The following handout presents visualization techniques for a single variable, categorical or discrete. While univariate visualizations are usually of little interest in reality, keeping things simple at one variable will help us get a feel for how Tableau thinks about different data types.

**Categorical data**

The first question we want to visualize in this handout is the following:

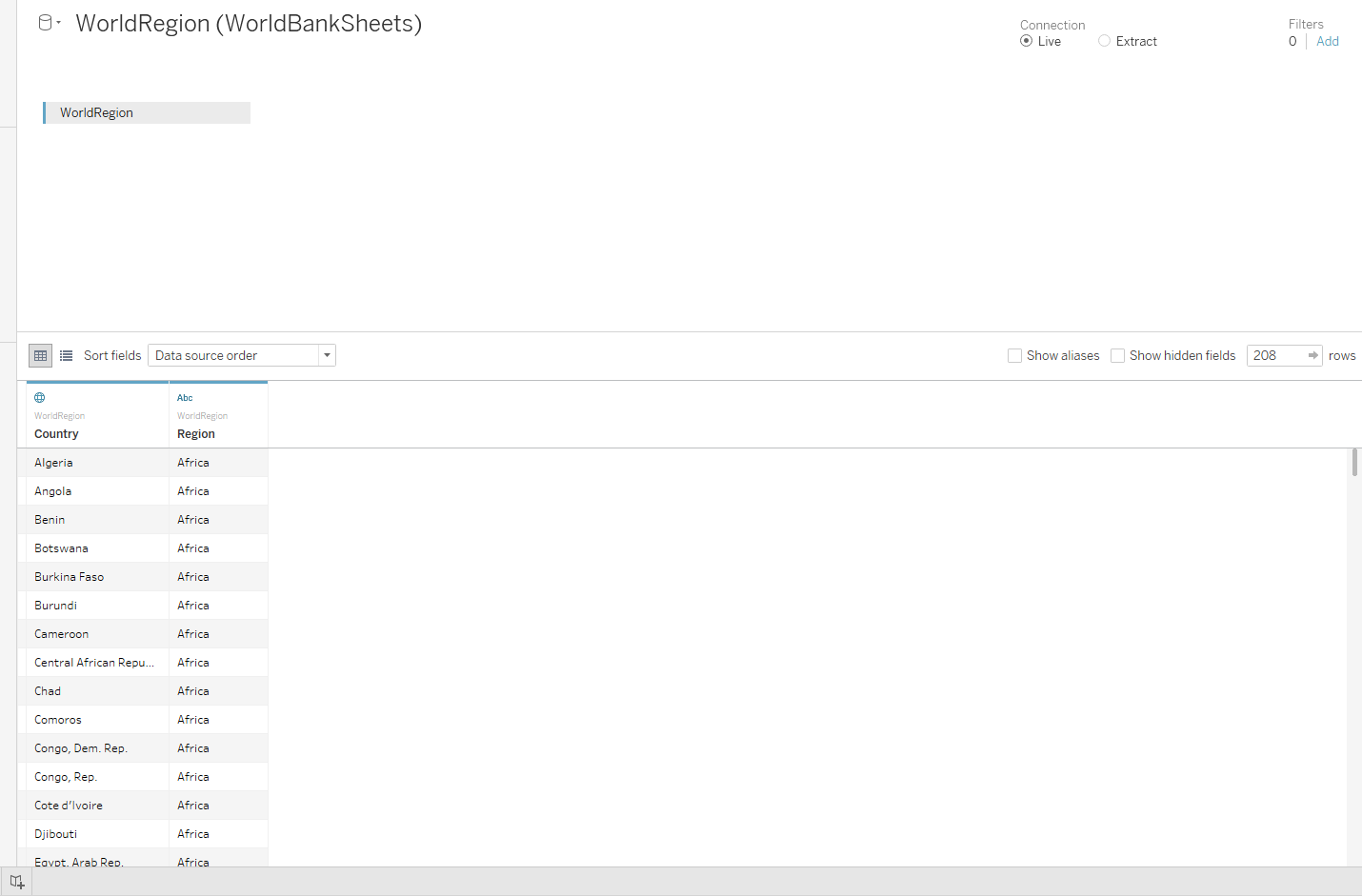
**How many countries are in each world region?**

We will answer this question using the “WorldRegion” sheet of the World Bank data set, which looks like this:

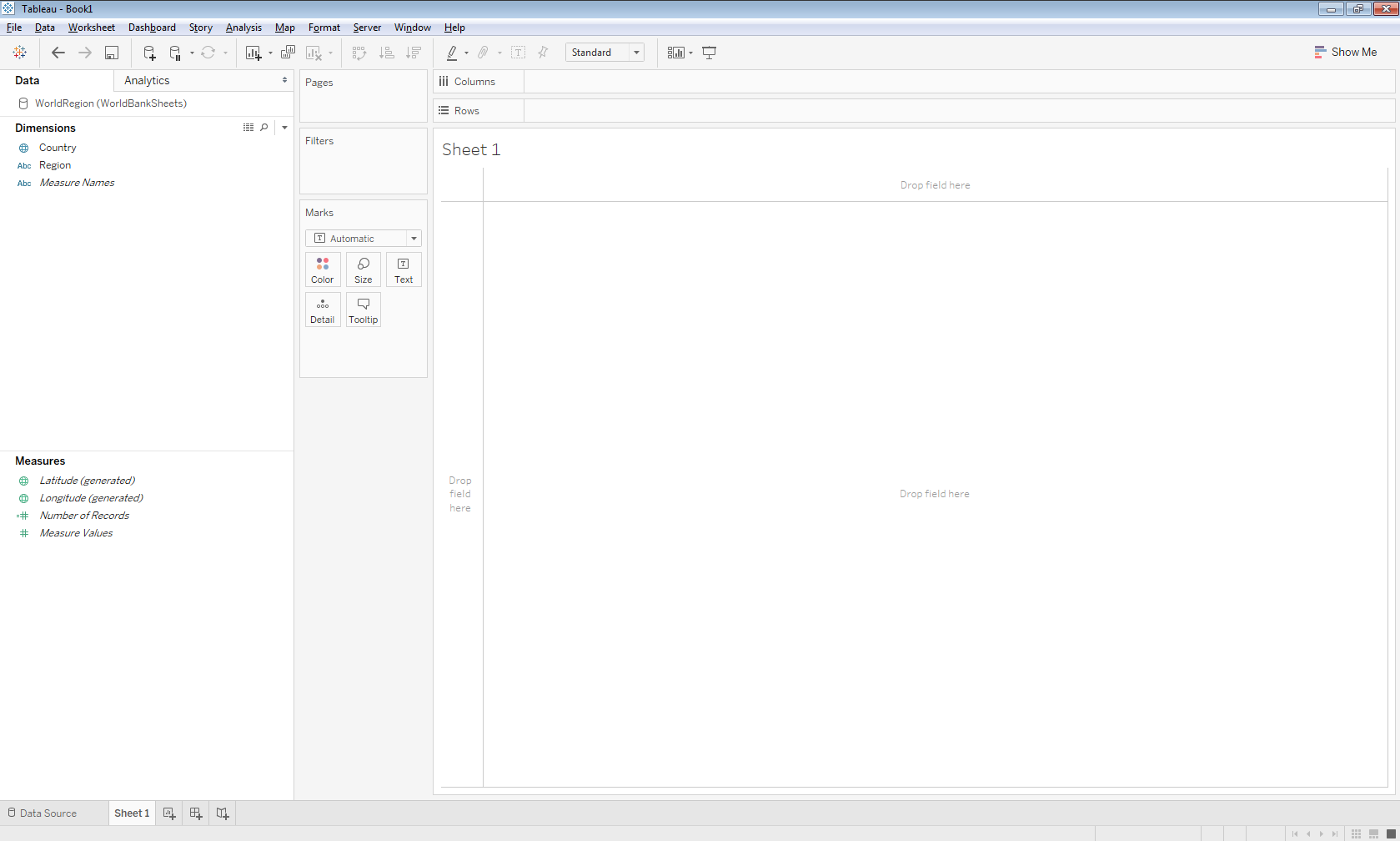


This is a **univariate** summary, because in essence it boils down to answer a question about a single field: **Region**. We don’t care about the actual data values in the **Country** column; only how many countries there are in each region.

* Connect to this data source:

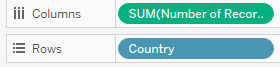


* Open a new worksheet. It should look like this:

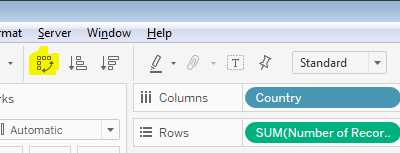


Now start visualizing! First, let’s get a feel for how Tableau operates here by carrying out the following tasks using the “**Country**” and “**Number of Records**” pills. The **Number of Records** pill is pretty critical for univariate visualizations. For the following tasks, the visualizations are not very informative, since there is only one record for each country; hence we will see a lot of 1’s!

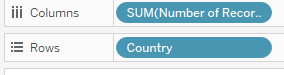
* Drag **Country** to **Rows**.
* Drag **Number of Records** to **Columns.** Note that it is automatically aggregated!



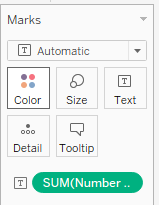
* Change the placement of the pills either by rearranging or clicking the “Swap” icon highlighted. What changes?



* Re-swap the pills to their original places. Right-click on the “Number of Records” and change it to “Discrete.” Note how the view changes. Why?



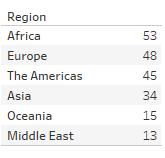
* Now move **Number of Records** off of columns, and instead drop it on the **Text** box on the **Marks** card. What results?



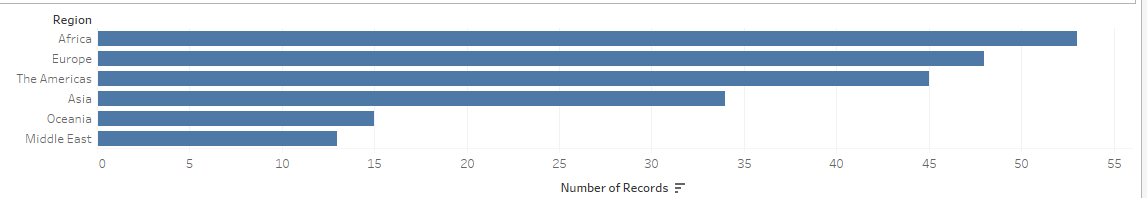
Now let’s return to the original question, using **Number of Records** and **Region** to visually summarize how many countries belong to each region. **Hint 1:** Look for these icons

 Hint 2: Recall that **aggregating** can be both **summing** and **averaging**. Try changing a measure aggregation from averaging to summing.

