STA035B Midterm 1, Winter 2024

Name _____

Student ID _____

Section _____

problem	points
1a	
1b	
2a	
2b	
3	
4	
5a	
5b	
total	

Problem 1

Consider the following code.

```
scores <- tribble(
  ~name, ~midterm1, ~midterm2,
  "Mary", 80, 90,
  "Jose", NA, 100,
  "Ali", 75, 95,
)
cleaned_scores <- scores %>%
  mutate(
    midterm1 = replace_na(midterm1, 100),
    midterm2 = replace_na(midterm2, 100)
)
```

For each of the following, draw the tibble which results from the following code. With words, describe how many rows and columns there are in the resulting tibble, and say whether or not there are missing values and, if there are any, where they appear in the tibble.

(a) 5 points:

```
scores %>%
  mutate(a = pmin(midterm1, midterm2))
```

(b) 5 points:

```
cleaned_scores %>%
  mutate(c = pmin(midterm1, midterm2))
```

Problem 2

Suppose we have a tibble `flights` whose first few rows look like this:

time_hour	carrier	flight	tailnum	origin	dest	air_time
2013-01-01 05:00:00	UA	1545	N14228	EWR	IAH	227
2013-01-01 05:00:00	UA	1714	N24211	LGA	IAH	227
2013-01-01 05:00:00	AA	1141	N619AA	JFK	MIA	160
2013-01-01 05:00:00	DL	725	N804JB	JFK	BQN	183
2013-01-01 06:00:00	DL	461	N668DN	LGA	ATL	116
2013-01-01 05:00:00	UA	1696	N39463	EWR	ORD	150

Describe the outputs of the following lines of code.

(a) 5 points:

```
flights %>%
  group_by(origin) %>%
  summarize(n = n())
```

(b) 5 points:

```
str_remove(flights$dest, '^[AEIOU]')
```

Problem 3 (3 points)

Consider the following tibbles:

```
df1 <- tribble(
  ~product, ~q1, ~q2,
  "A", 150, 200,
  "B", 120, 180
)

df2 <- tribble(
  ~product, ~quarter, ~sales,
  "A", "q1", 150,
  "A", "q2", 200,
  "B", "q1", 120,
  "B", "q2", 180
)
```

Which of the following code correctly transforms df2 into df1?

(A) `df2 %>% pivot_wider(id_cols = c(product, quarter), names_from = quarter, values_from = sales)`

(B) `df2 %>% pivot_wider(id_cols = product, names_from = quarter, values_from = sales)`

(C) `df2 %>% pivot_wider(id_cols = c(product, quarter), names_from = sales, values_from = quarter)`

(D) `df2 %>% pivot_wider(id_cols = quarter, names_from = product, values_from = sales)`

Problem 4 (5 points)

Consider the following vector of strings.

```
strings <- c("William;Order 1", "Jenny;order 2", "Alex;order 25")
```

Suppose we want to use regex to return the strings vector but where we erase the name preceding the semicolon and delete the semicolon. Which of the following options correctly does this task? Explain. (If you get the answer correct, you don't need an explanation. If you get it incorrect, any explanations for why some of the options are incorrect can get you partial points.)

(A) `str_remove(strings, "^\\b+;")`

(B) `str_remove(strings, "^\\w+;")`

(C) `str_remove(strings, "$[a-z]*;")`

(D) `str_remove(strings, "$[A-Za-z]+;")`

Problem 5

Consider the two following tibbles:

```
majors <- tribble(
  ~student_id, ~major,
  123, "Math",
  234, "Statistics",
  345, "Literature",
)

grades <- tribble(
  ~student_id, ~course, ~grade,
  345, "Machiavelli", "B",
  123, "Analysis", "A",
  456, "Organic Chemistry", "C"
)
```

For each of the following, draw the tibble which results from the following code. With words, describe how many rows and columns there are in the resulting tibble, and describe any missing values.

(a) 5 points:

```
majors %>% left_join(grades)
```

(b) 5 points:

```
grades %>% left_join(majors)
```