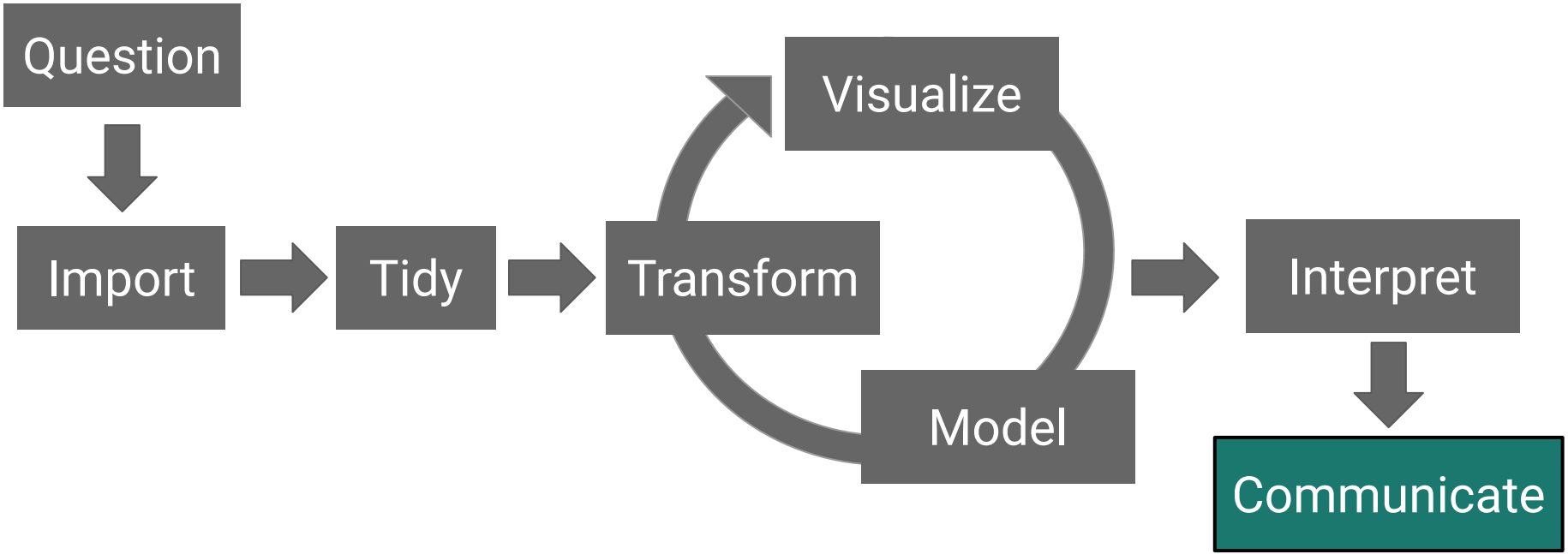
Data Communication

Lecture 14

Motivation



Modified from https://r4ds.had.co.nz/explore-intro.html#explore-intro

Objectives

- Markdown in Jupyter notebook
- Visual communication

Jupyter Notebook

- other platforms: R Markdown, Shiny
- provide a unified authoring framework for data analysis, combining code, results, comments, and interpretation of data
- necessary for reproducible research and support dozens of output formats (PDF, text, slides, Markdown, LaTeX, HTML)

Jupyter Notebook

- for communicating to your boss, who want to focus on the conclusions, not necessarily the code behind the analysis.
- for collaborating with your peers (including future you!), who are interested in both your conclusions, and how you reached them (i.e. the code).
- as an environment in which to do data science, as a modern day lab notebook where you can capture not only what you did, but also what you were thinking.

Markdown in Jupyter Notebook

Type	Or	... to Get
`*Italic*`	`_Italic_`	*Italic*
`**Bold**`	`__Bold__`	**Bold**
`# Heading 1`	`Heading 1` `=========`	Heading 1
`## Heading 2`	`Heading 2` `---------`	Heading 2

`[Link](http://a.com)`	`[Link][1]`
	⋮
	`[1]: http://b.org`
		Link

`![Image](http://url/a.png)`	`![Image][1]`
	⋮
	`[1]: http://url/b.jpg`
		

Markdown in Jupyter Notebook

```
> Blockquote
```

> Blockquote

```
* List
* List
* List
```

```
- List
- List
- List
```

- List
- List
- List

```
1. One
2. Two
3. Three
```

```
1) One
2) Two
3) Three
```

1. One
2. Two
3. Three

```
Horizontal rule:

---
```

```
Horizontal rule:

***
```

Horizontal rule:

Markdown in Jupyter Notebook

```
`Inline code` with backticks
```

`Inline code` with backticks

````
```
# code block
print '3 backticks or'
print 'indent 4 spaces'
```
````

```
    # code block
    print '3 backticks or'
    print 'indent 4 spaces'
```

```
# code block
print '3 backticks or'
print 'indent 4 spaces'
```

