

# Class Activity 10

Your name here

April 14 2024

## Your Turn 1

```
students <- tibble(
  id = 1:24,
  grade = sample(c("9th", "10th", "11th"), 24, replace = TRUE),
  region = sample(c("North America", "Europe", "Asia", "South America", "Middle East", "Africa"), 24, replace = TRUE),
  score = round(runif(24, 50, 100))
)
```

a. Create a new column `grade_fac` by converting the `grade` column into a factor. Reorder the levels of `grade_fac` to be “9th”, “10th”, and “11th”. Sort the dataset based on the `grade_fac` column.

*Answer:*

```
students_a <-
Error: <text>:2:0: unexpected end of input
1: students_a <-
  ^
```

b. Create a new column `region_fac` by converting the `region` column into a factor. Collapse the `region_fac` levels into three categories: “Americas”, “EMEA” and “Asia”. Count the number of students in each collapsed region category.

```
students_b <- students_a
Error in eval(expr, envir, enclos): object 'students_a' not found
```

c. Create a new column `grade_infreq` that is a copy of the `grade_fac` column. Reorder the levels of `grade_infreq` based on their frequency in the dataset. Print the levels of `grade_infreq` to check the ordering.

```
students_c <- students_a
Error in eval(expr, envir, enclos): object 'students_a' not found
```

d. Create a new column `grade_lumped` by lumping the least frequent level of the `grade_fac` column into an 'Others' category.

Count the number of students in each of the categories of the `grade_lumped` column.

```
students_d <-
Error: <text>:2:0: unexpected end of input
1: students_d <-
  ^
```

## Your Turn 2

Lets import the `gss_cat` dataset from the `forcats` library. This dataset contains a sample of categorical variables from the General Social survey.

```
# import gss_cat dataset from forcats library
forcats::gss_cat
# A tibble: 21,483 x 9
   year marital      age race rincome      partyid      relig denom tvhours
  <int> <fct>      <int> <fct> <fct>      <fct>      <fct> <fct>      <int>
1  2000 Never married    26 White $8000 to 9999 Ind,near ~ Prot~ Sout~      12
2  2000 Divorced        48 White $8000 to 9999 Not str r~ Prot~ Bapt~      NA
3  2000 Widowed         67 White Not applicable Indepe~ Prot~ No d~       2
4  2000 Never married    39 White Not applicable Ind,near ~ Orth~ Not ~       4
5  2000 Divorced        25 White Not applicable Not str d~ None Not ~       1
6  2000 Married         25 White $20000 - 24999 Strong de~ Prot~ Sout~      NA
7  2000 Never married    36 White $25000 or more Not str r~ Chri~ Not ~       3
8  2000 Divorced        44 White $7000 to 7999 Ind,near ~ Prot~ Luth~      NA
9  2000 Married         44 White $25000 or more Not str d~ Prot~ Other       0
10 2000 Married         47 White $25000 or more Strong re~ Prot~ Sout~       3
# i 21,473 more rows
```

Use `gss_cat` to answer the following questions.

a. Which religions watch the least TV?

```
# your r-code

gss_cat %>%

Error: <text>:5:0: unexpected end of input
3: gss_cat %>%
4:
  ^
```

b. Do married people watch more or less TV than single people?

```
# your r-code

gss_cat %>%
Error: <text>:5:0: unexpected end of input
3: gss_cat %>%
4:
  ^
```

c. Collapse the `marital` variable to have levels `Married`, `Not_married`, and `No_answer` . Include "Never

```
married", "Divorced", and "Widowed" in Not_married
# your r-code

gss_cat %>%

Error: <text>:6:0: unexpected end of input
4:
5:
  ^
```