

## Exploratory Graphical Analysis Using Tableau Public

### Challenging Tutorial to Create Tableau Dashboard

This week, we will work on a visualization inspired by [this Gapminder video](#) (Hans Rosling), using data from the World Bank's [World Development Indicators](#) (WDI).

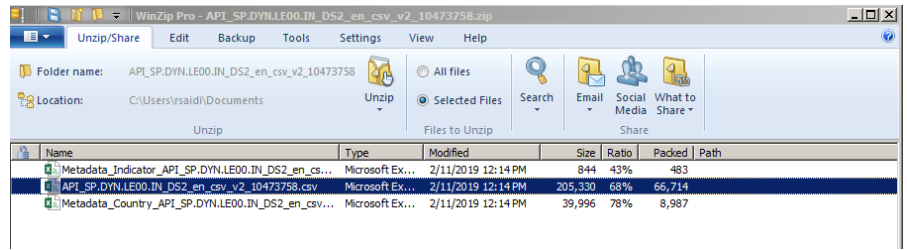
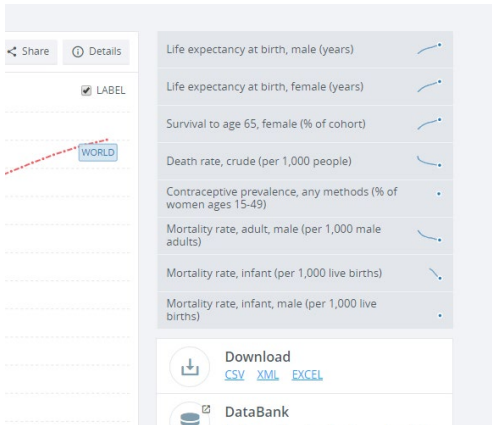
Use the `nations.csv` data from our class google drive. This dataset lacks the variable giving life expectancy at birth.

- `nations.csv` Data from the [World Bank Indicators](#) portal, which is an incredibly rich resource. Contains the following fields:
  - `iso2c iso3c` Two- and Three-letter [codes](#) for each country, assigned by the [International Organization for Standardization](#).
  - `country` Country name.
  - `year`
  - `population` Estimated [total population](#) at mid-year, including all residents apart from refugees.
  - `gdp_percap` [Gross Domestic Product per capita](#) in current international dollars, corrected for purchasing power in different territories.
  - `life_expect` [Life expectancy at birth](#), in years.
  - `population` Estimated [total population](#) at mid-year, including all residents apart from refugees.
  - `birth_rate` [Live births during the year per 1,000 people](#), based on mid-year population estimate.
  - `neonat_mortal_rate` [Neonatal mortality rate](#): babies dying before reaching 28 days of age, per 1,000 live births in a given year.
  - `region income` World Bank [regions and income groups](#), explained [here](#).

Now we will load the other data file we want to join with life expectancy at birth by nation from the World Bank. The steps for loading the dataset, `life_exp.csv` are as follows, but you will need to use R to convert from wide to long format. Alternatively, you can simply load the processed dataset `life_exp_long.csv` from the course datasets.

#### Step 1 (getting the data and cleaning it up):

- Download data in csv format for life expectancy at birth by nation over time from [here](#). Scroll down to **Download** then select CSV. You only need the API file (the others are the metadata files)



- Open this data in a spreadsheet and edit to remove the couple of rows above the header row, and remove columns to leave just the country names, country codes, and data from 1990 to 2016 (remove the two indicator columns and the empty 2017, 2018 columns). Save it as `life_expect.csv`
- Process this data using R to convert from wide to long format. When doing this, I would call the new column with life expectancy data `life_expect`, so that it can be used on other World Bank data in the same format. Use Open Refine to rename `Country Code` to `iso3c` and rename `Country Name` to `country`. Then the variable names will match those in `nations.csv`.
- Export the data in CSV format to your data folder as `life_exp_long.csv`.

## Step 2 (merge the 2 data sets using Tableau):

Download Tableau Public from <https://public.tableau.com/en-us/s/> (click the “Download the App” button) if you have not already done so.

