Capstone project. Research of correlation between inequality and poverty levels.

For my Capstone project I used data from World Bank that had around 1,000 of different economic development measurements for each country of the world.

1. ***What kind of cleaning steps did you perform?***
   Out of huge array of data I selected five inequality measurements (rows):
   - annualized average growth rate in per capita survey mean consumption or income, bottom 40% of population;
   - income share held by lowest 20%;
   - income share held by lowest 10%;
   - poverty headcount ratio at national poverty lines;
   - survey mean consumption or income per capita, total population

   I also selected one poverty measurement:
   - Poverty gap at national poverty levels.

I turned those rows into columns. Each of those columns had sub-columns by year - 2014, 2015, 2016, 2017, 2017, 2018 and 2019.

After exploring the data I realized there was no data for regions. So I deleted all regions and left separate countries only. There was not much data for 2018 and 2019 years as well, so I had to delete those rows.

I found average between years 2014 - 2017 and used that data to fill in columns of all measurements described above.

2. ***How did you deal with missing values, if any?***

Most countries had data for at least one year between 2014 and 2017. I found average between years 2014 - 2017 and used that data to fill in columns of all measurements described above. For countries with no data I used *None.*

3. ***Were there outliers, and how did you handle them?***

There were outliers, but I plan to explore them during my next stage that I plan to submit after completing Unit 7 of the curriculum (Data Storytelling).