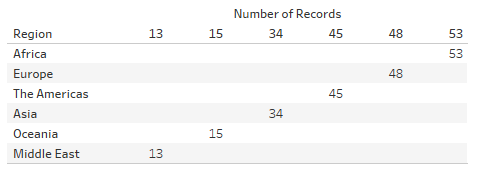
**Viz 1:**



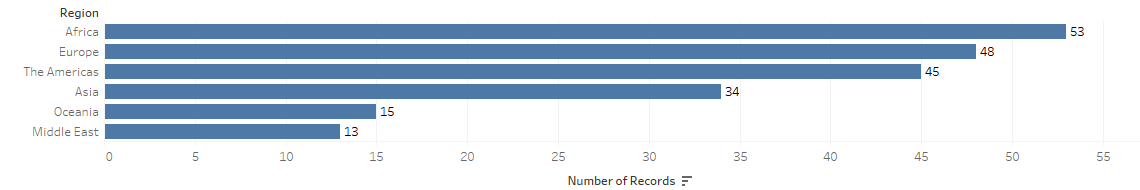
**Viz 2:**



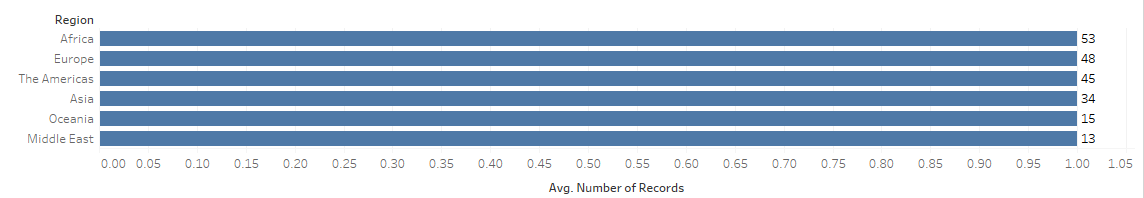
**Viz 3:**



**Viz 4:**

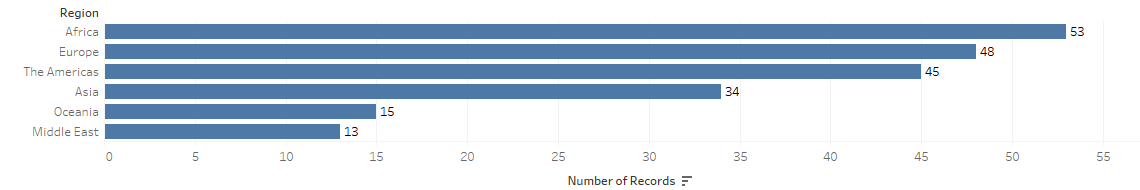


**Viz 5:**

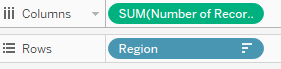


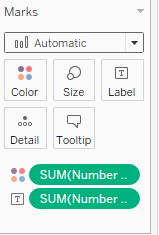
**Introducing color and size**

Let’s reconsider Viz 4. This is aka a **bar graph**, a common type of graph used for visualizing a single data field when the field is categorical in nature.

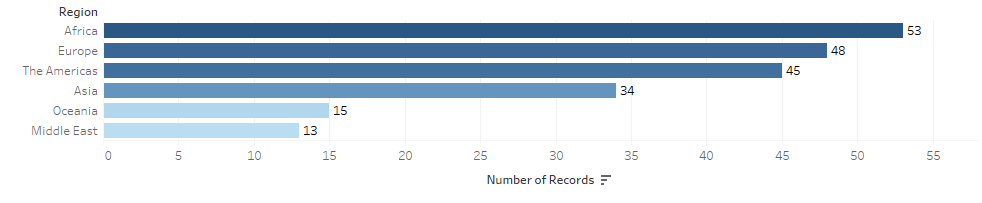


We can also encode the number of countries in each region using **color**. Arrange the pills as shown:

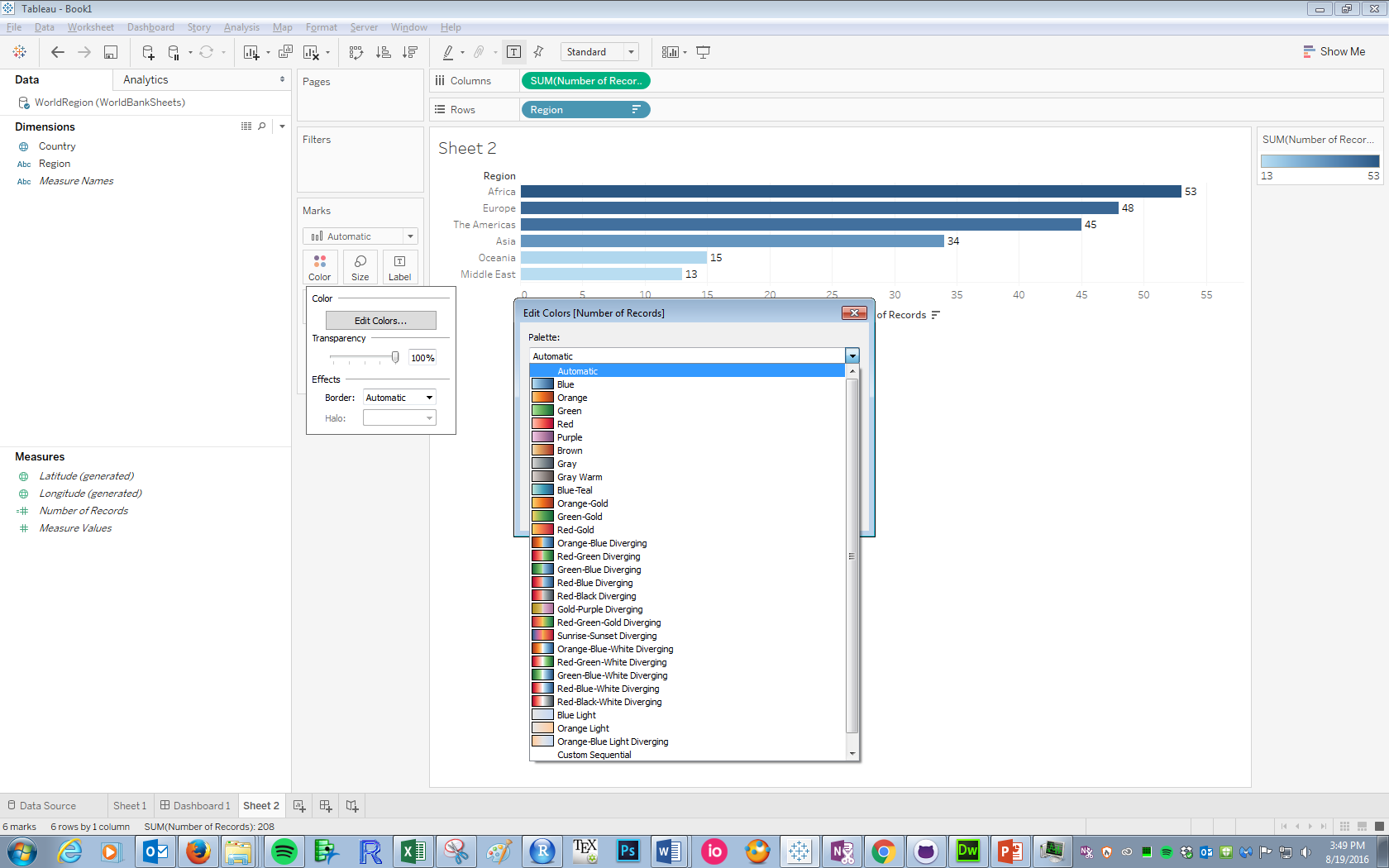




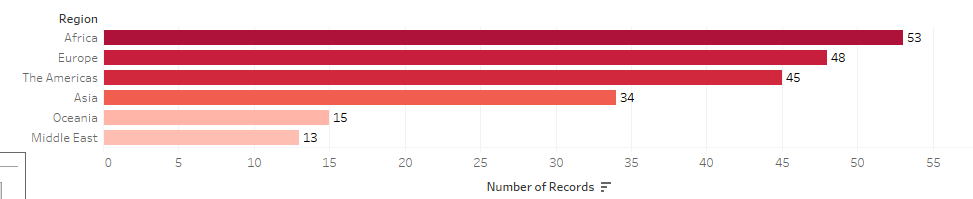
This will result in the following viz:

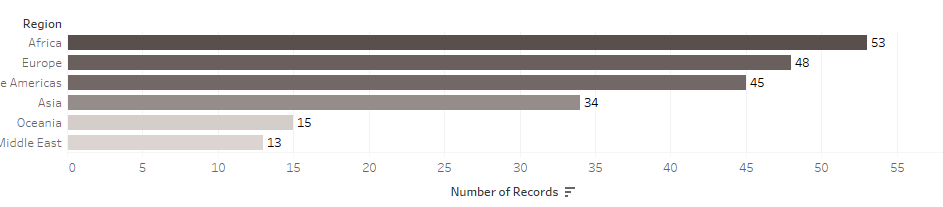


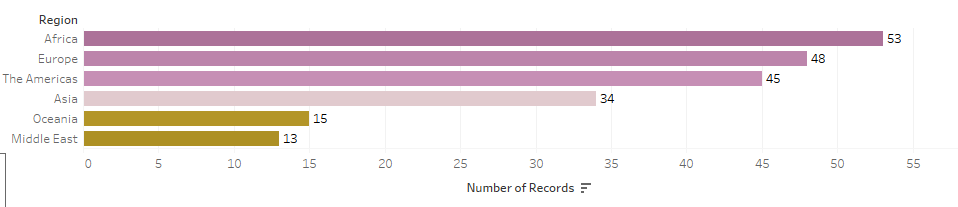
You can change the color palette by clicking on **Color** then **Edit Colors**:



Try creating the following:

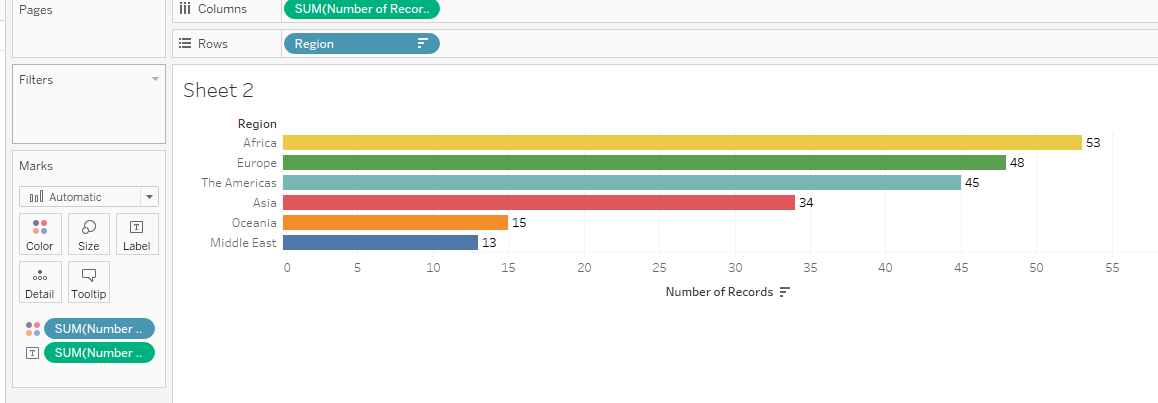




\

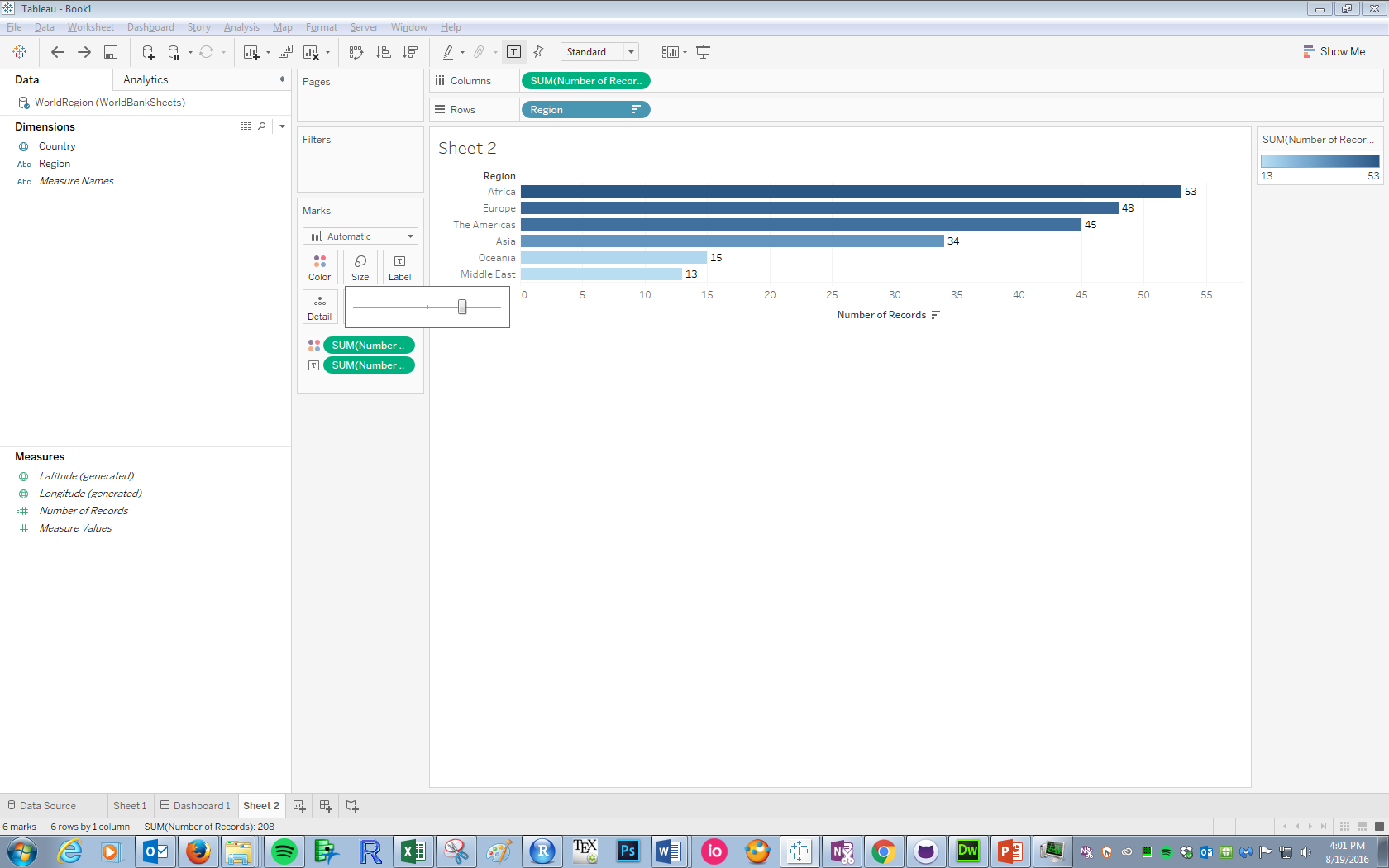
**Question: Is one of the above worse than the others?**

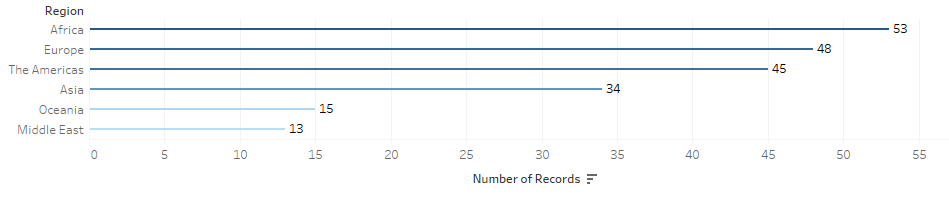
Now right-click on the **Number of Records** pill on **Color** and change it to discrete. You should see the following:



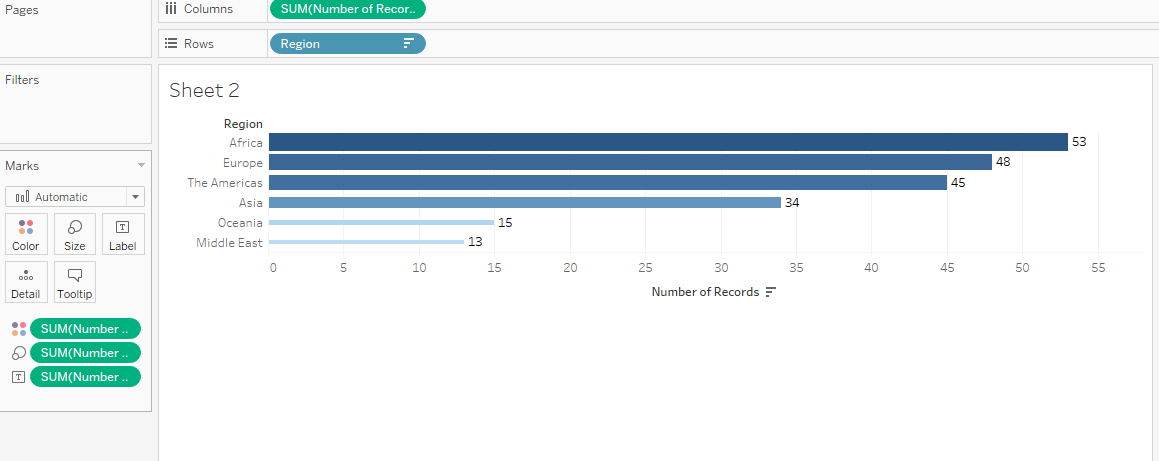
**QUESTION**: Are there any disadvantages of this approach?

Note that we can also encode number of countries with size. First, try clicking on the **Size** button and moving the slider:





Without a variable encoding size, all the bars change simultaneously. However, we can have the size of the bars be proportional to the size of the regions by dragging **Number of Records** to size, resulting in the following:

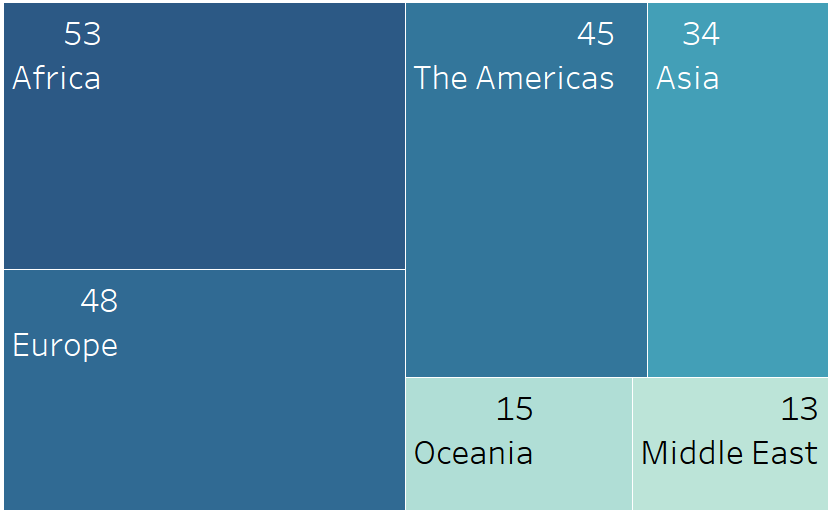


We now have the size of the region encoded by **Length**, **Color** (saturation), and **Size**. Whether or not to use so many redundant EPTs is kind of up to you as a designer!

