

Assignment 1

You are given 4 datasets of different forms of child malnutrition (Overweight, Underweight, Stunted and Wasted) and a Data Country Parameters dataset.

Note-

- Values may differ according to assumptions taken by you. Make sure to mention your reasons behind taking each assumption.
- Make sure to add plot titles and texts wherever required.

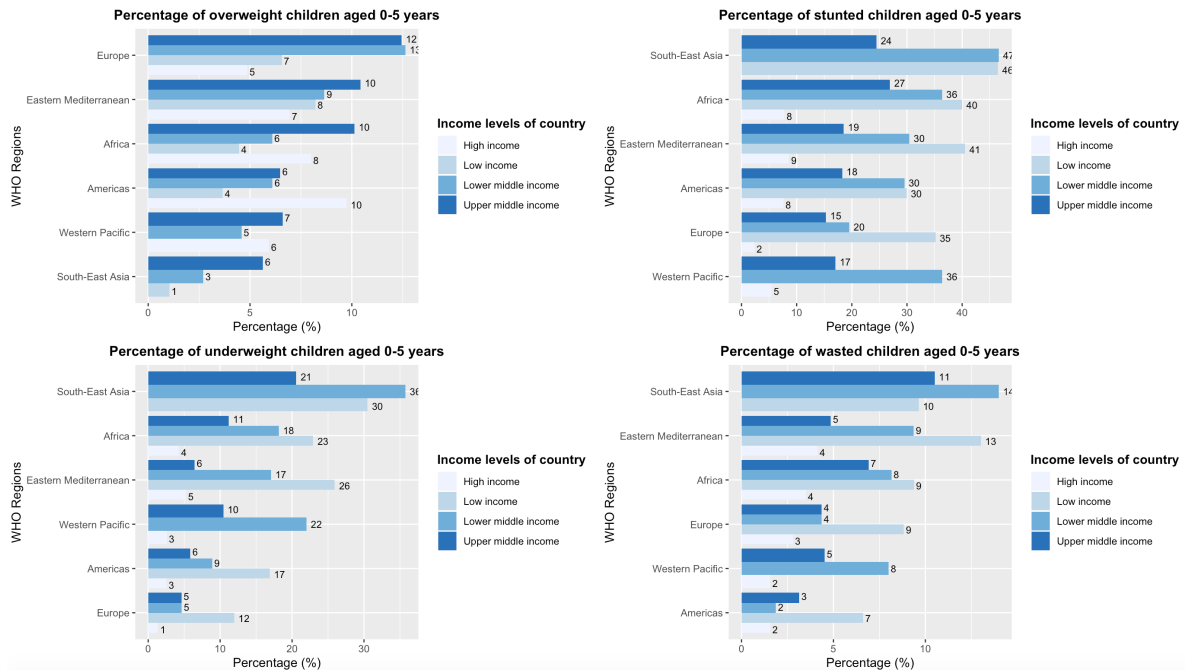
1. Perform the following steps to get the required table (15 points)

- Join all 4 forms of malnutrition tables using joins on Country and Year in R
- For each country, take the mean value for all the forms of malnutrition
- Import the Data country parameters dataset and using joins, add WHO Region and Income group from the sheet 8 of the dataset.

Your starting 15 rows of the result should look like:

	Country	AgerageOverWeightProportion	AgerageStuntedProportion	AgerageUnderWeightProportion	AgerageWastedProportion	WHO_REGION	World Bank Income group
1	Afghanistan	5.5000000	51.133333	34.266667	12.1000000	Eastern Mediterranean	Low income
2	Albania	20.7800000	24.160000	7.7000000	7.7600000	Europe	Upper middle income
3	Algeria	12.8333333	20.100000	7.2833333	6.2500000	Africa	Upper middle income
4	Angola	2.5000000	49.650000	28.0000000	6.7500000	Africa	Lower middle income
5	Argentina	11.5000000	10.733333	2.9000000	2.3333333	Americas	High income
6	Armenia	13.6200000	16.120000	3.4800000	3.9600000	Europe	Upper middle income
7	Australia	7.8000000	1.000000	0.1000000	0.0000000	Western Pacific	High income
8	Azerbaijan	9.1833333	21.816667	8.0833333	5.4333333	Europe	Upper middle income
9	Bahrain	7.5000000	13.900000	6.3000000	6.8000000	Eastern Mediterranean	High income
10	Bangladesh	0.9043478	55.243478	47.3565217	14.8521739	South-East Asia	Lower middle income
11	Barbados	12.2000000	7.700000	3.5000000	6.8000000	Americas	High income
12	Belarus	9.7000000	4.500000	1.3000000	2.2000000	Europe	Upper middle income
13	Belize	9.6000000	18.766667	5.2666667	2.3333333	Americas	Upper middle income
14	Benin	4.0600000	36.980000	20.1600000	8.0200000	Africa	Low income
15	Bhutan	4.8500000	44.250000	17.8000000	4.5750000	South-East Asia	Lower middle income

