SUBSETTING PRACTICE WITH ISLANDS

DSC 105 Intro to data science, Lyon College, Fall'24

[yourname] (pledged)

November 2, 2024

Readme

This exercise is very similar to the practice on the `Nile` data set. Here, we're using another built-in data set, `islands`. Instead of `time`, you need to use `names` to analyze the content of `islands`, and instead of a line plot, you will draw a dot plot.

10 Problems

Find out what `islands` contains and what its structure is. Take a look at `example(islands)`, too. Submit the completed file in Canvas for bonus points.

```
str(islands)
```

```
Named num [1:48] 11506 5500 16988 2968 16 ...
- attr(*, "names")= chr [1:48] "Africa" "Antarctica" "Asia" "Australia" ...
```

1. How many elements does the vector `islands` have?

   ```
   length(islands)
   ```

   ```
   [1] 48
   ```

2. Retrieve the third to fourth, the twelfth to thirty-fifth, and the 48th element of `islands` with one command, and check that you retrieved 28 elements altogether (with another command)

```
islands[c(3:5,12:35,48)]
length(islands[c(3:5,12:35,48)])
```

```
          Asia      Australia   Axel Heiberg           Cuba          Devon
         16988           2968             16             43             21
     Ellesmere         Europe      Greenland         Hainan     Hispaniola
            82           3745            840             13             30
      Hokkaido         Honshu        Iceland        Ireland           Java
            30             89             40             33             49
        Kyushu          Luzon     Madagascar       Melville       Mindanao
            14             42            227             16             36
      Moluccas    New Britain     New Guinea New Zealand (N) New Zealand (S)
            29             15            306             44             58
  Newfoundland   North America       Victoria
            43           9390             82
[1] 28
```

3. Create a vector `islands_unnamed` from `islands` that is not named and show its structure. Tip: to remove the `names` of a vector, assign `NULL` to it.

```
islands_unnamed <- islands
names(islands_unnamed) <- NULL
str(islands_unnamed)
```

```
num [1:48] 11506 5500 16988 2968 16 ...
```

4. Print the value of `islands` that belongs to the `names(islands)` element `"Iceland"`- which data science question could this answer?

```
islands[names(islands)=="Iceland"]
```

```
Iceland
     40
```

»What is the area of Iceland in thousands of square miles?«

5. Which element of `islands` corresponds to the greatest area? (Put differently: what is the greatest landmass on Earth?)

```
names(islands)[islands == max(islands)]
```

```
[1] "Asia"
```

6. What is the index of the next-to-last element of `islands`? Use `which` to answer this question and save the result as `index`.

```
index <- which(islands == islands[length(islands)-1])
index
```

```
Vancouver
       47
```

7. Check your answer to the last question by finding the value of `islands` that belongs to `index`.

```
islands[index]
```

```
Vancouver
       12
```

8. How many values of `islands` are larger than the next-to-last value of `islands`?

```
length(islands[islands > islands[index]])
```

```
[1] 47
```

9. What is the landmass of all areas listed in `islands`?

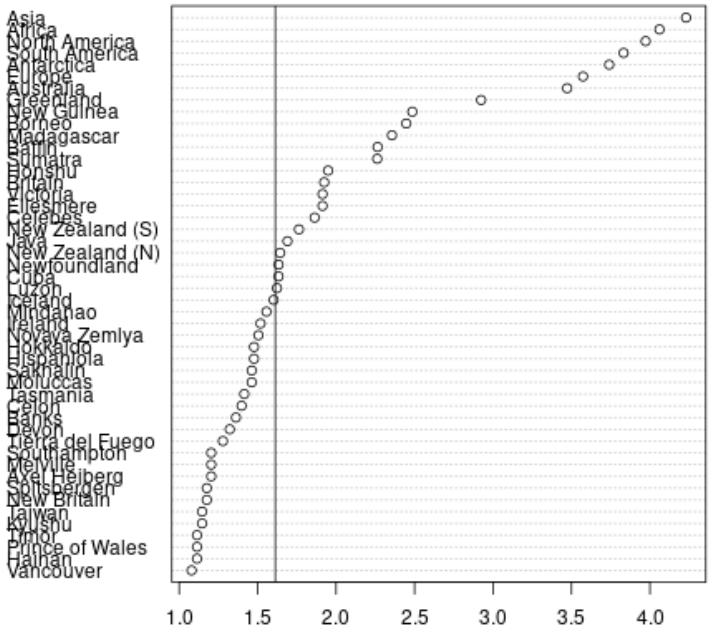
```
paste("Total landmass:", sum(islands), "thousands of square miles.")
```

```
[1] "Total landmass: 60131 thousands of square miles."
```

10. Make a dot plot of all entries in the data set `islands` using the `dotchart` function. As its only argument, use the `log10`-transformed, ordered data set `areas <- log10(sort(islands))`.

    Add a line for the median of this data set, `median(areas)` and add it to the plot with `abline(v=median(areas)`.

```
areas <- log10(sort(islands))
md_areas <- median(areas)
dotchart(areas)
abline(v=md_areas)
```



Tip: You can change the appearance of lines with the parameters `col`, `lty`, `lwd`. E.g. `col="red, lty=2, lwd=2` for a red, dashed, thick line.

Customized with line type, line width, title, labels and color:

```
dotchart(areas)
abline(v=md_areas, col="red",lty=2,lwd=2)
title("Area of Earth's landmasses (log-transformed)",
      xlab="(Log10 of) Landmass in thousands of miles")
```