Versions of multivariate visualizations abound, and there is no way we could touch on them all. In this handout, we will encounter several of the most important multivariate visualizations including:

* 2-variable bar charts (aggregation)
* Time-series charts (table calculations)
* Side-by-side and stacked bar charts and area charts

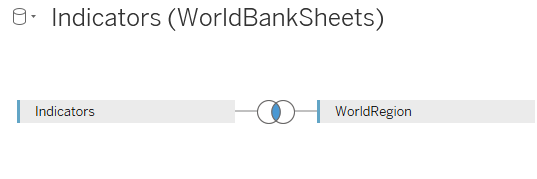
Data concepts include

* Aggregation
* Table calculation

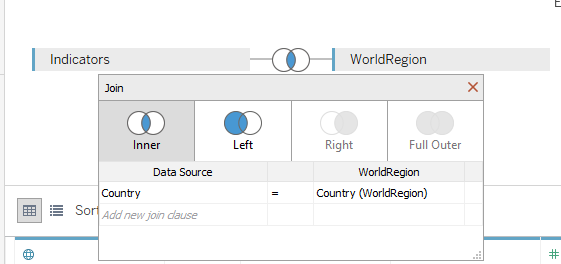
**TASK :** Use the World Bank data to investigate CO2 emissions per capita for each of the 6 regions.

Note that this task involves the use of two fields: Region and CO2 emissions. An immediate problem is we need to merge two of the sheets of the Excel file: the Indicators and the Regions sheets.

On the Data Source pane, drag both sheets onto the pane:

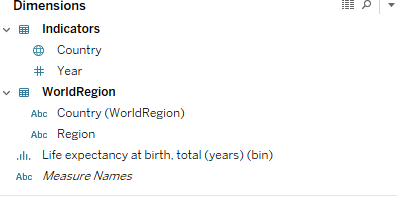


Note that a Venn diagram appears. Clicking on the join brings up the following:



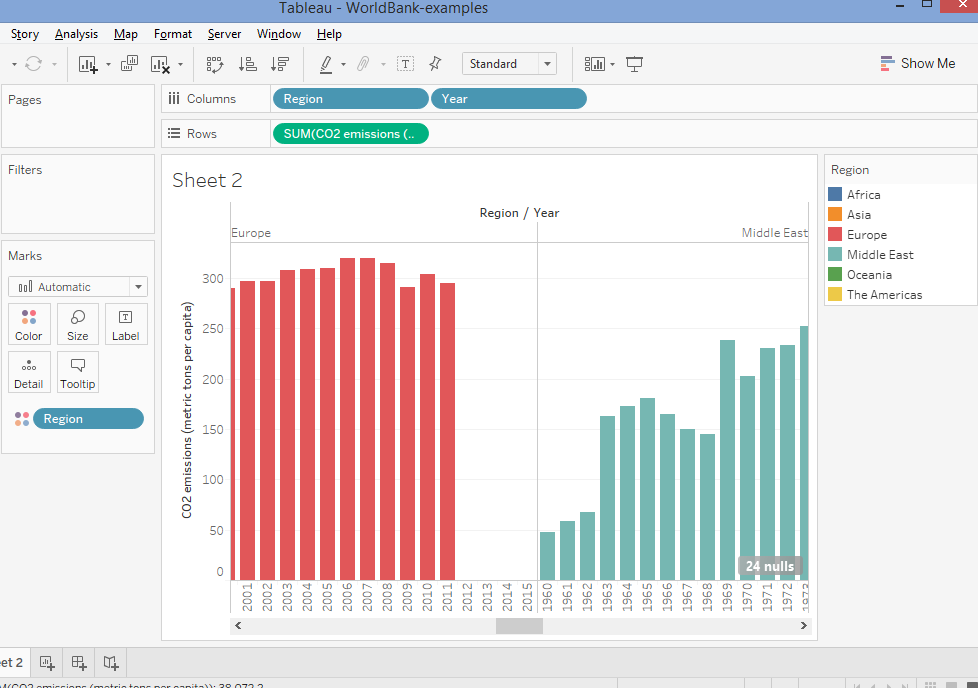
Note that Tableau has automatically merged the two sheets using “Country” as the column that is held in common by both data sources. Tableau guessed right here! If the common columns have different names (e.g. ID1 and ID2), the common column can be specified by clicking the dropdown arrows of the field names and choosing the fields held in common by both data sources.

Note the new view on the worksheet:

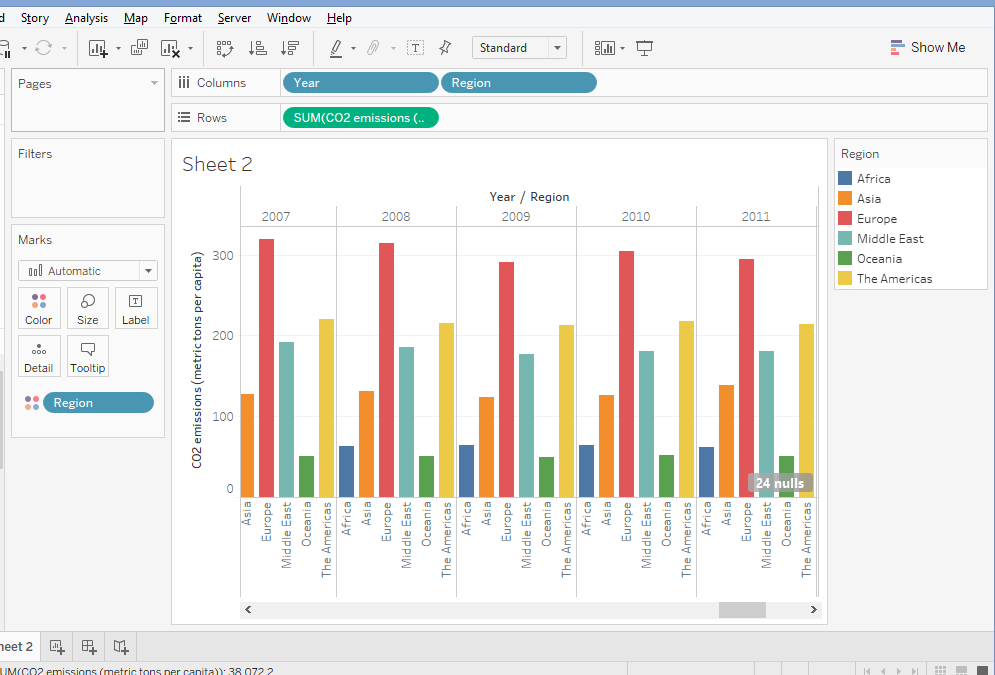


There are now two sources of fields, one from each spreadsheet.

Let’s create a view. Here’s one take:



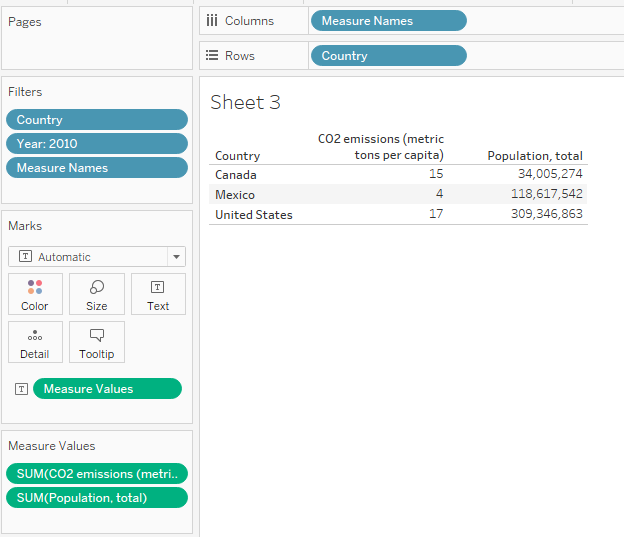
Not so great! Note what happens if you change the order of the “Year” and “Region” pills:



Better!

However, we’ve got a problem: ***Why are we using “SUM” to aggregate CO2 emissions across region for each year? Moreover, if we use “AVG” instead, we will still not get the appropriate averages!*  Populous countries would be weighted just as heavily as non-populous countries**

Here’s an example. What was the aggregated CO2 emissions per capita of the U.S., Mexico, and Canada in 2010? Some data:



It’s tempting to average 15, 4, and 17 to say the average emissions were 12 metric tons per capita! **THIS IS INCORRECT.** This would be weighting Canada as heavily as Mexico and the U.S., when it should be weighted at only about 10% as the U.S. since its population is much smaller.

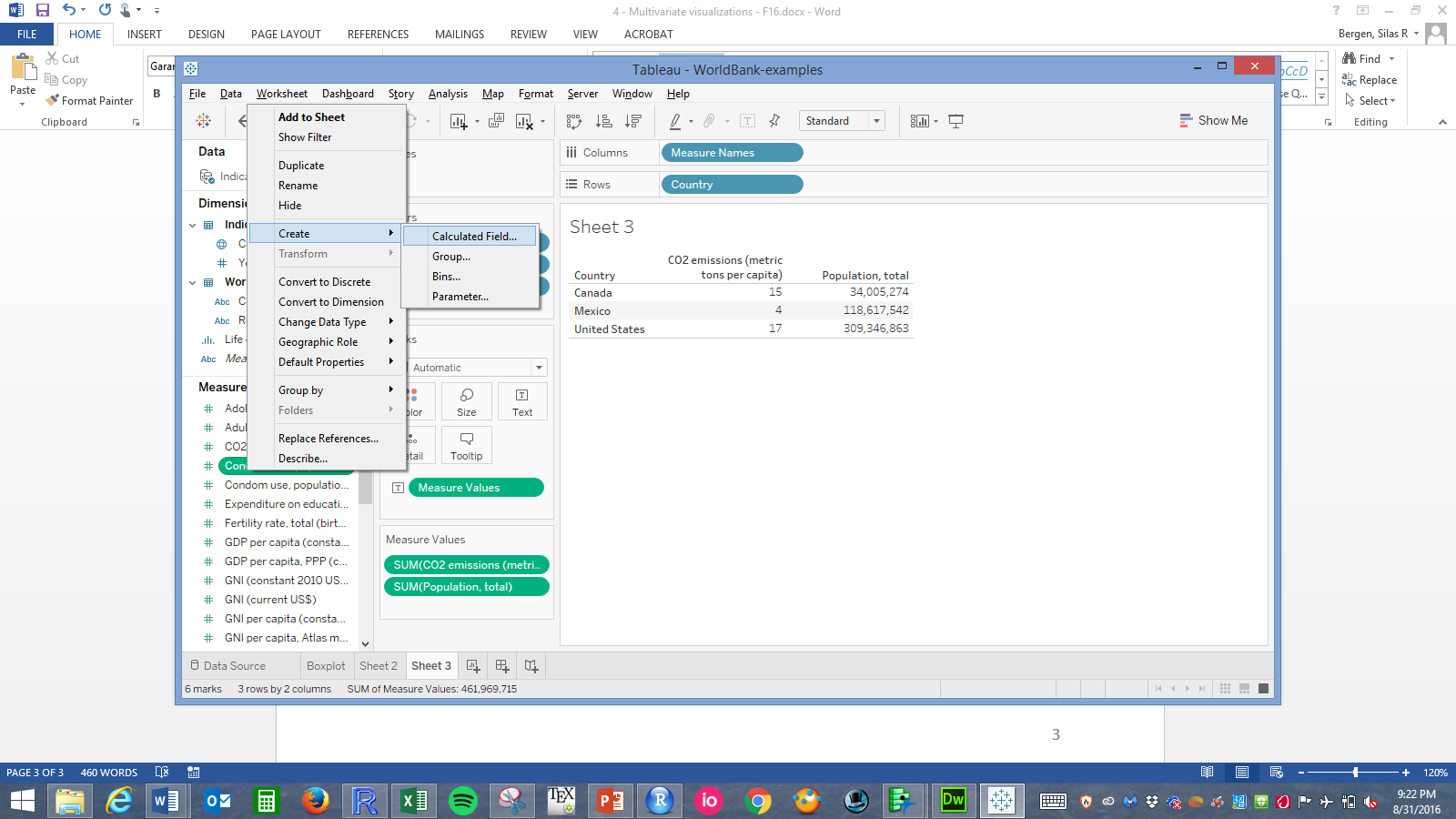
The APPROPRIATE way to proceed would be as follows:

* Find the TOTAL emissions for each country (15\*34,005,274 for Canada; 4\*118,617,542 for Mexico; 17\*309,346,863 for the U.S.)
* Sum up the total emissions
* Sum up the total populations
* Divide the total emissions by the total population

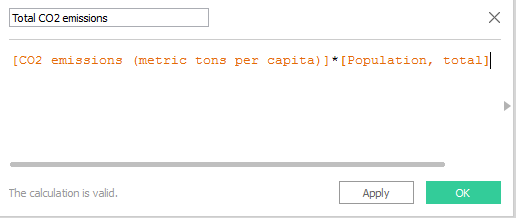
Doing this yields an aggregated CO2 emissions of 13.5 metric tons per capita, more in line with the emissions of the U.S. given its larger population with respect to Canada and Mexico.

