STA035B Midterm Practice

Spencer Frei

Instructions

Problem 1

Consider the following code.

```
scores <- tribble(
  ~name, ~wage, ~hours,
  "Mary", 25, 40,
  "Jose", 35, 38,
  "Ali", 37, NA,
  "Pat", NA, 42
)
cleaned_scores <- scores %>%
  mutate(
    wage = replace_na(wage, 0),
    hours = replace_na(hours, 0)
)
```

For each of the following, explain if the code is a valid command. If it is, describe what the output of the command is (if it is a tibble, draw the tibble). If it not, explain why.

a.

```
scores %>%
  mutate(a = wage * hours)
```

```
# A tibble: 4 x 4
  name   wage hours     a
  <chr> <dbl> <dbl> <dbl>
1 Mary     25    40  1000
2 Jose     35    38  1330
3 Ali      37    NA    NA
4 Pat      NA    42    NA
```

b.

```
scores %>%
  mutate(b = product(wage, hours, na.rm=TRUE))
```

c.

```
cleaned_scores %>%
  mutate(c = wage * hours)
```

```
# A tibble: 4 x 4
  name   wage hours     c
  <chr> <dbl> <dbl> <dbl>
1 Mary     25    40  1000
2 Jose     35    38  1330
3 Ali      37     0     0
4 Pat       0    42     0
```

Problem 2

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

origin	temp	dewp	humid	wind_dir	wind_speed	wind_gust	precip	pressure	visib	time_hour
EWR	39.02	26.06	59.37	270	10.35702	NA	0	1012.0	10	2013-01-01 01:00:00

origin	temp	dewp	humid	wind_dir	wind_speed	wind_gust	precip	pressure	visib	time_hour
EWR	39.02	26.96	61.63	250	8.05546	NA	0	1012.3	10	2013-01-01 02:00:00
EWR	39.02	28.04	64.43	240	11.50780	NA	0	1012.5	10	2013-01-01 03:00:00
EWR	39.92	28.04	62.21	250	12.65858	NA	0	1012.2	10	2013-01-01 04:00:00
EWR	39.02	28.04	64.43	260	12.65858	NA	0	1011.9	10	2013-01-01 05:00:00
EWR	37.94	28.04	67.21	240	11.50780	NA	0	1012.4	10	2013-01-01 06:00:00

For each of the following, explain if the code is a valid command. If it is, describe what the output of the command is. If it not, explain why.

a.

```
weather[3,]
```

b.

```
weather[,"EWR"]
```

c.

```
weather %>%
  summarize(a = mean(wind_gust))
```

```
# A tibble: 1 x 1
      a
  <dbl>
1    NA
```

d.

```
str_replace(weather$origin, '[ER]', 'X')
```

Problem 3

Consider the following vector of strings.

```
strings <- c("William; Grade: A", "Jenny; Grade: B-", "Alex; Grade: B+")
```

a. Suppose we want to use regex to return a vector containing strings that indicate only the student's grades (i.e., A, B-, B+ for the vector `strings`). Which of the following options correctly does this task?

(A) `str_replace(strings, "^[A-Za-z]+: (.*)", "\\1")`

(B) `str_replace(strings, ".*: .*", "\\2")`

(C) `str_replace(strings, "(.*): (.*)", "\\2")`

(D) `str_replace(strings, "^[A-Za-z]*: (.*)", "\\2")`