Visual Communication

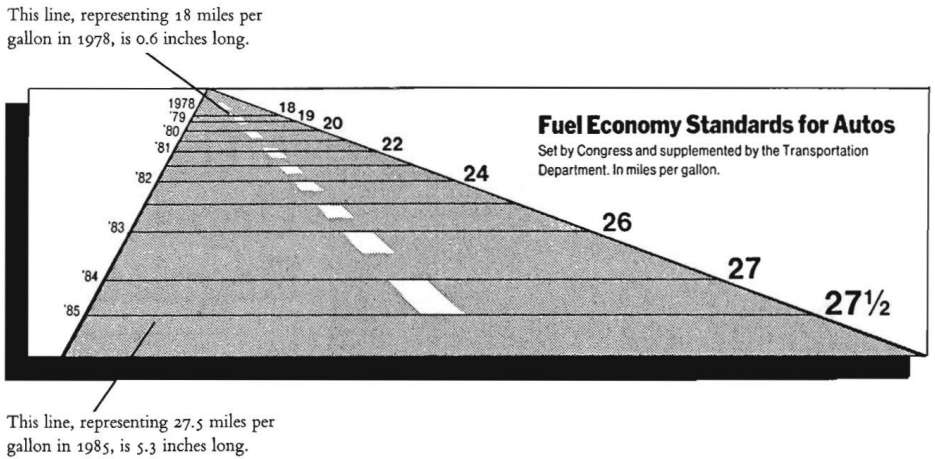
- in EDA, you prepare quick plots for personal use.
- after understanding your data, you need to communicate your understanding to others.
- you will need to invest considerable effort in making your plots as self-explanatory as possible.

Visual Communication

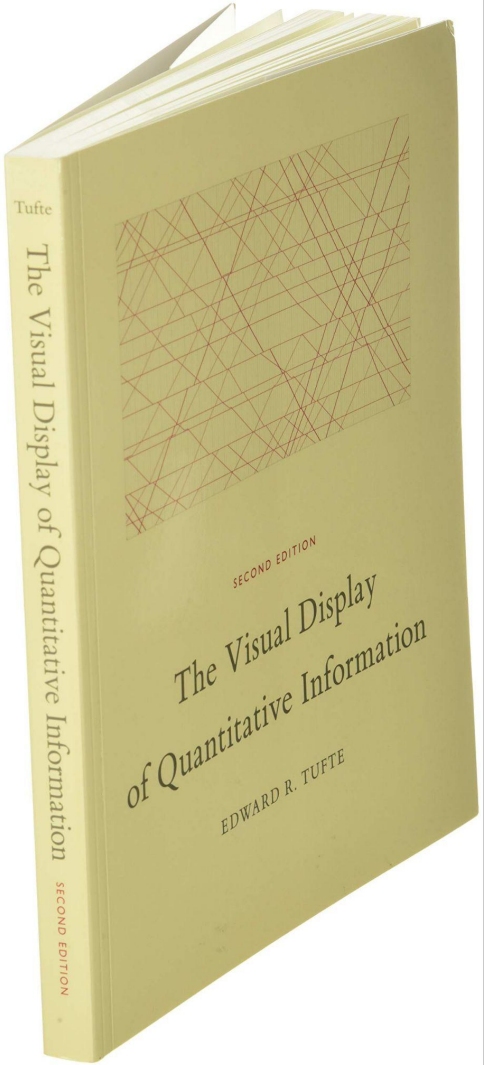
Here is an extreme example. A newspaper reported that the U.S. Congress and the Department of Transportation had set a series of fuel economy standards to be met by automobile manufacturers, beginning with 18 miles per gallon in 1978 and moving in steps up to 27.5 by 1985, an increase of 53 percent:

$$\frac{27.5 - 18.0}{18.0} \times 100 = 53\%$$

These standards and the dates for their attainment were shown:

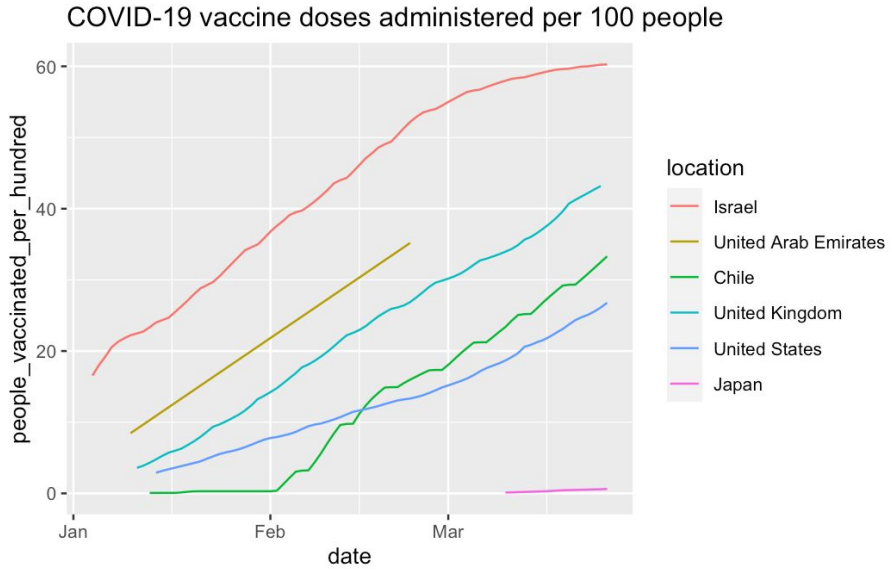


New York Times, August 9, 1978, p. D-2.



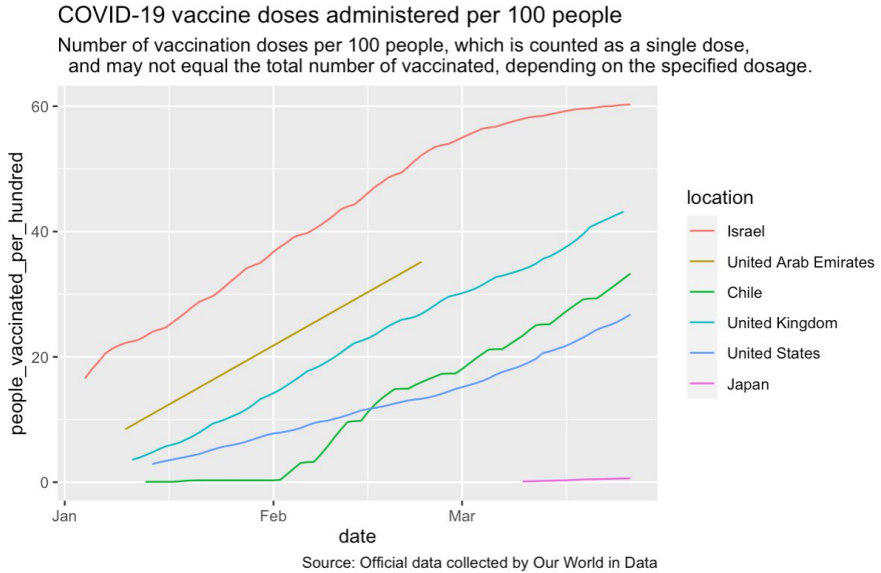
Labels

```
vaccine %>%
        ggplot(aes(date, people_vaccinated_per_hundred, color = location)) +
        geom_line() +
        labs(title = "COVID-19 vaccine doses administered per 100 people")
```



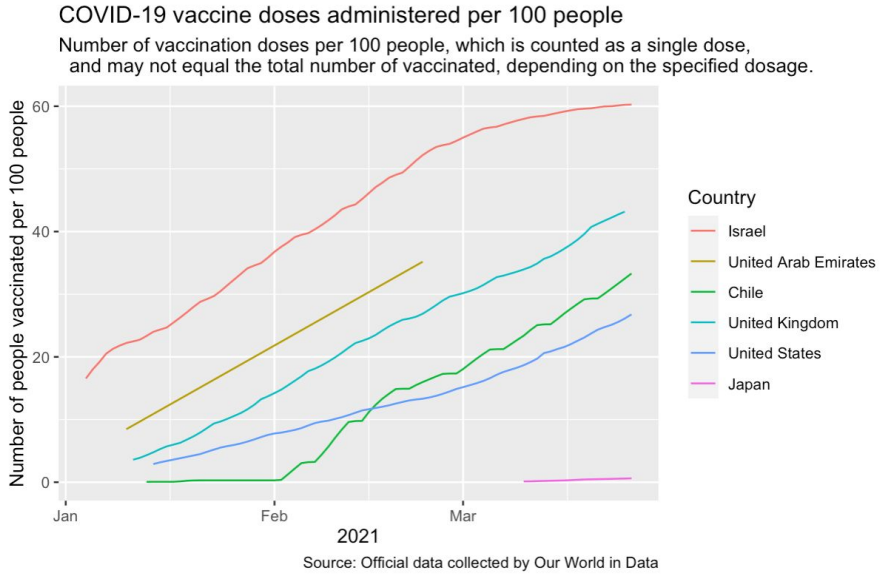
Labels

```
vaccine %>%
        ggplot(aes(date, people_vaccinated_per_hundred, color = location)) +
        geom_line() +
        labs(title = "COVID-19 vaccine doses administered per 100 people",
                subtitle = "Number of vaccination doses per 100 people, which is counted as a single dose,
                                and may not equal the total number of vaccinated, depending on the specified dosage.",
                caption = "Source: Official data collected by Our World in Data")
```



