

In-Class Exercise 3

⚠ This is a preview of the published version of the quiz

Started: Jul 21 at 3:50p.m.

Quiz Instructions

As long as you are online, your answers will be automatically saved while you complete the exercise. While you have unlimited attempts to complete this in-class-exercise before the due date, only the grade of the last submitted attempt will be recorded.

Note: For this all questions in the exercise, assume that the command library(tidyverse) has been successfully run.

Question 1

1.25 pts

How many missing values (NA's) are there **in total** across 288,158 observations in all 41 variables in the North American Stock Market dataset from 1994-2018? (If a single observation has missing values for three variables, this would count as three NAs, for example.)

Hint: is.na() also works with data frames.

- ☐ 288,158
- ☐ 1,103,773
- ☐ 1,423,872
- ☐ 1,637,332
- ☐ None of the above.

Question 2

1.25 pts

In the North America Stock Market 1994-2018 dataset, what is the **minimum** value of the sale variable? Note that the sale variable is in millions. The answers below are

rounded to one decimal place.

- ☐ -18,009.2
- ☐ -15,009.3
- ☐ 0.0
- ☐ 2.6
- ☐ None of the above.

Question 3

1.25 pts

Now, you want to keep all the observations where total liabilities (i.e., **lt**) is greater than or equal to \$10 million, and assets that are greater than or equal to \$10 million. (Remember the data is in millions so if it is equal to 10, that is \$10 million.)

What is the highest total assets (i.e., **at**) to total liabilities (i.e., **lt**) ratio on this screened sample?

The answer is **closest** to:

- ☐ 0
- ☐ 424
- ☐ 673
- ☐ 856
- ☐ 24680

Question 4

1.25 pts

Continue the previous question. What is the **gvkey** value of the company which has the highest ratio (**at/lt**) in question 2?

- ☐ 001072

☐ 001874☐ 013624☐ 015122☐ 151971

Question 5

1.25 pts

Now, go back to the **full companies** data frame you started with in Question 1. You are interested in the following question:

In fiscal year (fyear) 2007, what fraction of companies had total assets greater than \$1 billion (at > 1000)?

You then want to round this result to the 3rd decimal place.

Which of the following blocks of code below would do this?

☐

```
numerator1 <- companies %>% filter(fyear == 2007, at>1000)
denominator <- companies %>% filter(fyear == 2007)
round(numerator1/denominator,3)
```

☐

```
numerator1 <- companies %>%
  select(fyear == 2007, at>1000) %>%
  summarise(n())

denominator <- companies %>%
  select(fyear == 2007) %>%
  summarise(n())

round(numerator1/denominator,3)
```

☐

```
numerator1 <- companies %>%
  filter(fyear == 2007, at>1000) %>%
  summarise(sum())

denominator <- companies %>%
  filter(fyear == 2007) %>%
  summarise(sum())

round(numerator1/denominator,3)
```

☐ numerator1 <- companies %>%
filter(fyear == 2007, at>1000) %>%
summarise(n())

denominator <- companies %>%
filter(fyear == 2007) %>%
summarise(n())

round(numerator1/denominator,3)

☐ None of the above.

Question 6

1.25 pts

For this question, use the full North American Stock Market 1994-2018 dataset, from the data frame called *companies*.

Now that you have done this, you would like to create a new data frame with **only** the following columns:

company name (i.e., conm)
employment (i.e., emp), and
fiscal year (i.e., fyear).

You would also like to remove the observations for which employment (i.e., emp) is missing (that is, where it is equal to *NA*). Finally, you would like your new data frame to contain observations from the 2010 fiscal year only.

Which of the following commands below will let you do this?

- ☐ df1 <- companies %>% filter(conm, emp, fyear) %>% select(!is.na(emp), fyear==2010)
- ☐ df1 <- companies %>% filter(conm, emp, fyear) %>% select(is.na(emp), fyear==2010)
- ☐ df1 <- companies %>% select(conm, emp, fyear) %>% filter(!is.na(emp), fyear==2010)
- ☐ df1 <- companies %>% select(conm, emp, fyear) %>% filter(emp, fyear==2010, na.rm=TRUE)
- ☐ None of the above.

Question 7**1.25 pts**

Continuing the previous question, you now want to find the difference between the *maximum* value and the *minimum* value of the number of employees recorded in the **df1 dataset** in fiscal year 2010. Employment is listed in thousands, and so you found that the difference (in thousands) is:

- ☐ 2,093
- ☐ 2,100
- ☐ 2,384
- ☐ 2,490
- ☐ None of the above.

Question 8**1.25 pts**

Continue the previous question, and use **df1**. Since employment is listed in thousands, you'd like to create a new dataset called **df2**, which has all the same variables as **df1**, but another variable called `emp_actual` which lists actual employment.

For example, if the value of `emp` for an observation is 1,000, you'd like `emp_actual` to be equal to 1,000,000.

The code to do this is:

- ☐ `df2 <- companies %>%
mutate(emp_actual = 1000*emp)`
- ☐ `df2 <- df1 %>%
mutate(emp_actual = 1000*emp)`
- ☐ `df2 <- df1 %>%
mutate(emp_actual == 1000*emp)`
- ☐ `df2 <- df1 %>%
mutate(emp_actual = 1000000*emp)`

☐ None of the above.

Not saved

Submit Quiz