Data Structures
FOR 128: Lab 4

Insert Your Name Here

2024-09-25

Welcome

Welcome to Lab 4! Today, we'll focus on data structures (such as vectors and dataframes), and logical subsetting of data structures.

Learning objectives

- Gain more understanding of important data structures in `R`.
- Apply logical subsetting to data.

Deliverables (i.e., what to put in the lab drop box)

Upload your rendered PDF (`lab_04.pdf`) **and** Quarto (`lab_04.qmd`) document to the lab drop box. Make sure the Quarto document properly renders to PDF.

Collaborator(s)

List any collaborators you worked with below.

Part 1: Six Trees

We'll first look at six trees measured across three forest stands as described in IFDAR.

Exercise 1

Do Exercise 4.1 from IFDAR

Exercise 2

Do Exercise 4.2 from IFDAR

Exercise 3

Do Exercise 4.3 from IFDAR

 a.

 b.

 c.

 d.

Exercise 4

Do Exercise 4.4 from IFDAR

Exercise 5

Do Exercise 4.5 from IFDAR

 a.

 b.

 c.

 d.

 e.

 f.

 g.

Exercise 6

Do Exercise 4.6 from IFDAR

Exercise 7

Do Exercise 4.7 from IFDAR

Exercise 8

Do Exercise 4.8 from IFDAR

Exercise 9

Do Exercise 4.9 from IFDAR

Part 2: the `forested` package

This section of the lab will use the `forested` R package. If you haven't already, you'll need to install it with the `install.packages()` function.

Data Overview

The `forested` R package contains a dataset, also called `forested`, that contains plot data collected by the U.S. Department of Agriculture, Forest Service, Forest Inventory and Analysis (FIA) Program in Washington state. Each row represents a plot and each column represents a value measured at that plot. Take a look at the help file for the `forested` dataset to learn more about the data.

Exercise 10

Load the `forested` R package and its data with the code below. How many rows does the `forested` dataframe have? How many columns?

```
library(forested)
```

```
Warning: package 'forested' was built under R version 4.3.3
```

```
data(forested)
str(forested)
```

```
Classes 'tbl_df', 'tbl' and 'data.frame':   7107 obs. of  19 variables:
 $ forested        : Factor w/ 2 levels "Yes","No": 1 1 2 1 1 1 1 1 1 1 ...
 $ year            : num  2005 2005 2005 2005 2005 ...
 $ elevation       : num  881 113 164 299 806 736 636 224 52 2240 ...
 $ eastness        : num  90 -25 -84 93 47 -27 -48 -65 -62 -67 ...
 $ northness       : num  43 96 53 34 -88 -96 87 -75 78 -74 ...
 $ roughness       : num  63 30 13 6 35 53 3 9 42 99 ...
 $ tree_no_tree    : Factor w/ 2 levels "Tree","No tree": 1 1 1 2 1 1 2 1 1 2 ...
 $ dew_temp        : num  0.04 6.4 6.06 4.43 1.06 1.35 1.42 6.39 6.5 -5.63 ...
 $ precip_annual   : num  466 1710 1297 2545 609 ...
 $ temp_annual_mean: num  6.42 10.64 10.07 9.86 7.72 ...
 $ temp_annual_min : num  -8.32 1.4 0.19 -1.2 -5.98 ...
 $ temp_annual_max : num  12.9 15.8 14.4 15.8 13.8 ...
 $ temp_january_min: num  -0.08 5.44 5.72 3.95 1.6 1.12 0.99 5.54 6.2 -4.54 ...
 $ vapor_min       : num  78 34 49 67 114 67 67 31 60 79 ...
 $ vapor_max       : num  1194 938 754 1164 1254 ...
 $ canopy_cover    : num  50 79 47 42 59 36 14 27 82 12 ...
 $ lon             : num  -119 -123 -122 -122 -118 ...
 $ lat             : num  48.7 47.1 48.8 45.8 48.1 ...
 $ land_type       : Factor w/ 3 levels "Barren","Non-tree vegetation",..: 3 3 3 3 3 3 2 2 3
```

Exercise 11

Tell me about three columns of your choice in the forested dataset. In particular: - What are they measuring? - Do they have units? If so, what are they? - Do they have factor levels? If so, what are they?

Exercise 12

Create a new dataframe by subsetting the forested R package to only include rows that are forested. Call this dataframe only_forested.

Exercise 13

Similarly to Exercise 11, create a new dataframe by subsetting the forested R package to only include rows that are not forested. Call this dataframe not_forested.

Exercise 14

    a. What is the average (mean) percent tree canopy cover on the forested plots?

    b. What is the average (mean) percent tree canopy cover on the unforested plots?

Exercise 15

Print the row with the highest elevation for each dataset using logical subsetting: `only_forested` and `not_forested`. Which dataset contains the highest elevation plot?

Hint: you might want to use the `which.max()` function.

Exercise 16

Use logical subsetting to find the plot in the `forested` dataset with the highest annual precipitation. Is this plot forested or not? What is it's elevation?

Exercise 17

Use logical subsetting to print the elevation and percent canopy cover of all plots in the `forested` dataset with elevation greater than 2,500 meters.

Exercise 18

Use logical subsetting to print the elevation and percent canopy cover of all plots in the `forested` dataset with elevation greater than 2,500 meters and percent canopy cover greater than 12%.

Exercise 19

Use logical subsetting to print all plots in the `forested` dataset that have either barren or non-tree vegetation land cover types.

Wrap up

Congratulations! You've made it to the end of Lab 4. Make sure to render your final document and submit both the .pdf and .qmd file to D2L.