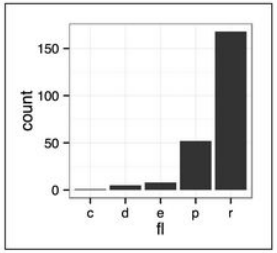
Labels

```
vaccine %>%
        ggplot(aes(date, people_vaccinated_per_hundred, color = location)) +
        geom_line() +
        labs(...
                x = "2021",
                y = "Number of people vaccinated per 100 people",
                color = "Country")
```

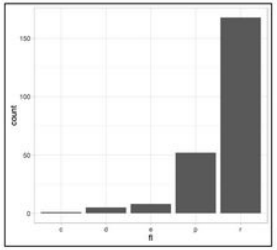


Themes

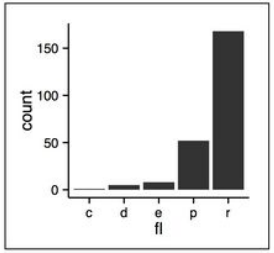
Theme functions change the appearance of your plot.



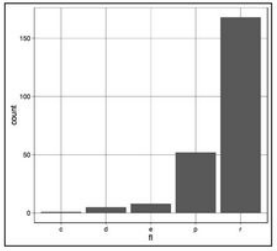
theme_bw()
White background with grid lines



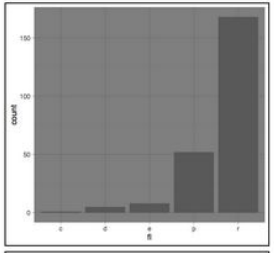
theme_light()
Light axes and grid lines



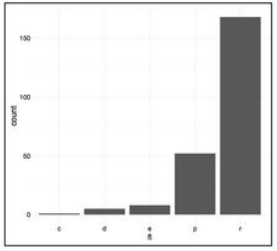
theme_classic()
Classic theme, axes but no grid lines



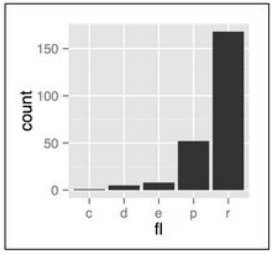
theme_linedraw()
Only black lines



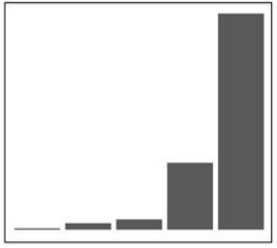
theme_dark()
Dark background for contrast