## Launch Tableau Public Desktop

Under the Connect heading at top left, select Text File, navigate to the file `nations.csv` and Open. At this point, you can view the data, which will be labeled as follows:

- Text: Abc
- Numbers: #
- Dates: calendar symbol
- Geography: globe symbol

You can edit fields to give them the correct data type if there are any problems.

Once you load `nations` into Tableau, double click on `nations` at the top and the Venn diagram will appear.

nations

Then add life\_exp\_long (csv)

nations.csv is made of 1 table. ⓘ

nations.csv

nations.csv is made of 2 tables. ⓘ

nations.csv

life\_exp\_long.csv

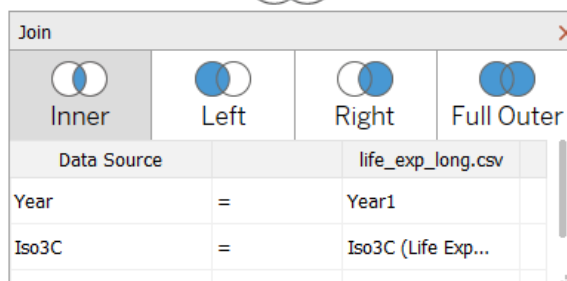
Click on the Venn diagram for the options for variables you want to merge to appear.

sv is made of 2 tables. ⓘ

sv life\_exp\_long.csv



Add “year” and “iso3c” for each of the variables to merge.



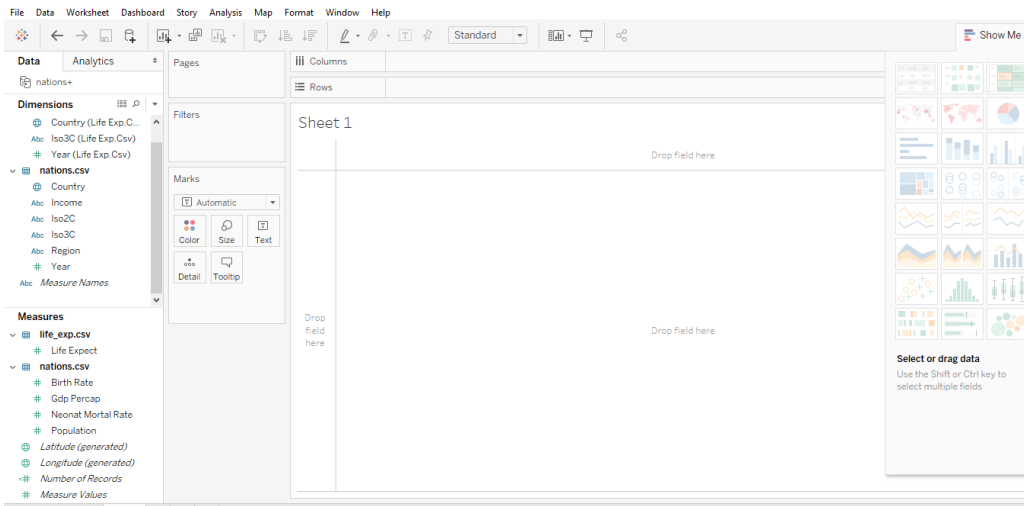
To control how the join is performed, click on the circles with the blue area of overlap. Remember that there is one row in this data for each country in each year. So to join the data appropriately, we need to join by **year** and the **iso3c** country code (joining by unambiguous codes is generally a better idea than joining by names). Choosing an **inner join** will retain rows from each file only where there are matches.

The dataset is ready to be used.

