Exploratory data analysis

Tips for effective data visualization

Prof. Dr. Jan Kirenz

Alphabetical order is rarely ideal

Plot Code

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

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

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

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"
    )
  )
```

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"
    )
  )
```

Move them to the y-axis

Plot Code

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

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

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

Segmented bar plots can be hard to read

Plot Code

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

Use facets

Plot Code

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

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

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

Use informative labels

Plot Code

```
ggplot(brexit, aes(y = opinion, fill = opinion)) +
  geom_bar() +
  facet_wrap(~region, nrow = 1) +
  guides(fill = FALSE) +
  labs(
    title = "Was Britain right/wrong to vote to leave EU?",
    x = NULL, y = NULL
  )
```

A bit more info

Plot Code

```
ggplot(brexit, aes(y = opinion, fill = opinion)) +
  geom_bar() +
  facet_wrap(~region, nrow = 1) +
  guides(fill = FALSE) +
  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/x0m
    x = NULL, y = NULL
  )
```

Let's do better

Plot Code

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

```
ggplot(brexit, aes(y = opinion, fill = opinion)) +
  geom_bar() +
  facet_wrap(~region,
    nrow = 1,
    labeller = label_wrap_gen(width = 12)
  ) +
  guides(fill = FALSE) +
  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
  )
```

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 = FALSE) +
  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"
  ))
```

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 = FALSE) +
  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

```
ggplot(brexit, aes(y = opinion, fill = opinion)) +
  geom_bar() +
  facet_wrap(~region, nrow = 1, labeller = label_wrap_gen(width = 12)) +
  guides(fill = FALSE) +
  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()
```