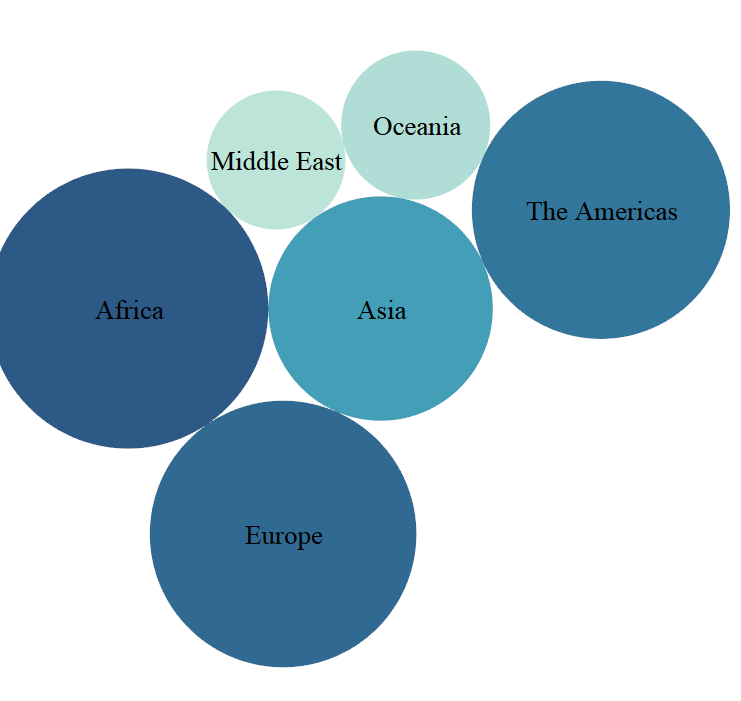
**Other univariate displays**

Although bar charts are the most common way (and, as we know, most effective way) of visualizing a single categorical data field, here are a couple other ways. See if you can recreate these visualizations by noting whether the visualization has AXES or HEADERS or NEITHER, what EPTs are being used, and how.

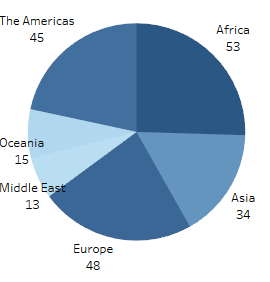
**Tree map:**



**Bubble Chart:**



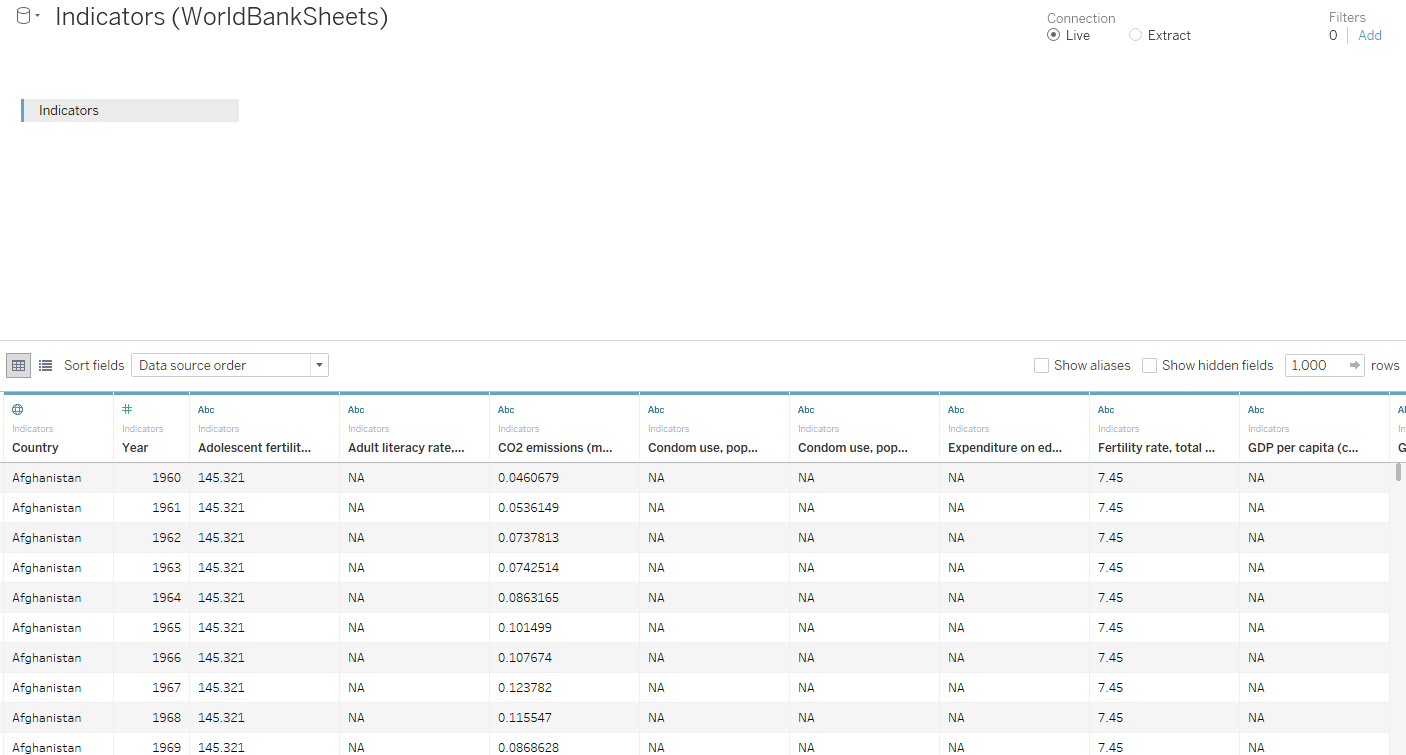
**Word cloud: Pie chart:**



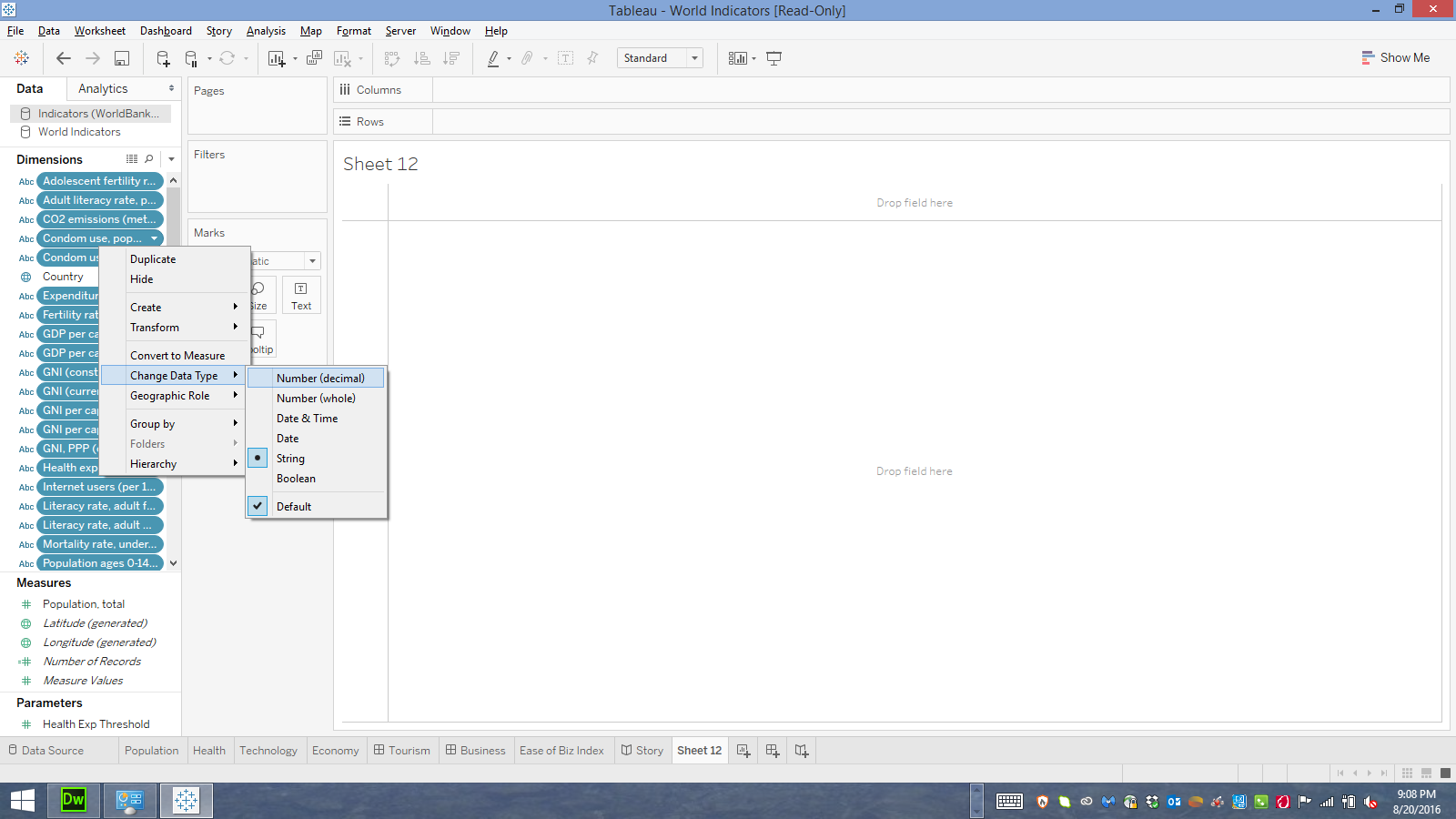
**Visualizing a single quantitative variable**

We just discussed visualization methods for a single categorical variable. What if the variable is quantitative? Let’s consider visualizing the 2014 life expectancies of countries in the World Bank Indicators sheet.

