**2.2 HW – Working with Data**

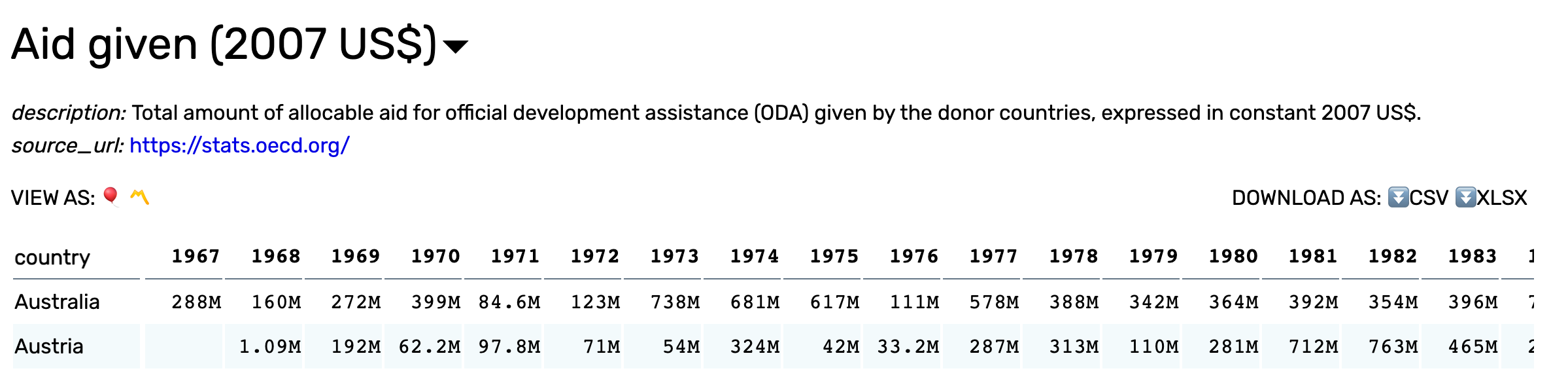
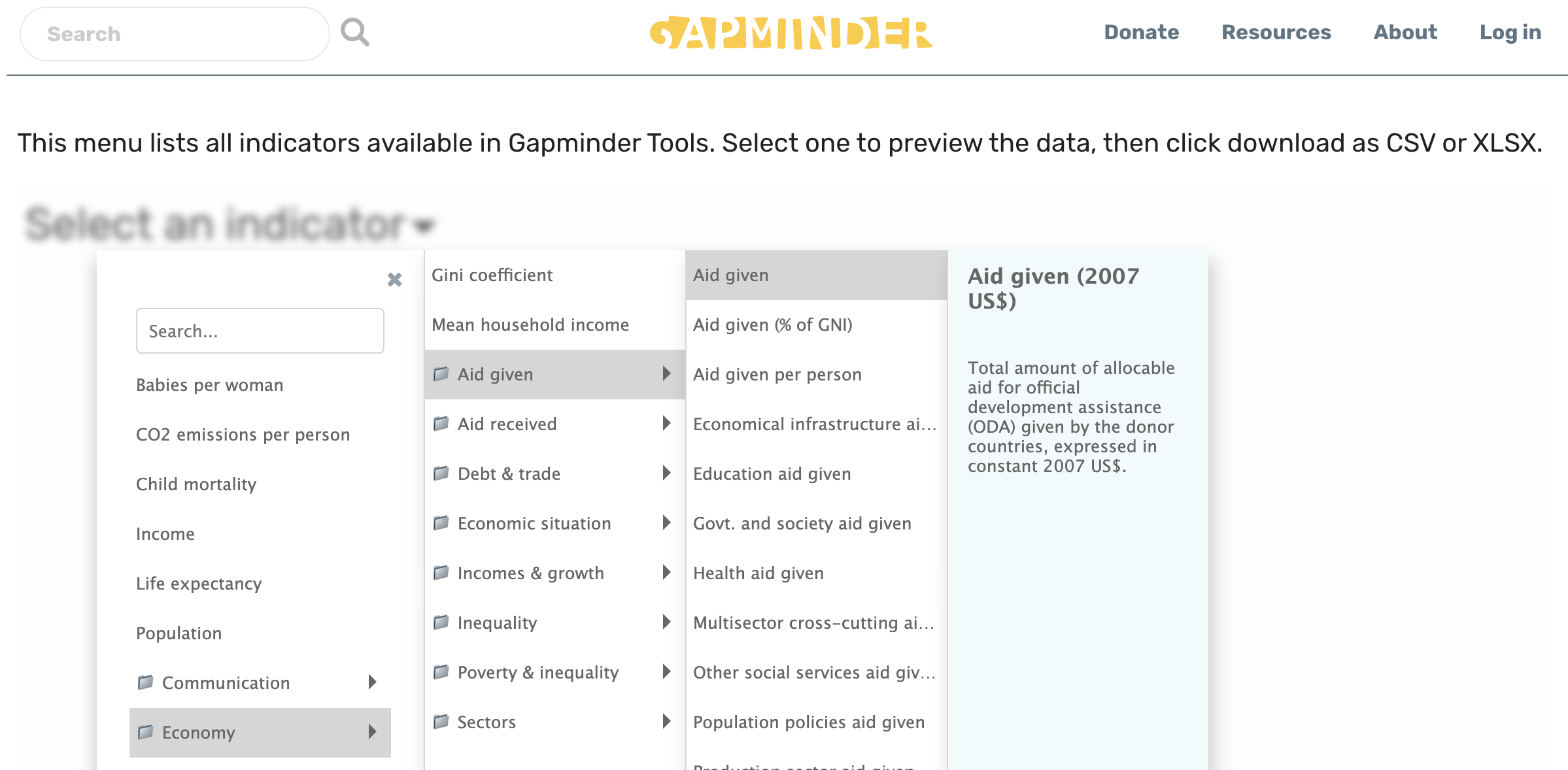
Submission: Completed Tableau workbook via Canvas.

