Theme 2 – Data Visualisation: ggplot2 Essentials

Eemil Mikkka

Last updated: 22nd of March, 2023

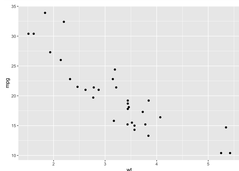
README

For this week's exercise, you will be given only the R code that was used to create visualisations in Theme 2 lecture slides' examples. Other than that, you have to figure out how to perform the tasks independently.

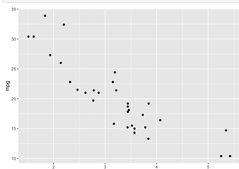
💡 TIPS for Theme 2 assignments

- You can google for help for how to create plots with `ggplot()` function. Some examples of keywords in Google search:
  - "r ggplot how to draw scatterplot" (in case you need to know how to draw scatterplot, ofcourse)
  - "r ggplot geom_point()" (again, if you want to draw scatterplot. To google examples for bar charts, for instance, you should write "r ggplot geom_col()" or "r ggplot geom_bar")
  - **ggplot2** package has also quite comprehensive website with examples of how to draw different kind of plots using the `ggplot()` function. This specific link directs to (again) to a page instructing how to create a scatterplot with `geom_point()`, but there are examples dealing with other ***geometries*** (e.g. `geom_col()`, `geom_boxplot()`, and so on) too.
  - Below, you may find an example on how to draw a very basic scatterplot with `ggplot()` using the build-in `mtcars` dataset. The "build-in" simply means that you don't have to import the data to R with the `here()` function or similar methods, but it already exists in R environment for reproducible examples and tutorials on how to use functions in R. Here's the example from ggplot2 package's website:

```
# Activate tidyverse package (ggplot2 is part of this meta package)
library(tidyverse)

# Raw example taken from the ggplot2 website
p <- ggplot(mtcars, aes(wt, mpg))
p + geom_point()
```




```
# This can be drawn also with the lecture slides logic like this:
ggplot(mtcars) +
  aes(x = wt, y = mpg) +
  geom_point()
```



- Use Moodle's Discussion Forum, so that I or other students can help with potential issues.

Support Material: Visualisation Examples

Packages

```
library(tidyverse)
library(here)
```

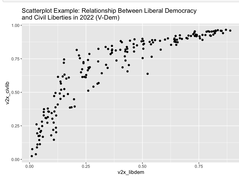
Import the data

In the Theme 2 lecture slides' plotting examples, I used the same `vdem_subset.csv` data you're familiar with. However, I created one extra category for visualising mean levels of democracy within ex-USSR political unions in 2022 (plot created with `ggplot()` and `geom_boxplot()` combo). Hence, the `vdem_subset.csv` is updated and named as `vdem_subset_upd.csv` for this example. We will look at how to create this kind of new variables with several categories in upcoming lecture dealing with more complex data wrangling.

```
# Import the updated V-Dem data
vdem <- here("data/vdem_subset_upd.csv") %>%
  read.csv()
```

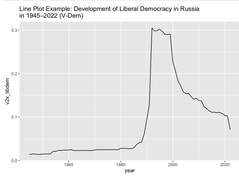
Scatterplot, `geom_point()`

```
vdem_2022 <- vdem %>%
  filter(year == "2022")

# Graph scatterplot
ggplot(vdem_2022) +
  aes(x = v2x_libdem, y = v2x_civlib) +
  geom_point() +
  labs(title = "Scatterplot Example: Relationship Between Liberal Democracy\nand Civil Liberties in 2022 (V-Dem)")
```



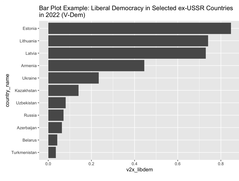
Line plot, `geom_line()`

```
vdem_rus_after_1945 <- vdem %>%
  filter(country_name == "Russia",
         year >= 1945)

# Graph line plot
ggplot(vdem_rus_after_1945) +
  aes(x = year, y = v2x_libdem) +
  geom_line() +
  labs(title = "Line Plot Example: Development of Liberal Democracy in Russia\nin 1945-2022 (V-Dem)")
```



Bar plot, `geom_col()`

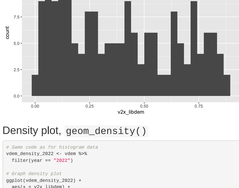
```
vdem_libdem_selected_ex_ussr_in_2022 <- vdem %>%
  filter(country_name %in% c("Russia",
                             "Estonia",
                             "Latvia",
                             "Lithuania",
                             "Uzbekistan",
                             "Kazakhstan",
                             "Turkmenistan",
                             "Russia",
                             "Azerbaijan",
                             "Belarus",
                             "Ukraine"),
         year == "2022")

# Graph bar chart
ggplot(vdem_libdem_selected_ex_ussr_in_2022) +
  aes(x = v2x_libdem, y = reorder(country_name, +v2x_libdem)) +
  geom_col() +
  labs(title = "Bar Plot Example: Liberal Democracy in Selected ex-USSR Countries\nin 2022 (V-Dem)",
       y = "country_name")
```



Histogram, `geom_histogram()`

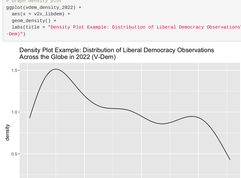
```
vdem_histogram_2022 <- vdem %>%
  filter(year == "2022")

ggplot(vdem_histogram_2022) +
  aes(x = v2x_libdem) +
  geom_histogram() +
  labs(title = "Histogram Example: Distribution of Liberal Democracy Observations\nAcross the Globe in 2022 (V-Dem)")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Density plot, `geom_density()`

```
# Same code as for histogram data
vdem_density_2022 <- vdem %>%
  filter(year == "2022")

# Graph density plot
ggplot(vdem_density_2022) +
  aes(x = v2x_libdem) +
  geom_density() +
  labs(title = "Density Plot Example: Distribution of Liberal Democracy Observations\nAcross the Globe in 2022 (V-Dem)")
```



Box plots, `geom_boxplot()`

```
vdem_boxplot <- vdem %>%
    filter(year == "2022",
           !is.na(country_group)) # countries other than ex-USSR have NA values (i.e. missing data) with this var
lable

# Graph box plot
ggplot(vdem_boxplot) +
  aes(x = reorder(country_group, -v2x_libdem), y = v2x_libdem) +
  geom_boxplot() +
  labs(title = "Box Plot Example: Comparing Mean Levels of Democracy within ex-USSR\nPolitical Unions in 2022 (V-Dem, political unions categories created 'manually')",
       x = "country_group")
```