Click on Sheet 1 at the bottom:

United Arab Emirates	ARE	2001	74.7129	AE	ARE	United Arab Emirates	2001	85.1
United Arab Emirates	ARE	1992	72.2964	AE	ARE	United Arab Emirates	1992	73.1
United Arab Emirates	ARE	1994	72.8456	AE	ARE	United Arab Emirates	1994	74.1
United Arab Emirates	ARE	2007	76.1112	AE	ARE	United Arab Emirates	2007	75.1
United Arab Emirates	ARE	2004	75.4573	AE	ARE	United Arab Emirates	2004	87.1
United Arab Emirates	ARE	1996	73.3839	AE	ARE	United Arab Emirates	1996	79.1
United Arab Emirates	ARE	2006	75.9045	AE	ARE	United Arab Emirates	2006	82.1

The joined data should look like this in a Tableau worksheet. You can see which variables came from which file:

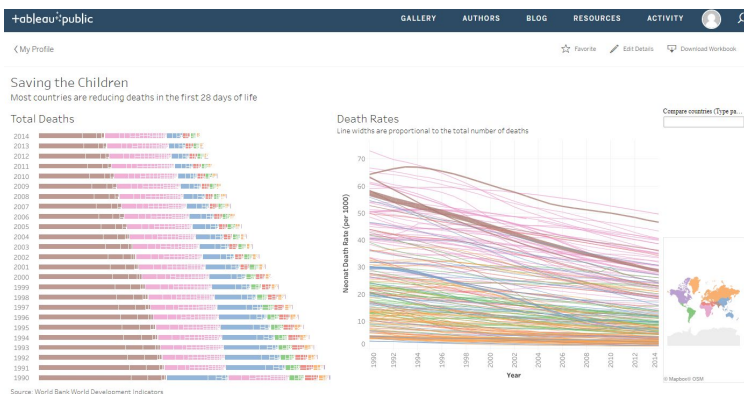


## Back to Tableau Public

Here is an [introductory training video on beginning to create a visualization in Tableau](#) you can watch at your leisure.

*The following is a thorough introductory [Lesson from Peter Aldhous](#) to create a data viz in Tableau.*

Here is what you will be creating in Tableau “*Saving the Children*”



## Dimensions and measures: categorical and continuous

The fields should appear in the Data panel at left. Notice that Tableau has divided the fields into Dimensions and Measures. These broadly correspond to categorical and continuous variables. Dimensions are fields containing text or dates, while measures contain numbers.

If any field appears in the wrong place, click the small downward-pointing triangle that appears when it is highlighted and select **Convert to Dimension** or **Convert to Measure** as required.

## Shelves and Show Me

Notice that the main panel contains a series of “shelves,” called `Pages`, `Columns`, `Rows`, `Filters` and so on. Tableau charts and maps are made by dragging and dropping fields from the data into these shelves.

Over to the right you should see the `Show Me` panel, which will highlight chart types you can make from the data currently loaded into the `Columns` and `Rows` shelves. It is your go-to resource when experimenting with different visualization possibilities. You can open and close this panel by clicking on its title bar.

## Columns and rows: X and Y axes

The starting point for creating any chart or map in Tableau is to place fields into `Columns` and `Rows`, which for most charts correspond to the X and Y axes, respectively. When making maps, longitude goes in `Columns` and latitude in `Rows`. If you display the data as a table, then these labels are self-explanatory.

## Some questions to ask this data

- How has the **total number of neonatal deaths** changed over time, globally, regionally, and nationally?
- How has the **neonatal death rate** for each country changed over time?

## Create new calculated variables

The data contains fields on birth and neonatal death rates, but not the total numbers of births and deaths, which must be calculated. From the top menu, select `Analysis > Create Calculated Field`. Fill in the dialog box as follows (just start typing the field name `Births` to select it for use in a formula):

Births

$[Birth Rate] * [Population] / 1000$

The calculation is valid.

Apply OK

Notice that calculated fields appear in the `Data` panel preceded by an `=` symbol.

Now create a second calculated field `Neonatal deaths` giving the total number of neonatal deaths:



In the second formula, we have rounded the number of neonatal deaths to the nearest thousand using -3 (-2 would round to the nearest hundred, -1 to the nearest ten, 1 to one decimal place, 2 to two decimal places, and so on.). Check that `Neonatal deaths` ends up in your **Measures** as well as the **Births**.