We need to create a variable in Tableau that does the correct aggregation automatically when summarizing across region. First, we need to create a new field for **Total CO2 emissions.**

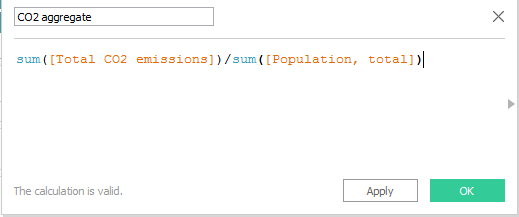
* Right-click anywhere in the data fields area. Hover over Create 🡪 Calculated Field:



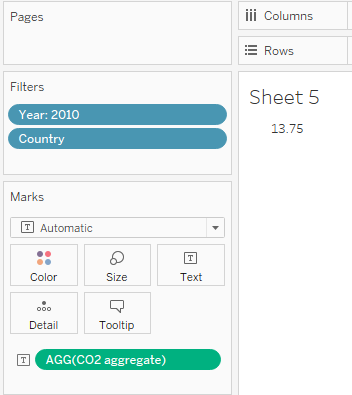
* Create total emissions as shown. Note that if the expression is okay, you will see “The calculation is valid” in grey. Click OK.



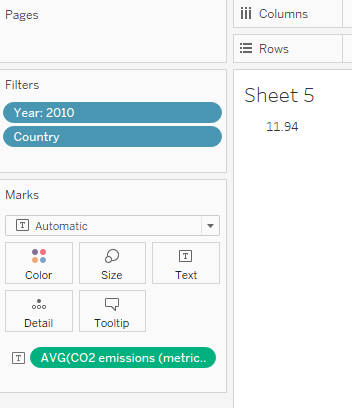
* Now create the **aggregated** version of CO2 per capita. **VERY IMPORTANT:** note that we SUM over total CO2 and SUM or Total population, THEN divide. This will happen anytime you summarize portions of this new, aggregated variable:



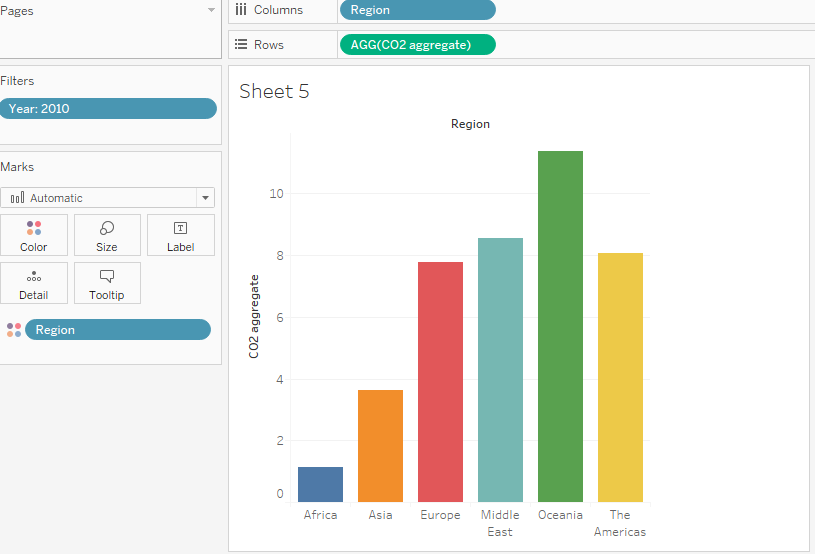
* Note the correct, aggregated CO2 emissions for these three countries (different from 13.5 due to rounding):



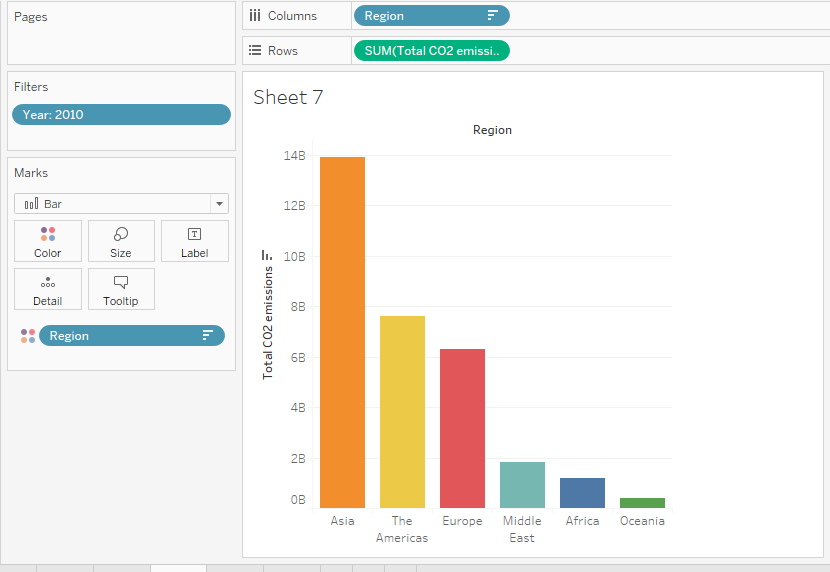
* Compare to the INCORRECT approach of simply averaging the CO2 emissions for these three countries:



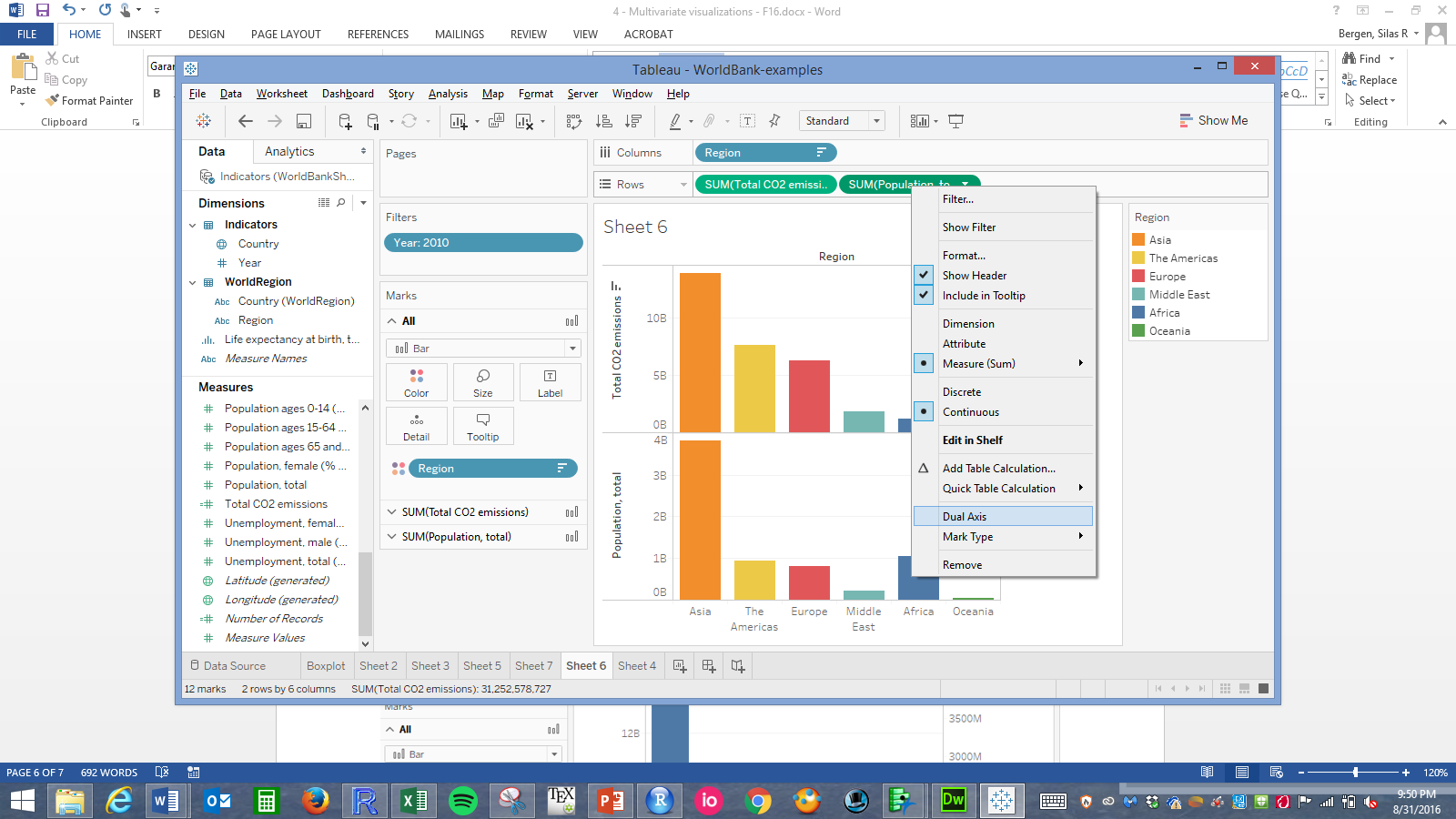
Now that we have an appropriate defined aggregated variable, let’s try creating some views:

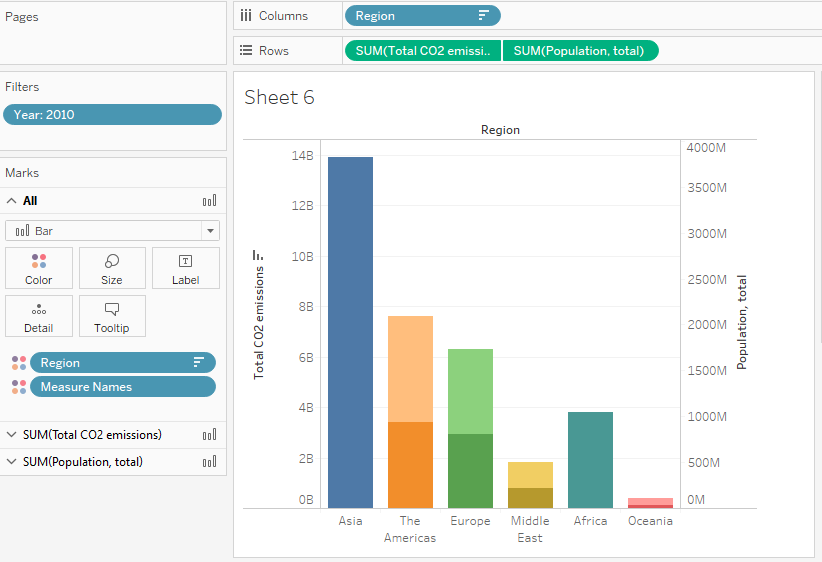


Interestingly, Oceania has the highest CO2 emissions per capita in 2010 of all regions. This is probably due to its small population. Looking at total CO2 emissions tells quite a different (and more expected) story:



We can see this phenomenon somewhat more clearly via creation of a dual axis chart. Note how the two green pills have flat edges that meet: to get this, right-click on one of the pills and click “Dual Axis Chart:

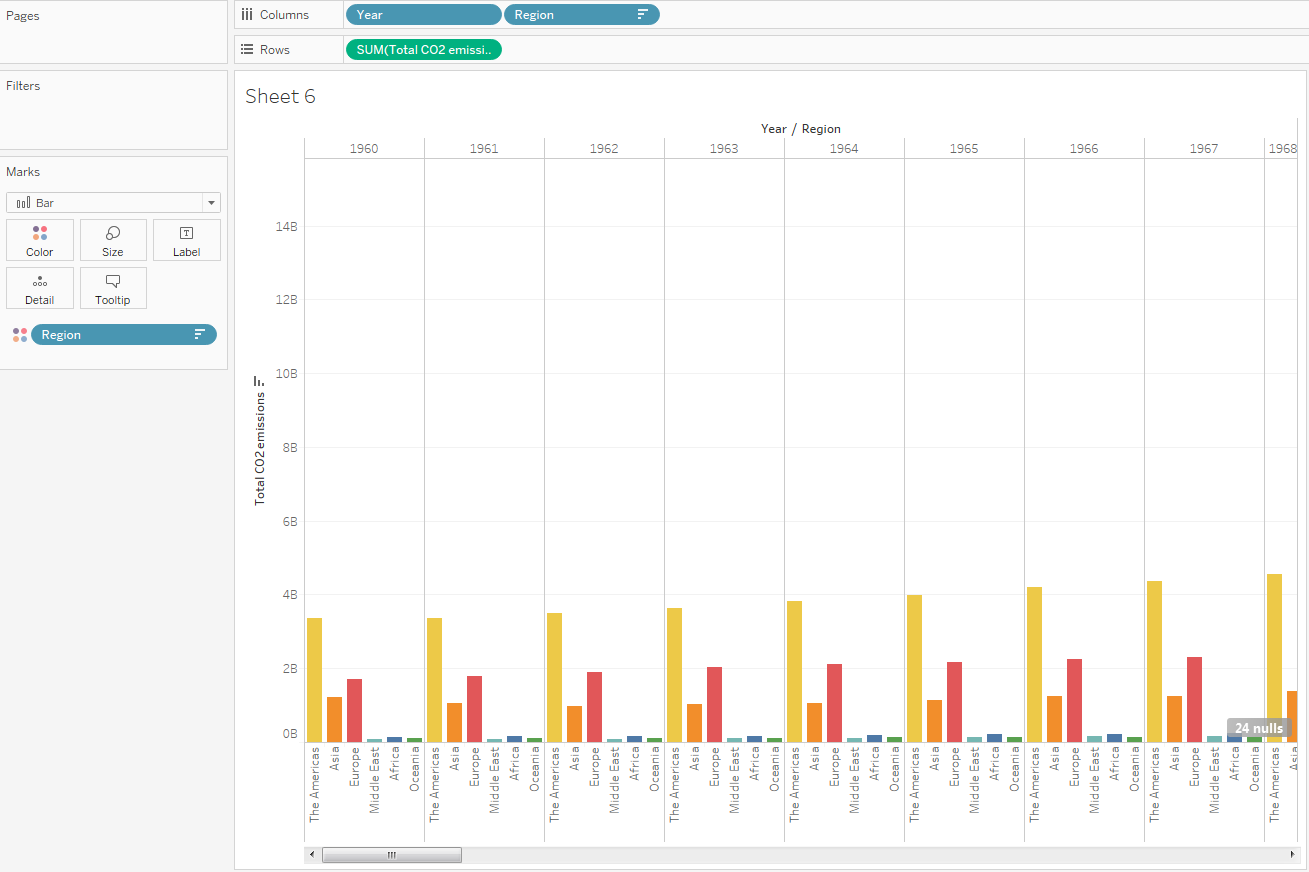




Note that Oceania does appear to have the largest CO2-to-population ratio, but this is mostly due to the fact that their population is so small. (**Note that this dual axis chart has many flaws in and of itself, but is useful as an exploratory tool here!!**)

**Task: Investigate total CO2 emission trends over time by region and country**

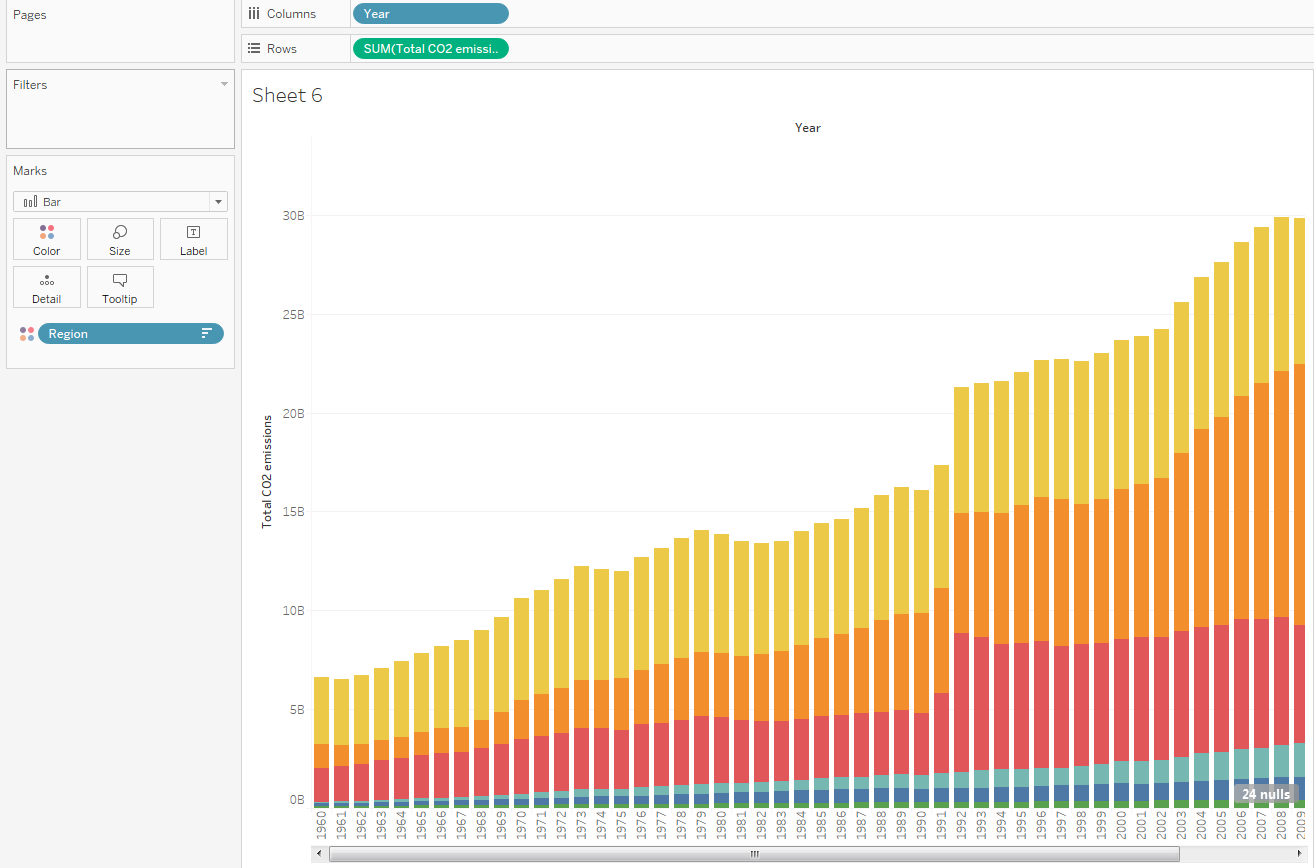
One way to visualize this by region alone would be to simply add “Year” as a pill to the Columns shelf of a previous visualization:



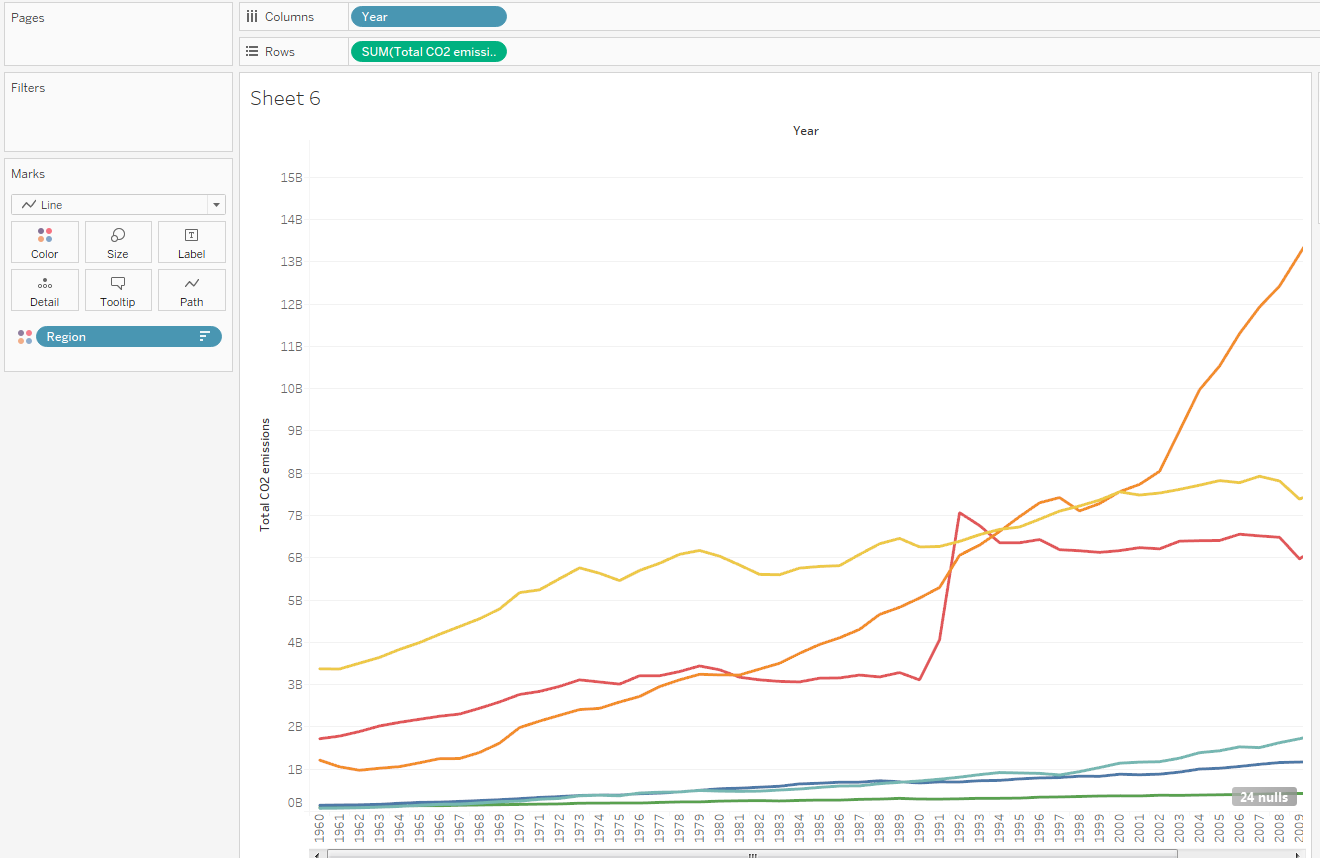
There are several disadvantages of this approach, including:

* It’s very difficult to see what we actually want to see: the trend over time.
* It requires SCROLLING! Scrolling = bad, in general. You want the viewer to be able to see everything at once.

