methodology to tackle assignment 3

Subscribe for email updates.

No tags yet. + Add Tag

Sort replies by: Oldest first Newest first Most popular

 Xin Wang · 6 days ago

Hi all,

I have a hard time tackling the assignment 3. Can anyone share how to approach the questions in a way that is similar to assignment 1 practice? Or resources that I need to turn to review before addressing them? I have checked slides but still have no clues where to start.

Thanks.

Serena

4 · flag

Clarke · 5 days ago

Here is how I approached **best**, hopefully this isn't giving away too much. Let me know if you want **rankhospital** and **rankall**.

Things to do:
1. Read outcome data
Easy enough, just follow the instructions given on the assignment.

2. Check that state and outcome are valid.
We will first need to create vectors of each unique state and outcome so that we can check the input values against. We know that we can extract each state from the State column of the outcome data. We'll probably have to create a vector of the three unique outcome values.

Once you have the vectors, use if loops to check if the input values match one of the unique values. There are some suggestions for that here: https://class.coursera.org/rprog-005/forum/thread?thread_id=1124.

3. Next refer to the documentation to find the columns for 30 day mortality for heart attack, heart failure, and pneumonia. What I did was create three different if loops for **if(outcome == "heart attack"), if(outcome == "heart failure"){** and **..."pneumonia"**. If it matched, I found it useful to grab the applicable column, rename that column to **x** and stick it into a completely new data frame, along with the other variables we will need. The forums are full of questions on problems with the data changing from factor to numeric to character, etc. Read the help page for **?data.frame**, specifically **stringsAsFactors.**

4. Now that you have a data frame with the required data, do some subsetting to return the hospital name in the input state with the lowest 30 day death rate (you can also subset beforehand). But make sure you attend to the "Handling ties" instructions, there's a step for this.

↑ 10 ↓ · flag

Fabio Adamo · 21 hours ago

can you add the link for this handling ties, please

↑ 0 ↓ · flag

Clarke · 12 hours ago

You can use `order` to order the data according first to mortality rate and second to hospital name, then when you return the best hospital, the first in alphabetical order will be returned if there is a tie, which is what the assignment instructions ask for. Read ?order and also http://www.ats.ucla.edu/stat/r/faq/sort.htm, specifically the sort2.hsb2 variable.

↑ 0 ↓ · flag

Victor Stoddard Signature Track · 12 hours ago

Order can take more than one argument, so you can order by a numeric column then by a character column. If you do this right after you create your data frame, you don't need to handle ties later on.

↑ 0 ↓ · flag

+ Comment


Susan Mol · 5 days ago

Hi Serena,

For point 2 from Clarke, you don't really need loops, because you can check if the argument is 'in' the vector.
The most confusing part about R programming for me, (coming from a very loopy background) is that

you can do so much without using loops. It feels wrong and I have to get past that! Vectorised operations are a different paradigm and important in R.
The points about what to do are ok, but think about the 'how' part and see if you can do it without loops.

regards,
Sue

↑ 2 ↓ · flag

A post was deleted

+ Comment

 Xin Wang · 5 days ago

Thank you Susan and Clarke.

To Clarke: this has been helpful though I have to figure out what the steps mean specifically :)

I had a weird submit which is correct for part 3 and then got this repeatedly:

Error in sys.call(sys.parent()) : node stack overflow
Error during wrapup: node stack overflow

What does this mean?

Thanks!

Serena

↑ 0 ↓ · flag

+ Comment

Sofie De Meyer Signature Track · 5 days ago

Hi Clarke,
Could you provide info for rankhospital and rankall as well?
Thanks!
Sofie

↑ 1 ↓ · flag

+ Comment

Clarke · 5 days ago

Sure.

For rank hospital, follow the same first steps as for best. As I said, I subset for the state, then run if loops for outcome to pull the necessary data into a new dataframe, then order it . So going into it, if the input is "heart attack" and "TX," I'll have a dataframe **hosp** that looks like this, with the left column rowname, the **hosp$x** mortality data (on second thought I should have called this **Rate** to match the given output), and the **hosp$Hospital.Name** hospital name:

```
> head(hosp)
      x                     Hospital.Name
42   8.1       FORT DUNCAN MEDICAL CENTER
187  8.5  TOMBALL REGIONAL MEDICAL CENTER
204  8.7 CYPRESS FAIRBANKS MEDICAL CENTER
61   8.7           DETAR HOSPITAL NAVARRO
116  8.8           METHODIST HOSPITAL,THE
69   8.8  MISSION REGIONAL MEDICAL CENTER
```

Once these are ordered, you can cbind a Rank column with nrows of a number representing the rank of each hospital.

Then you just have to write three if loops for the num input.

```
  if(num == "best"){
```

Return the number one ranking hospital. Should be easy if you've already ordered the data.

```
  if(num == "worst"){
```

Return the worst ranking hospital.

```
  if(is.numeric(num)){
```

Now you have to attend to the instructions to return NA if num is greater than the number of hospitals in that state. So first you need a vector that is nrows of the state. If num is greater than that, return NA, else proceed.