Here we run some simple arithmetic, but it's possible to use a wide variety of functions to manipulate data in Tableau in many ways. To see all of the available functions, click on the little gray triangle at the right of the dialog boxes.

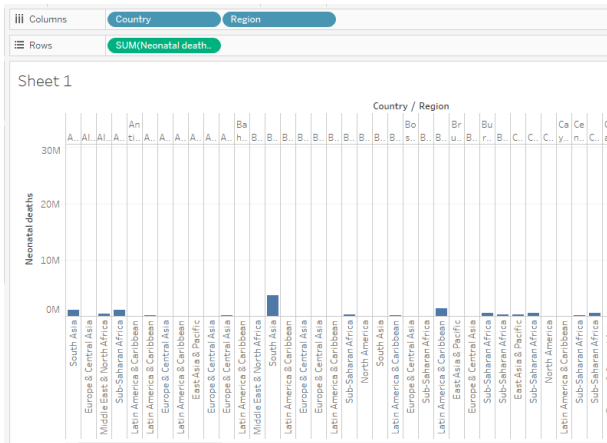
## Understand that Tableau's default behavior is to summarize/aggregate data

As we work through today's exercise, notice that Tableau routinely summarizes or aggregates measures that are dropped into Columns and Rows, calculating a SUM or AVG (mean), for example. This behavior *can* be turned off by selecting **Analysis** from the top menu and unchecking **Aggregate Measures**. However, it is not a good idea to do this, as it will disable some Tableau functions. Instead, if you don't want to summarize *all* of the data, drop categorical variables into the **Detail** shelf so that any summary statistic will be calculated at the correct level for your analysis. If necessary, you can set the aggregation so it is being performed on a single data point, and therefore has no effect.

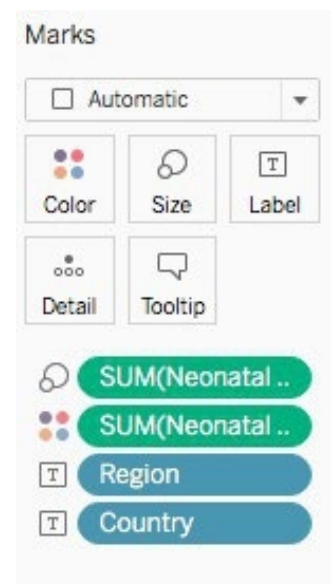
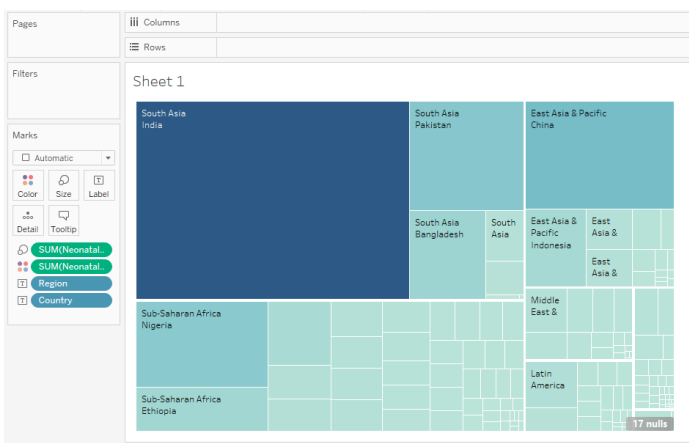
## Make a series of treemaps showing neonatal deaths over time

Recall making treemaps in R at the beginning of the course. A treemap allows us to directly compare the neonatal deaths in each country, nested by region.

Drag `Country` and `Region` onto Columns and `Neonatal deaths` onto Rows.