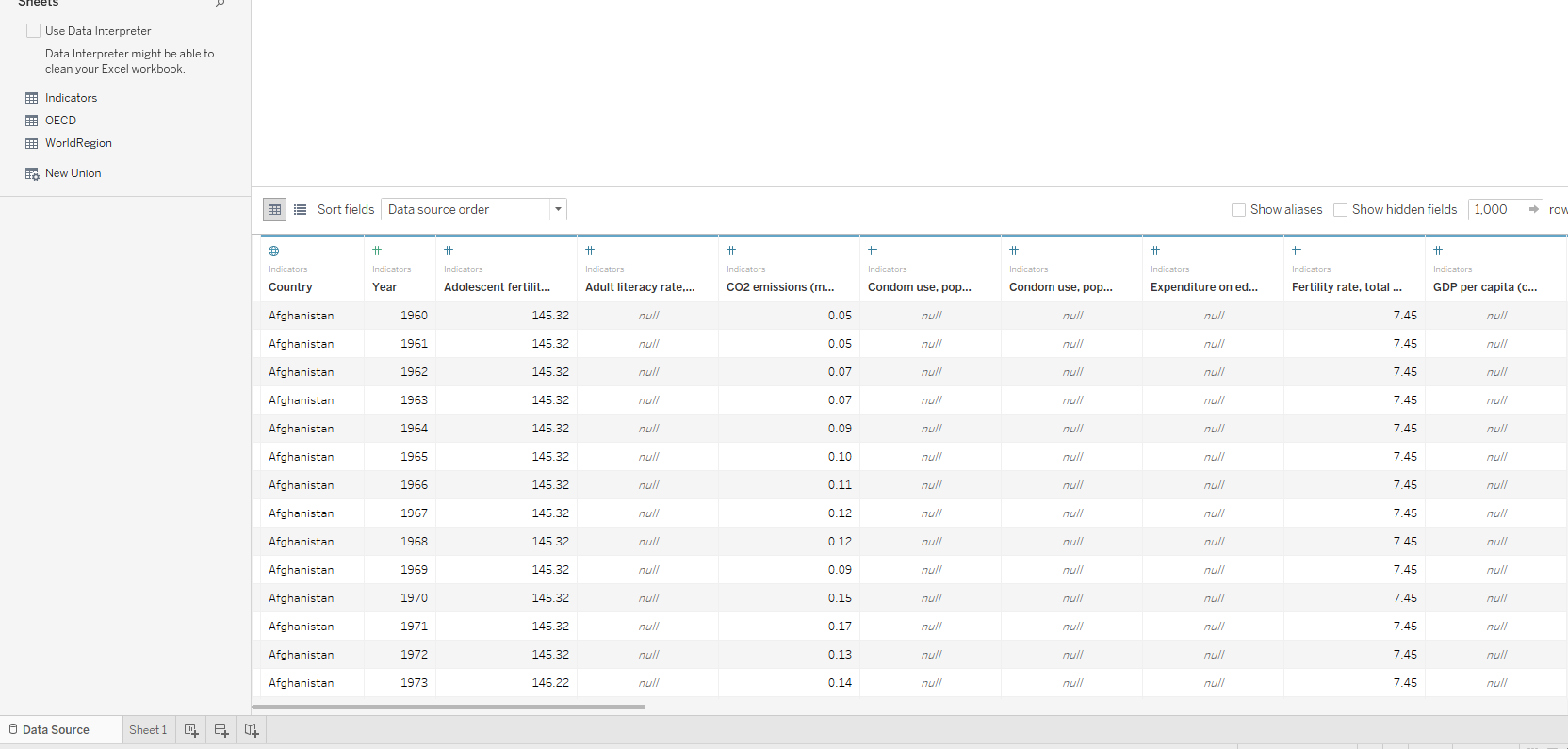
Open a new workbook. Connect to the WorldBankSheets excel file and drag Indicators to the connection panel:



First, note that many of the variables that should be quantitative are in fact currently thought of as strings by Tableau. Change them to numeric data by highlighting these fields, right-clicking, and changing the data type as shown:



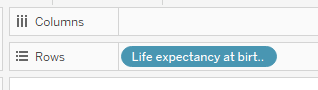
Note how this changes the missing values in the data view on the **Data Source** tab:



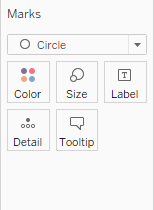
The most straightforward visualization would be to show each individual life expectancy as its own data point. Let’s create this:

**Tasks:**

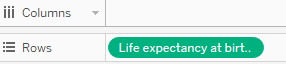
* Drag **Life Expectancy** to **Rows**. Note the view.



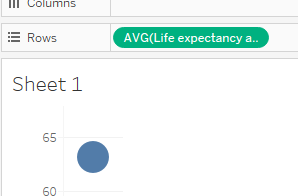
* Change the **Marks** to **Circle**. Again note the change in view.



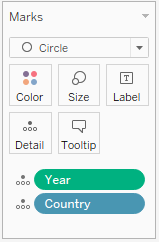
* **Life Expectancy** is currently modeled as **Discrete** as indicated by the blue pill, and as confirmed by seeing the life expectancies as headers. This isn’t what we want! Right-click on the pill and change it to **Continuous:**



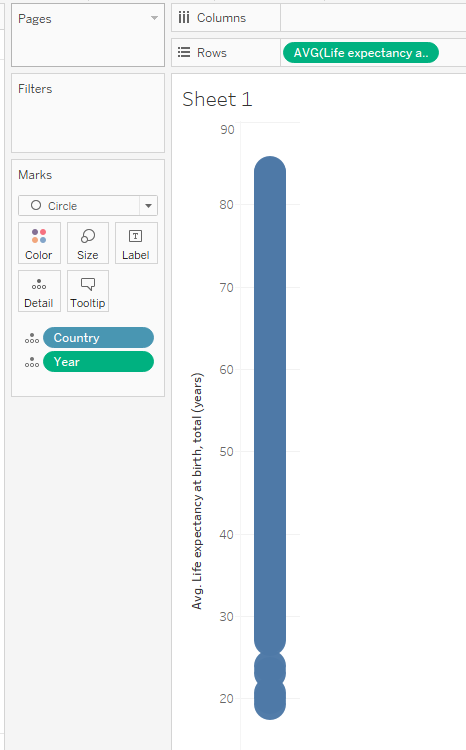
* Right-click again, and convert the field from a dimension to a measured (either **Sum** or **Avg**). Note the aggregation that now occurs: **do we want this**?



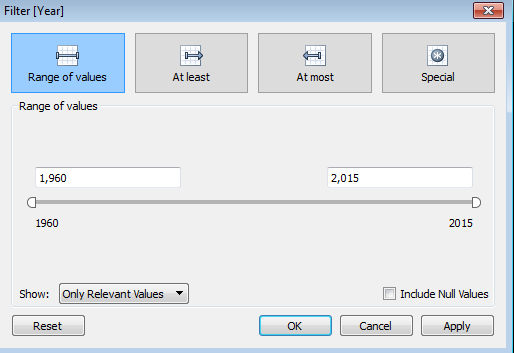
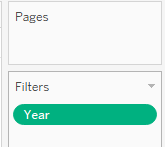
* Change the field back to a dimension. Note that if we hover over each circle, we see only the life expectancy. It might be nice to see some more information about each circle, for instance the country and year. Drag these two fields to **Detail,** and note the new information is automatically added to the tooltip.



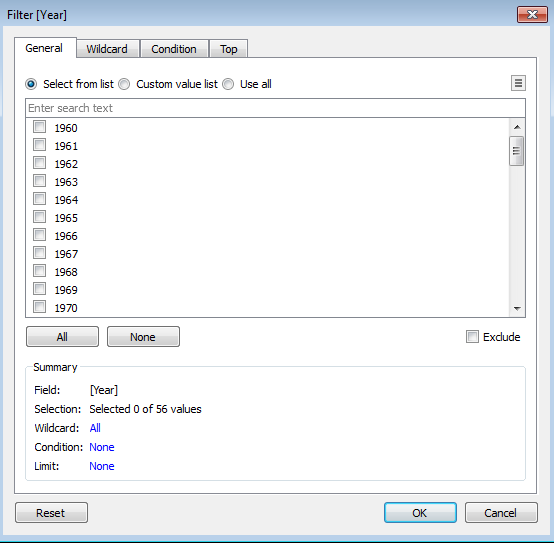
* Once again, convert the field from a dimension to a measure. **Why does the view not change as it did before?**



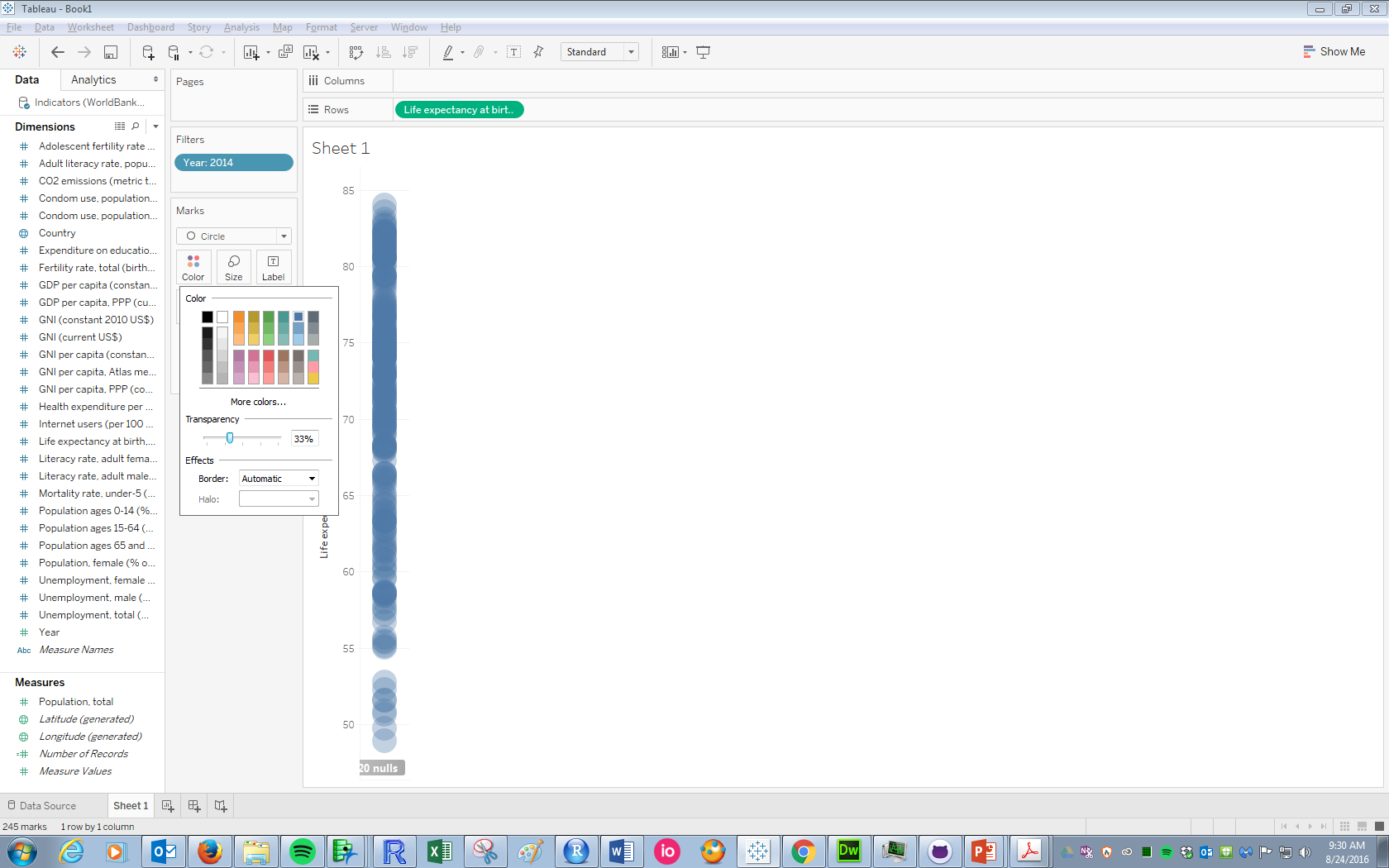
* Change the field back to a dimension. We don’t want to visualize the life expectancies from all years however, just 2014. This will be our first use of a **Filter**. Drag **Year** to **Filters.** This is the filter view you see:



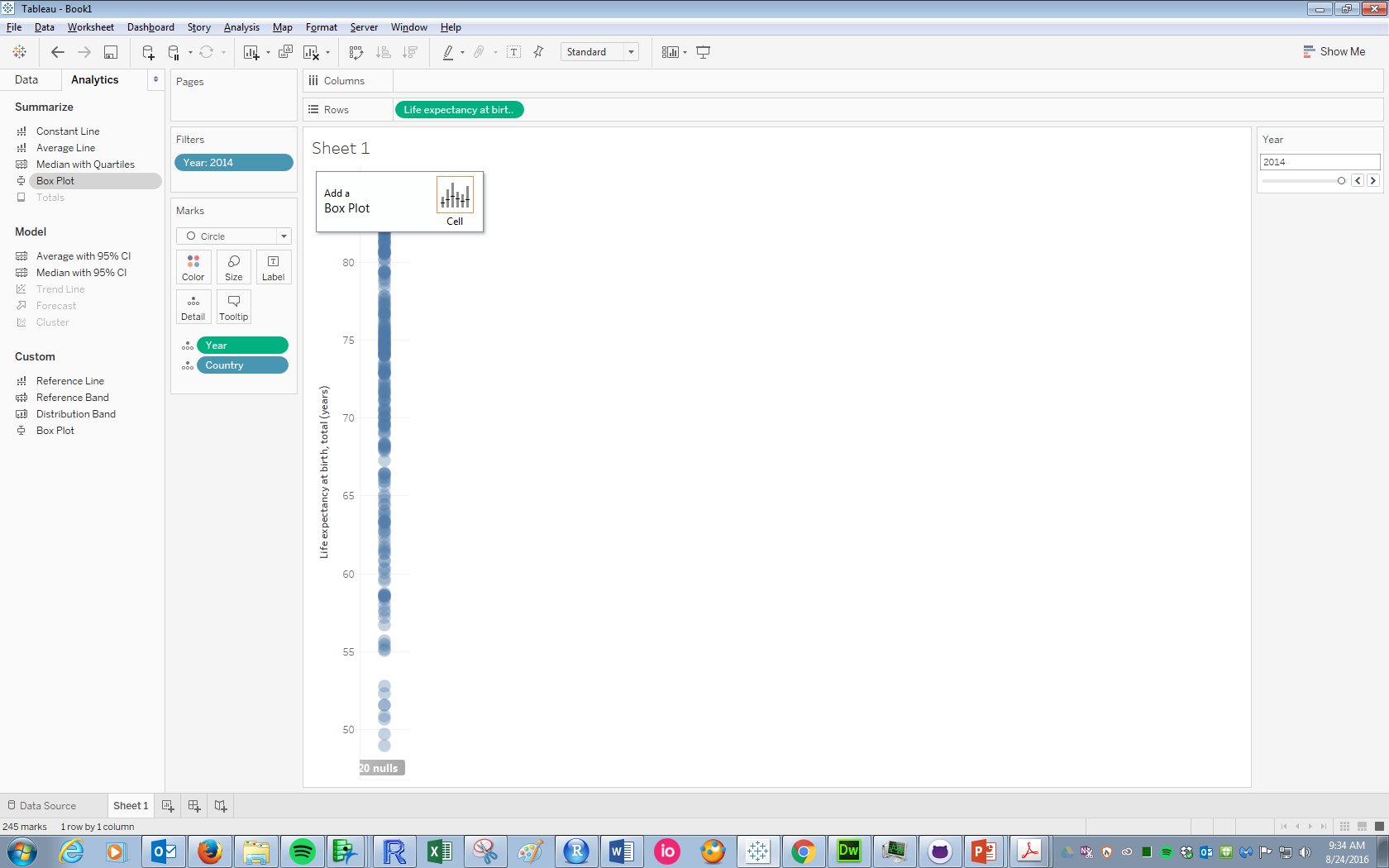
* **Year** is more of a discrete variable in our case, and Tableau has a different format for filtering on discrete fields. Hit **OK** on the original filter, and right-click on the **Year** pill to change it to **Discrete**. Note how the filter view changes:



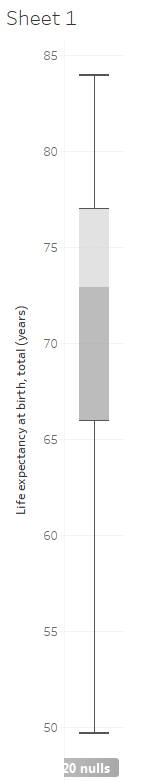
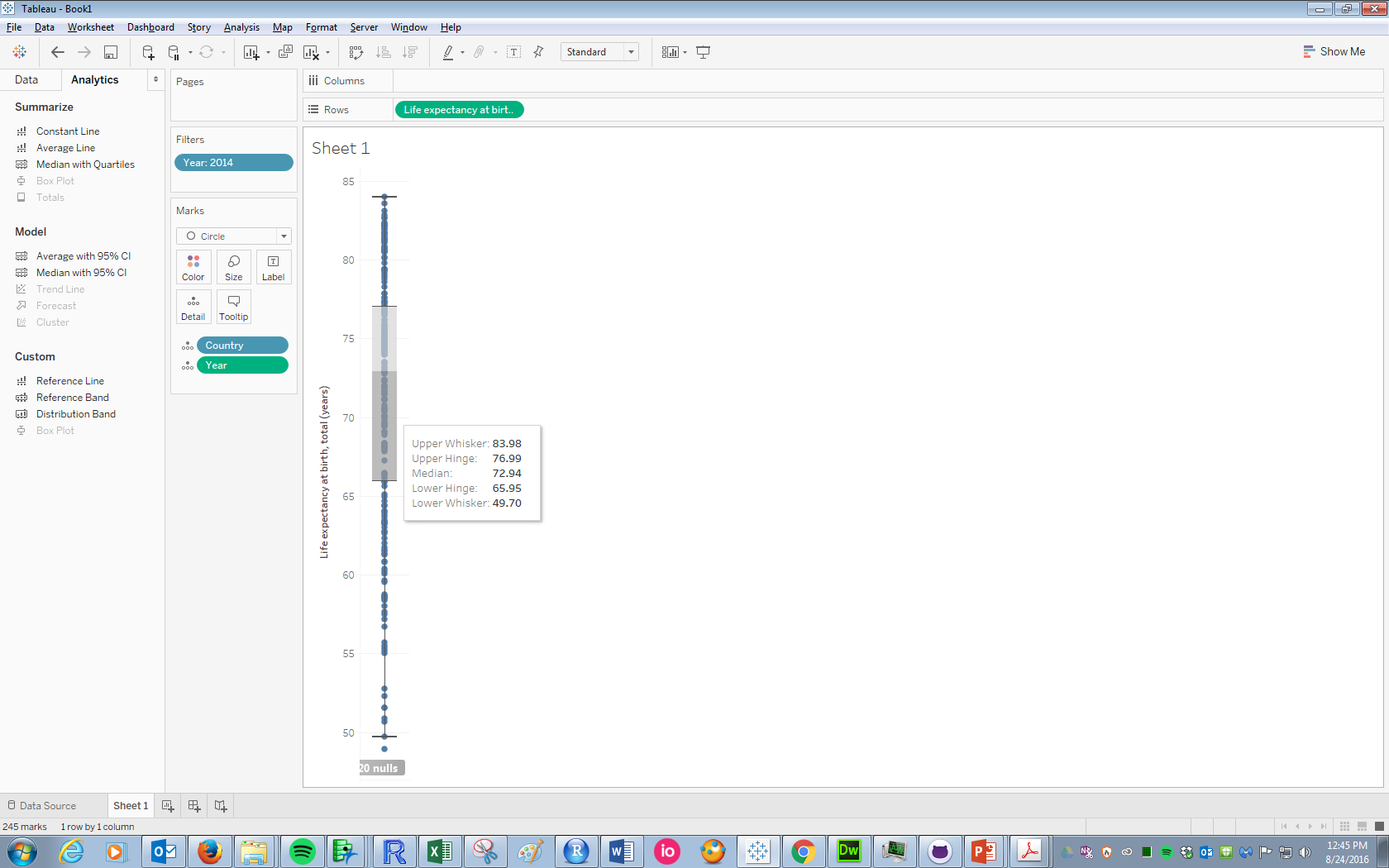
* Scroll down and check 2014, then click **OK**. We now see fewer (but still a lot!) of points! One way to visually improve overlapping points is to add transparency. Click on the **Color** button and drag the transparency lever to something less than 100%:



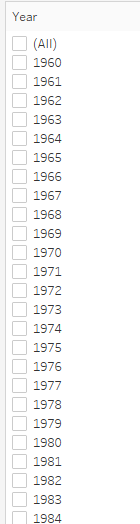
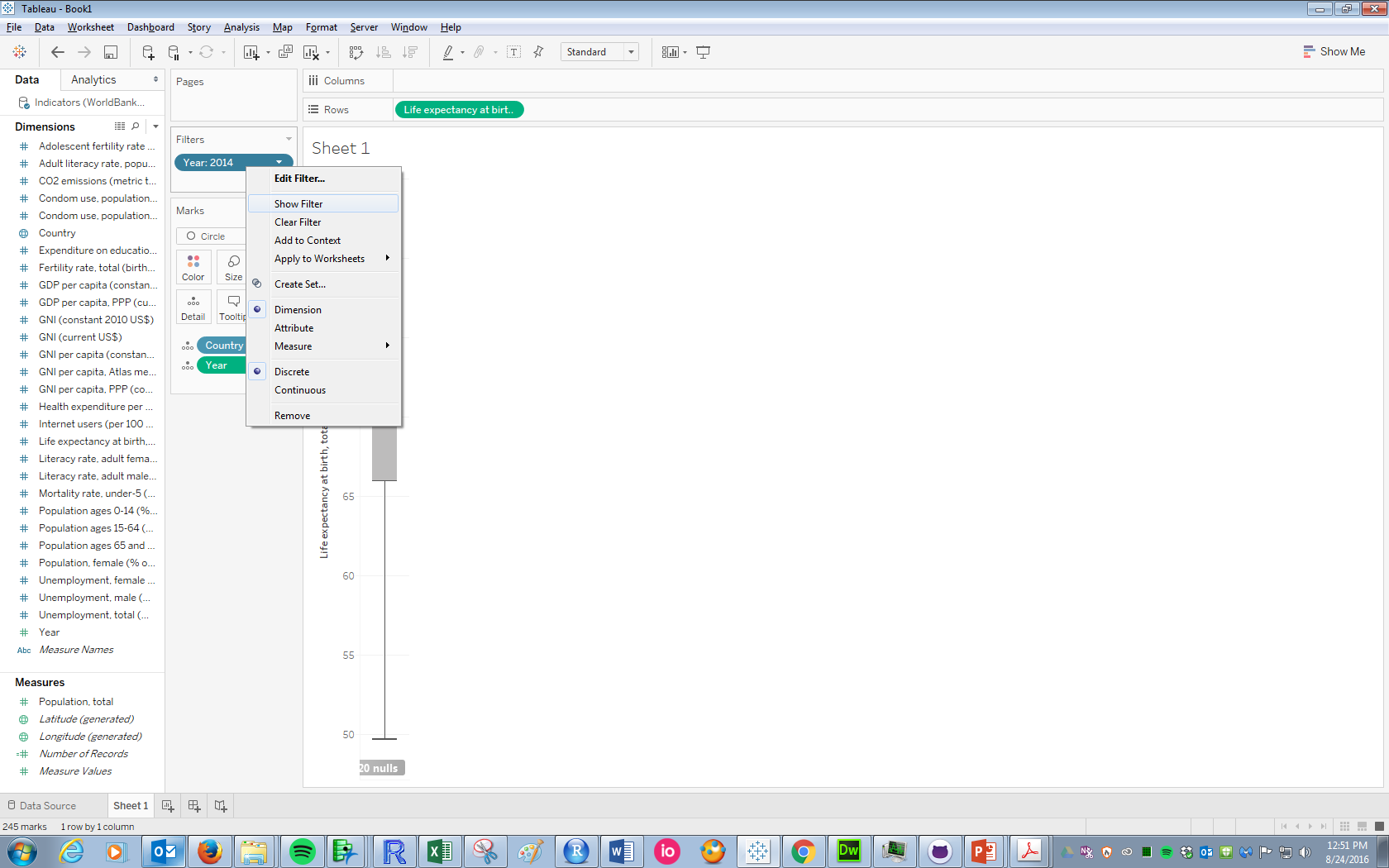
* This is better: we can sort of see that there are more countries in the 70-80 range than the 60-70 range, for example. But viewing overlapping points like this is a rather poor way of visualizing a single quantitative variable. A better, more common way is to create a **boxplot**. Click the **Analytics** tab, click **Box Plot** and drag it to **Cell**:



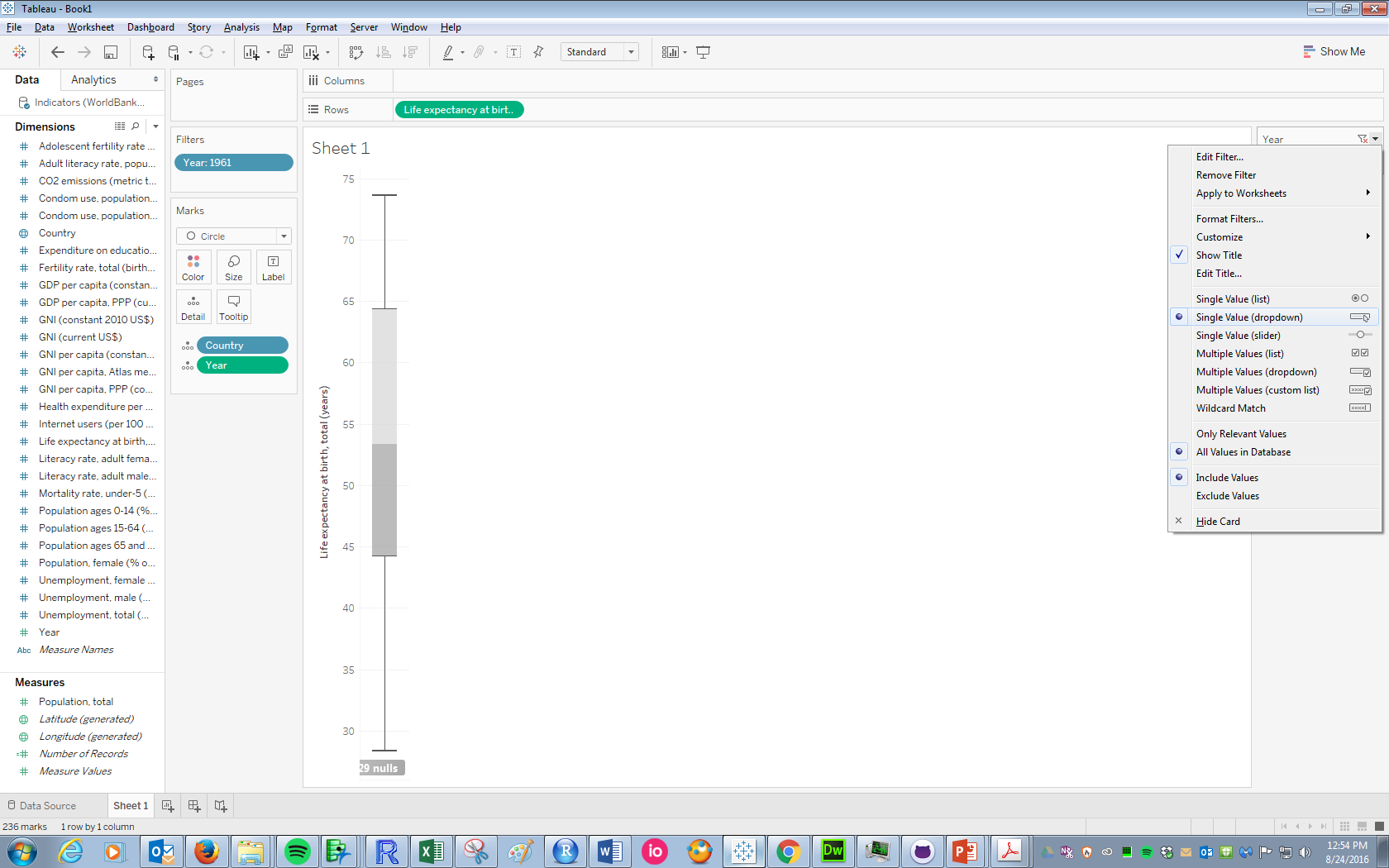
* Decrease the size of the points, and the final image looks like the plot on the left. Note that hovering over the boxplot gives information about the max, min, and quartiles. Changing the color to white leaves just the boxplot, which in some ways is more efficient as it presents only the most pertinent information:



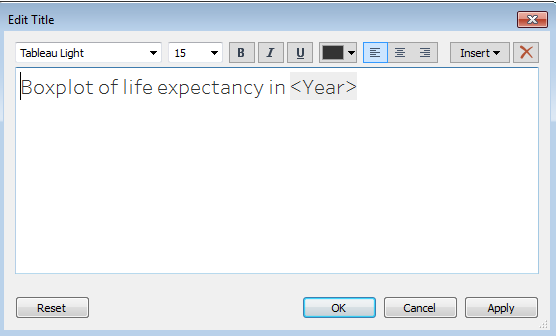
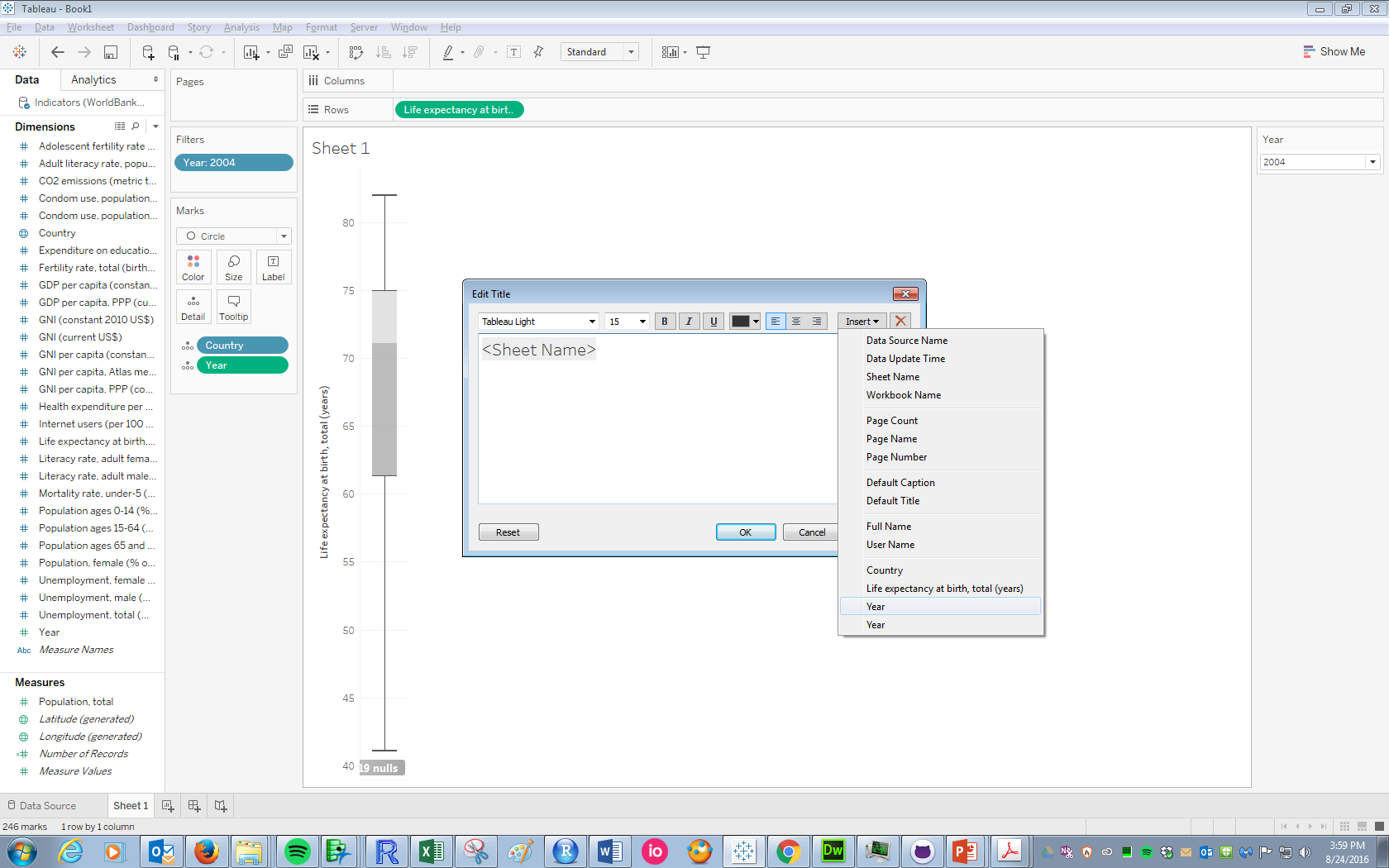
* Why 2014? Maybe we want to let the user choose which year’s life expectancies to display. Right-click on the **Year** pill in **Filters** and click **Show Filter.** This produces a list of years to the right; now the user can specify which year to display:

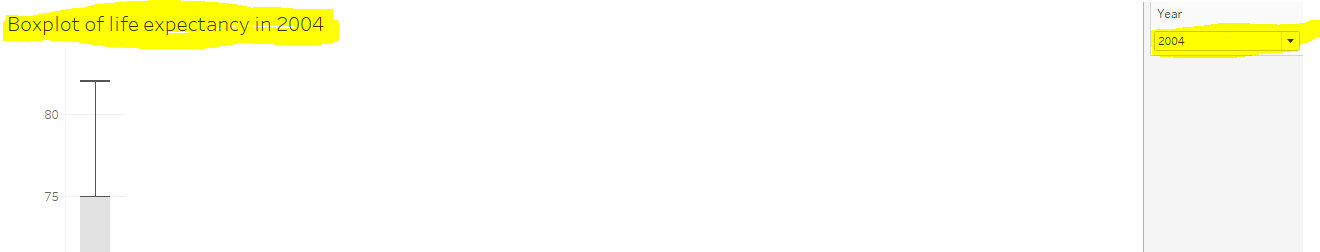


* This filter is in poor taste. First, the list of options are displayed somewhat overwhelmingly. Second, it allows the viewer to click multiple years to display at once: probably not a good idea! Click the dropdown arrow in the upper-right of the filter to change how it is displayed. In this case, a single-selection, dropdown menu would work well. The user now has the ability to display a boxplot for a given year.



* Note the name of this pane, “Sheet 1”, is not very informative. We could at least have it tell us which year we are looking at! Double-click on the name of the sheet. Click on the “Insert” button, and click **Year**. Then edit it to be something more informative. Note the name of the sheet will now update to reflect the year the user has selected:





**Histograms: an alternative way to visualize a single quantitative variable**

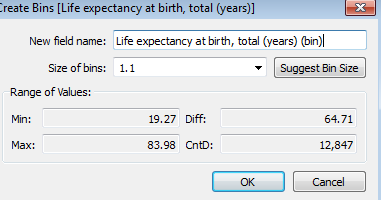
You’ve probably seen histograms before. In Tableau, they are created by making **bins** of a quantitative variable: essentially creating a new, **discrete** version of the variable and making the heights of the bars equal to the count of observations in each bin. In that sense, it turns a univariate visualization task into a bivariate one.

To create a histogram in Tableau:

* Make sure **Life expectancy** is continuous under the data tab. Right-click on it to change if need be.

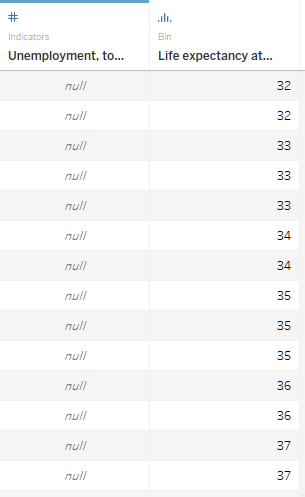


* Right-click 🡪 Create 🡪 Bins. We see the following dialogue box. Tableau suggests a default bin width of 1.1 years; we’ll let that be for now.



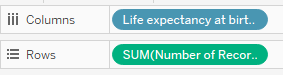
* Click OK, and note a new field has been created that shows up alongside the original Life Expectancy field. You can also see it in the Data Source pane:

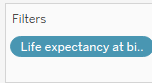


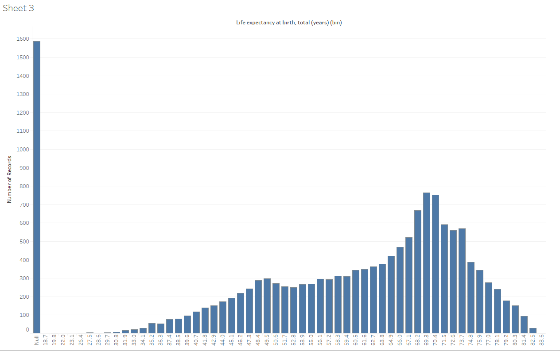


* Now create the view:

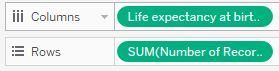
1. Arrange the pills as follows:



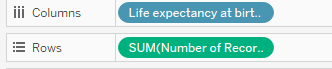
1. Get rid of the “nulls” by right-clicking on that bin and clicking “Exclude”. Note what happens to the filter:



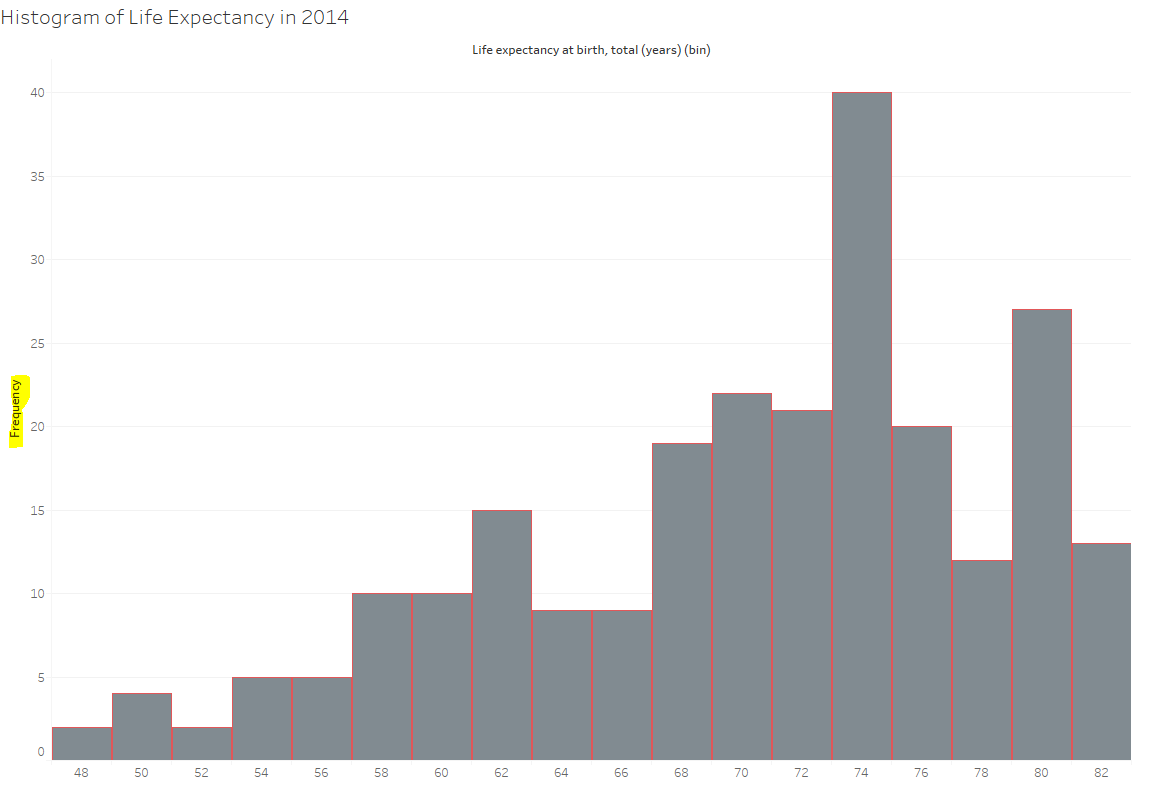
1. Filter on year. Note that you might need to change the bin width to make the histogram look “smooth” when filtered by a given year! Try alternating between this look:



And this one:



1. Play with some different colors:



**Moral of story: Histograms in Tableau are way more a pain than they should be.**