Overview

* This problem is an application of the methods from the notes ‘2.2 – Working with Data’, where we learned how to join multiple datasets in Table Prep Builder and create animated visualizations.
* **It is designed to be open-ended, allowing you to explore different relationships and plot types.**

Data source

* The Gapminder data that Hans Rosling’s visualization is built from a dataset that that combines data from multiple sources into unique coherent time-series about numerous health and socio-economic variables.
  + There are lots of other variables with data setup in a similar fashion available for download, which we will explore in this assignment.
  + Our goal is to find other factors that could have a potentially significant relationship with life expectancy over time.
* Navigate to the [Gapminder site](https://www.gapminder.org/data/) and find TWO other variables to join with the already merged regions, population and life expectancy data accompanying this assignment.
  + Once on the website, scroll down to “Select an Indicator”, select a context, then download the csv file. For example:



* **DATA QUALITY NOTE: We must choose variables that DO NOT HAVE values formatted like the red box above. These are in the data as text and therefore will not ever get treated as numbers.**
* **So think about the scale of the variable and be sure to preview your data to make sure everything is formatted as a number in the .csv download.**

Assignment

* The goal is to work with data from multiple sources in Tableau Prep Builder and practice creating plots to glean trends / relationships in data in an attempt to tell a story.

1. Use Table Prep Builder to join data.
   * Before visualizing the data, we need to join the data just as we did in the video notes.
   * Create a flow where you perform the following steps as necessary:
     + Connect to data sources (csv files).
     + Rename fields (ensure header are correctly setup).
     + Pivot and rename resulting pivoted fields.
     + Add a clean step to convert Year to number (whole) so that it joins correctly.
     + LEFT join new pivoted data to the existing gapminder.csv data (this way if the variable you chose doesn’t have much historical data we can still keep all the records from the original Gapminder and just have nulls for non-matched country-year combos).
     + Setup Applied Join Clauses.
     + Remove duplicated columns.
     + Output final data.
   * A single merged dataset saved as a .hyper file with no duplicated columns is necessary before moving onto the next part.
2. Recreate Hans Rosling’s visualization using one of your two variables instead of average income per person.

* Open a new Tableau workbook and load the .hyper file created from Tableau Prep Builder. Click on data source and covert to extract. Then save the workbook as a .twbx file so that the data source is included when you submit the workbook on Canvas.
* This is just recreating the steps from the notes to see if your variable has a significant relationship with life expectancy throughout time.
* Add reference lines with formatting for the X and Y variables like we did in the notes.

1. Create at least TWO more polished plots visualizing the relationships between any of the variables.

* You may not (/ should not) use all of the plots that you create when exploring.
* Polished plots means that there are neat / informative titles and axis labels, well-chosen colors, visuals that make statistical sense (i.e. using the plots correctly based on the data types and context). Can you make a story from the plots and context of the data?
* The plots you create can be as zoomed in on specific subsets of the data or very macro pictures like visual in part (a).
  + Examples include: line plot of foreign aid given for countries in North America from 2010-18, boxplots of country population by region over time, etc.
  + Be creative, find a trend!

1. Create a dashboard combining your polished plots with a Big Idea.

* Organize your polished plots into a neat dashboard, where the title and/or subtitle is a nicely formatted sentence containing a “Big Idea” of your plots.
* Recall that the Big Idea from Storytelling with Data is the “so what” of your narrative that is boiled down to a single sentence. It has three components
  + It must 1) articulate your unique point of view 2) convey what’s at stake and 3) be a complete sentence.
* Apply these components to the context of the data and your visuals.

What will be looked at for grading

* Single joined dataset correctly saved in the .twbx file, correct recreation of the Hans Rosling visualization with a new variable, organization of the final dashboard, attempted narrative via Big Idea statement, quality of the polished plots.