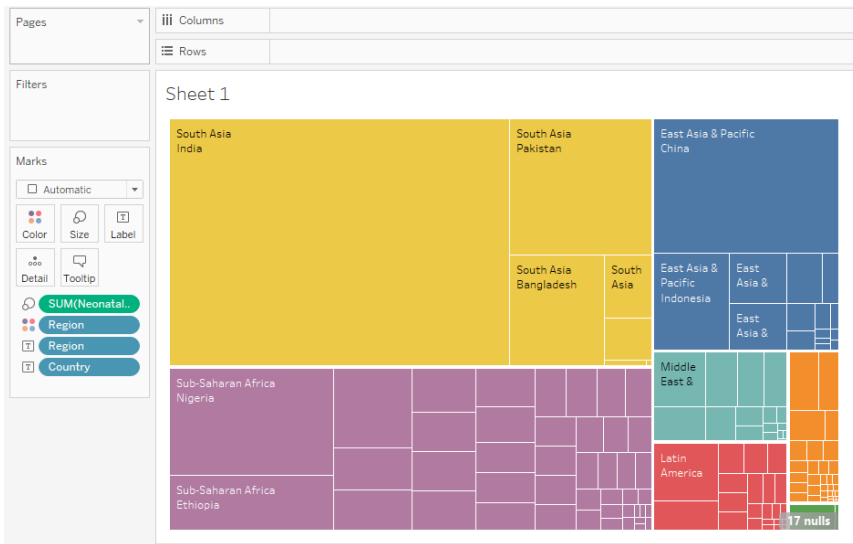


Then open Show Me and select the treemap option. The initial chart should look like this:

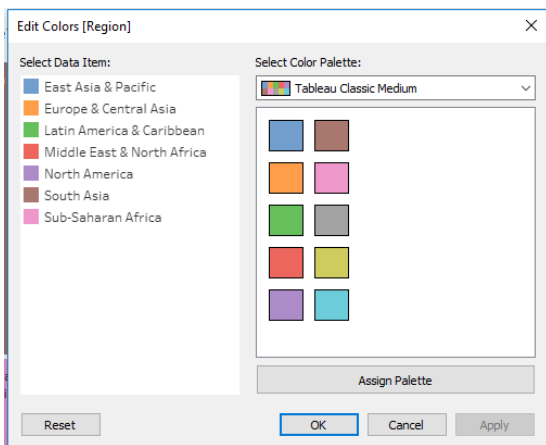


Look at the Marks shelf and see that the size and color of the rectangles reflect the SUM of Neonatal deaths for each country, while each rectangle is labeled with Region and Country:

Now drag Region (from Dimensions) to Color to remove it from the Label and color the rectangles by region, using Tableau's default qualitative color scheme for categorical data:



For a more subtle color scheme, click on **Color**, select **Edit Colors...** and at the dialog box select the **Tableau Classic Medium** qualitative color scheme, then click **Assign Palette** and **OK**.



(Tableau's qualitative color schemes are well designed, so there is no need to adopt a ColorBrewer scheme. However, it is possible to edit colors individually as you wish.)

Click on **Color** and set transparency to 75%. (For your assignment you will create a chart with overlapping circles, which will benefit from using some transparency to allow all circles to be seen. So we are setting transparency now for consistency.)

The treemap should now look like this

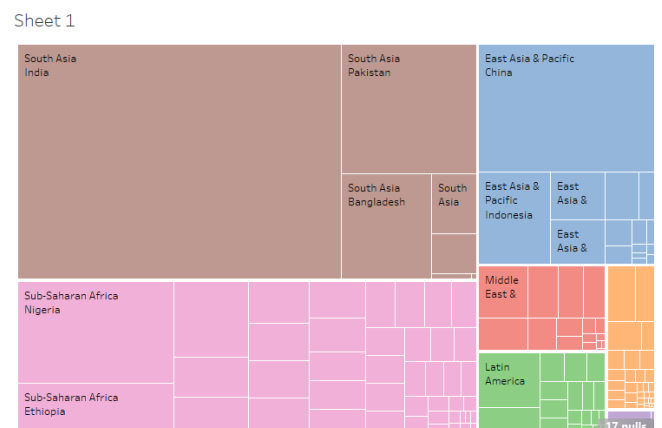
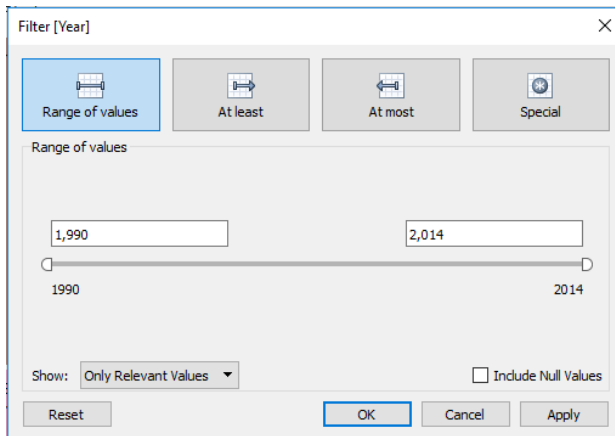




