Data Science and R – Lab 16

Exercises with dplyr II. Learning outcomes:

`min_rank(), dense_rank(), row_number(), filter(), select(), arrange(), group_by(), summarise(), lag(), inner_join(), left_join(), full_join(), piping (%>%)`

0) Ensure the 'dplyr', 'magrittr' and 'gapminder' packages are installed and loaded. We'll be working with the gapminder dataset.

1) Create a vector c(1, 2, 3, 3, 2, 1) and run min_rank(), dense_rank() and row_number() on it. What are the output for each?

Answer:____

Answer:____

Answer:____

2) Load a copy of gapminder by assigning it to my_gap

3) Use piping to select all rows for the country "Korea, Rep." How many rows were returned?

Code:____

Answer:____

4) How many observations do we have per continent?

Code:____

Answer:____

5) Find the countries with the lowest and highest life expectancies. Use min_rank()

Code:____

Answer:____

6) Find the countries with the lowest and highest life expectancies over time (grouped by year). Use min_rank(). Which countries had the lowest and highest life expectancies in 1977?

Code:____

Answer:____

7) How many countries are there in each continent? You may want to summarise and and use n_distinct()

Code:____

Answer:____

8) Working only with countries in Africa, get the countries with the lowest life expectancies over time (grouped by year). Which country had the lowest life expectancy in 1987?

Code:____

Answer:____

9) Again working only with countries in `Africa`, get the countries with the highest and lowest life expectancies over time. Use `min_rank()` again

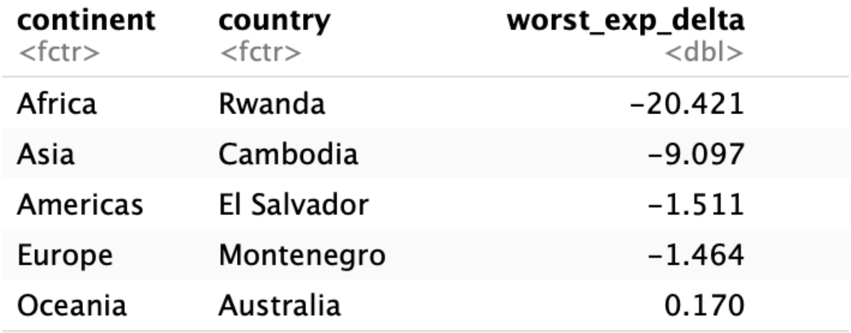
Code:____

10) Working with Africa, create a new column with the ranking of the life expectancy (`lifeRank`), 1 for the lowest life expectancy within Africa. Then, filter only "Gambia", "Sierra Leone", "Reunion", “Rwanda" after the year 1960. You may need to create a separate data for the african continent only for this to get accurate ranking. How many times is Sierra Leone ranked 1?

Code:____

11) Which country experienced the worst 5-year drop in life expectancy in each continent? For each country per continent, you may want to create a new column containing the difference between the current and previous life expectancy given by `lag()` . Then, summarise on this value to get the smallest difference for each country and only take the country with the smallest difference per continent.

Code:____



continent <fctr>	country <fctr>	worst_exp_delta <dbl>
Africa	Rwanda	-20.421
Asia	Cambodia	-9.097
Americas	El Salvador	-1.511
Europe	Montenegro	-1.464
Oceania	Australia	0.170

Joins

12) Inspect the `country_codes` dataset. We want to join this data with the `gapminder` data. Only taking the country and continent columns, perform a `left_join` between `gapminder` and `country_codes`. Return only one row per country, i.e. there should not be duplicate rows. (Hint: Try using `distinct()` or `slice()`)

Code:____

13) Try other join functions such as `inner_join()` and `full_join()` to see what the differences are. You may also want to try to join with `country_colors`

Code:____