Popquiz 1

Without any other resources, complete the following tasks in R for Solar Eclipse data from NASA/kaggle.com for a total of 5,000 years.

1. Store URL: Create an R object `url` for the web address http://tinyurl.com/solar-csv and check that the object is there:

```
url <- 'http://tinyurl.com/solar-csv'
url  # or ls()
ls()
```

2. Data import: The CSV data do have a header row. Import the CSV file from the web into a dataframe `solar` and display its data structure.

```
solar <- read.csv(url,header=TRUE)
str(solar)
```

3. Index extraction: Write a command to extract the number of the column named `Calendar.Date` from the `solar` dataframe.

```
which(names(solar)=='Calendar.Date')
which(names(solar)=='Eclipse.Type')
which(names(solar)=='Latitude')
which(names(solar)=='Longitude')
which(names(solar)=='Central.Duration')
```

```
[1] 2
[1] 7
[1] 10
[1] 11
[1] 15
```

4. Transformation: Write the following steps in a code block:

    1. Store the indices of some columns (given in a table below) in a vector `cols`.
    2. Copy these columns from the dataframe `solar` to a new dataframe `sol`.
    3. Change the names of the columns from the old to the new name shown in the table.
    4. Display the structure of `sol`.

Column	Old name	New name
2	Calendar.Date	date
7	Eclipse.Type	type
10	Latitude	lat
11	Longitude	lon
15	Central.Duration	tot

```
cols <- c(2,7,10,11,15)
sol <- solar[cols]
names(sol) <- c('date','type','lat','lon','tot')
str(sol)
```

5. Display the contingency table of the solar eclipse types (in `type`) and store it in an R object called `tbl`. Then print the table.

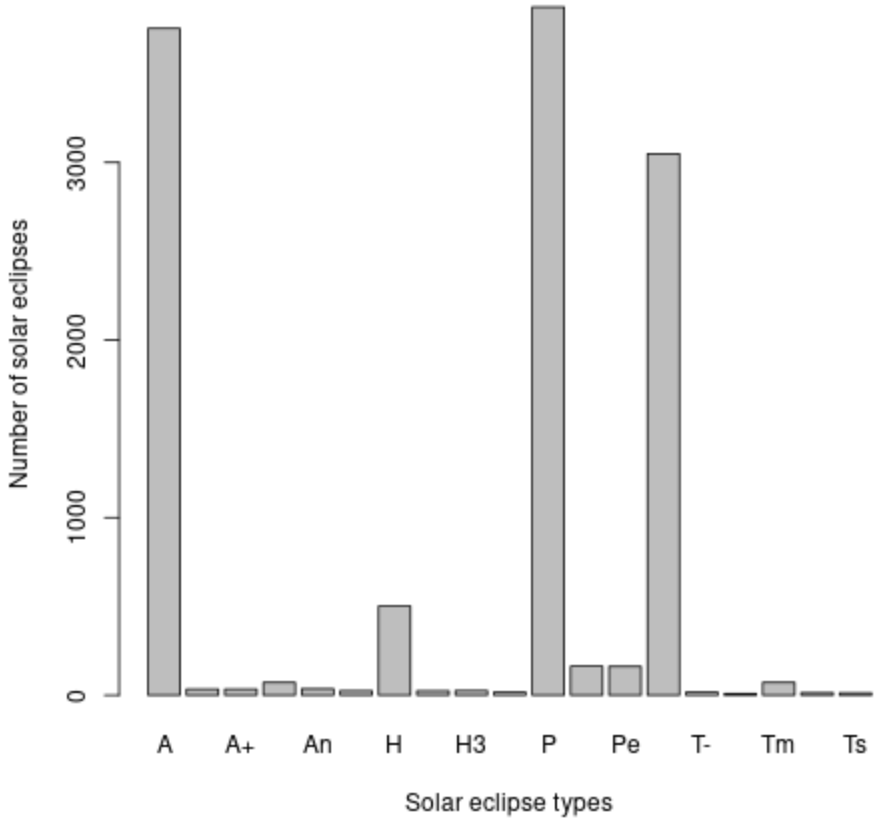
```
tbl <- table(sol$type)
tbl
```

6. How many types of solar eclipses are there? Write a command that returns the number of types.

```
length(tbl)
```

7. The table `tbl` has `names` and frequencies. Make a `barplot` of the table names. Label the x- and y-axis appropriately.

```
barplot(tbl,
        xlab="Solar eclipse types",
        ylab="Number of solar eclipses")
```

