**Assignment 3 - Data Handling**

**Dataset: bikedata**

**Import the Dataset into R**

**Research Question: How many of the cyclists were students, how often did they ride, and what was the average distance they rode?**

|  |  |  |  |
| --- | --- | --- | --- |
| **#** |  | **Question** | **Answer** |
| 1. |  | **Examining the dataset** |  |
|  | a) | What is the age of the 7th rider in the dataset? | BikeData[7,2] |
|  | b) | How many of the first 10 riders in the dataset ride daily? | nrow(subset(BikeData[1:10,],cyc\_freq =='Daily')) |
|  | c) | What is the speed of the first female who cycles less than one time per month (in miles/hour)? | subset(BikeData, BikeData$gender == 'F' & BikeData$cyc\_freq == 'Less than once a month')[1,'speed'] |
| 2. | a) | What type of variable is *student*? | Quantitative - interval |
|  | b) | What type of variable is *cyc\_freq*? | Qualitative - nominal |
|  | c) | What type of variable is *distance*? | Quantitative - ratio |
| 3. |  | **table(bike$student)** |  |
|  | a) | What is the name of the dataframe? | bike |
|  | b) | What is the name of the variable? | student |
| 4. |  | **student <- bike[bike$student==1,]**  What is the role of the comma?  [1] It tells R to include all the variables (columns) for the riders that are students  [2] It tells R to include all the students (rows), beginning with the first one | 1 |
|  |  | What does "**student**" refer to in this line of code?: **table(student$cyc\_freq)**  [1] the original variable called "student"  [2] the new dataframe called "student" | 2 |
| 5. |  | **Research Question:**  **How many of the cyclists were students, how often did they ride, and what was the average distance they rode?**  Write R code to conduct analysis |  |
|  | a) | Find the number of students in the dataset. | nrow(student) - 14 |
|  | b) | Pull out the student data into a separate dataframe for analysis. | result = student |
|  | c) | Make a table to find how often the students ride. | table(result$cyc\_freq)  Daily Several times per week  8 6 |
|  | d) | Find the average distance ridden | mean(result$distance) - 6.257857 |
| 6. |  | Find answers for the following from the code you have executed |  |
|  | a) | How many **students** are in the dataset? | nrow(result) - 14 |
|  |  | How many variables are in the new data frame "**student**"? | ncol(result) - 9 |
|  | b) | We want to know **how often** the students ride. What is the most frequently observed response? | max(table(result$cyc\_freq)) - 8 |
|  |  | How is the vector "distance" described in the **workspace**? | Ratio – numeric with range 0.52 – 13.95 |
|  | c) | How far do students ride **on average**? | mean(result$distance) |
| 7. |  | **Conclusion** |  |
|  |  | We examined data on ------- student riders. Most of the student riders ( a total of ----- out of ------) rode their bikes -------. On average, the students rode about -------miles on each trip. | We examined data on 14 student riders. Most of the student riders (a total of 8 out of 14) rode their bikes daily. On average, the students rode about 6.257857 miles on each trip. |