Faculty of Computers and Artificial Intelligence
Bioinformatics Department
2022/2023
BioStatistics Course

Assignment One

Delivery Instructions:

1. Cheaters will be graded by *–ve points, Don't copy any code from anywhere.*
2. Due Date: **December 05, 2022**
3. No late submission will be accepted.
4. Each student should solve this assignment **individually**.
5. upload your R code as well as screenshots of the outcomes in zip file **YourGroup-Student1ID-Assignment1.zip**

Load the medical data from the given file data.csv, which is a dataset of a patient demographic containing standard information regarding individuals from a variety of ancestral lines.

Attributes

- id
- gender: male or female
- dob: Date of Birth
- zipcode
- employment_status
- education
- marital_status
- children: Number of children
- ancestry: Ancestry refers to a person's ethnic origin or descent
- avg_commute: average commuting time
- daily_internet_use
- disease

Faculty of Computers and Artificial Intelligence
Bioinformatics Department
2022/2023
BioStatistics Course

Then Perform the following tasks using R:

1. Show the first 10 rows and the last 10 rows.
2. Using Date of Birth attribute, extracts the gender, average commuting time, and ancestry data for the oldest three.
3. Identifies the gender, daily internet use, average commute time, ancestry, and diseases among those with more than two children.
4. Using a table , indicate the number of rows that have any missing values and the number that do not.
5. Provide a summary of the data for each column, showing "Min, 1st Qu, Median Mean, 3rd Qu and Max" for each numerical column and the Number of each Category for categorical data.
6. Identify the columns that are having any missing values, and then remove any rows where all of the columns have missing values.
7. Show the average daily usage of the internet for each level of education.
8. Show the distribution of the children count using a histogram.
9. Utilizing line graphs, compare how men and women's avg commute distributions differ.
10. Make a histogram to show the gender distribution.
11. Use a histogram to show gender distribution for each disease.
12. Use a chart to demonstrate whether there is a relationship between age and the type of disease.
13. Make a chart to show the total number of children per disease.
14. Make a chart to show the ancestry distribution.