theme_minimal()
Minimal theme, no background

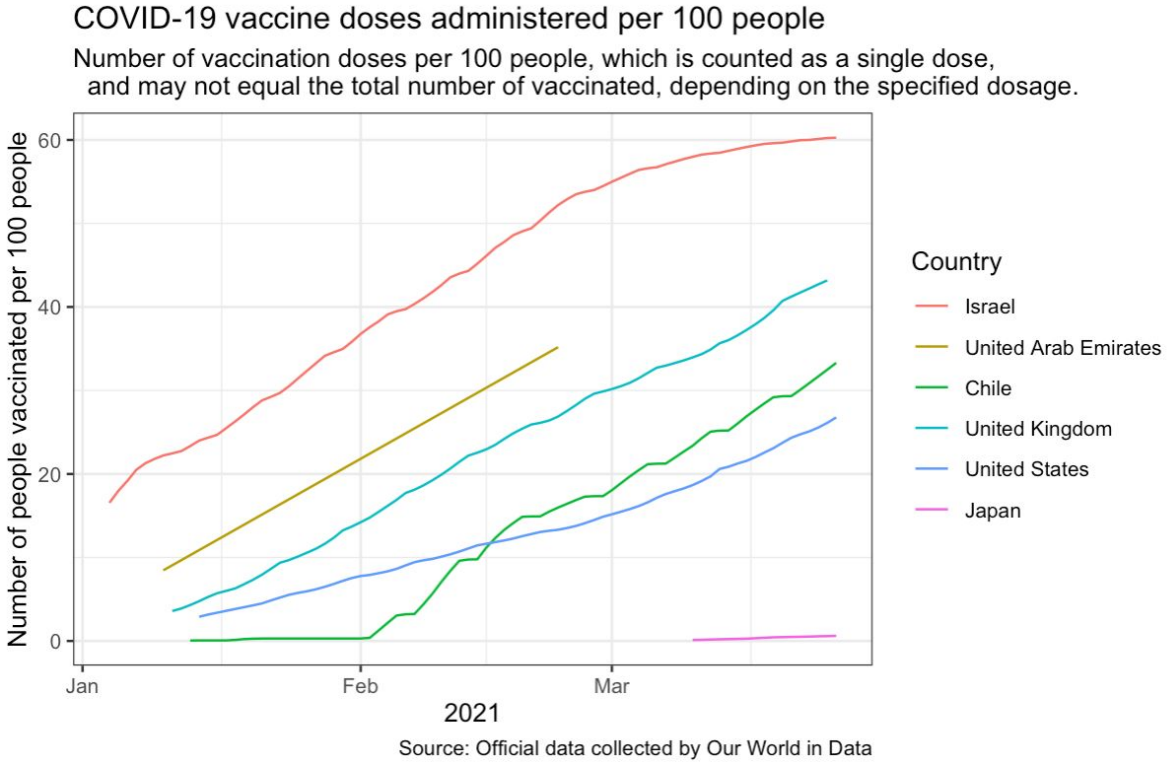


theme_gray()
Grey background (default theme)



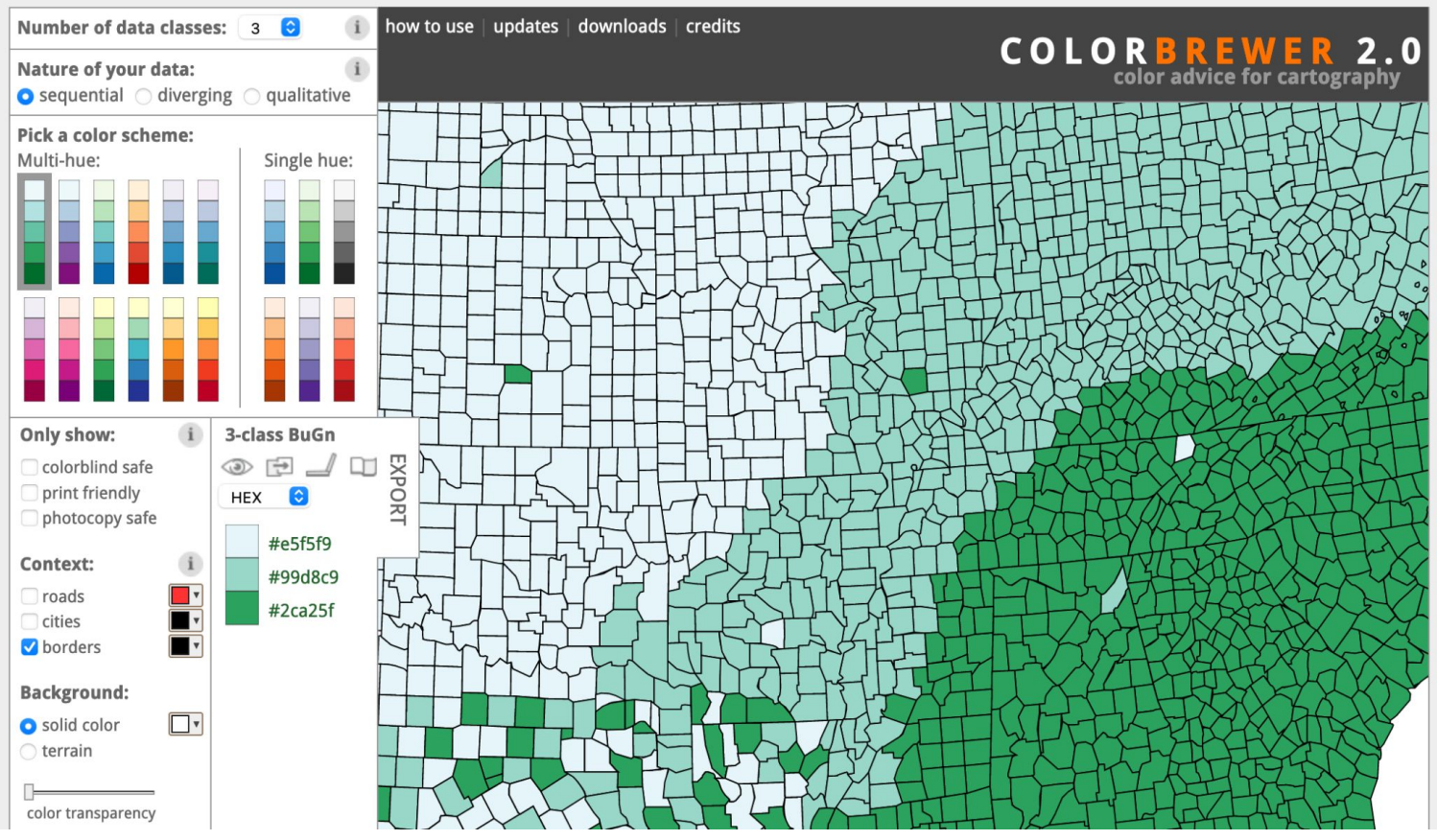
theme_void()
Empty theme, only geoms are visible

```
library(ggthemes)
vaccine %>%
        ggplot(aes(date, people_vaccinated_per_hundred, color = location)) +
        geom_line() +
        labs(...) +
        theme_bw()
```



