UTAustinX: UT.7.10x Foundations of Data Analysis - Part 1

Bookmark

Reflect on the Question | Analyze the Data | Draw Conclusions

Welcome! This week's lab is intended to be more guided compared to the following weeks as this is your first experience with the lab format for the course.

We will go over several skills that will be important for successfully working with R scripts and data sets for the remainder of the semester. Let's begin by revisiting some of the basic practices you will want to follow when you start each lab.

Lab 1: Cycling in Austin

In 2011, researchers at the Texas Transportation Institute and the Center for Transportation Research at UT Austin ran an advertising campaign aimed at recruiting Austin cyclists to join the South Congress Bike Mapping Project. As members of the project, cyclists downloaded and used *Cycle Tracks*, a smartphone app developed by the San Francisco County Transportation Authority to track where people are riding their bikes based on their GPS points. The goal was to gain new information about bike commuting patterns and this data set is based on the results of the study: 3600 trips tracked from 315 users over a 6 month period. Data includes distances traveled, speed of travel, and reasons for travel among other variables.

Directions:

1. Open R Studio
2. Click the "Import Dataset" button at the top of the workspace window. Choose "from text file."

- Week 3: Bivariate Distributions
- Week 4: Bivariate Distributions (Categorical Data)

3. Click on the Data folder.
4. Select the BikeData file. Make sure "Yes" is checked beside Header. Then click Upload.

(1/1 point)
When we first load a dataset, it is a good idea to immediately **rename** it something simple and familiar to you.

1. In this example, what is the new name for "BikeData?"
`bike<-BikeData`

bike ✔ **Answer:** bike

You have used 1 of 2 submissions

(1/1 point)
In many of our labs, we only want to work with a **subset** of the data.

2. In this example, which variable is being used to identify the cases that will be included in a new dataframe called "males?"
`males <- bike [bike$gender == 'M',]`

gender ✔ **Answer:** gender

You have used 1 of 2 submissions

(1/1 point)
When we want to find an average, we will pull out a **vector** of the the variable and then take its **mean**.

3. In this example, a vector called "male_times" will include a string of values from which dataframe?
`male_times <- males$time`

males ✔ **Answer:** males

You have used 1 of 2 submissions

(1/1 point)
4. Assuming we have run the lines of code from the previous questions, if we want to find the mean ride time for the males, what should we place in the parentheses?
`mean( )`

- time
- bike$time
- male_times ✔
- males

You have used 1 of 2 submissions

(1/1 point)

5. One of the following questions will be answered in lab using basic R functions. Select the question that will require **subsetting** the data.

- How many daily riders are in the dataset and what is their average age? ✔
- What was the average ride time for all the riders in the dataset?
- How many bike riders are in the dataset?

You have used 1 of 2 submissions

© All Rights Reserved

© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY OPENedX

Download on the App Store

ANDROID APP ON Google play