2. Using the datasets provided, plot the graphs as shown below with titles and labels. (15 points)

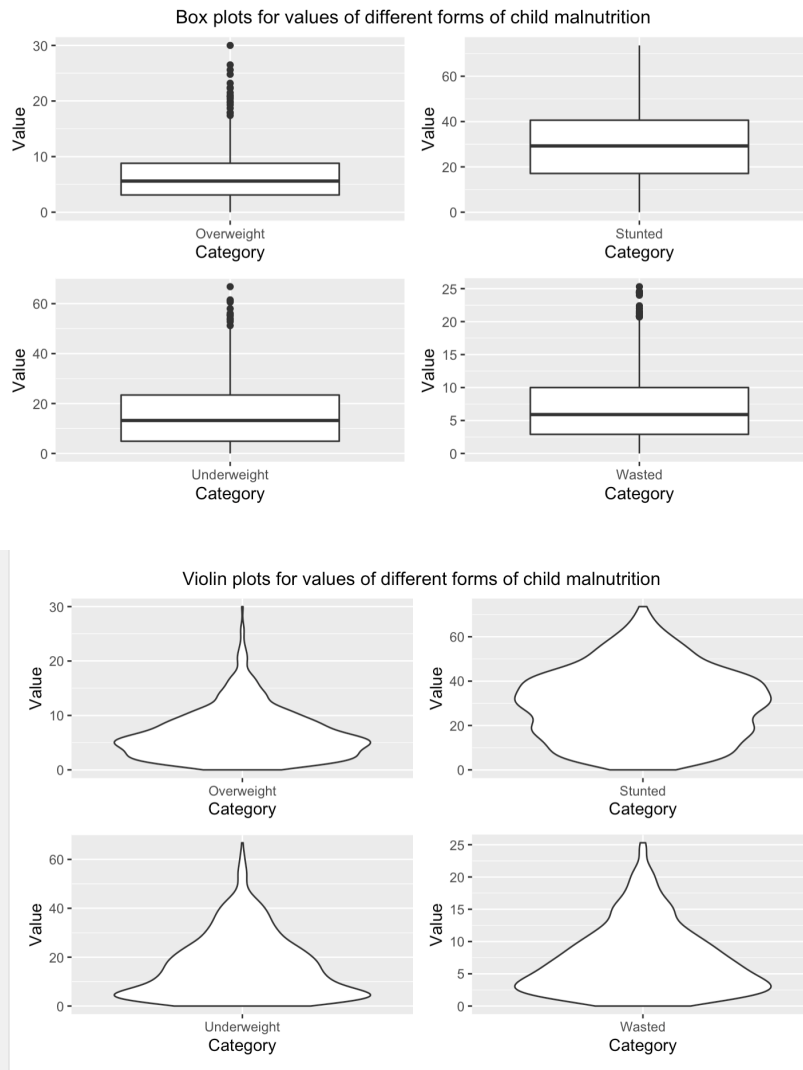


3. Using the datasets, fill in the correlation values between different forms of malnutrition in the table. (15 points)

	Overweight	Stunted	Underweight	Wasted
Overweight				
Stunted				
Underweight				
Wasted				

What does positive, negative and zero values of correlation between 2 variables imply?

4. Using the datasets, plot the following Violin and Box plots. (20 points)



Write 2 or more points about the inferences you get from these plots.

The directory of farmers market across the US is given in the file named “fm.csv”. Answer the following questions from the dataset.

5. Write a code to compute the number of farmers markets by state and arrange them in descending order of number of farmers market. (10)
6. Write a code to compute the number of farmers market by cities in Massachusetts and display top five cities. (10)
7. Generate the following table using pivot function. First column should contain the states. Second column should have Payment System. For Payment System consider the columns, “Credit”, “WIC”, “WICcash”, and “SNAP” from the original farmers market data. Third column should have the number of farmers market offering the payment services. (15)

States	Payment System	#Farmers Market
New York	Credit	10
New York	WIC	20
New York	WICcash	30
New York	SNAP	40

The above table is only for explaining the problem. Students need to generate the entire long form table.

Submission Format

1. Submit .RMD file via Canvas
2. The file should be named as follows, firstname_homework1.R
3. The code should follow tidyverse style guide (<https://style.tidyverse.org/index.html>)
4. The tidyverse style guide has style standards for naming objects, indentation and how to write long lines of codes to name a few