Color

https://colorbrewer2.org/



RColorBrewer package

```
library(RColorBrewer)
vaccine %>%
        ggplot(aes(date, people_vaccinated_per_hundred, color = location)) +
        geom_line() +
        … +
        scale_color_brewer(palette = "Dark2")
```

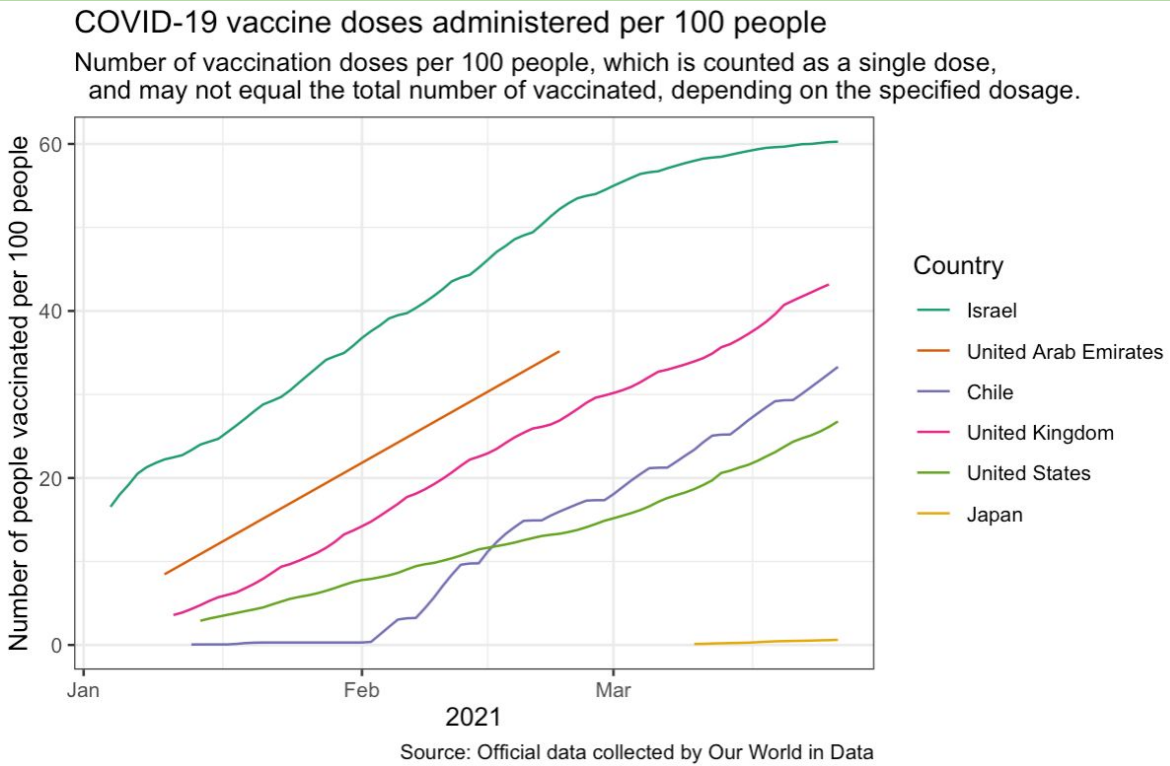
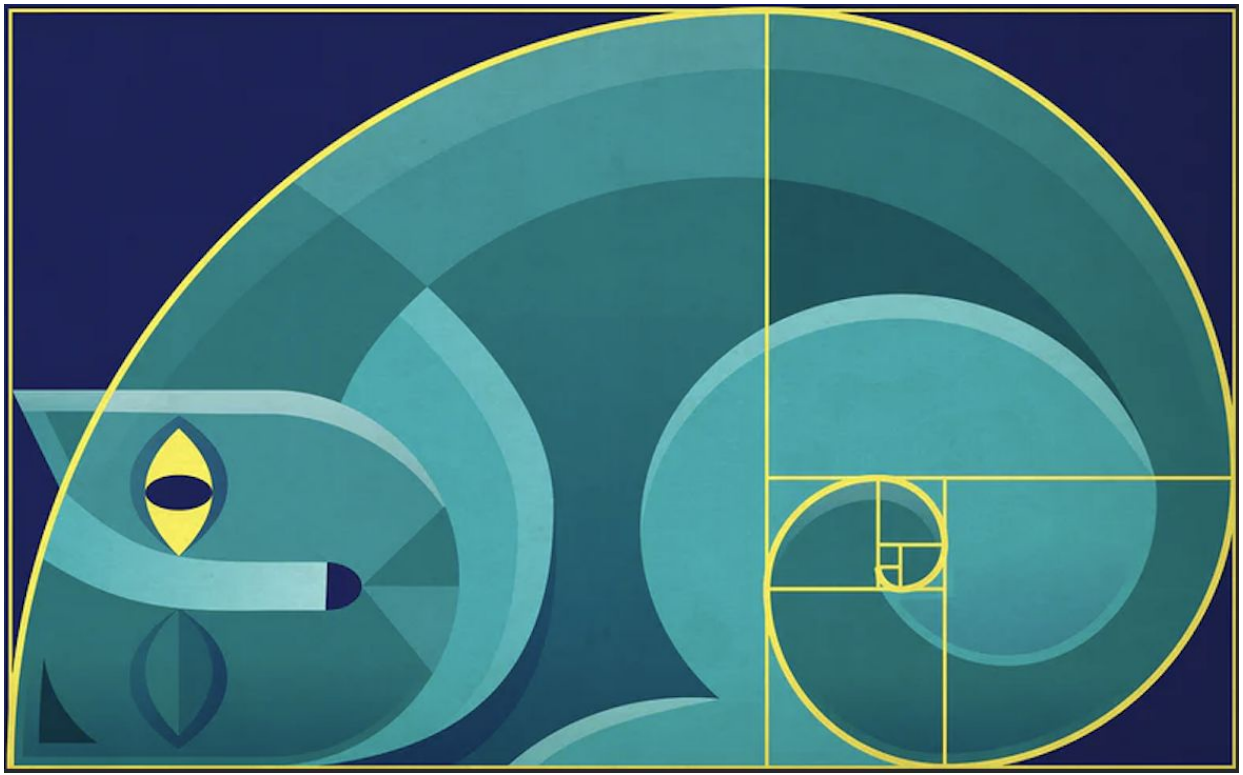


