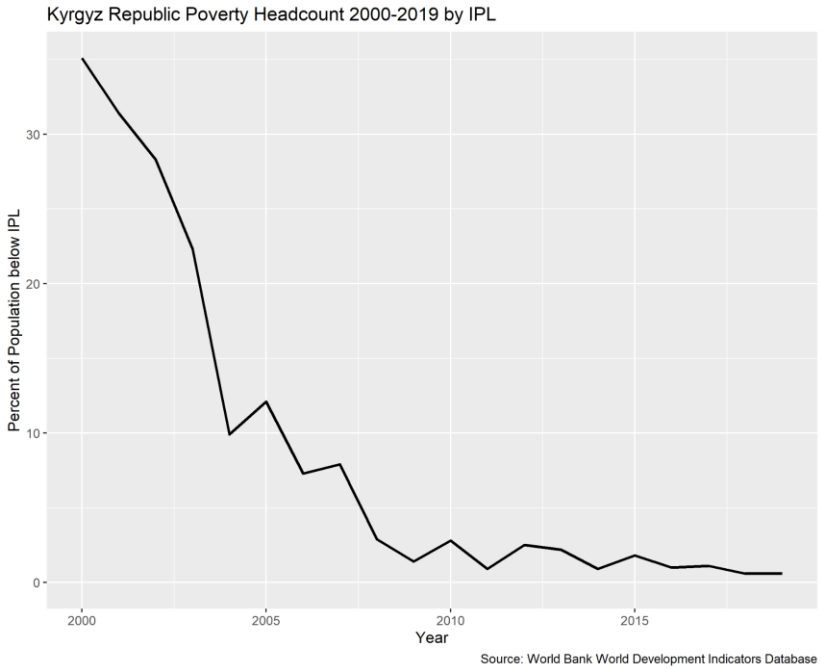
Data Module Answer Keys for Exercise #1 and #2 of Poverty Repository

Teacher Answer Key: Answers **Bold** and *Italicized*

Exercise #1: The Line of Best Fit

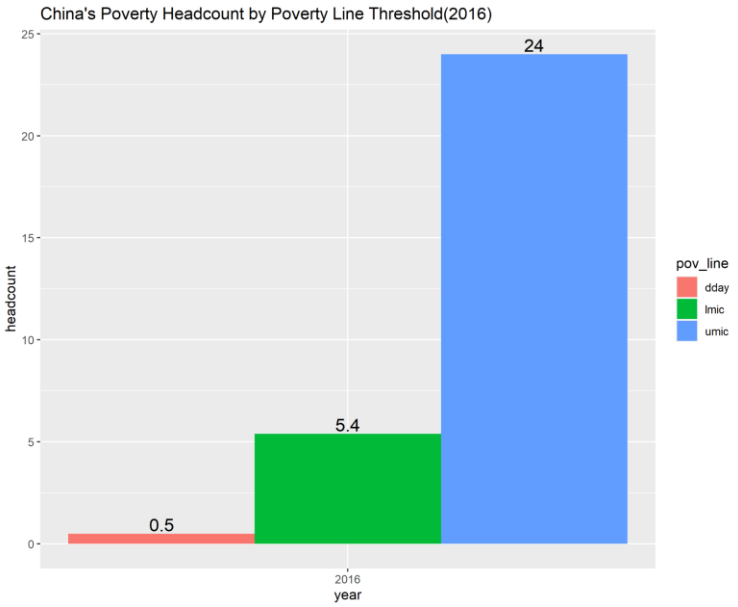
a. For this exercise you will be using a pre-tidied dataset to visualize the poverty headcount ratios of different income-bracket countries using multiple Poverty Lines. To access the dataset go to this link https://github.com/thomasboswellstudent/Poverty.git and click the green "code" button and "download zip." Open the zip file and go to the data folder to access "pov_hc_countries.csv".
    i. ***Answer: Students should be able to locate the file using the instructions above, no answer for subquestion a).***
b. In the Dataset, how many observations do you see? How many variables does the dataset have? How many NAs appear in the dataset? What do the NAs tell us about the poverty headcount data?
    i. ***Answer: The Dataset has 19350 observations***
    ii. ***Answer: The dataset has 5 variables***
    iii. ***Answer: The dataset has 14443 NAs***
    iv. ***Answer: A majority of the data is missing***
c. Imagine the Kyrgyz government is preparing to make a statement on the progress they have made in addressing poverty since 1990. Look over the data for the Kyrgz Republic from 1990 (hint: use the filter function) to the most recent dataset. How many observations do you see? Are there any years with unavailable data?
    i. ***Answer: 90 observations for the Kyrgyz republic***
    ii. ***Answer: 1990-1999 have no available data***
d. Think about the 2008 global financial crisis. Do we have data for the poverty headcount in the year this event occurred for the Kyrgyz republic? Why would this data be important to have?
    i. ***Answer: Yes***
    ii. ***Answer: Something along the lines of wanting to be able to assess the impacts of the financial crisis on the Kyrgyz Republic***
e. Create a visualization to demonstrate the progress the Kyrgyz Republic has made. Do this using dday measure which represents the World Banks international poverty line of $1.90. What visualization did you use? Why did you choose that visualization over another?
    i. ***Answer: They can create a trendline graph or possible a bar plot. The idea is for them to present a trajectory and consider what that tells us about poverty in the Kyrgyz republic***

