Data Documentation

* More in-depth documentation found on our website
  + <https://stm7631.github.io/AA-08-Washington-Data-Wizards/doc.html>
* Shape Data from ESPEN Website or from <https://data.humdata.org/organization/ocha-nigeria>
* Drug data from Nigeria JRSM 2018, downloaded from the ESPEN Data Portal
  + All LF Data (for the drug ALB and IVM) was disregarded because it wasn’t complete
  + PZQ Data was pulled from the PZQ sheet
    - Total Drugs Procured comes from PZQ -> Tablets to be procured
    - Total People Requiring Treatment comes from Population Requiring Treatment with PZQ -> Totals
  + IVM Data was pulled from the IVM Sheet
    - Total people requiring treatment comes from Target Population -> Oncho -> Total
    - Total drugs procured comes from IVM -> Oncho only
      * It also comes from IVM -> LF+Oncho, only if the corresponding row in the IVM -> Oncho only column is 0.
        + This is due to the fact that the Oncho Only and LF+Oncho columns are the same when they are both filled in.
  + MBD Data was pulled from the ALB\_MBD Sheet
    - Total people requiring treatment comes from Target Population -> STH(MBD) -> Total
    - Total drugs procured comes from MBD -> Tablets to be Procured
* Disease MDA Data is from ESPEN Portal
  + Link TBD (ESPEN is down at the time of writing)
* GBD Data from IHME 2017
  + YLD Rate pulled from the GBD tool, for each disease, and for all genders, and ages.
  + <http://ghdx.healthdata.org/gbd-results-tool>
* Nigeria State Population Data
  + Summed up from district population field in JRSM
* Additional Data Sources for the future
  + Disease by country CSV from ESPEN
  + <http://espen.afro.who.int/countries/nigeria> (replace Nigeria with country name)
    - Navigate to the disease you are looking at, and then the data option.
* Formulas
  + In the data, when we talk about **normalized variables**, it means that we just divided a value in a column by the mean of the column. We do this to put things in terms of scale - how much bigger is a state’s value compared to the mean.
  + When we talk about **standardized variables**, we are converting them to a measure of standard deviation. What this tells us is how far out from the mean a value is. If a value is negative, that means it is under the mean. If it is positive, it is above the mean.
  + **Drug Ratio**: The sum of all the drugs procured in each area divided by the sum of all the people requiring treatment in each area, as shown in the JRSM.
  + **Standardized Drug Procurement vs Treated:** The number of drugs procured in each area (from JRSM) divided by the number of people treated in the area (from IU Data). This value is then standardized.
  + **Drug Procurement vs Treated:** The number of drugs procured in each area (from JRSM) divided by the number of people treated in the area (from IU Data).
  + **Disease MDA Coverage:** The number of people treated for a disease (from IU Data) divided by the total number of people who require treatment (from IU Data)
  + **Disease YLD Value:** The population of an area (from the JRSM) divided by the YLD rate for the disease (calculated in the IHME Data set)
  + **Disease Measure:** Normalized Drug Ratio in each area divided (calculated) by Normalized YLD value (calculated) for the disease in each area