrong to vote to leave EU?",  
    subtitle = "YouGov Survey Results, 2-3 September 2019",  
    caption = "Source: bit.ly/2lCJZVg",  
    x = NULL, y = NULL  
  )
```



Fix up facet labels

Plot

Code



Fix up facet labels

Plot

Code

```
ggplot(brexit, aes(y = opinion, fill = opinion)) +  
  geom_bar() +  
  facet_wrap(~region,  
    nrow = 1,  
    labeller = label_wrap_gen(width = 12)  
  ) +  
  guides(fill = "none") +  
  labs(  
    title = "Was Britain right/wrong to vote to leave EU?",  
    subtitle = "YouGov Survey Results, 2-3 September 2019",  
    caption = "Source: bit.ly/2lCJZVg",  
    x = NULL, y = NULL  
  )
```



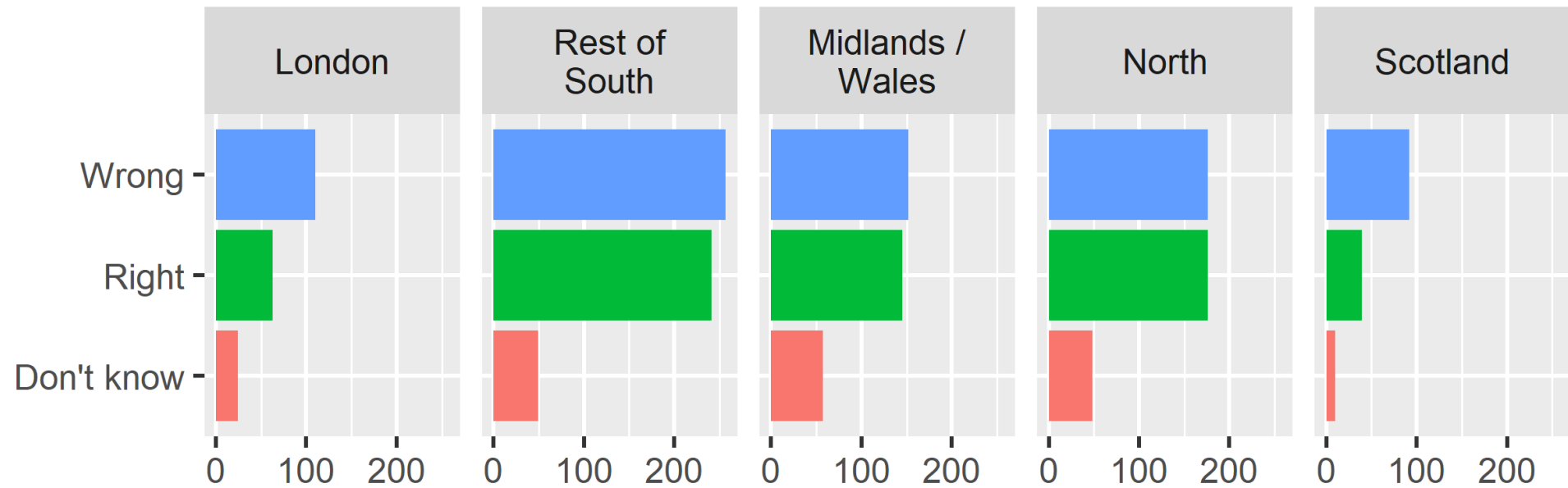
Select meaningful colors



Rainbow colors not always the right choice

Was Britain right/wrong to vote to leave EU?