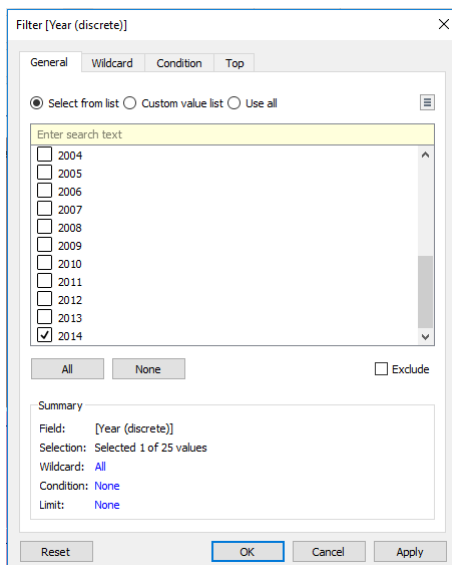
Tableau has by default aggregated `Neonatal deaths` using the `SUM` function, so what we are seeing is the number for each country added up across the years.

To see one year at a time, we need to filter by year. If you drag the existing `Year` variable to the `Filters` shelf, you will get the option to filter by a range of numbers, which is not what we need:



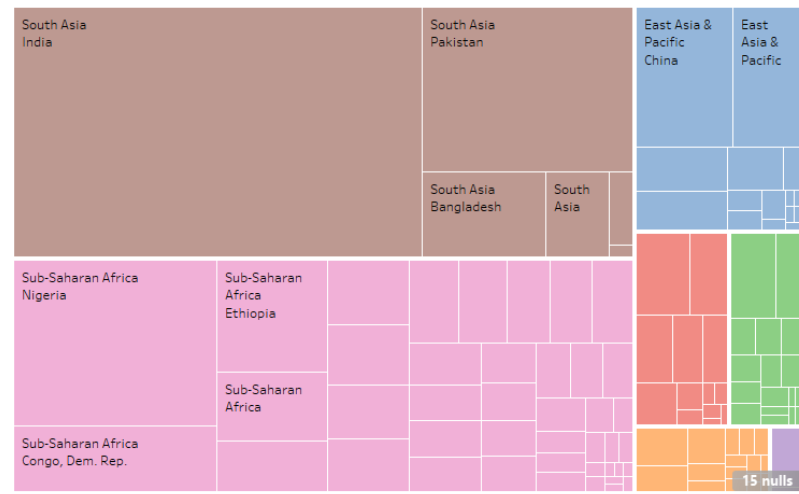
Instead, we need to be able check individual years, and draw a treemap for each one. To do that, select `Year` in the `Dimensions` panel and `Duplicate`.

