

Due : September 21, 2017

Name:
PUID:

Instruction: Please submit your R code along with a brief write-up of the solutions (do not submit raw output with ERRORS:). Some of the questions below can be answered with very little or no programming. However, write R code that outputs the final answer and does not require any additional paper calculations.

Q.N. 1) The data frame *Rat* from the *PASWR* package has the survival time in weeks of 20 male rats exposed to high levels of radiation. Draw a histogram of the survival times of the rats.

Q.N. 2) A data set 'Gapminder' is available in the attached file with this assignment. Gapminder contains data on life expectancy, population and GDP for 142 countries from 1952 to 2007. Below are the variables:

country = Name of the country

continent = Name of five continent

year = ranges from 1952 to 2007 in increments of 5 years

lifeExp = life expectancy at birth (in years)

pop = population

gdpPercap = GDP per capita

- Import the data in R and print first 5 rows (observations)
- Install and load the library ggplot2.
- How many unique countries are represented per continent?
- Map 'gdpPercap' to the x-axis and 'lifeExp' to the y-axis.
- Add points to the plot, make the points size 3 and map continent onto the aesthetics of the point and print the graph. Change the scale of x-axis.

Hint: See section 6.1 and 6.2 in the link below and follow the instructions.

<http://nagraj.net/bims8382-textbook/data-visualization-with-ggplot2.html>

Q.N. 3) The number of visits to a website on each day by visitors is recorded. If a user accesses the site after 30 minutes of inactivity, that will be logged as a new visit. The data is available in the Blackboard as "website traffic".

- Create a chart (side-by-side box plot) that shows the variability in website traffic for each day of the week.
- Recreate the graph to display the box plot in the order of the days of a week.
- Calculate the numerical summary of the website traffic data for each day of the week.

Q.N. 4) Table 1 and Table 2 below are the test scores of 10 students in Test 1 and Test 2

| Name | Test 1 |
|---------|--------|
| Ana | 56 |
| Brian | 78 |
| Cathy | 87 |
| Dough | 89 |
| John | 95 |
| Lucas | 98 |
| Marcus | 59 |
| Nabin | 78 |
| William | 87 |
| Zoe | 98 |

Table 1: Test 1 Scores

| Name | Test2 |
|---------|-------|
| Ana | 86 |
| Brian | 67 |
| Cathy | 78 |
| Dough | 89 |
| John | 87 |
| Lucas | 67 |
| Marcus | 94 |
| Nabin | 78 |
| William | 81 |
| Zoe | 83 |

Table 2: test 2 scores

- a) Use ***merge(..)*** to create a single table containing the student's test 1 and test 2 scores.
- b) How many students did better in the second test?
- c) How many students did better in the first test?
- d) How many students have the same score in both tests?
- e) Calculate the average and standard deviation of both tests.

Q.N. 5) The data set provided below concerns the Auto-MPG.

<https://archive.ics.uci.edu/ml/machine-learning-databases/auto-mpg/auto-mpg.data>

The list of the variables are provided below

- | | | |
|---------------|--------------|-----------------|
| 1. mpg: | 2. cylinders | 3. displacement |
| 4. horsepower | 5. weight | 6. acceleration |
| 7. model year | 8. origin | 9. car name |

- Import the data in R
- Please replace the variables V_1, V_2, \dots, V_9 by the names provided in the table above
- There are some missing values marked as "?". Please remove these missing value and identify the dimension of complete data.
- Create a parallel box-plot for (complete data) to display the mpg by number of cylinders.

Q.N. 6) Access the data from url [http : //www.stat.berkeley.edu/users/statlabs/data/vote.data](http://www.stat.berkeley.edu/users/statlabs/data/vote.data) and store the information in an object named **vote** using the function `read.table()`. This includes the 1988 Stockton Primary Exit Poll Survey:

- How many variables are included in the survey? Please print the variables.
- One of the variable included is the voter's race. Note that following code are used.

0 = missing, 1 = White, 2 = Hispanic, 3 = Black, 4 = Asian, 5 = Other

Display the distribution of the voter's race graphically.