YouGov Survey Results, 2-3 September 2019



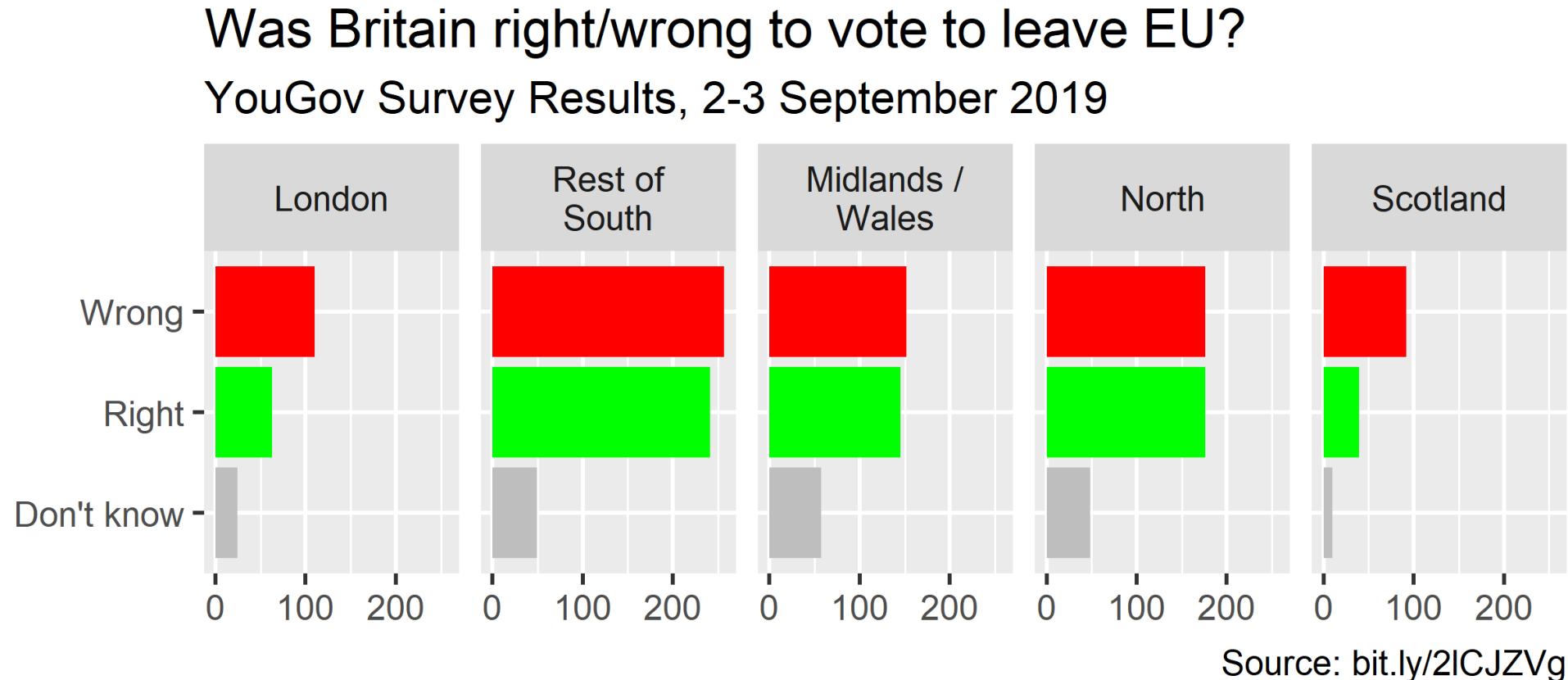
Source: bit.ly/2ICJZVg



Manually choose colors when needed

Plot

Code



Manually choose colors when needed

Plot

Code

```
ggplot(brexit, aes(y = opinion, fill = opinion)) +  
  geom_bar() +  
  facet_wrap(~region, nrow = 1, labeller = label_wrap_gen(width = 12)) +  
  guides(fill = "none") +  
  labs(title = "Was Britain right/wrong to vote to leave EU?",  
        subtitle = "YouGov Survey Results, 2-3 September 2019",  
        caption = "Source: bit.ly/2lCJZVg",  
        x = NULL, y = NULL) +  
  scale_fill_manual(values = c(  
    "Wrong" = "red",  
    "Right" = "green",  
    "Don't know" = "gray"  
  ))
```



Choosing better colors

colorbrewer2.org

The screenshot shows the ColorBrewer 2.0 web application interface. The browser address bar displays the URL: `colorbrewer2.org/#type=diverging&scheme=RdBu&n=3`. The interface includes the following elements:

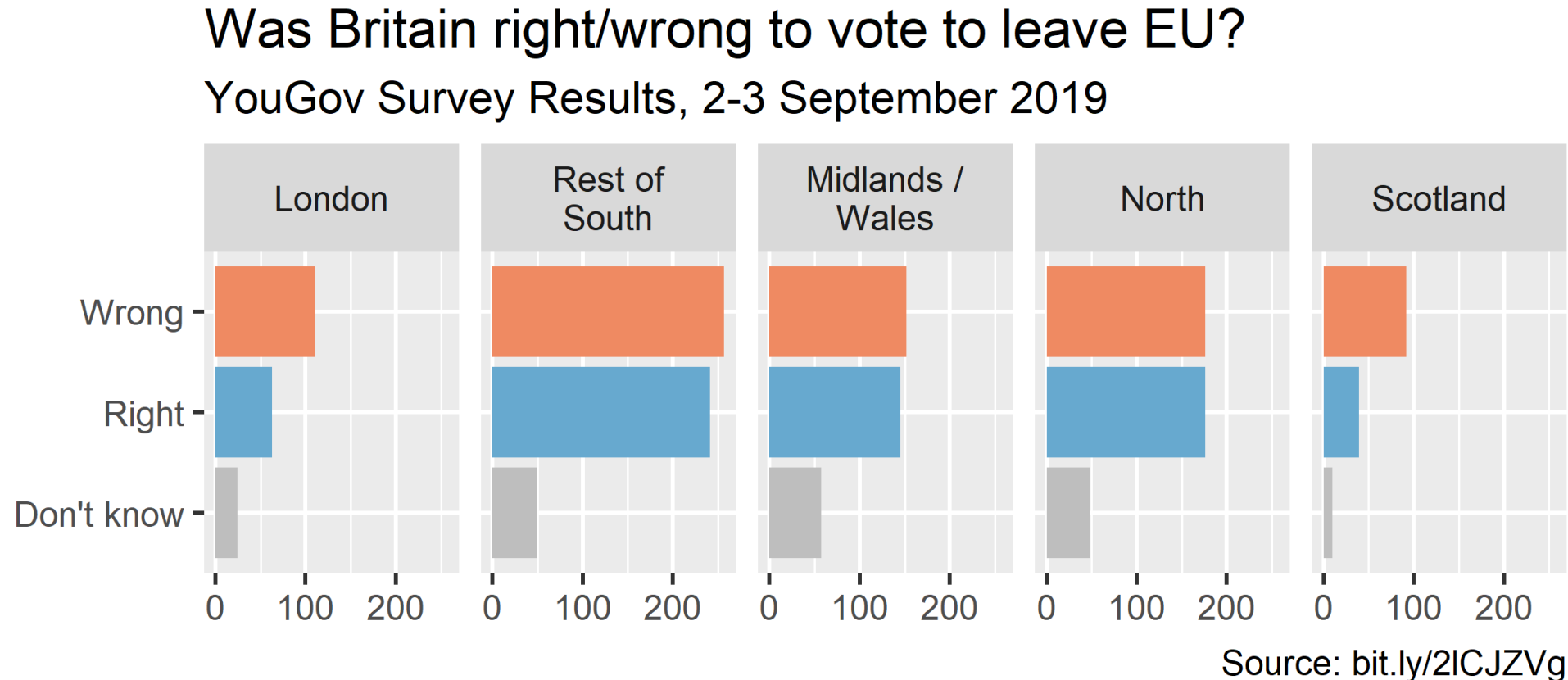
- Number of data classes:** A dropdown menu set to 3.
- Nature of your data:** Radio buttons for sequential, **diverging** (selected), and qualitative.
- Pick a color scheme:** A grid of 12 color scheme thumbnails.
- Only show:** Checkboxes for colorblind safe (checked), print friendly, and photocopy safe.
- Context:** Checkboxes for roads, cities, and **borders** (checked).
- Background:** Radio buttons for solid color (selected) and terrain, with a color transparency slider below.
- 3-class RdBu:** A legend showing three color swatches: orange (#ef8a62), white (#f7f7f7), and blue (#67a9cf).
- EXPORT:** A button for exporting the selected scheme.
- Map:** A map of the United States with a 3-class diverging color scheme applied to the borders.
- Footer:** Copyright information for Cynthia Brewer, Mark Harrower, and The Pennsylvania State University, along with links for source code, feedback, and previous versions. The **axismaps** logo is also present.



