Assignment01

Zhiteng Ma

2022-09-19

My Git settings:

user.name=Zhiteng1000 user.email=zm75@duke.edu

My Git remote settings:

origin https://github.com/Zhiteng1000/EDA-Fall2022 (fetch) origin https://github.com/Zhiteng1000/EDA-Fall2022 (push) upstream https://github.com/ENV872/EDA-Fall2022 (fetch) upstream DISABLE (push)

My Git changes:

fa4e72b (HEAD -> main, origin/main, origin/HEAD) Add Lesson 3 Rmd files c2de398 Add Assignment 2 b94f9ee uploaded Lab2 file 2342d5a Create 02_Reproducibility_CodingBasics.Rmd d74b397 Adds 01_gitExercise.pdf 20a7b70 Add data files f992081 Create Assignments folder 972c063 Create Lessons folder 34b6cc8 Create .gitignore a07d95c Initial commit

Creating code and output:

Basics Day 1

#Generate a sequence of numbers from one to 100, increasing by fours. Assign this sequence a name. seq(from=1, to=100, by=4)

#Compute the mean and median of this sequence. x <- seq(from=1, to=100, by=4) result.mean <- mean(x) print(result.mean)

#Ask R to determine whether the mean is greater than the median. median.result <- median(x) print(median.result)

#Insert comments in your code to describe what you are doing. result.mean != median.result

Basics Day 2

Label each vector with a comment on what type of vector it is.

name <- c("Ma", "Zhi", "teng", "shuai")
##character

scores <- c(49, 70, 51, 55)
##integer

pass <- c(“FALSE”, “TRUE”, “TRUE”, “TRUE”)
##character

studentData <- data.frame(name = c(“Ma”, “Zhi”, “teng”, “shuai”), scores = c(49, 70, 51, 55), pass = c(“FALSE”, “TRUE”, “TRUE”, “TRUE”)) print(studentData)

var.labels <- c(name = “name in Class”, scores = “student scores”, pass = “pass or not”)

##QUESTION: How is this data frame different from a matrix? ##Both represent ‘rectangular’ data types, meaning that they are used to store tabular data, with rows and columns. The main difference, as you’ll see, is that matrices can only contain a single class of data, while data frames can consist of many different classes of data.

ifelse(scores >= 50,“TRUE”,“FALSE”)

ifelse worked, because we should compare the scores with 50, if not above 50, it will print FALSE.