

Final exam R practice

Principles of Data Science with R

Dr. Uma Ravat

PSTAT 10

R Programming Review

- Review of R by conducting an analysis of a dataset.
- Use the `titanic` dataset available in `data` folder.

Do not use the Titanic dataset from RStudio datasets. This is a different dataset.

titanic data set description: 12 variables

- PassengerId: Serial Number
- Survived: Contains binary Values of 0 & 1. Passenger did not survive — 0, Passenger Survived — 1.
- Pclass — Ticket Class | 1st Class, 2nd Class or 3rd Class Ticket
- Name — Name of the passenger
- Sex — Male or Female
- Age — Age in years — Integer
- SibSp — No. of Siblings / Spouses — brothers, sisters and/or husband/wife
- Parch — No. of parents/children — mother/father and/or daughter, son
- Ticket — Serial Number
- Fare — Passenger fare
- Cabin — Cabin Number
- Embarked — Port of Embarkment | C- Cherbourg, Q —Queenstown, S — Southampton

R Programming Review: Analyzing the 'titanic' dataset

- load the titanic dataset to your R session
- Is titanic a data frame?
- Write code to return the first 2 observations.
- Write code to return the last 2 observations
- List all the variable names

Analyzing the 'titanic' dataset (cont): Summary Statistics

- Write code to summarize passenger fares.
- How many people survived compared with how many did not survive.
- Produce a barplot to show the number of survivors and the number of non-survivors.

Use the `airquality` dataset.

- Find the mean temperature.
- Find the median temperature.
- Determine the quantiles for temperature

Continue to use the airquality dataset.

- Compute the mean value for each variable. Write only one line of code.
- Produce a histogram of airquality temperatures. Include a heading and choose a suitable color.
- Is Temp normally distributed?
- Construct side-by-side vertical boxplots showing temperature grouped by Month.

Good luck!

- Read the questions carefully
- Be well reseted before the exam
- You'll do well.