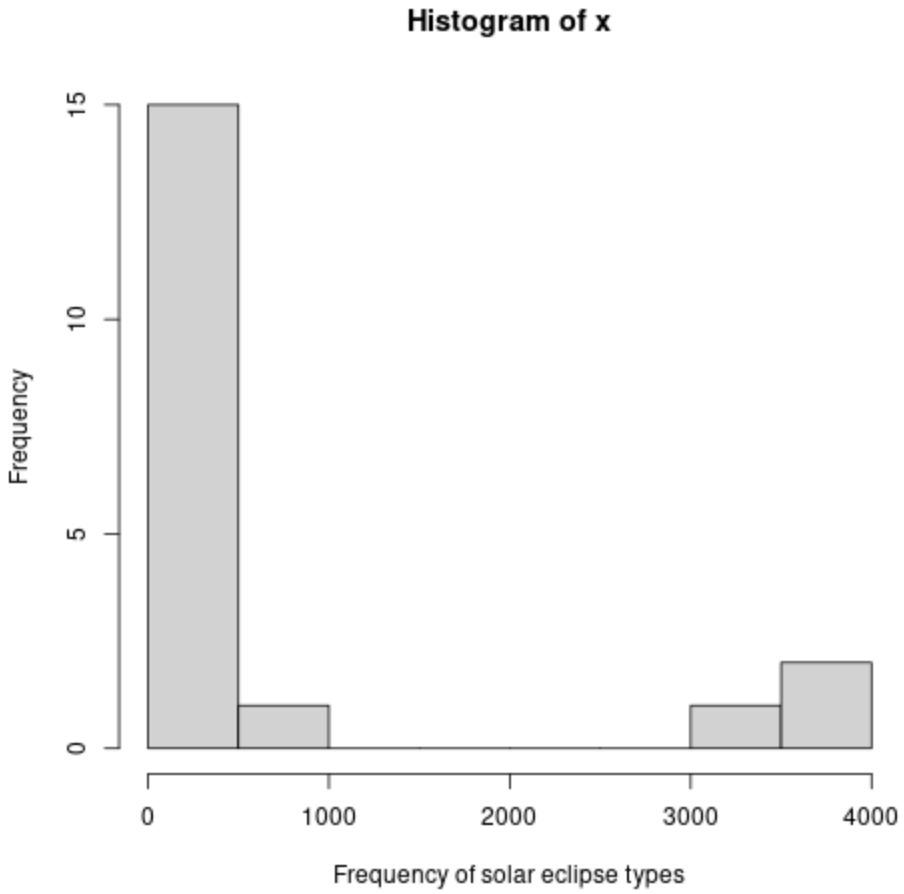


8. Convert the table `tbl` to a dataframe `types` and name the first column `Type`, then print the dataframe.

```
types <- as.data.frame(tbl)
names(types)[1] <- "Type"
types
```

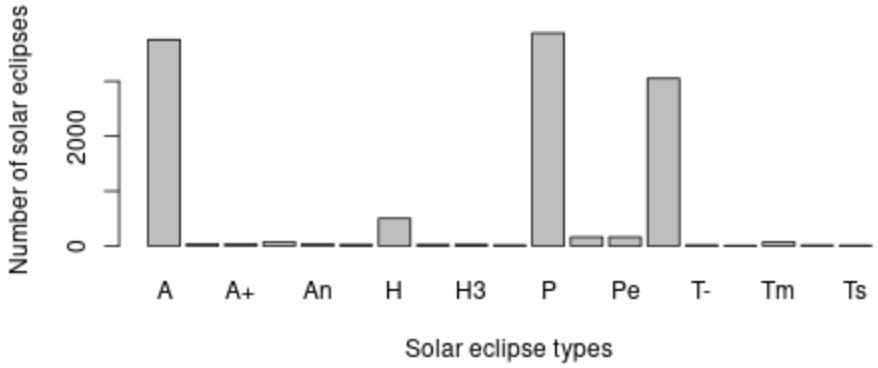
9. Make a histogram of the frequencies in `types` in decreasing order, and name the x-axis appropriately. Tip: to sort a vector `x` in decreasing order, run `sort(x, decreasing=TRUE)`.

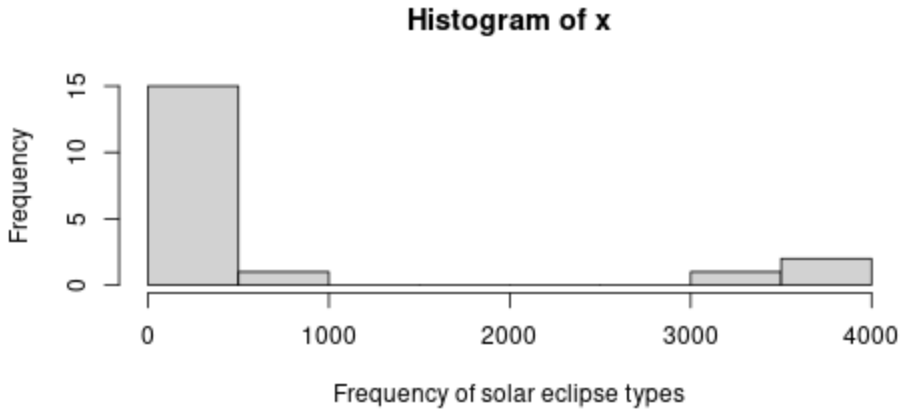
```
x <- sort(types$Freq,decreasing=TRUE)
hist(x,
     xlab="Frequency of solar eclipse types")
```



10. Place both the barplot and histogram you just made in one panel on top of one another (barplot on top, histogram below).

```
par(mfrow=c(2,1))
barplot(tbl,
        xlab="Solar eclipse types",
        ylab="Number of solar eclipses")
x <- sort(types$Freq,decreasing=TRUE)
hist(x,
     xlab="Frequency of solar eclipse types")
par()
```





11. When you're finished, tell me your confidence that your code is error-free.
12. Create an Emacs Org-mode file and enter your code.
13. Enter all code blocks first, then start at the top and run them.
14. Grade yourself based on the percentage of your code that worked.
15. Hand in your graded pop quiz if you like, for bonus points.
16. Fix your code based on the solutions file.

Author: Marcus Birkenkrahe

Created: 2024-03-04 Mon 13:32

Validate