Week 11 Tutorial Worksheet

AY23/24 Semester 1

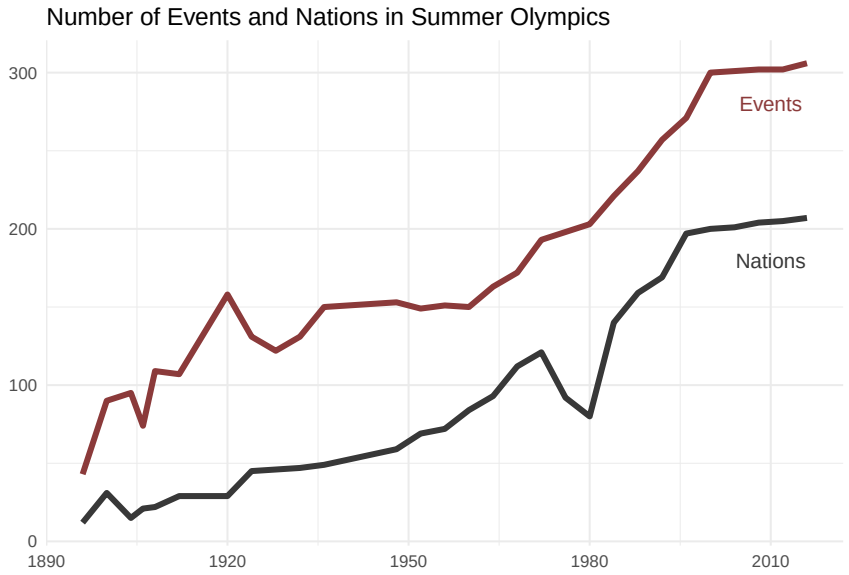
No submission required

Question 1

In this tutorial, we continue to work with the Olympics data from `TidyTuesday`. We will use the following two data sets:

- `olympics.csv` contains information on Olympic games from 1896 to 2016.
- `regions.csv` maps the 3-digit NOC codes to the country/region names.

1. Create a graph on the number of nations and events in each Summer Olympics from 1896 to 2016 using `ggplot2` plotting. Your graph can be similar to the following.



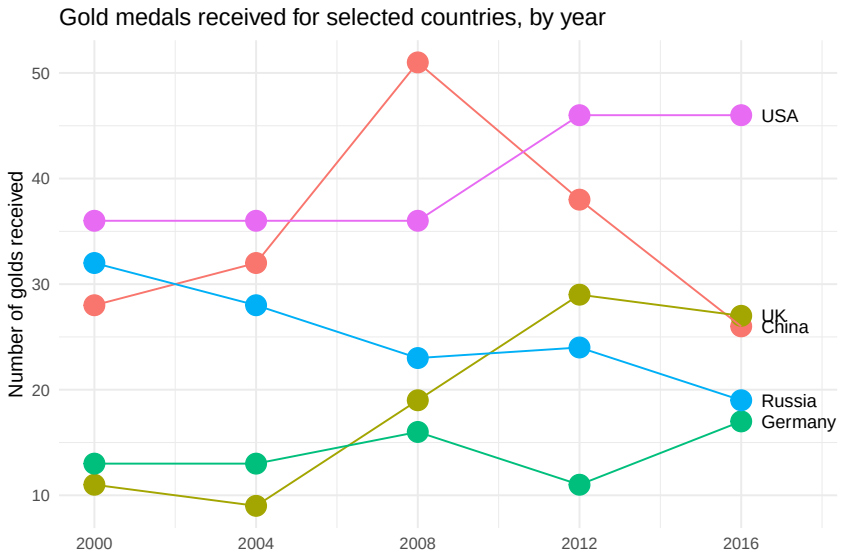
2. Which five countries won the most gold medals in 2016? Use your `dplyr` skills answer this question.

   After that, extract the names (in `noc`) of the top five countries as a vector named `summer_top_5`. The vector should take the following value:

```
summer_top_5
```

```
## [1] "USA" "GBR" "CHN" "RUS" "GER"
```

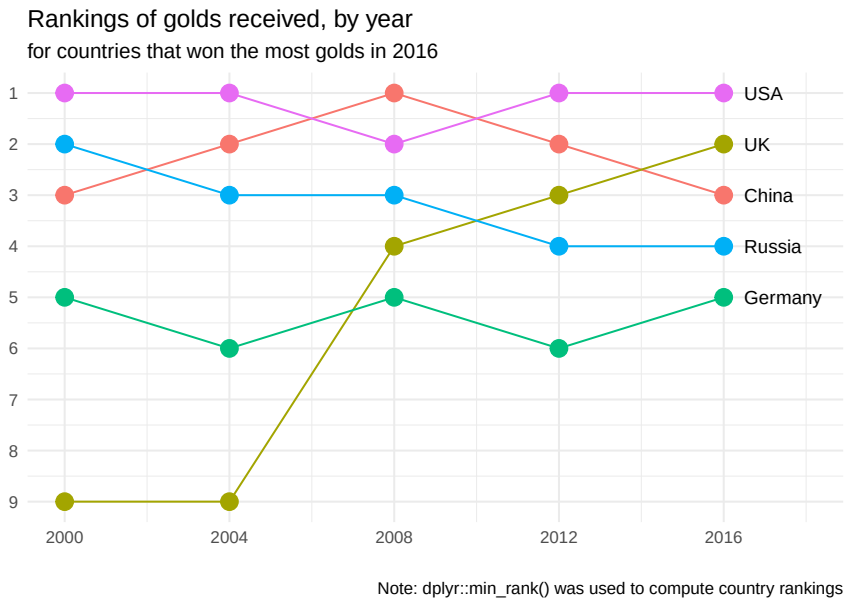
3. Compute the number of gold medals received for the `summer_top_5` countries since the year of 2000. Use it to re-create the plot below, as closely as you can.



4. Instead of the total number of golds, let us now examine the countries' **rankings** in golds received across years. Prepare the data to obtain the rankings in golds received for the `summer_top_5` countries from 2000 to 2016. Save your resulting data frame as an object named `ranks_top_5`.

   Hint: Ranking functions in `R` vary in how they handle tied values. `dplyr` provides two handy functions: `min_rank()` and `dense_rank()`. For sports data, the former is typically the conventional choice. More information is available in their documentation.

5. Use `ranks_top_5` to re-create, as much as you can, the plot below.



Requirements

- You code should generate a vector named `summer_top_5` and a data frame named `ranks_top_5`.
- The knitted HTML should contain three plots, one each for Question 1.1, Question 1.3, and Question 1.5.