Use better colors

Plot

Code



Use better colors

Plot

Code

```
ggplot(brexit, aes(y = opinion, fill = opinion)) +  
  geom_bar() +  
  facet_wrap(~region, nrow = 1, labeller = label_wrap_gen(width = 12)) +  
  guides(fill = "none") +  
  labs(title = "Was Britain right/wrong to vote to leave EU?",  
        subtitle = "YouGov Survey Results, 2-3 September 2019",  
        caption = "Source: bit.ly/2lCJZVg",  
        x = NULL, y = NULL) +  
  scale_fill_manual(values = c(  
    "Wrong" = "#ef8a62",  
    "Right" = "#67a9cf",  
    "Don't know" = "gray"  
  ))
```



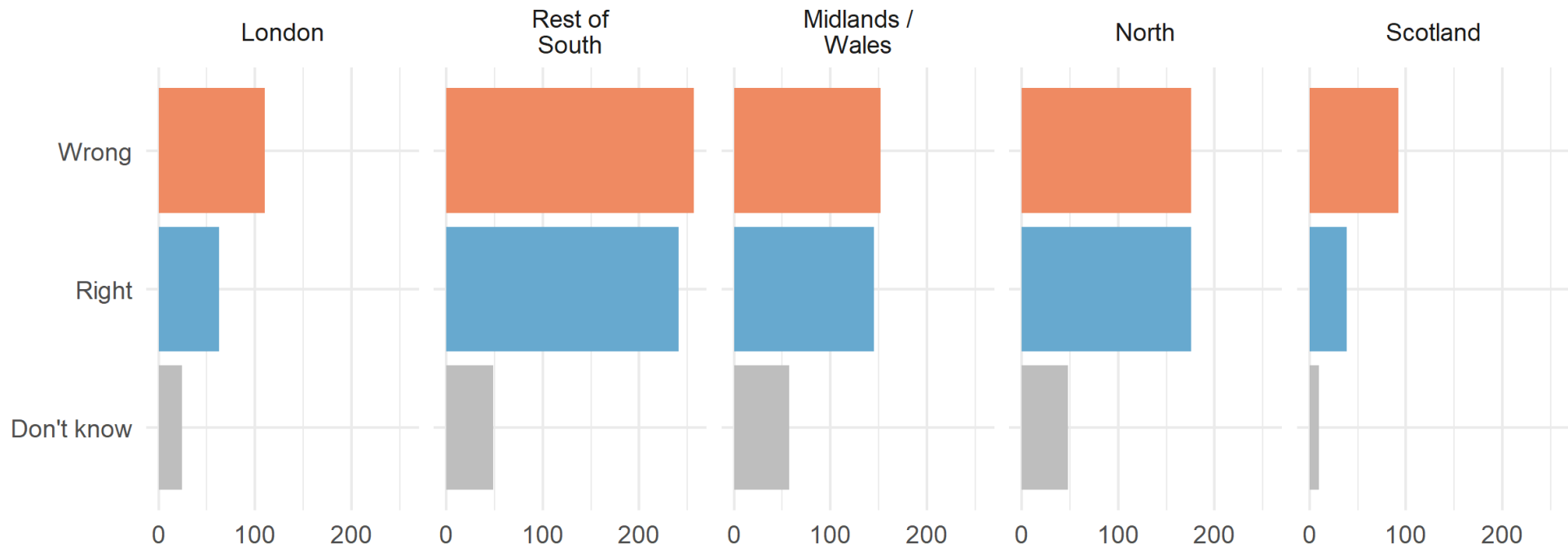
Select theme

Plot

Code

Was Britain right/wrong to vote to leave EU?

YouGov Survey Results, 2-3 September 2019



Source: bit.ly/2ICJZVg



Select theme

Plot

Code

```
ggplot(brexit, aes(y = opinion, fill = opinion)) +  
  geom_bar() +  
  facet_wrap(~region, nrow = 1, labeller = label_wrap_gen(width = 12)) +  
  guides(fill = "none") +  
  labs(title = "Was Britain right/wrong to vote to leave EU?",  
        subtitle = "YouGov Survey Results, 2-3 September 2019",  
        caption = "Source: bit.ly/2lCJZVg",  
        x = NULL, y = NULL) +  
  scale_fill_manual(values = c("Wrong" = "#ef8a62",  
                               "Right" = "#67a9cf",  
                               "Don't know" = "gray")) +  
  theme_minimal()
```



Your turn!

- RStudio Cloud > AE 07 - Brexit + Telling stories with dataviz > brexit.Rmd.
- Change the visualisation in three different ways to tell slightly different stories with it each time.