Figure size

Golden ratio:
fig.width = 6
fig.asp = 0.618

Specified:
fig.width = 6
fig.height = 3.7



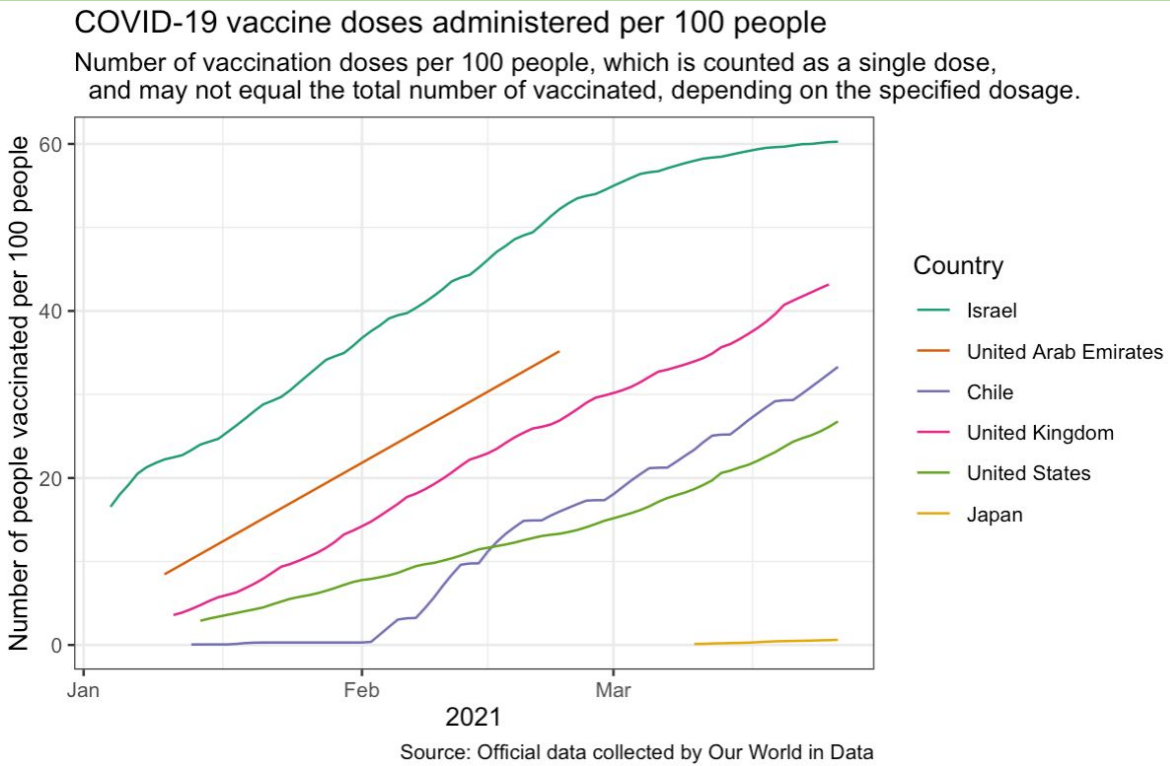
https://en.99designs.jp/blog/tips/the-golden-ratio/