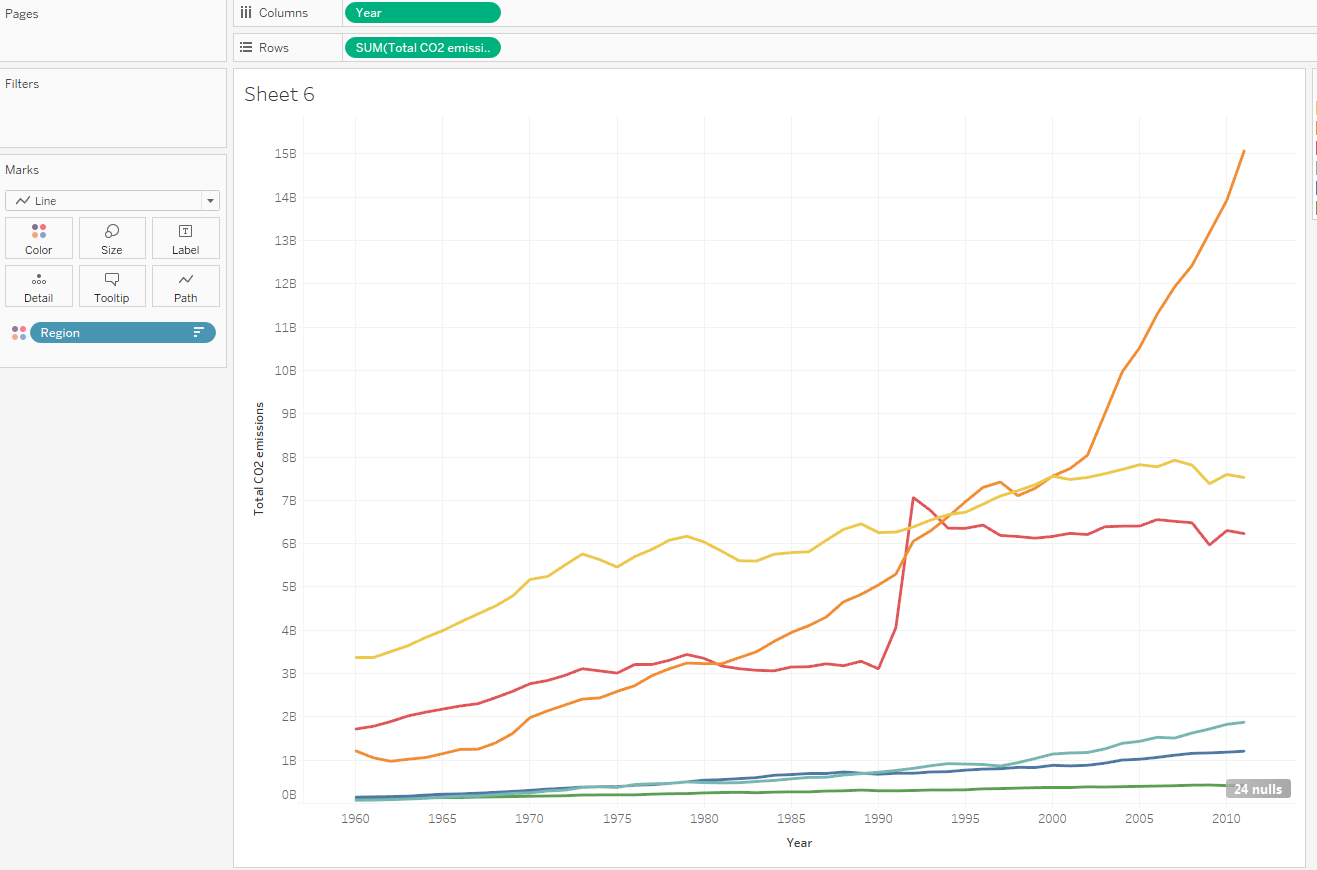
One alternative:



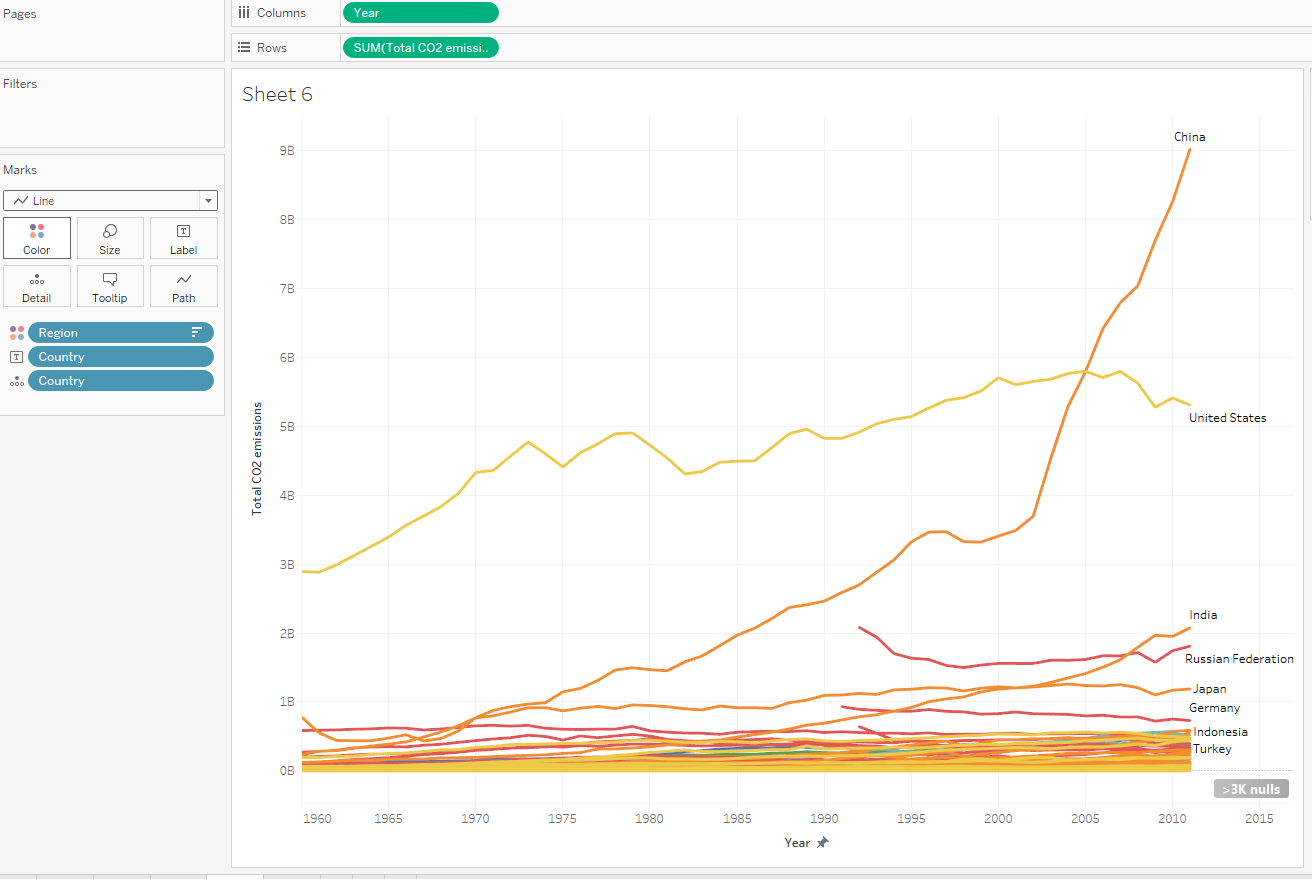
Another alternative:



And another (this only changes the axis):



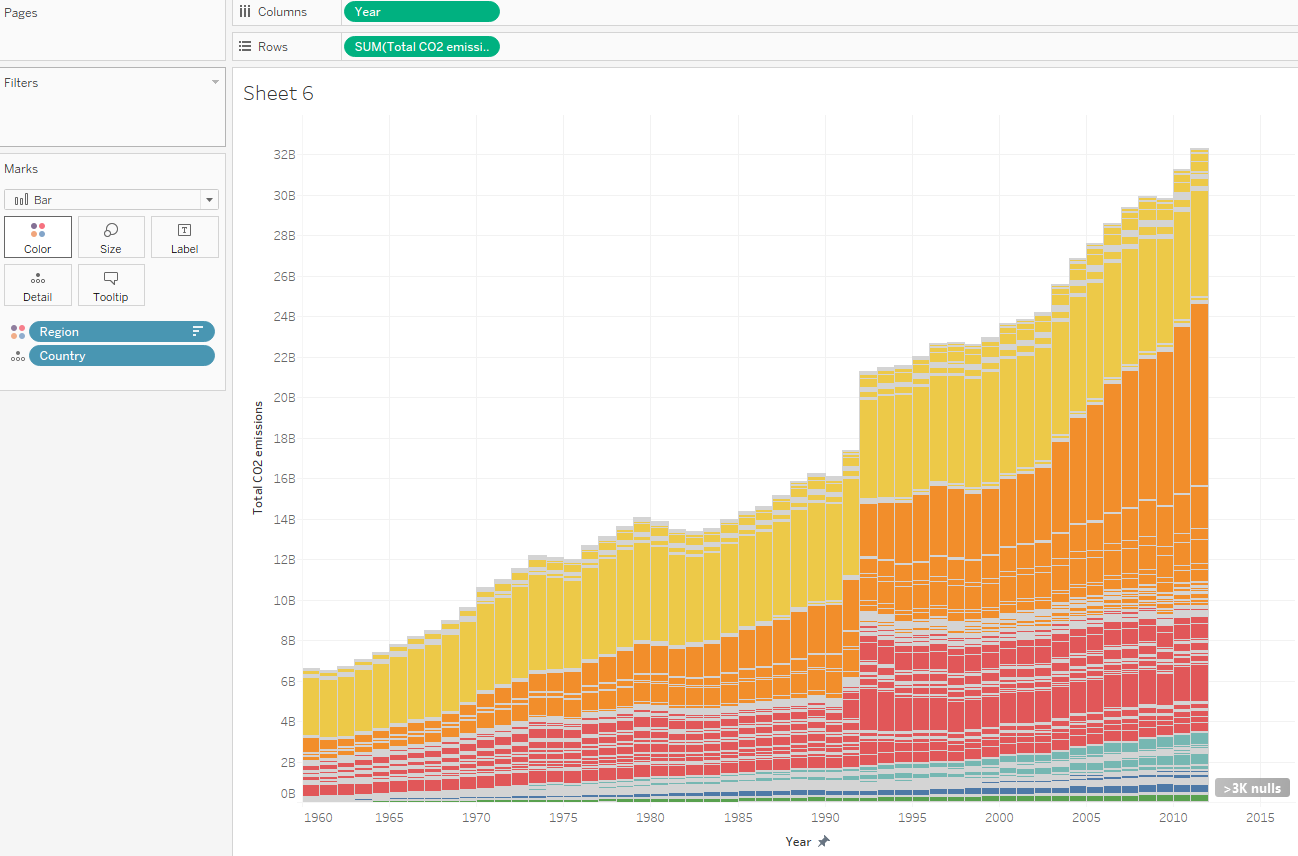
And another:



The previous graph is adding some nice detail about the countries, but has the following disadvantages:

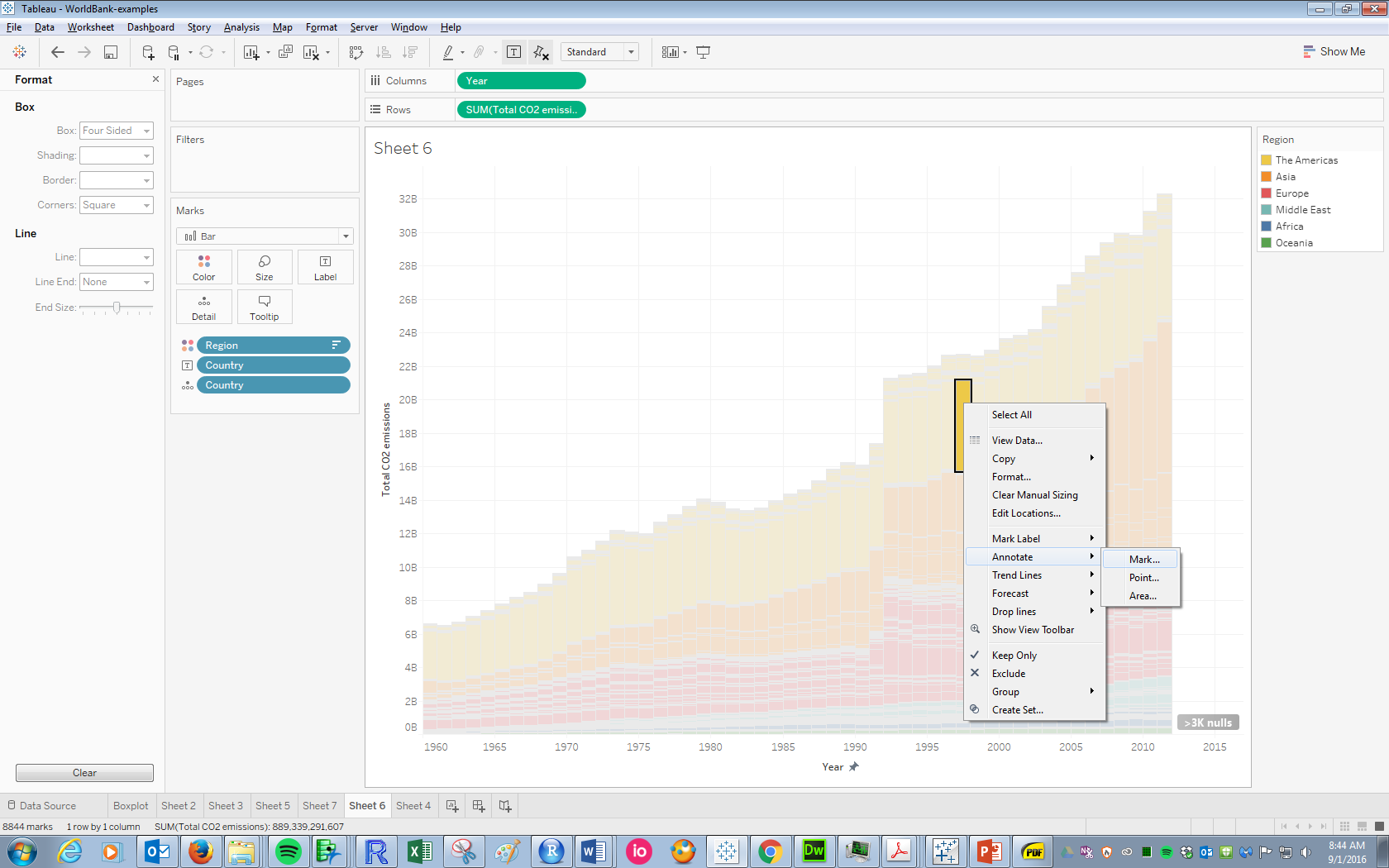
* This graph is basically about China, the U.S., and a few other heavy emitters. So many countries don’t emit as much as the big emitters China and the U.S. that they are all “scrunched” down at the bottom of the graph and are impossible to view.
* This graph is color-coded by region, but it is extremely difficult to pick out any regional patterns. We need to group countries in the same region together somehow.

