```
# Load package(s) ----
## If you get an error trying to run the library() functions below, it means
## that the package is not installed. Execute install.packages("packagename")
## in the console, replacing "packagename" with
## "tidyverse", for instance, if you want to install the "tidyverse" package.
## Do that for all packages.

# install.packages("tidyverse")
# install.packages("scales")
# install.packages("ggrepel")

library(tidyverse)
library(scales)
library(ggrepel)

# Get urban population data between 2013-2022 from Eurostat ----
## Source:
https://ec.europa.eu/eurostat/databrowser/view/URB_LPOP1__custom_5120358/default/table?
lang=en
pop_13_22_f <- read_csv("data/raw/urb_lpop1__custom_5120358_page_linear.csv")
pop_13_22_c <- read_csv("data/raw/urb_cpop1_page_linear.csv")
ua <- read_csv("data/raw/URAU_LB_2021_3035_CITIES_within_frame.csv")

# Examine the data ----
# View(pop_13_22_f)
glimpse(pop_13_22_f)

# Clean the data ----
pop_13_22_long <- pop_13_22_f |>
  # Remove unnecessary columns
  select(-c(DATAFLOW, `LAST UPDATE`, freq, indic_ur, OBS_FLAG)) |>
  # Remove countries
  filter(nchar(cities) > 2)

# Show the population trend of all capital cities in the region
pop_13_22_long |>
  # filter(cities %in% c("BE001F", "DE001F", "NL001F", "UK001F")) |>
  filter(substr(cities, 3, 6) == "001F") |>
  left_join(ua |> filter(CITY_CPTL == "Y"), by = c("cities" = "FUA_CODE")) |>
  group_by(cities) |>
  mutate(label = if_else(TIME_PERIOD == max(TIME_PERIOD), as.character(URAU_NAME),
NA_character_)) |>
  ungroup() |>
  ggplot(aes(x = TIME_PERIOD, y = OBS_VALUE, color = cities, group = cities)) +
  geom_line() +
  geom_label_repel(aes(label = label),
                   nudge_x = 1,
                   na.rm = TRUE) +
  scale_x_continuous(breaks = seq(from = 2013,
                                  to = 2021,
                                  by = 1),
                     limits = c(2013, 2023)) +
  xlab("Year") +
  ylab("Number of inhabitants") +
  labs(title = "Population dynamics in the region",
       subtitle = "Eurostat population data on FUA level between 2013-2021") +
  scale_color_discrete(name = "Muncipality") +
  theme(legend.position = "none")

# Generate table in wide format
pop_13_22_c_wide <- pop_13_22_c |>
  pivot_wider(names_from = TIME_PERIOD, values_from = OBS_VALUE) |>
  select(-c(DATAFLOW, `LAST UPDATE`, freq, indic_ur, OBS_FLAG, `2022`)) |>
  filter(nchar(cities) > 2) |>
  group_by(cities) |>
  summarise_all(.funs = c(mean="sum")) |>
  ungroup() |>
  mutate(across(where(is.double), ~na_if(., 0)))
```

```
# Write wide table with time-series data
write_csv(pop_13_22_c_wide, "data/processed/pop_13_22_c_wide.csv", na = "")
```