Next I found it useful to use the `which` function. Read ?which. If you've added a **hosp$Rank** column, you can ask which hosp$Rank equals num, and subset **hosp** to return that value.

I'll do rankall later, I have to actually go work!

↑ 2 ↓ · flag

Victor Stoddard Signature Track · 11 hours ago

I think it may be simpler to just use a residual (else) condition at the end:

```
If ( num == "worst" ){
    #do something...
} else if ( num == "best") {
    #do something...
} else {
    #residual. do something...
}
```

It seems tidier than having multiple, independent *if* conditions, and you don't have to use is.numeric().

↑ 0 ↓ · flag

+ Comment

Àurea Signature Track · 5 days ago

Hi Clarke!
First of all, thank you very much for sharing with us the steps you followed, it has really helped me a lot. I'm very lost at the last step of "best function".

It is what I have done, and don't know how to continue..

best <- function(state, outcome) {

##Fix the working directory and read the outcome-of-... file (named file).
##Condition of invalid outcome or state (it runs)
##dat is a subset with only the hospitals in the state
##Step 3 (loops with every outcome). I can determine the minimum of each one: it is an example

if (outcome=="heart attack"){
m<- min(dat[, 11], na.rm=T)
}
#second loop
#third loop
}

But then, I dont know how to subset... I would really appreciate if anybody could help me please. Thank you in advanced!

↑ 0 ↓ · flag

Clarke · 5 days ago

Hi Aurea,

First, make sure you are ordering the data so as to fulfill the handling ties instructions. You can order the data by mortality rate and hospital name in one function.

I did not use the if loops to get the minimum value, but rather to get the data I needed. Then after the if loop closed, I ordered it and returned the value required by the function.

```
if (outcome=="heart attack"){
  dat.compl <- subset dat to be complete cases of data from the mortal
ity
  hosp <- data.frame("Rate"= mortality data from dat.compl as numeric,
  "Hospital.Name" = hospital name from dat.compl , stringsAsFactors =
FALSE)
```

* If you did not read in the CSV with na.strings = "Not Available" then you'll need to find a way to exclude the Not Available. There is advice for that at https://class.coursera.org/rprog-005/forum/thread?thread_id=796

After running the loop, now hosp contains all data you will need to fulfill the function. After you've ordered it, you no longer need `min` because the best value should be first. Return the hospital name.

0 · flag

+ Comment

argyn kuketayev Signature Track · 5 days ago

On SE they ask posters to show they did an honest effort to solve the problem before helping with home work. You simply asked to show you the solution without even trying it yourself. This is ridiculous.

-7 · flag

Clarke · 5 days ago

Eh, relax and cut her some slack, she actually just asked for methodology/approach, not the solution. This isn't SE, it is a classroom. The lectures give very little direction or starting point. If people are genuinely here to learn they'll put forth the effort intrinsically after they've been given a little guidance. I think the lectures should have given a few application exercises, but they were chunks of content rushed through and not really tied together. The guidance I am giving--I do hope it's not too much--but I think it is what should have been included as part of the assignment. I would bet the attrition rate in this class is way too high because the assignments have so little connection to the lecture.

10 · flag

Qamaruddin Shaikh Signature Track · 16 hours ago

I agree with you, the problem with these lectures are , that it is being taught by PH.Ds, and they think that the student of the course should have the same knowledge as they have. they dont understand we are from the mainstream workforce, handling tonnes of applications, and we are the folks who need guidance like step-by-step tutorial.I think I am better learn R from youtube then these prestigious university.Imagine they found a better way of earning money, $49*500000 is a good money for them.I know the instructor will fail me. But thats OK. I am better off learning from youtube and contribute towards my community.

-1 · flag

Victor Stoddard Signature Track · 12 hours ago

These problems encourage independent study and exploration. View the videos or slides again, search the web, and use Debug, use this forum (constructively) and you will arrive at the solution.

0 · flag

+ Comment

 Xin Wang · 5 days ago

Thanks for the understanding Clarke.

We oftentimes talk about learning with enough input, but I am struggling with limited chucks of codes to understand how to construct. As a newbie, assignment 1 practice is very helpful. I do hope we could have a document similar to that for assignment 3. As even with that, a portion of us still take time and struggle to figure out how to apply the learning to the actual assignment.

Serena

2 · flag

+ Comment

Anonymous · 21 hours ago

Hi,
this has been really helpful. I think it is only with the help of fellow students that one can work on

assignments. the lectures and information provided by the professor is not useful at all. I am not all satisfied.

1 · flag

+ Comment

New post

To ensure a positive and productive discussion, please read our forum posting policies before posting.

B | I | Link | <code> | Pic | Math | Edit: Rich | Preview

Make this post anonymous to other students

Subscribe to this thread at the same time

Add post