Here’s another take that addresses some of the issues of the line graph, using a stacked bar chart:

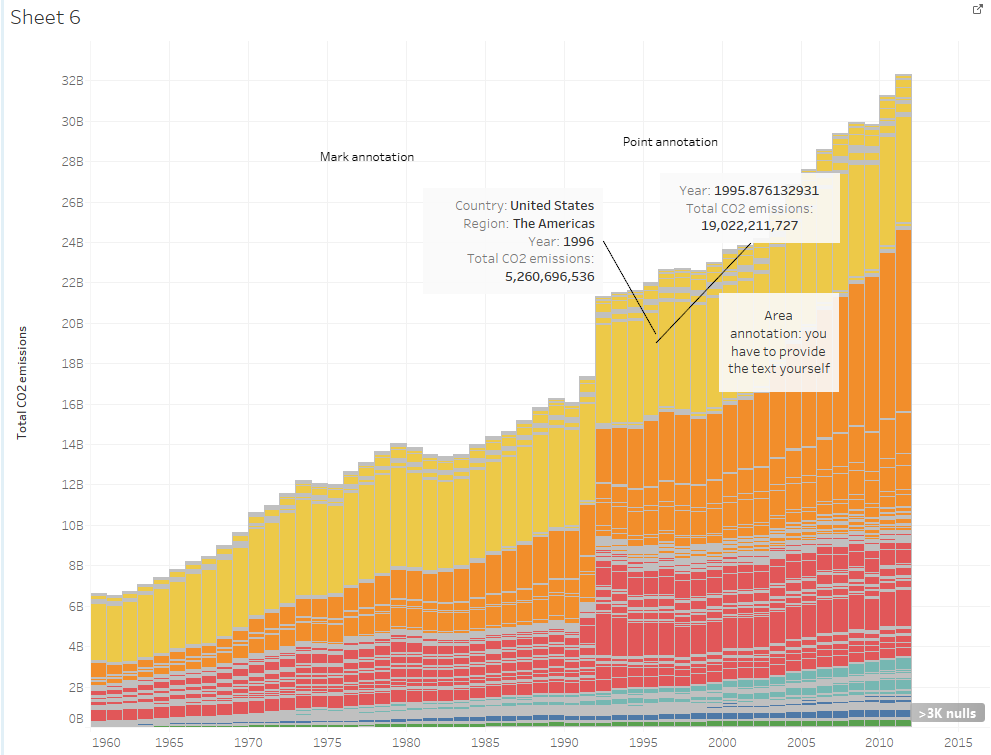


This one is nice, but it might be helpful to still point out some of the heavy emitters without having to mouse over them and rely on the tooltip. We can do this by way of ***ANNOTATIONS***.

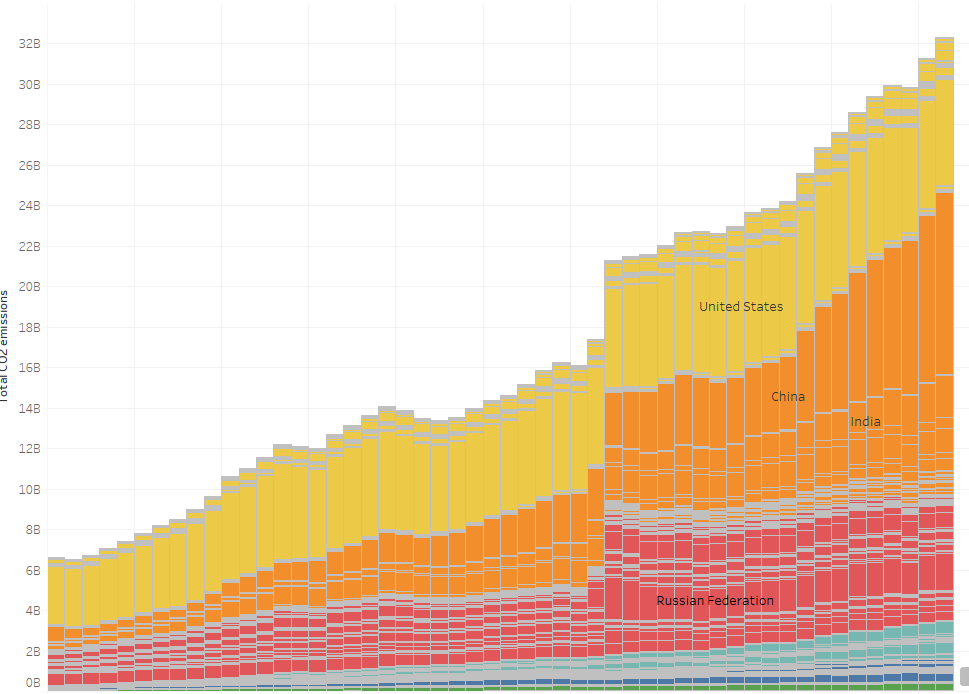
* Right-click on one of the bars for the U.S., then click Annotate. Note we have three options: Mark, Point, and Area. Mark is for the most detailed annotations; Area provides the least detail.



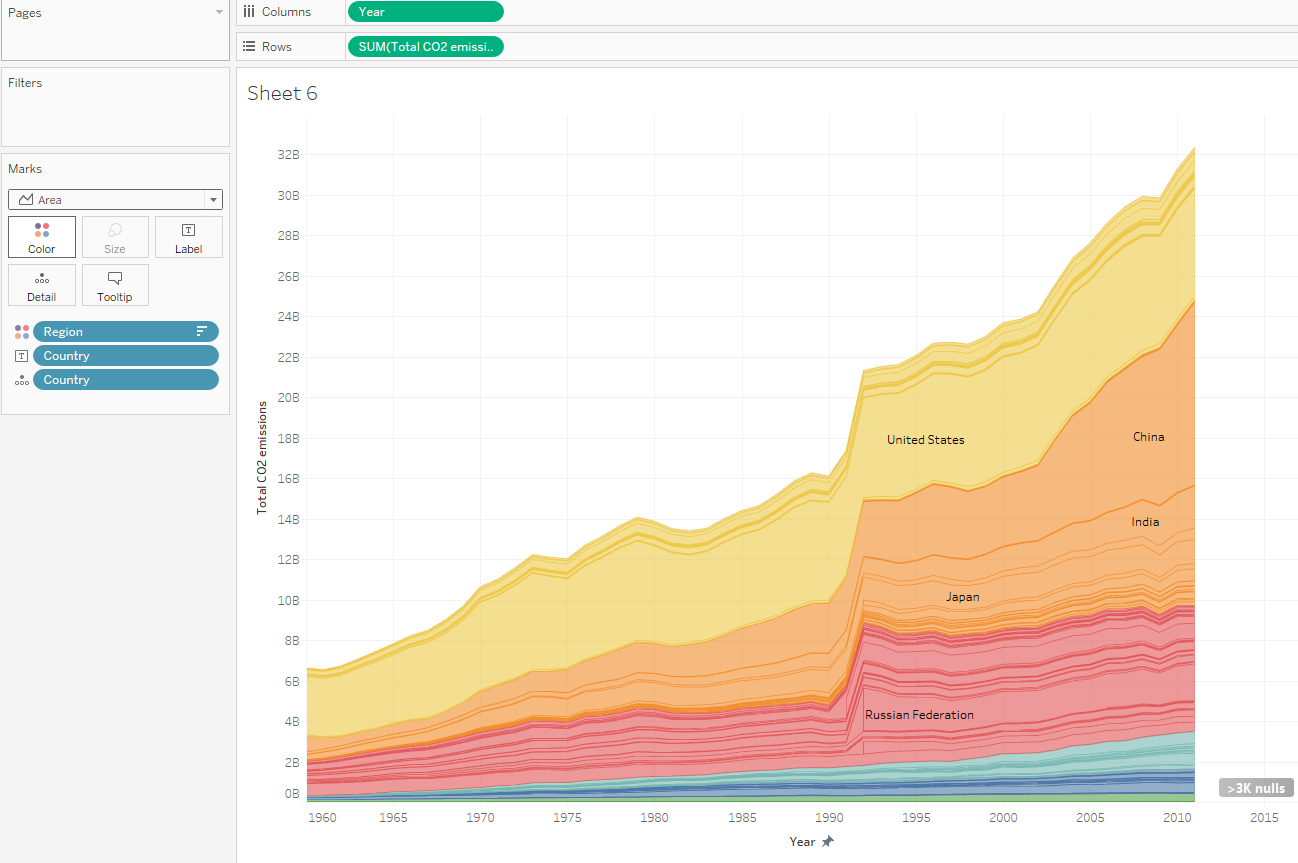
Examples of different annotations (note the different flexibilities of the mark and point annotation arrows, and the absence of the area annotation arrow):



Here’s a nicely annotated version of the above visualization, using Area annotation. To format background and other elements of the annotation, right-click on the annotation and click Format.