ii. ***They should be able to explain that this visualization allows them to show a trend from the past to the most recent data.***



f. Now consider the application of the DDAY poverty line and the LMIC and UMIC lines, when would it be appropriate to use the DDAY over the LMIC, why?
   i. ***Answer: The student should be able to identify the different poverty line thresholds and convey the fact the the LMIC and UMIC lines are meant for countries in the Upper middle income bracket and lower middle income bracket.***
g. Choose another country in the dataset, pick one from each income bracket (low income, lower-middle income, and upper-middle income) designated by the World Bank (you might have to go online to look this data up). Choose a visualization to show the

difference in poverty headcounts based on the DDAY and the line that is appropriate for their country's income level.



i. ***Answer: Here they should be able to choose three countries and depict either a Bar plot or a trendline graph to show the differences in poverty headcount ratio.***

ii. ***Ex viz:***

Exercise #2: Manage My Dataset

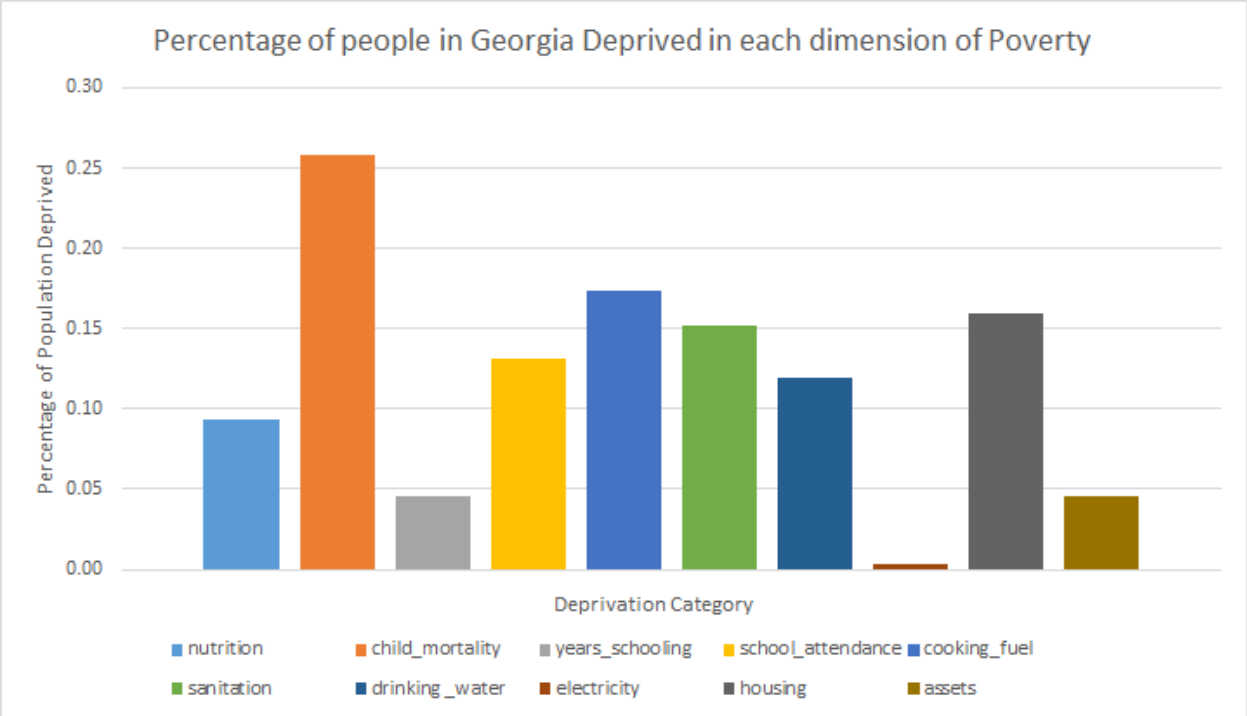
Objective: In this exercise you will work on locating, analyzing, cleaning, and presenting your findings using the Multi-Dimensional Poverty Index dataset. The goal is to understand where data comes from in its raw form and what simple practices are available to create a dataset that is usable and intuitive for other users in your network. All parts of this exercise can be accomplished in excel but you are invited to incorporate other softwares like R or Python to complete the assignment.

