Quiz 2

02/04/2019

Instructions:

Write your name at the top right. You are to work on this quiz alone without any help from any other resource *except for a single* 8.5×11 *inch page of handwritten notes.*

Problems:

1 We have mostly seen two kinds of data structures so far, "lists" and "vectors". The dataframe "*cats*" looks like the following:

	coat	*weight*	*likes_string*
1	*calico*	2.1	1
2	*black*	5.0	0
3	*tabby*	3.2	1

1.1 What kind of data structure is the variable $cats[1]$?

    Columns of dataframes are vectors of uniform data types. The variable $cats[1]$ is the first column of *cats*. Because this is a specialized type of vector of "Factors" this will also be accepted.

1.2 What kind of data structure is the variable $cats[1,]$?

    Rows of dataframes are lists because they contain in-homogeneous data types. The variable $cats[1,]$ is the first row of the dataframe cats.

1.3 What kind of data structure is the variable $cats[1:2,]$?

    Dataframes are lists. The variable $cats[1:2,]$ contains the first two rows and all columns of cats and is itself a dataframe.

1.4 What is the output of the statement $print(length(cats))$?

    Lists have length equal to their number of entries. Because *cats* contains 3 entries, each equal to one of the vectors, the output is 3.

1.5 What is the output of the statement $print(dim(cats))$?

    Dataframes are also recognized to have vertical directions, with length equal to the length of each of the list entries. Because the dataframe has 3 rows and three columns, the output is "3 3".

1.6 Recall that *cats$coat* represents *categorical labels* for the data. Suppose the third cat in *cats* has gone bald and we want to update its coat. Will the command,

$$cats[3, 1] \leftarrow \text{`}bald\text{'}$$

successfully assign the character string '*bald*' to cat number 3? Explain why or why not.

> When a dataframe has a vector of "factors" or categorical labels, we can only add entries or add rows that contain the same "levels", i.e., entries that are one of the existing labels. In this case, direct entry will result in a null value.

2 Suppose we are plotting data from a variable called "*gapminder*". The library "ggplot2" requires *three* layers to describe a complete graph or plot. Explain what layer is missing in the command below:

$$ggplot(data = gapminder, aes(x = gdpPercap, y = lifeExp))$$

> The library ggplot2 uses three layers to plot data: the data itself, the axes or canvas for the plot and *the representation of the data* as lines, bars, dots etc... The representation of the data is missign from the command.