

RMarkdown Group Activity

Write your Group Name

2001

Goal Group Activity

To apply the principles of reproducible data science methods to a given dataset.

To perform exploratory analysis including data importing, cleaning, wrangling and visualization using `tidy` methods.

To document the analysis of the given dataset and prepare a report (in HTML and Word) using RMarkdown.

About the Dataset

The dataset contains information on the deaths due to COVID-19 in 14 districts of Kerala state. This information is available for download from the Government of Kerala COVID-19 Dashboard (<https://dashboard.kerala.gov.in/covid/>).

Variables of the dataset

Variable Name	Description
sl_no	Serial Number
name	Name of the Deceased
age	Age of the Deceased
sex	Gender of the Deceased
district	District of Residence of the Deceased
date_reported	Date of Reporting
date_of_death	Date of Death

Assignments

1. Load the required packages

```
# library()
```

2. Import the Data

```
# Hint: before you import the data, you check the extension of the file.  
# This may help you to chose between the read_csv(), read_excel, read_rds() functions
```

3, Check the class and structure of the dataset

Hint: Hint is in the question

4. Describing the dataset

Write an R code to look at the number of rows and columns.

Bonus points if you are using two different ways

Give quick summary of the dataset

Bonus points if you are using two different ways

What are the different types of variables in the dataset

Provide your answer below:

How many districts are there?

Bonus points if you are using two different ways

How many deaths occurred in each district?

Hint: you can use the count() function

According to this dataset, when did the maximum number of deaths occur?

Which are the top five districts in COVID-19 deaths?

Hint: arrange?

Is there a delay between death and reporting of death?

If yes, how many days is the delay?

Hint: use mutate() to subtract the relevant variables from each other, you can try the mean() function

Is this delay same across all the districts?

Hint: group_by()

Create a new categorical variable representing age as a dichotomous variable

Hint: ifelse(), case_when()

Create a new categorical variable representing the Wave of COVID

Hint: Use a cut off of 2021-04-01

Hint: ifelse(), case_when()

ASSIGNMENTS

Group A & B

What is the mean age of the deceased?

Answer:

What is the difference in the mean age between both the genders?

Answer:

Formulate a hypothesis to check the association between age and gender?

Answer:

Prepare an appropriate plot to visualise this relationships.

Answer:

Which statistical tests should be applied here?

Answer:

Please make a summary table using gtsummary for the same?

Answer:

Interpret the findings

Answer:

Group C & D

What are distribution of deaths across time?

Answer:

Visualize mortality of COVID-19 in Kerala in different districts

Answer:

Which are the high burden district? Does these districts remain the same in both the COVID waves? Comment.

Answer:

Is there an association between COVID-19 deaths and the waves of COVID?

Answer:

Prepare an appropriate plot to visualise this relationships.

Answer:

Which statistical tests should be applied here?

Answer:

Please make a summary table using gtsummary for the same?

Answer:

Interpret the findings

Answer:

MEGA BONUS points if you use inline code to achieve this task!
