Module 1 Study Guide

Keaton Wilson

12/21/2019

Overview

This study guide is designed to cover expectations for student knowledge and skills in Module 1, and should be used as a tool prior to Exam 1. It's a series of questions that cover the bulk of the material we went through in this module, and closely mirrors the kinds of questions and topics students should expect on Exam 1.

It is broken up by topic for easier organization.

Data Science in the World

During the first class we talked about the field of data science and how it relates to students' experiences and how it might be beneficial to students in this course.

Questions

1. Give a concise definition of Data Science
2. List two reasons that data literacy are important to someone whose profession isn't in the sciences.

Introduction to RStudio

We spent a lot of time in this module becoming familiar with how to use the RStudio IDE. Below are questions that cover some knowledge about the different basic functions that it provide.

Questions

1. Describe the functions of all of the different panes of RStudio (remember that some panes have multiple functions!). 2. List two useful features of RStudio that aren't related to panes (think about navigating file structure, and pop-ups!)

2. Describe one feature of RStudio that you found on your own during this module that we didn't discuss in class.

Programming 101 in R

We covered a lot of ground in a short amount of time on introductory programming. This includes basic R syntax such as assignments, operators, data classes, commenting and vectors.
Questions
1. Explain the difference between what is happening in the two code blocks below.

```r
x = 2
y = 2
x+y
```

```
## [1] 4
```

```r
x = 2
y = 2
z = x+y
print(z)
```

```
## [1] 4
```

2. Name and describe three different data classes in R.

3. What is the character used to comment out lines in R?
4. Concisely give the definition of a vector in R, as you understand it.
5. Thinking forward to the work and classes to come, why do you think these particular peices of syntax/knowledge are important?

Programming 102 in R

We used previously-built skills to delve into more relastic workflows of exploring data.
Questions
1. How do you create a brand new dataframe in R?
2. Can you create a dataframe with two columns, one 22 rows long and one 18 rows long?
3. In base R, how do you select just a certain column of a dataframe?
4. How would you select just rows 1:8?
5. What kind of data class is the following vector: vec = c("TRUE", "FALSE", "TRUE", "TRUE", "TRUE")

Experimental Design

We did a whole day working on an a group-based experimental design module - students designed their own experiment, received critique and discussed it, and finally revised their work to incorporate the discussion and critique.

Questions

Can you describe what a control is? Why is it important? If I gave you an experiment example, could you critique the design?

2d-data, libraries, tidyverse

We worked through some of the core concepts and tools of the tidyverse, including the pipe, and core functions (select, filter, glimpse, slice).

Questions

1. Be comfortable with the workflow and chaining things together - creating a new dataframe that is a subset of an existing dataframe that filters rows and columns.

2. What does the pipe (%>%) do?

3. How do the select() and filter() functions differ?
4. Why do have to use library(tidyverse) before running the filter() function?
5. What does the following code do:

```
data %>%
  filter(data$species == "Gulo gulo") %>%
  select(ID, sex, num_of_attacks, rabies_titer)
```