Scenario: Your superior is producing a report for an international poverty reduction non-profit that uses data to determine where they should invest in new programs to alleviate multi-dimensional poverty. She has asked you to assist her in the data wrangling and cleaning process.

a. Open the mpi_table_raw.xlsx file found in the For Students folder of this github repository -> https://github.com/thomasboswellstudent/Poverty.git
   i. ***Answer: Student should be able to locate the file in the appropriate poverty repository***
b. Your superior needs you to tidy the dataset so that she can easily import the data into different softwares. She asks that you remove any excess information so that the first row only contains variable names.
   i. Open the mpi_raw and create a copy of the file with the file name mpi_tidy.
      1. ***Answer: Student must follow the directions***
   ii. In the mpi_tidy dataset, create a new sheet and name it codebook.
      1. ***Answer: Student must follow the directions***
   iii. Identify the information in the first several rows. If it's explanatory and not related to the column titles, move this information to the codebook sheet. How did you store this data in the codebook?

1. ***Answer: The student must create a codebook and move metadate information to the Codebook as a way to tidy the data***

c. Look at the column headers. These headers should now populate the first row of your dataset with names like “country” etc.
    i. Rename these variables so that the variables are lowercase. How can you use your codebook to store information on what each of the variable names stands for in your tidied dataset? Explain why you chose to name the variables the way you did. Why would this tidying process be useful for your superior?
        1. ***Answer: The student can choose whatever naming convention they want to name the variables in row 9 so that it is a tidy dataset. However, they should choose variable names that are lowercase, intuitive, and document the variable names in the codebook on the following page***
d. After reviewing the tidy dataset, your superior asks you to identify the MPI headcounts of certain countries in 2020
    i. Which country had the highest MPI headcount in 2020 and what is it?
        1. ***Answer: Madagascar with 69.08% of of the country living in Multidimensional Poverty***
    ii. Which country had the lowest MPI headcount in 2020 and what is it?
        1. ***Answer: Georgia has the lowest MPI headcount ratio in 2020 with 0.34% of the country living in Multidimensional Poverty***
    iii. Create a visualization using the two countries to determine what dimension of poverty these two countries suffer from the most in 2020
        1. ***Answer: The student should be able to create two simple bar plots that show the percent of the population in Georgia that lives in each dimension of poverty. They can do the same for Madgascar. Viz example:***



e. Your superior tells you that the non-profit she is preparing the report for is interested in investing in poverty alleviation projects in Afghanistan, Bangladesh, Zimbabwe. They would like to know which dimension of poverty to target to make the most impact on the MPI.
   i. Visualize the 10 dimensions of poverty for each country.
      1. ***Answer: The student will create visualizations for these three countries similar to the visualizations provided above***
   ii. Which aspect of poverty should they target? Explain your reasoning? Is there any unavailable data in these countries? How might this affect your analysis and the non-profit's decision on where to implement their projects?
      1. ***Answer: Students should provide an answer that considers how the MPI is calculated. If their aim is to reduce the MPI, they can try to game to index by choosing the one that is weighted the most. However, if they are just looking to consider the area where the most people are deprived,***

then they have more options to choose from. These are the considerations we would like them to make here.

f. Consider the two exercises you've conducted. One uses the Poverty Headcount Ratio from a unidimensional monetary measure. The data in this exercise uses the MPI and calculates a poverty headcount ratio based on the MPI itself and the different dimensions of poverty. What are the benefits to using one of these measures over the other? If you used a poverty headcount ratio using the unidimensional measure, what would you policy prescription be to alleviate poverty? How would your reasoning change if you were using a multi-dimensional measure?

   i. ***Answer: Students here should identify the tradeoffs between unidimensional and multidimensional measures. They might note that there is more data in the unidimensional measure dataset so it can be used to help more countries. They may note that unidimensional measures lead to unidimensional policy solutions like money where multidimensional measures offer opportunities to target certain segments of the population.***