In-class Exercise 13 Results for Simran Mander

Score for this attempt: **10** out of 10
Submitted Nov 22 at 9:17pm
This attempt took 1,789 minutes.

Question 1

2 / 2 pts

Start with the our usual North American Stock Market dataset. Note that naicsh is a numerical variable which indicates the industry in which a given firm operates. First, create a new data frame by dropping any `NA` values in fyear or naicsh and keeping any observation if its naicsh is larger than `111111`. There are 187148 observations in the new data frame.

Now you will count how many observations there are for each naicsh in each fyear.

You can see within each combination of `naicsh` and `fyear`, there can be more than one observation. As part of this question, we want to find how many combinations of `naicsh` and `fyear` only have exactly **TWO** observations.

Take the example of IMPERIAL GINSENG PRODS LTD (gvkey 027301) and CALGENCE INC (gvkey 012349). In fiscal year `1994`, both of them were the only two observations classified under the industry with the naicsh classification number of 111219.

What is the number of combinations of `naicsh` and fyear which have only two observations.

2455

Answer 1:

Correct!

187148

Answer 2:

Correct!

2455

```
dat <- companies %>% select(gvkey, conm, fyear,
naicsh) %>% filter(!is.na(fyear), !is.na(naicsh),
naicsh>111111)
dat1 <- dat %>% group_by(fyear, naicsh) %>%
summarize(count = n())
dim(dat) # you could use View(n) of course # will
give you 187148
sum(dat1$count == 2) # should produce 2455
```

Question 2

2 / 2 pts

Please complete the following code which will create a column named `total_emp` that indicates the total employment (i.e., total number of employees) for a given location (i.e., loc) and a given naicsh industry code.

```
dat2 <- companies %>%
```

select (gvkey, conm, fyear, naicsh, loc, emp) %>%

group_by (naicsh, loc) %>%

mutate (total_emp = sum (emp, na.rm = TRUE))

Note that you want to retain all the observations.

Answer 1:

Correct!

select

Answer 2:

Correct!

group_by

Answer 3:

Correct!

mutate

Answer 4:

Correct!

sum

```
dat2 <- companies %>%
select(gvkey, conm, fyear, naicsh, loc, emp) %>%
group_by(naicsh, loc) %>%
mutate(total_emp = sum(emp, na.rm = TRUE))
```

Question 3

2 / 2 pts

Based on our usual North American Stock Market data frame (i.e., `companies`), create a new data frame where you only keep firms with `naicsh` of `335228` or `naicsh` of `339112` and with headquarter location (i.e., `loc`) of `ANT`. And then, create a new column called total_emp.

```
dat3 <- companies %>%
        filter(naicsh == 335228 | naicsh == 339112, loc == "ANT") %>%
        group_by(naicsh, loc) %>%
```

```
mutate(total_emp = sum(emp, na.rm = TRUE))
```

After this step, you run the following code:

```
dat3$emp[dat3$naicsh == 335228 & dat3$fyear == 2000] <-
3
```

By using `dat3`, What would be the mean value of `total_emp` for the firms classified under the `naicsh` code of `335228` in financial year (`fyear`) of `2000`?

71.7

Answer 1:

Correct!

companies

Answer 2:

Correct!

==

Answer 3:

Correct!

|

Answer 4:

Correct!

==

Answer 5:

Correct!

"ANT"

Answer 6:

Correct!

71.7

```
dat3 <- companies %>%
    filter(naicsh == 335228 | naicsh == 339112, loc == "ANT") %>%
    group_by(naicsh, loc) %>%
    mutate(total_emp = sum(emp, na.rm = TRUE))

dat3$emp[dat3$naicsh == 335228 & dat3$fyear == 2000] <- 3  # note that this assignment would will not affect total_emp

dat3 %>%
  filter(naicsh == 335228, fyear == 2000) %>%
  summarise(mean(total_emp, na.rm=TRUE))
```

Question 4

2 / 2 pts

Again, use our usual North American Stock Market dataset. First, remove any observations with `NA` in the `loc, ni` or `fyear` columns.

Note that for each `fyear` and `loc` combination, there can be multiple observations recorded in the dataset. it is possible that in a given combination of `fyear` and loc, `ni` values of the all the observations are *negative*. You are asked to identify how many of those combinations (i.e., `fyear-loc` combinations) in which the `ni` values of all the observations are *negative*.

There are 113 such `fyear-loc` combinations.

Answer 1:

Correct!

113

```
dat4<- companies %>%
     select(fyear, loc, ni) %>%
    filter(!is.na(fyear), !is.na(loc), !is.na(ni)) %>%

   group_by(fyear, loc) %>%

   summarise(marked = ifelse(max(ni) < 0, TRUE, FALSE)) %>%
   filter(marked==TRUE)
# View(dat4) to see the number of rows
```

Question 5

2 / 2 pts

Please start with our usual North American Stock Market dataset. You are asked to create a new summary data frame.

First, remove any observations with `NA` in the `at, emp` or `fyear` columns.

Then, for each fyear, you are asked to calculate `ave_at` which is the average of `at` values recorded in that fyear and `ave_emp` which is the average of emp values recorded in that year. You are asked not to retain all the observations of the screened data.

Finally, you are asked to order the data in the descending order of fyear.

What is the `ave_at` value of the last row of the new data frame? Please round the value to its second decimal place.

2238.99

Answer 1:

Correct!

2238.99

```
dat5 <- companies %>%
  filter(!is.na(fyear), !is.na(at), !is.na(emp)) %>%
  group_by(fyear) %>%
   summarise(ave_at=mean(at), ave_emp=mean(emp)) %>%
   arrange(desc(fyear))

round(dat5$ave_at[20],2)
```

Quiz Score: **10** out of 10