Midterm II

2022-10-26

Your name:

Questions

Q1

Given below are the monthly deaths from bronchitis, emphysema and asthma in the UK from 1974 to 1979.

```
knitr::kable(mydata)
```

year	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
1974	3035	1721	2933	1607	2787	1489	3102	1498	2815	1529	3084	1461
1975	2552	1524	2889	1545	3891	1300	2294	1361	3137	1366	2605	1354
1976	2704	1596	2938	1396	3179	1356	2385	1346	2679	1357	2573	1333
1977	2554	2074	2497	1787	2011	1653	2444	1564	1969	1570	2143	1492
1978	2014	2199	1870	2076	1636	2013	1748	1640	1870	1535	1693	1781
1979	1655	2512	1726	2837	1580	2823	1554	2293	1633	2491	1504	1915

a.

Write a for loop that will return a vector of the ratio of the mean value to the median value for columns 2-13 in `mydata` (shown above).

b.

Describe what is returned by the code below, including the type of R object produced, the length or dimension of the object, and the information contained in the object.

```
map_dbl(mydata %>% select(-1), mean) %>% mean()
```

c.

Describe what is returned by the code below, including the type of R object produced, the length or dimension of the object, and the information contained in the object.

```
ratio_fun <- function(x) quantile(x, probs= c(0.25, 0.5, 0.75))
map_dfr(mydata %>% select(-1), ratio_fun, .id = "month")
```

d.

Describe what is returned by the code below, including the type of R object produced, the length or dimension of the object, and the information contained in the object.

```
map_dfc(mydata %>% select(-1), ratio_fun)
```

e.

Describe what is returned by the code below, including the type of R object produced, the length or dimension of the object, and the information contained in the object.

```
lapply(mydata %>% select(-1), ratio_fun) %>%  unlist()
```

f.

Describe what is done by the code below. What type of R object is `mydata_long`? What is the length or dimension of the `mydata_long`?

```
mydata_long <- mydata %>% pivot_longer(cols = 2:13,
                        names_to = "months",
                        values_to = "deaths")
```

g.

Write a function that takes the R object from `f` and year as arguments and returns the median deaths for that particular year. Additionally, your function should give a warning if the user enters a year that is not in the dataset.

Q2

a.

Consider the function below. Give the output produced by `myfun(3)` and explain how you arrived at your answer.

```
myfun <- function(x) {
  if (x < 3 | x > 3){
  rep("hi", x)
  } else{
  rep("bye", x)
  }
}
```

Q3

Consider the following data frames for this question.

```
df1 = data.frame(Id = c(1:6), Item = c(rep("Lake1", 3), rep("Lake2", 3)))
df2 = data.frame(Id = c(1, 3, 5), PH = c(rep("Acidic", 2), rep("Basic", 1)))
df1
##   Id  Item
## 1  1 Lake1
## 2  2 Lake1
## 3  3 Lake1
## 4  4 Lake2
## 5  5 Lake2
## 6  6 Lake2
df2
##   Id     PH
## 1  1 Acidic
## 2  3 Acidic
## 3  5  Basic
```

What do the following codes do? Provide a sketch of the resulting data frame.

a.

```
semi_join(df1, df2)
```

b.

```
right_join(df1, df2)
```

c.

```
full_join(df1, df2)
```

Q4

Consider the following string:

```
my_string <- "A 5.1 magnitude earthquake strikes near San Jose, US Geological Survey 10 reports"
```

a.

Carefully explain what `str_view_all` will highlight in `my_string` for the given pattern.

```
pattern <- "[\\d?.\\d]"
str_view_all(my_string, pattern)
```

b.

Explain how you would use regular expression to count the number of words in `my_string`. Remember, 5.1 and 10 do not count as words. Include the actual commands you would need to accomplish this task.

c.

Carefully explain what the following code chunk will do.

```
pattern <- "[\\d+]"
str_replace_all(my_string, pattern, "X")
```

d.

Suppose you want to place a period at the end of `my_string`. Fill in the appropriate pattern and `stringr` function below that can place a period at the end of `my_string`.

```
pattern <- _________________
_____________(my_string, pattern, ".")
```