There’s still one aesthetic problem with the stacked bar chart, which is the messiness of the borders between the bars. An area chart does away with this. Note that it also automatically annotates some of the heaviest emitters:



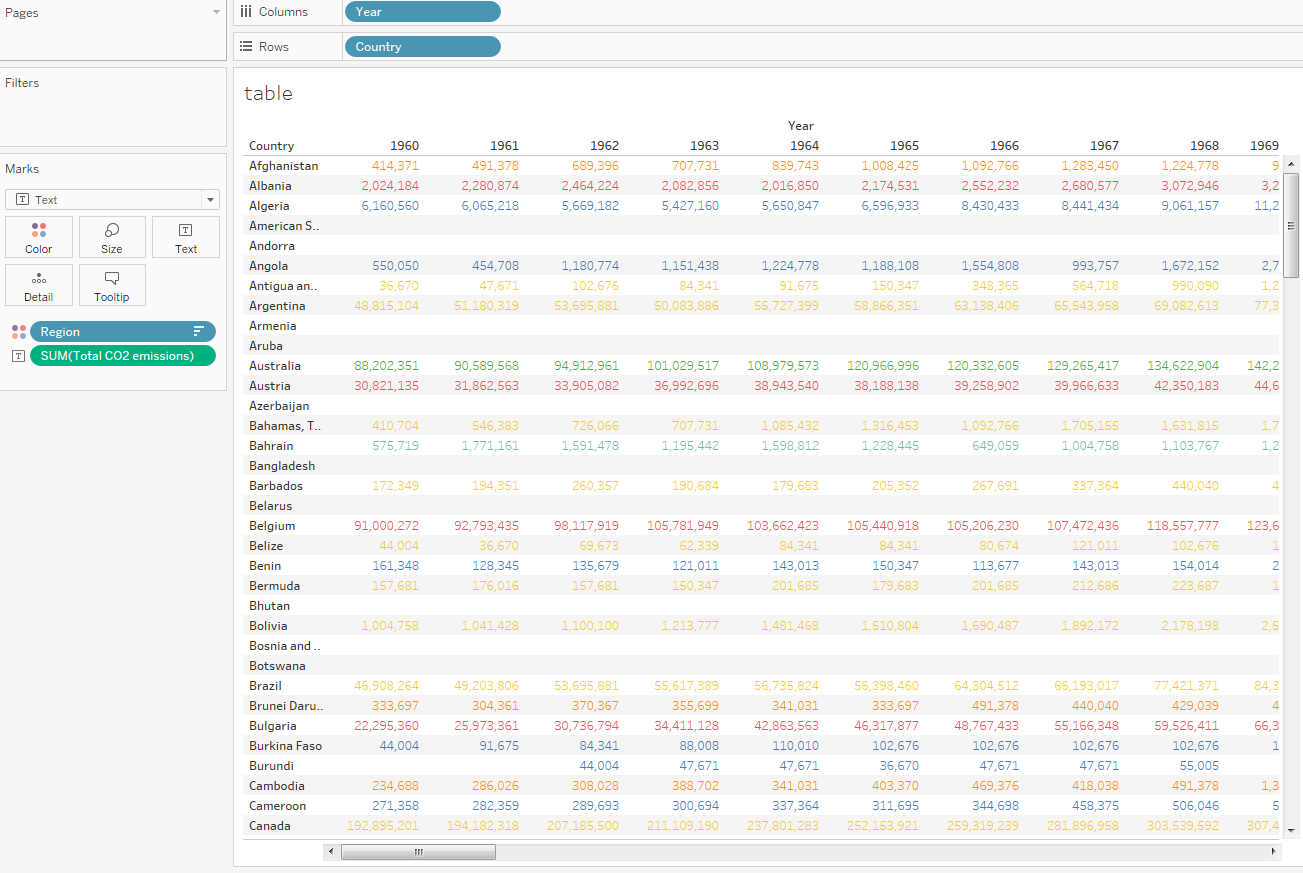
**Table calculations**

The area chart is nice, and begs another view of these data:

**What percent of total yearly emissions is each country responsible for?**

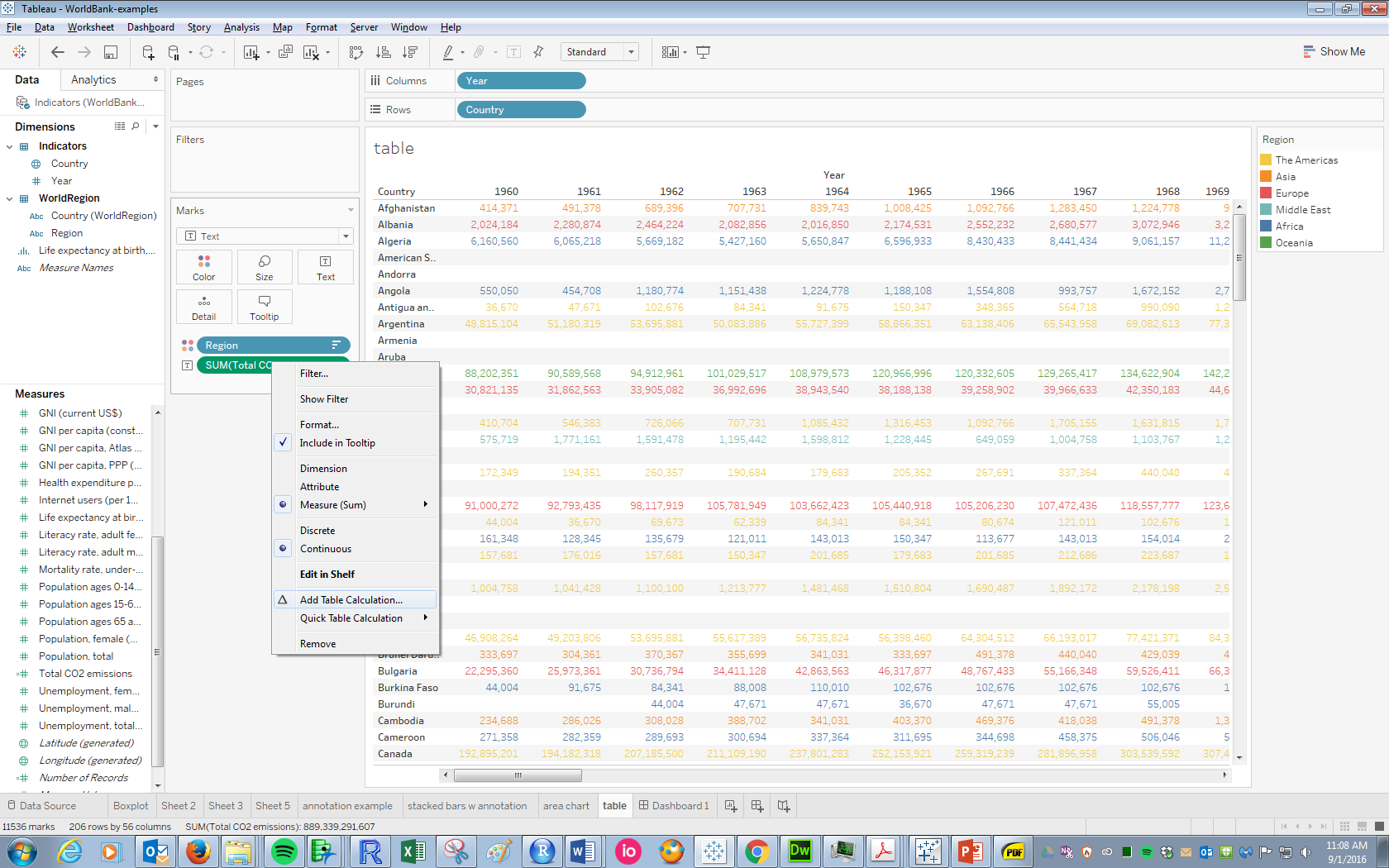
You can imagine what such an area chart would look like: a perfect rectangle, with the vertical axis going from 0% to 100%. To create these percentages, we need to create a *Table calculation*.

To thoroughly understand table calculations, let’s create the tabular representation of the data in the area chart shown above:

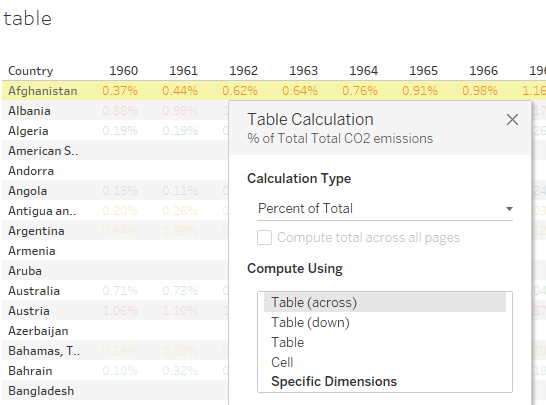


The entries in the table denote the total CO2 emissions for each country in each year. What if want to change these to percentages instead? To do this we need to add a table calculation.

Right-click on the green CO2 emissions pill 🡪 Add Table Calculation:



This yields the following dialogue box, where we can choose both the type of calculation (Percent of Total, here) and the method of computing percentages: across the table, down the table, on the entire table, or for each cell. Try out the different methods to see what happens, and note that Tableau helpfully highlights the portion of the table you are totaling to create the percentages (this feature is new in Tableau 10.0, and one of the revisions I like a lot):

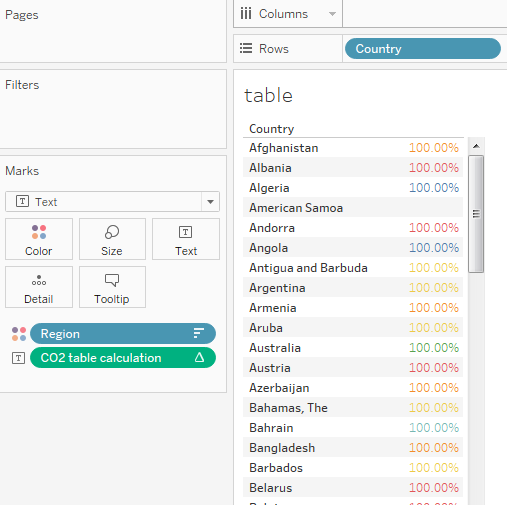


After selecting a method, note the triangle on the CO2 pill. This indicates a Table calculation is being performed on this field:

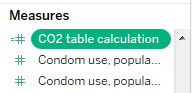


**Some important facts about table calculations**

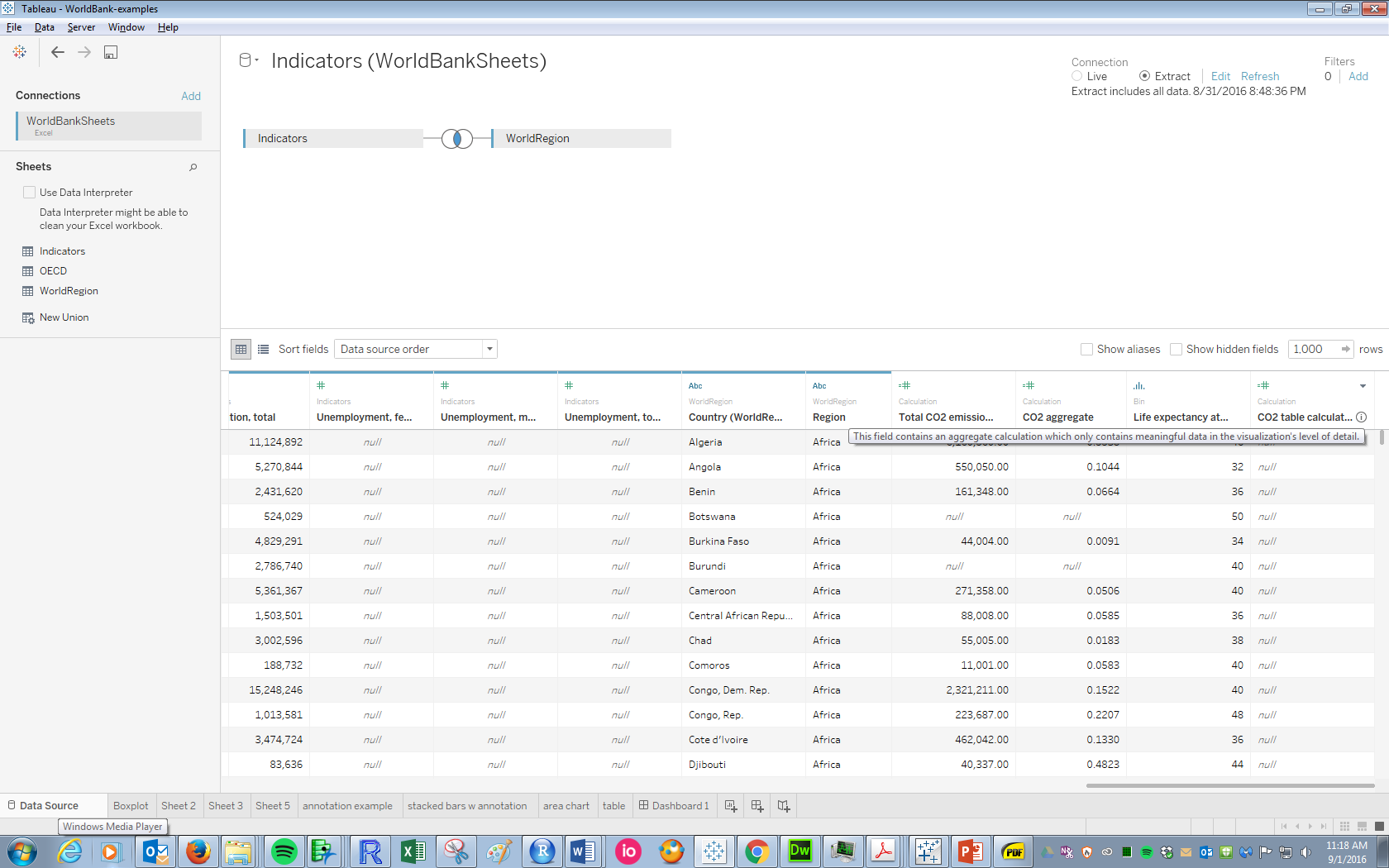
* They depend on the level of detail in specific view you have created. Note what happens to the table when you remove either the Year or the Country pill; e.g.:



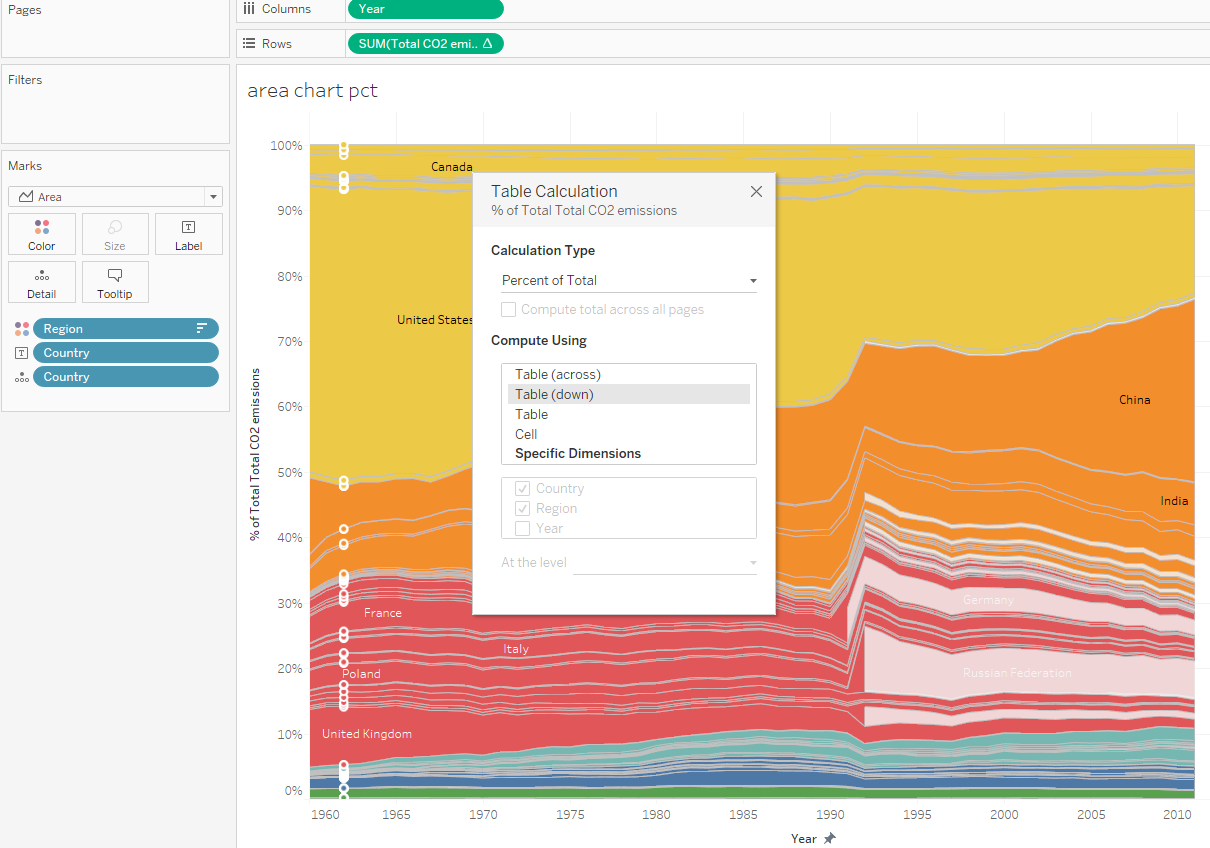
* Because of the previous point, they can NOT be treated as a new field. Note what happens if you hold CTRL while dragging this pill over to the Measures pane and dropping it, then renaming it CO2 table calculation:



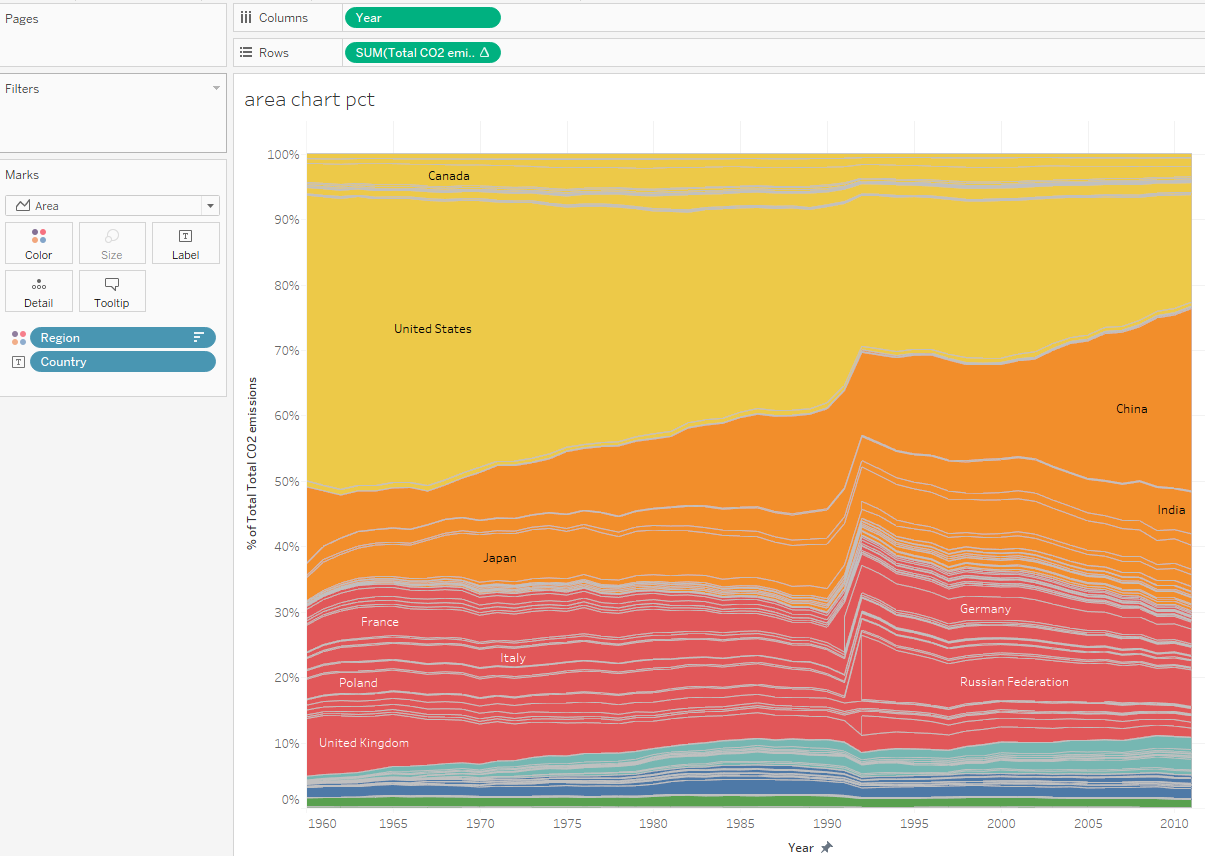
Head back to the data source, and notice the message that appears next to the field header:



Let’s apply a table calculation to the Area Chart (note how the view changes as you hover over different calculation options):

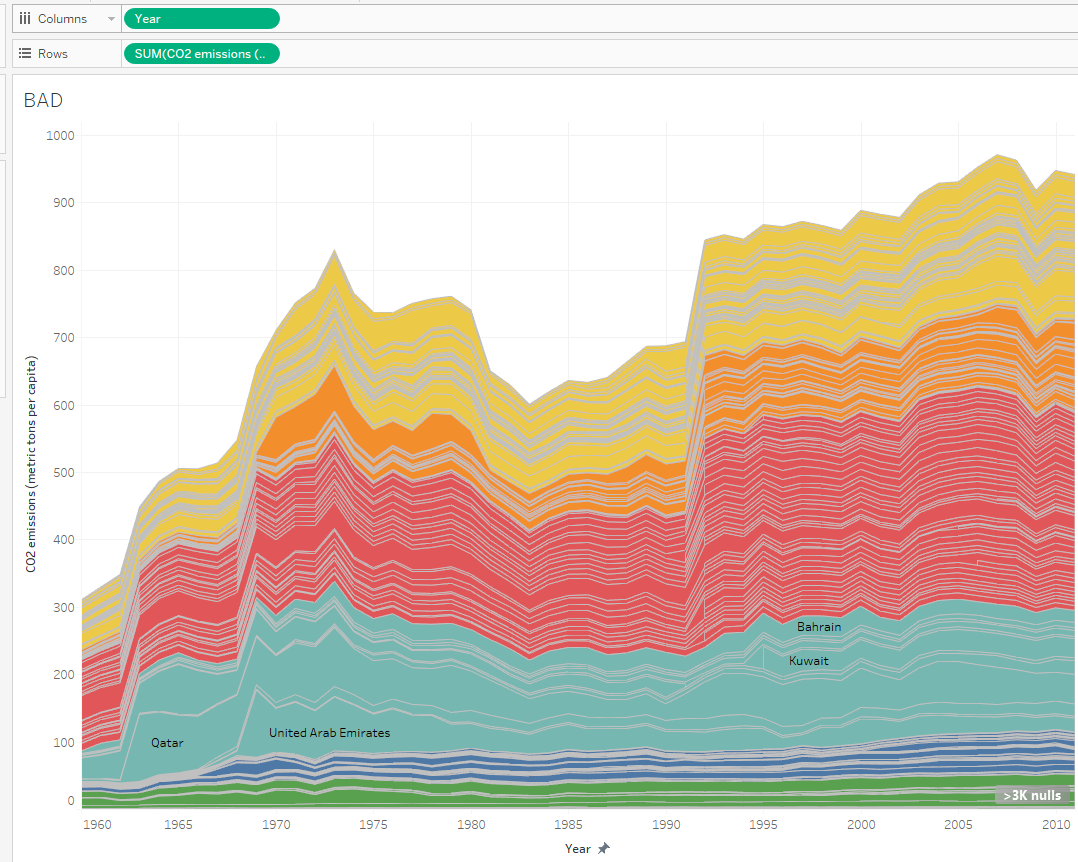


We now have what seems to be a decently “final” visualization:

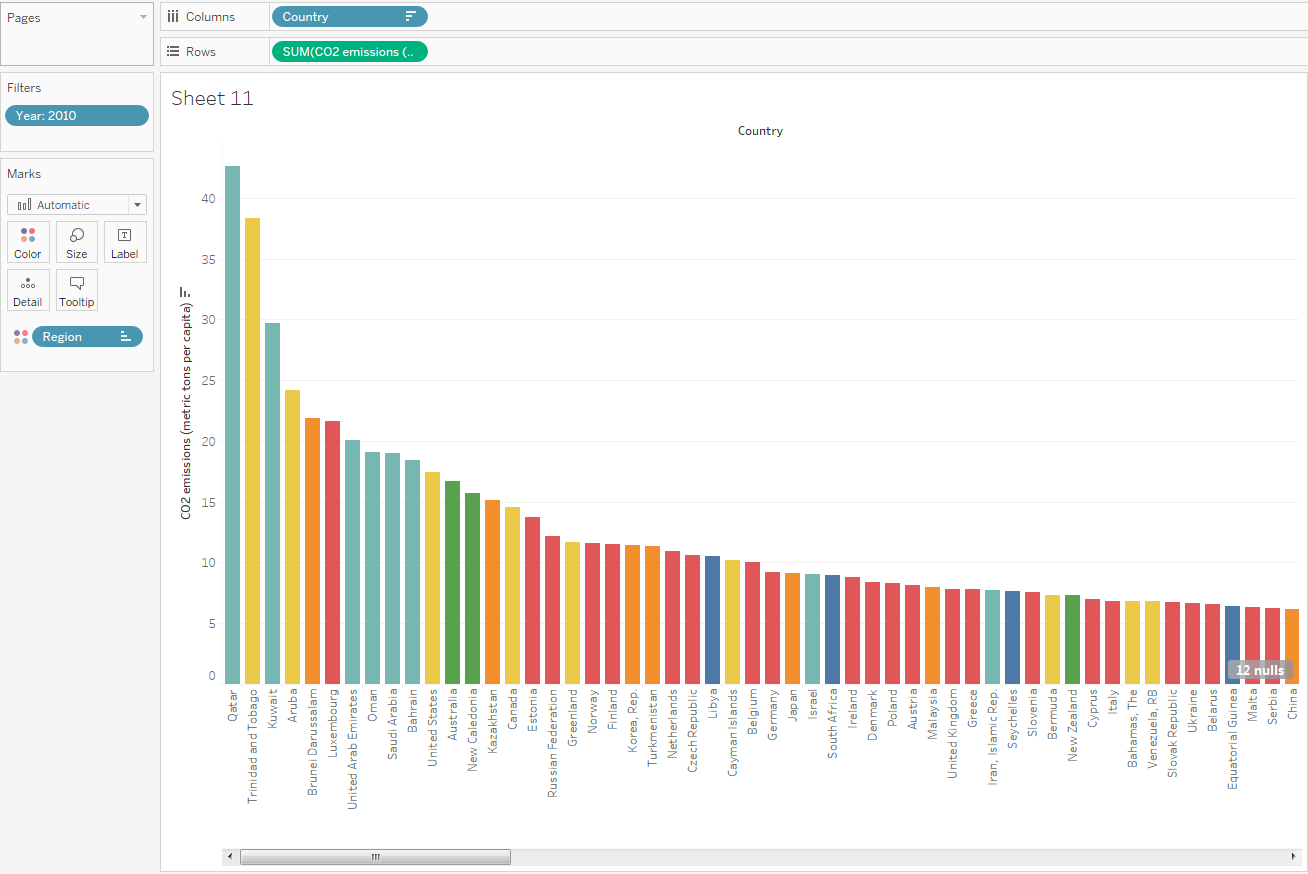


**IMPORTANT WORD OF CAUTION REGARDING AREA AND STACKED BAR CHARTS:** They are appropriate if it makes sense to add the heights of the individuals slices to get the total, e.g. if each slice represents a percent of total, or part of a total emission like above. An area chart is NOT appropriate if we were visualizing metric tons PER CAPITA of CO2, because TOTAL metric tons per capita is NOT obtained by simply summing across the individual country-level metric tons per capita, as discussed previously. ***Do NOT use area/stacked bar charts when visualizing rates or ratios!!***

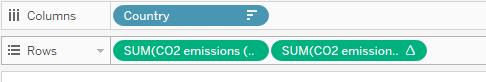
**BAD BAD BAD:**

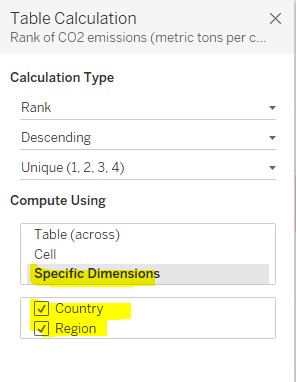


Although total emissions yields some interesting insights, standardizing by population does provide some additional further insight. For example:



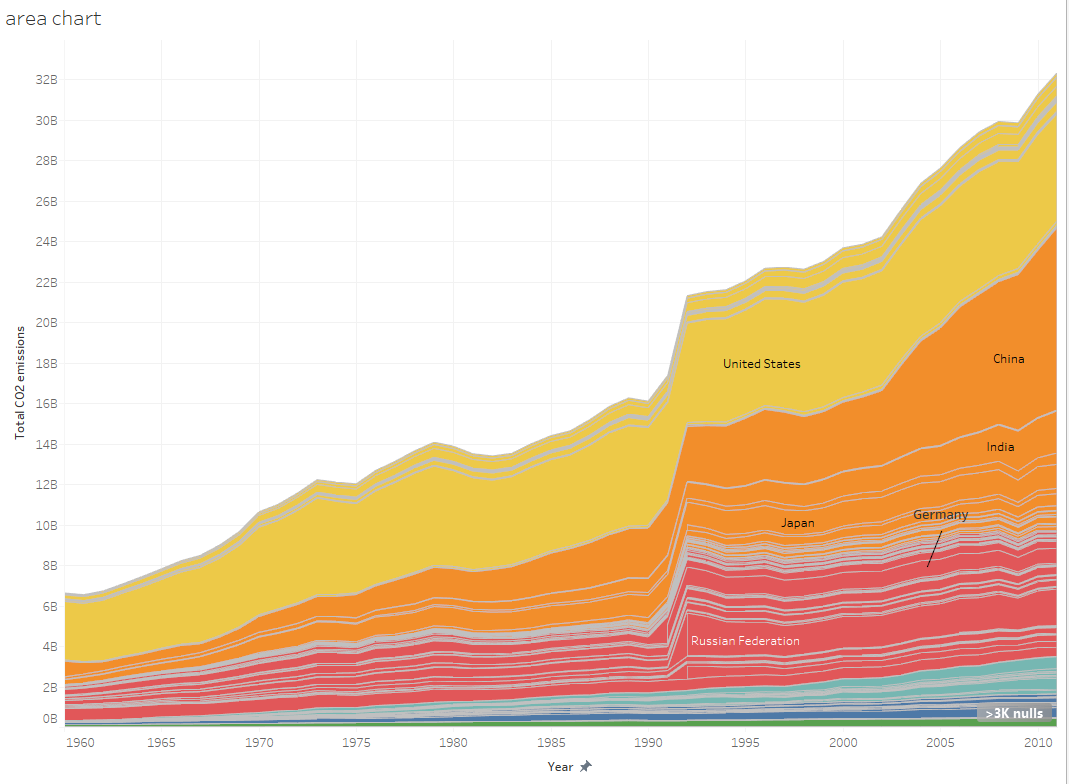
The Middle Eastern, oil-producing countries stand out. It might be nice to create a “refined” version of this bar chart that only shows the top-10 emitters. To do this, we need to create a table calculation on CO2 emissions, then filter on this new variable. Create another instance of this pill, and add a Rank table calculation. Note the need to specify the dimensions we want to rank across; what happens if we don’t specify this?





We only really want this variable for filtering purposes. Bringing it over to filter provides the following:

Let’s return to the earlier image with total CO2 emissions:

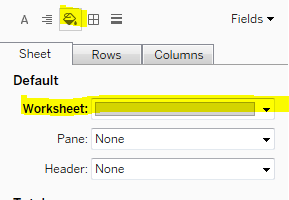
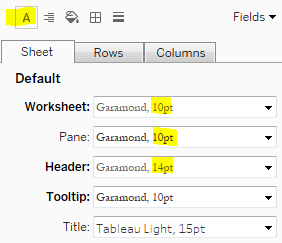


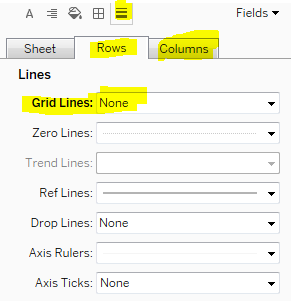
Insights:

* Asia has rapidly grown to be the largest emitter of CO2 of the world regions, with China, India, and Japan leading the way.
* The U.S. has historically been a disproportionate CO2 emitter with respect to the rest of the world, surpassed only recently by China.
* Germany wasn’t included until 1992, nor The Russian Federation until 1992, shortly after the fall of the Iron Curtain and Berlin Wall. Both have been significant emitters, for as long as there has been data emitting ~4% of world’s total emissions. Total emissions probably significantly underestimated prior to 1991?

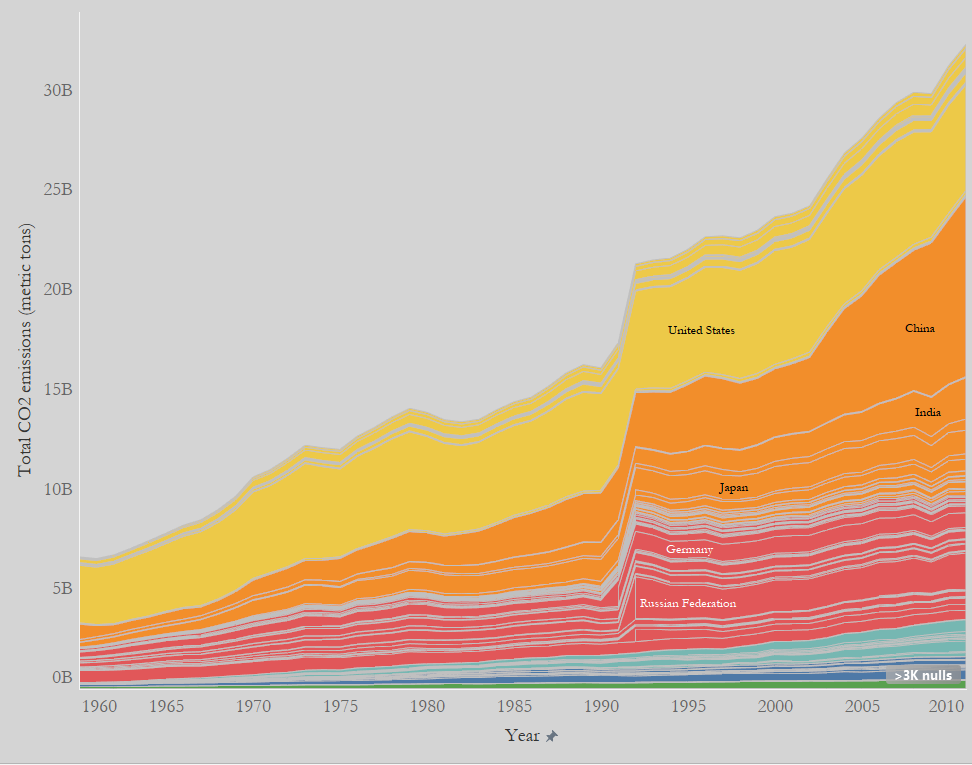
Let’s format this area chart a bit, then add it to a dashboard with these insights.

* Edit Y-axis to indicate the units (metric tons); double-click on the axis to edit.
* Hide the title of the chart by right-clicking “area chart” and clicking “Hide Title”
* Right-click anywhere in the area chart and click “Format.” Note the formatting pane appears at left. Increase the axis font size (and change it if you wish), and get rid of the grid bars using the “Font”, “Shading,” and “Lines” tools as shown:

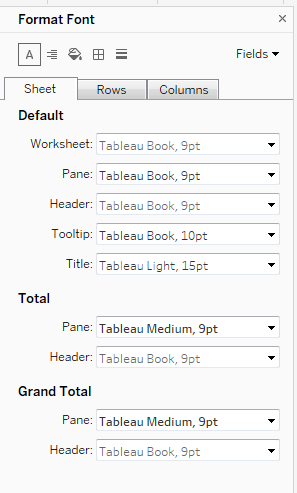




Formatted area chart:



**Brief Aside: The formatting pane**



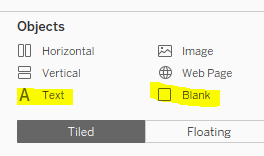
The formatting pane has 5 tools:

* Text 
* Alignment 
* Shading 
* Borders 
* Lines 

They are pretty self-explanatory, but a few components bear emphasis:

* Worksheet formats will be applied to the *entire worksheet* (the pane *and* the headers)
* Pane formats will be applied to the pane only (e.g., the country labels in the area chart)
* Header formats will be applied to the headers only (axis titles/tick mark labels)

Final dashboard is shown below. Note that dashboard shading is accomplished by clicking **Dashboard 🡪 Format Dashboard**; this brings up a special “Format Dashboard” menu that applies to the entire dashboard rather than individual worksheets. I have also used a tiled text object for the new title, and tiled blank objects to space the “Insights” text object the way I want it:



Blank object

Text object

Blank object

Text object

