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).

```
store_ratio <- rep(NA, 12)
for (i in 1:12){
  store_ratio[i] <- mean(mydata[[i+1]])/median(mydata[[i+1]])
}
store_ratio
##  [1] 0.9475127 1.0211682 0.9192351 1.1046946 1.0479366 1.1281562 0.9636675
##  [8] 1.0561724 1.0114028 1.0713664 0.9614080 1.0538435
```

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()
## [1] 2056.625
```

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")
## # A tibble: 12 x 4
##    month `25%` `50%` `75%`
##    <chr> <dbl> <dbl> <dbl>
##  1 jan   2148. 2553  2666.
##  2 feb   1627. 1898. 2168.
##  3 mar   2027. 2693  2922
##  4 apr   1560. 1697  2004.
##  5 may   1730. 2399  3081
##  6 jun   1389. 1571  1923
##  7 jul   1884. 2340. 2429.
##  8 aug   1395. 1531  1621
##  9 sep   1895. 2324  2781
## 10 oct   1407. 1532  1561.
## 11 nov   1806. 2358  2597
## 12 dec   1381. 1476. 1709.
```

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)
## # A tibble: 3 x 12
##     jan   feb   mar   apr   may   jun   jul   aug   sep   oct   nov   dec
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2148. 1627. 2027. 1560. 1730. 1389. 1884. 1395. 1895. 1407. 1806. 1381.
## 2 2553  1898. 2693  1697  2399  1571  2340. 1531  2324  1532  2358  1476.
## 3 2666. 2168. 2922  2004. 3081  1923  2429. 1621  2781  1561. 2597  1709.
```

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()
## jan.25% jan.50% jan.75% feb.25% feb.50% feb.75% mar.25% mar.50% mar.75% apr.25%
## 2148.50 2553.00 2666.50 1627.25 1897.50 2167.75 2026.75 2693.00 2922.00 1560.50
## apr.50% apr.75% may.25% may.50% may.75% jun.25% jun.50% jun.75% jul.25% jul.50%
## 1697.00 2003.75 1729.75 2399.00 3081.00 1389.25 1571.00 1923.00 1884.50 2339.50
## jul.75% aug.25% aug.50% aug.75% sep.25% sep.50% sep.75% oct.25% oct.50% oct.75%
## 2429.25 1395.25 1531.00 1621.00 1894.75 2324.00 2781.00 1406.75 1532.00 1561.25
## nov.25% nov.50% nov.75% dec.25% dec.50% dec.75%
## 1805.50 2358.00 2597.00 1380.75 1476.50 1708.75
```

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 data object from `f` and year as an argument and returns the median deaths for that particular year. Additionally, your function should give a warning if the user enters a year that not in the dataset.

```
iqr_func <- function(data, year){
  if (year >= 1974 & year <= 1979){
    data %>% filter(year == year) %>%
    summarize(med = median(deaths))
  } else {"Please enter a year from 1974 to 1979"}
}
iqr_func(mydata_long, 1975)
## # A tibble: 1 x 1
##     med
##   <dbl>
## 1  1870
```

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)
##   Id  Item
## 1  1 Lake1
## 2  3 Lake1
## 3  5 Lake2
```

b.

```
right_join(df1, df2)
##   Id  Item     PH
## 1  1 Lake1 Acidic
## 2  3 Lake1 Acidic
## 3  5 Lake2  basic
```

c.

```
full_join(df1, df2)
##   Id  Item     PH
## 1  1 Lake1 Acidic
## 2  2 Lake1   <NA>
## 3  3 Lake1 Acidic
## 4  4 Lake2   <NA>
## 5  5 Lake2  basic
## 6  6 Lake2   <NA>
```

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 `x` 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 x. Include the actual commands you would need to accomplish this task.

```
str_replace_all(my_string, pattern = "[\\d?.\\d]", "") %>%
  str_extract_all(pattern = "\\w+") %>%
  unlist() %>%
  length()
## [1] 11
```

c.

Carefully explain what does the following code chunk will do.

```
pattern <- "[\\d+]"
str_replace_all(my_string, pattern, "X")
## [1] "A X.X magnitude earthquake strikes near San Jose, US Geological Survey XX reports"
```

d.

Suppose you want to place a periods at the end of the `my_string`. The replacement attempt below contains one mistake. Write down the correct command that can place a periods at the end of `my_string`.

```
pattern <- "$"
str_replace_all(my_string, pattern, ".")
## [1] "A 5.1 magnitude earthquake strikes near San Jose, US Geological Survey 10 reports."
```