Lab 14 *Some of these problems may be more challenging than others. Please feel free to work with others, attend office hours, or post on the course discussion forum if you need help. While collaboration with other students is encouraged, each student is responsible for submitting his or her own work. This assignment should be submitted in one well-commented SAS program. For any questions that require a written answer, do so in the SAS comments. Be sure to re-name the uploaded SAS scripts according to the naming convention* `LastnameFirstinitial_Lab#.sas` (*e.g.,* `PileggiS_Lab14.sas`).

1. Create a SAS library to access the `O2012.sas7bdat` data set. This data set is about Olympic medalists from the 2012 Olympics. Recall that this data set starts with an "oh" and not a "zero".
2. Write a couple of SAS procedures to help you familiarize yourself with the data and summarize the data. How many observations are there? What does an observation represent? Note your findings as a comment in your SAS code.
3. Create a new, temporary data set that sorts the olympics data by participating countries.
4. Create a second new, temporary data set. In this data set, create the following variables. It may be helpful if you create them in the chunks presented below, then carefully check your data to make sure the variables were created correctly.
   (a) medal variables

   `num_medalists` - total number of medalists for each country

   `num_gold` - total number of gold medals for each country

   `num_silver` - total number of silver medals for each country

   `num_bronze` - total number of bronze medals for each country

   `num_total` - total number of medals (gold, silver, and bronze) for each country

   (b) functions of medals

   `ave_medals` - total number of medals divided by total number of medalists for each country (represents average medals earned per medalist)

   `max_medals` - maximum number of total medals earned by an *individual* athlete for each country

   (c) athlete variables

   `ave_age` - average age of all medalists for each country

   `ave_wt` - average weight of all medalists for each country

   `ave_ht` - average height of all medalists for each country

   `prop_m` - proportion of all medalists that are male for each country

5. Carefully examine the `ave_medals` variable.

   (a) In a comment in your SAS code, explain why it doesn't make sense for a country to have a `ave_medals` less than one.

   (b) Identify the three countries that have `ave_medals` less than one. Note your results in a comment in your SAS code.

   (c) Print the `country`, `name`, `total`, `num_total`, `num_medalists` and `ave_medals` variables for these three countries. Explain how they ended up having a `ave_medals` less than one. Note your findings in a comment in your SAS code. *Note: you do not need to fix or create a cleaned version of this variable.*

6. Similarly to the `ave_medals` variable, the `ave_wt` variable was also incorrectly calculated.

   (a) Use a SAS procedure to identify the six *sports* in which athletes have missing values for weights. Note your findings as a comment in your SAS code.

   (b) Use a SAS procedure to identify which sport has the largest number of missing values for weight. Note your findings as a comment in your SAS code.

   (c) Go back to your data step from question 4 and create a new variable that represents the total number of medalists without missing weight values (`num_nomisswt`).

   (d) Calculate a corrected average weight of the athletes for each country (ave_wt_fixed) using the new variable from (b).

   Note: Other variables that have missing values were also calculated incorrectly. You are not required to fix these other variables.

7. Create country level information.

   (a) Create a third temporary data set that only has one observation per country. This data set should only have the variable `country` and the variables that you created in questions 4 and 6.

   (b) Print your data using an option such that a maximum of two decimal places are displayed in the output. So that you may check your work, a print out for the first 15 countries is provided on the next page (note the output is *wrapping* due to the number of variables).

The SAS System

Obs	Country	max_medals	num_medalists	num_gold	num_silver	num_bronze	num_total
1	Australia	4	25	7	17	8	32
2	Azerbaijan	1	1	0	0	1	1
3	Belarus	1	4	1	1	2	4
4	Belgium	1	2	0	1	1	2
5	Brazil	1	5	1	1	3	5
6	Canada	1	25	0	18	7	25
7	Colombia	1	3	0	2	1	3
8	Croatia	1	4	0	4	0	4
9	Cuba	1	5	2	2	1	5
10	Czech Republic	1	2	0	2	0	2
11	Democratic People's Republic of Korea	1	5	4	0	1	5
12	Denmark	1	7	0	1	6	7
13	Egypt	1	1	0	1	0	1
14	France	3	27	11	11	11	33
15	Georgia	1	1	1	0	0	1

Obs	num_nomisswt	ave_medals	ave_age	ave_wt	ave_wt_fixed	ave_ht	prop_m
1	25	1.28	24.44	73.16	73.16	179.68	0.24
2	1	1.00	18.00	56.00	56.00	156.00	1.00
3	4	1.00	29.50	68.25	68.25	167.75	0.25
4	2	1.00	26.00	61.00	61.00	171.00	0.50
5	5	1.00	23.40	81.60	81.60	176.40	0.60
6	25	1.00	28.24	77.76	77.76	180.32	0.44
7	3	1.00	26.67	64.67	64.67	170.67	0.67
8	3	1.00	23.50	70.50	94.00	140.50	1.00
9	4	1.00	26.40	62.40	78.00	0.00	0.60
10	2	1.00	27.00	85.50	85.50	183.50	1.00
11	5	1.00	23.00	57.20	57.20	156.00	0.40
12	7	1.00	29.14	76.71	76.71	184.71	0.86
13	1	1.00	21.00	82.00	82.00	188.00	1.00
14	27	1.22	25.11	76.96	76.96	182.15	0.52
15	1	1.00	20.00	66.00	66.00	167.00	1.00