

INFO-332 Assignment 1

In this assignment, you will use your newly acquired R skills to manipulate datasets. We will work with the COVID-19 world dataset. The assignment contains 10 questions. For each question, you are asked to generate a certain view from the dataset (e.g. a piece of information, a statistic, a plot, etc.).

An example is provided below to highlight how an acceptable solution is structured.

Lastly, a helper script is provided for you to get started. Rename the script with your last and first name as part of the file name and submit the same file to Moodle once you complete the solution.

Load dataset and Libraries

First, let's load the COVID dataset from the given csv file. Notice, we shall use *read_csv* rather than *read.csv*.

```
library(readr)
df_covid <- read_csv("owid-covid-data.csv")
names(df_covid)

## [1] "iso_code"
## [2] "continent"
## [3] "location"
## [4] "date"
## [5] "total_cases"
## [6] "new_cases"
## [7] "new_cases_smoothed"
## [8] "total_deaths"
## [9] "new_deaths"
## [10] "new_deaths_smoothed"
## [11] "total_cases_per_million"
## [12] "new_cases_per_million"
## [13] "new_cases_smoothed_per_million"
## [14] "total_deaths_per_million"
## [15] "new_deaths_per_million"
## [16] "new_deaths_smoothed_per_million"
## [17] "reproduction_rate"
## [18] "icu_patients"
## [19] "icu_patients_per_million"
## [20] "hosp_patients"
## [21] "hosp_patients_per_million"
## [22] "weekly_icu_admissions"
## [23] "weekly_icu_admissions_per_million"
## [24] "weekly_hosp_admissions"
## [25] "weekly_hosp_admissions_per_million"
## [26] "new_tests"
## [27] "total_tests"
## [28] "total_tests_per_thousand"
## [29] "new_tests_per_thousand"
## [30] "new_tests_smoothed"
```

```
## [31] "new_tests_smoothed_per_thousand"
## [32] "positive_rate"
## [33] "tests_per_case"
## [34] "tests_units"
## [35] "total_vaccinations"
## [36] "people_vaccinated"
## [37] "people_fully_vaccinated"
## [38] "new_vaccinations"
## [39] "new_vaccinations_smoothed"
## [40] "total_vaccinations_per_hundred"
## [41] "people_vaccinated_per_hundred"
## [42] "people_fully_vaccinated_per_hundred"
## [43] "new_vaccinations_smoothed_per_million"
## [44] "stringency_index"
## [45] "population"
## [46] "population_density"
## [47] "median_age"
## [48] "aged_65_older"
## [49] "aged_70_older"
## [50] "gdp_per_capita"
## [51] "extreme_poverty"
## [52] "cardiovasc_death_rate"
## [53] "diabetes_prevalence"
## [54] "female_smokers"
## [55] "male_smokers"
## [56] "handwashing_facilities"
## [57] "hospital_beds_per_thousand"
## [58] "life_expectancy"
## [59] "human_development_index"
```

Example with solution shown

Say we're interested in the total number of new covid cases in Italy in the month of March 2020. Here are 2 approaches (among many other possible) to getting the correct answer.

A compact (but somewhat hard to follow) approach:

```
sum(df_covid$new_cases[df_covid$location=="Italy" &
                        df_covid$date>="2020-03-01" &
                        df_covid$date<"2020-04-01"])
```

```
## [1] 104664
```

A less compact (but easier to follow) approach:

```
df_italy <- df_covid[df_covid$location=="Italy",]
df_italy_march <- df_italy[df_italy$date>="2020-03-01" &
                          df_italy$date<"2020-04-01",]
sum(df_italy_march$new_cases)
```

```
## [1] 104664
```

Both solutions are acceptable!

Note: I reveal the final answers for questions 1&2 on purpose to give you help validate your solutions.

Question 1

Find the total number of COVID deaths in the USA in December 2020. Display the answer to console.

```
## [1] 80998
```

Question 2

Generate a dataset that shows how total covid deaths compare on dec 31 2020 for Mexico, Canada, and USA. Display the answer to console.

```
## # A tibble: 3 x 2
##   location      total_deaths
##   <chr>          <dbl>
## 1 Canada          15762
## 2 Mexico         125807
## 3 United States   351861
```

Question 3

Generate a dataset that shows the top 5 **countries** by total covid cases as of dec 31 2020. Display the answer to console.

Question 4

Generate a dataset that shows the top 5 **countries** by total covid deaths relative to the population as of dec 31 2020. Display the answer to console.

Question 5

Find the date and number of new cases on the day the USA had its highest number of new cases in a single day. Display the answer to console.

Question 6

Find the total number of fully vaccinated people in the USA as of the most recent date in the dataset. Display the answer to console.

Question 7

Calculate the average number of vaccinations per day in the USA in the month of March 2021. Display the answer to console.

Question 8

Plot the daily covid cases in the World as function of time. Hint: use the `plot` function which takes to arguments x and y. Optionaly use the *type* argument to make it a line polt.

Question 9

Find the country with population greater than 50 million and lowest number of total covid cases as of March 1 2021. Display the country and total cases to date.

Question 10

Generate a dataset that shows the total covid cases as of dec 31 in all the countries that their name end with “land”. Display the answer to console. Hint: modifying the search phrase to “land\$” ensures that “land” falls at the end of the word.