Saving figures

```
ggsave(
  filename,
  plot = last_plot(),
  device = NULL,
  path = NULL,
  scale = 1,
  width = NA,
  height = NA,
  units = c("in", "cm", "mm"),
  dpi = 300,
  limitsize = TRUE,
  ...
)
```

device Device to use. Can either be a device function (e.g. png()), or one of "eps", "ps", "tex" (pictex), "pdf", "jpeg", "tiff", "png", "bmp", "svg"

```
vaccine %>%
        ggplot(aes(date, people_vaccinated_per_hundred, color = location)) +
        geom_line() +
        … +
        scale_color_brewer(palette = "Dark2")
ggsave("vaccine_plot.pdf")
```

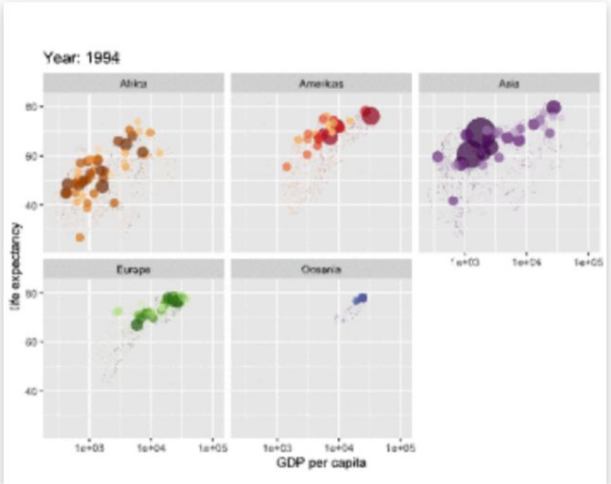


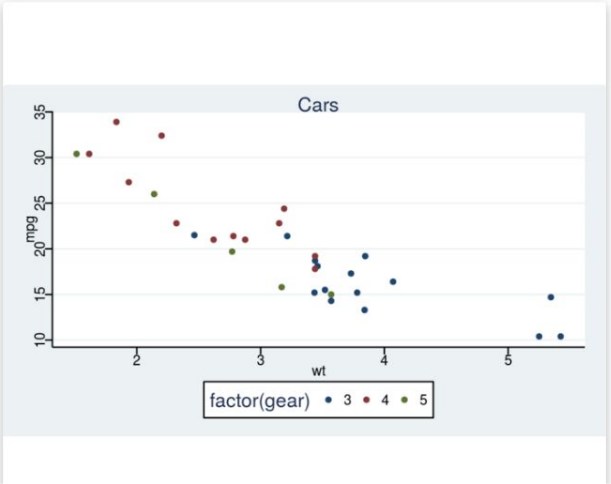
https://exts.ggplot2.tidyverse.org/gallery/

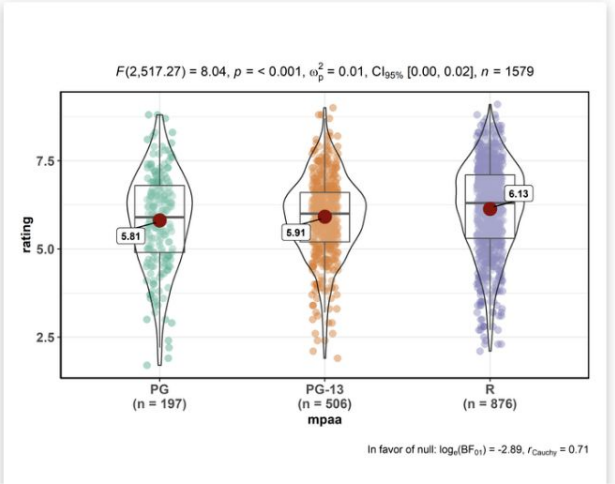
82 registered extensions available to explore

Sort: Github stars ▾ Text Filter: search name, auth Author Filter ▾ Tag Filter ▾ CRAN Only

Showing 66 of 82







Take-away message

- It does not matter how great your analysis is, unless you can communicate it to others.
- Plots should stand alone, properly labeled and annotated.