Question 1 [50 points]

The data for this question comes from the `STAR` dataset from the `AER` library. Below is a summary and five sample rows of a modified version of that dataset containing information from a study examining the effect of reducing class size on student performance in primary school.

```
str(STAR_data)
```

```
'data.frame':   3114 obs. of  6 variables:
 $ student_ID: int  1 2 3 4 5 6 7 8 9 10 ...
 $ stark     : Factor w/ 3 levels "regular","small",..: 2 2 1 2 1 1 2 2 1 3 ...
 $ star1     : Factor w/ 3 levels "regular","small",..: 2 2 1 2 1 1 2 2 1 3 ...
 $ readk     : int  447 450 448 447 431 451 478 455 430 437 ...
 $ read1     : int  507 579 651 533 558 548 514 530 490 503 ...
 $ read2     : int  568 588 614 608 608 596 569 608 622 552 ...
```

```
STAR_data %>% slice(sample(1:n(), 5))
```

```
  student_ID        stark        star1 readk read1 read2
1       1868 regular+aide regular+aide   421   495   545
2       2730        small        small   503   590   637
3        703      regular regular+aide   427   530   614
4       1541 regular+aide        small   495   501   536
5       1769 regular+aide regular+aide   436   503   599
```

Besides the Student ID, we will focus on four other measures from the data: `stark` and `star1`, which indicate the type of class in kindergarten and grade 1, respectively ("regular", "small", or "regular+aide"); and `readk`, `read1`, and `read2` which are reading scores from kindergarten, grade 1 and grade 2 respectively.

(a) [**5 pts**] Write a line of code that will generate the following `tibble` (or `data.frame`) with the total number of students who were in each type of class in kindergarten:

```
# A tibble: 3 x 2
# Groups:   stark [3]
  stark            n
  <fct>        <int>
1 regular       1067
2 small          987
3 regular+aide  1060
```

Answer

CONTINUED ON NEXT PAGE

(b) [**5 pts**] Write a line of code that will generate the following `tibble` (or `data.frame`) with the total number of students who were in each combination of type of class in kindergarten and grade 1, as below:

```
count_table
```

```
# A tibble: 9 x 3
# Groups:   stark, star1 [9]
  stark        star1             n
  <fct>        <fct>         <int>
1 regular      regular         518
2 regular      small            85
3 regular      regular+aide    464
4 small        regular          29
5 small        small           924
6 small        regular+aide     34
7 regular+aide regular         491
8 regular+aide small            85
9 regular+aide regular+aide    484
```

Answer:

(c) [**5 pts**] Assume the tibble from part (b) is called `count_table` as above. Now write a line of code that produces a tibble which gives, for each class type in kindergarten, the *proportion* of students in each class type in grade 1:

Answer:

Here is some code which creates an object `STAR_what`.

```
STAR_what <- STAR_data %>%
  pivot_longer(cols=readk:read2,names_to="Test",values_to="Score") %>%
  select(-student_ID)
```

(d) [**5 pts**] What class of object is STAR_what?

Answer:

CONTINUED ON NEXT PAGE

2

In class we used `xtabs` to create contingency tables of counts of combinations of qualitative variables, as in this example:

```
STAR_who_denom <- xtabs(~star1+Test+stark,data=STAR_what)
STAR_who_denom
```

```
, , stark = regular

             Test
star1          read1 read2 readk
  regular        518   518   518
  small           85    85    85
  regular+aide   464   464   464

, , stark = small

             Test
star1          read1 read2 readk
  regular         29    29    29
  small          924   924   924
  regular+aide    34    34    34

, , stark = regular+aide

             Test
star1          read1 read2 readk
  regular        491   491   491
  small           85    85    85
  regular+aide   484   484   484
```

(e) **[5 pts]** What will the code `STAR_who_denom[1,3,2]` return as output?

Answer:

CONTINUED ON NEXT PAGE

`xtabs` can also be used to sum up values of another variable for different combinations of `star1`, `Test` and `stark` by putting the variable name in front of the `~`. For example, we can find the **total** of all scores by using

```
STAR_who_num <- xtabs(Score~star1+Test+stark,data=STAR_what)
STAR_who_num
```

```
, , stark = regular

            Test
star1          read1  read2  readk
  regular     273728 306238 228798
  small        45797  50785  37660
  regular+aide 249580 276710 205622

, , stark = small

            Test
star1          read1  read2  readk
  regular      15396  17009  12617
  small       500773 552478 413608
  regular+aide 18338  20488  14927

, , stark = regular+aide

            Test
star1          read1  read2  readk
  regular     261220 290488 218272
  small        44596  49270  37070
  regular+aide 258514 286343 212980
```

CONTINUED ON NEXT PAGE

(f) [**5 pts**] Using `STAR_who_num` and `STAR_who_denom`, write a single line of code that assigns the average score for each `star1` by `Test` by `stark` combination to an object called `STAR_avg` as seen below:

`STAR_avg`

```
, , stark = regular

            Test
star1              read1     read2     readk
  regular         528.4324 591.1931 441.6950
  small           538.7882 597.4706 443.0588
  regular+aide    537.8879 596.3578 443.1509

, , stark = small

            Test
star1              read1     read2     readk
  regular         530.8966 586.5172 435.0690
  small           541.9621 597.9199 447.6277
  regular+aide    539.3529 602.5882 439.0294

, , stark = regular+aide

            Test
star1              read1     read2     readk
  regular         532.0163 591.6253 444.5458
  small           524.6588 579.6471 436.1176
  regular+aide    534.1198 591.6178 440.0413
```

Answer:

(g) [**10 pts**] Write a line of code that creates an array that contains the difference between the average `read2` and `readk` scores for each `stark` by `star1` combination using `STAR_avg` above.

```
            stark
star1              regular     small regular+aide
  regular          149.4981 151.4483      147.0794
  small            154.4118 150.2922      143.5294
  regular+aide     153.2069 163.5588      151.5764
```

Answer:

CONTINUED ON NEXT PAGE

(h) [**10 pts**] Write code (possibly multiple lines) using the original `STAR_what` to produce a tibble containing the same rows and columns as the object in part (g).

```
# A tibble: 3 x 4
# Groups:    star1 [3]
  star1         regular small `regular+aide`
  <fct>          <dbl> <dbl>          <dbl>
1 regular         149.  151.           147.
2 small           154.  150.           144.
3 regular+aide    153.  164.           152.
```

Answer:

END OF QUESTION 1