Exercises: R Markdown

YSC2210 - DAVis with R

Michael T. Gastner

1 Typeset a section from Wickham et al.'s (2022) ggplot2 book

Typeset section 2.3 ('Key Components') from Wickham et al.'s (2022) online book 'ggplot2: Elegant Graphics for Data Analysis' as an R Markdown document and generate HTML output. This section from the book can be found at https://ggplot2-book.org/getting-started.html#basic-use. Please also typeset the exercise text in section 2.3.1, but you do not need to solve the exercises.

The result does not need to be a pixel-perfect copy. We would need to obtain additional files from the authors if we wanted to do that because they do not use plain-vanilla R Markdown. However, mimic the appearance of R code chunks and inline code in Wickham et al.'s text. Also pay attention to inline text that is bold or in monospaced font. Scale the figure so that it appears similar to the figure in Wickham et al.'s book.

Feel free to copy and paste the text from the book's website.

2 Report about the New York City mayoral election 2017

(a) Create a new R Markdown document with

- the title 'New York Mayoral Election 2017',
- your team's name as the author and
- today's date.

(b) Edit the YAML header to create a table of contents.

(c) Add a section header 'Introduction'.

(d) In the introduction section, embed the second sentence from the Wikipedia page about the election (https://en.wikipedia.org/wiki/2017_New_York_City_mayoral_election)[1] as a block quote:

> 'Incumbent Democrat Bill de Blasio won reelection to a second term with 66.2% of the vote against Republican Nicole Malliotakis.'

Add a little bit more text of your own before and after the quote.

(e) In a footnote, add a hyperlink to the Wikipedia page and the date when you accessed the page.

(f) Add a section header 'Data presented as a table'.

(g) Make an HTML table of the votes by borough as shown in figure 1.[2] The data are in the link to `ny_mayor.csv` on the Canvas assignment page. See https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome_table_in_html.html to give your table some pizzazz using the **kableExtra** package. You may also search the World Wide Web for help.

- Add the caption 'New York Mayoral Election 2017' above the table.

[1] Accessed on 19 January 2022.
[2] Data from https://edition.cnn.com/election/2017/results/new-york-city-mayor, accessed on 19 January 2022.

- Add commas to the numeric values (e.g. '117,712') for easier readability. Do not hard-code the commas. Instead, use R to insert the commas for you.
- Use alternating background colours for the rows for easier readability.
- Because the table does not need to cover the full width of the page, set the argument `full_width` in `kable_styling()` equal to `FALSE`.
- Group the columns for the candidates (de Blasio, Mallotakis, Other) and give the group the multi-column heading 'Candidate'.
- Add a bottom row with the total votes for each candidate. Draw a black horizontal line above this row.
- Add the footnote 'Source: CNN' in slightly smaller font below the table.

New York Mayoral Election 2017

	Candidate		
Borough	**de Blasio**	**Malliotakis**	**Other**
Bronx	117,712	23,715	3,138
Brooklyn	254,755	74,343	10,602
Manhattan	190,312	53,853	10,186
Queens	171,867	94,911	8,974
Staten Island	25,466	70,125	2,747
Total	760,112	316,947	35,647

Source: CNN

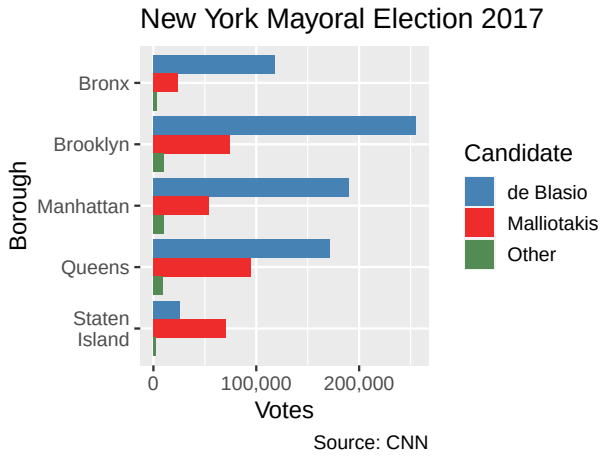
Figure 1: In problem 2(g), you are asked to make an HTML table as shown above. The green frame is only shown for clarity; you do not need to create it in your solution.

(h) Add a section header 'Data presented as a bar plot'.

(i) Make a dodged bar plot with the code below. Centre the plot. Adjust its width and height so that the plot is easily legible and the typical font size in the figure is comparable to the font size in the text.

```
# Put calls to library() at the start of your RMD file
library(tidyverse)
library(scales)
ny_mayor_long <- read_csv("ny_mayor.csv") |>
  mutate(borough = str_replace(borough, " ", "\n")) |>
  pivot_longer(-borough) |>
  mutate(name = fct_rev(name))
ggplot(ny_mayor_long, aes(borough, value, fill = name)) +
  geom_col(position = "dodge") +
  scale_x_discrete(limits = rev) +
  coord_flip() +
  scale_y_continuous(labels = label_comma()) +
  scale_fill_manual(
    labels = c("Other", "Malliotakis", "de Blasio"),
    values = c("palegreen4", "firebrick2", "steelblue")
  ) +
  labs(
    title = "New York Mayoral Election 2017",
    caption = "Source: CNN",
    x = "Borough",
    y = "Votes",
```

```
    fill = "Candidate"
) +
guides(fill = guide_legend(reverse = TRUE)) +
theme(legend.key.size = unit(1, "lines"))
```



(j) Add a section header 'What is better: table or bar plot?'

(k) Write one or two short paragraphs (maximum 10 sentences) about the advantages and disadvantages of presenting the data as a table rather than as a bar plot.

(l) Hide all R code.