Select the new variable and `Convert to Discrete` and then `Rename` it `Year (discrete)`. Now drag this new variable to `Filters`, select 2014, and click `OK`:



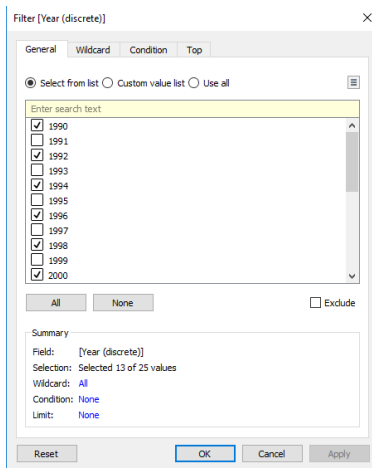
The treemap now displays the data for 2014:

Sheet 1

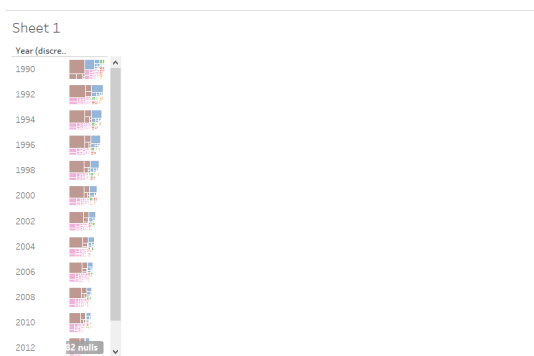


That is good for a snapshot of the data, but with a little tinkering, we can adapt this visualization to show change in the number of neonatal deaths over time at the national, regional and global levels.

Select **Year (discrete)** in the **Filters** shelf and **Filter . . .** to edit the filter. Select all the years with even numbers and click **OK**:



Now drag **Year (discrete)** onto **Rows** and the chart should look like this:

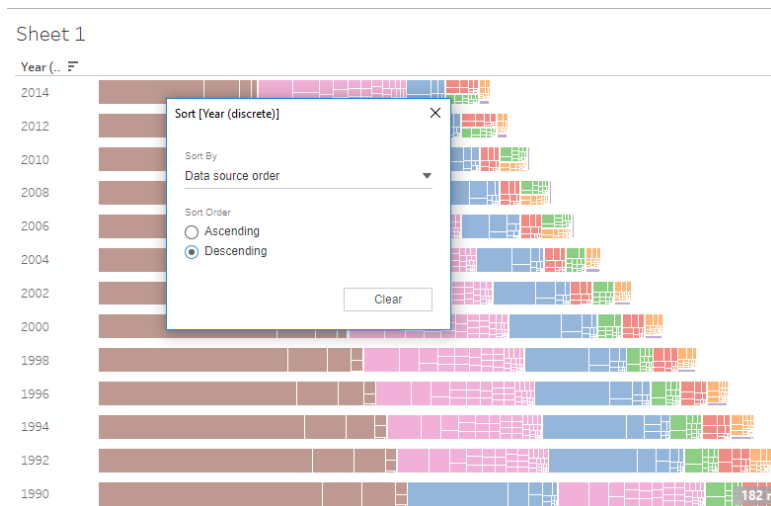


The formatting needs work, but notice that we now have a bar chart made out of treemaps.

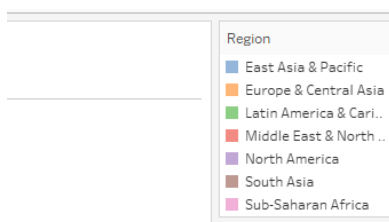
Extend the chart area to the right by changing from Standard to Entire View on the dropdown menu in the top ribbon:



I find it more intuitive to have the most recent year at the top, so select `Year (discrete)` in the `Rows` shelf, select `Sort` and fill in the dialog box so that the years are sorted in `Descending` order:



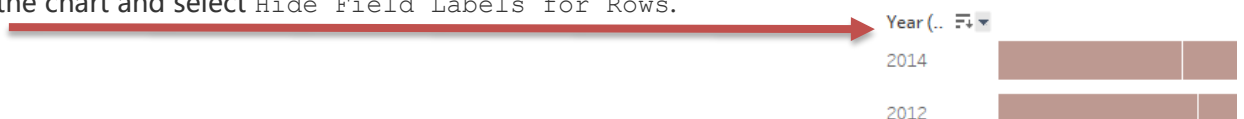
We will create a map to serve as a legend for the regions, so click on the title bar for the color legend and select `Hide Card` to remove it from the visualization.



On the down-arrow to the right, select **Hide Card**

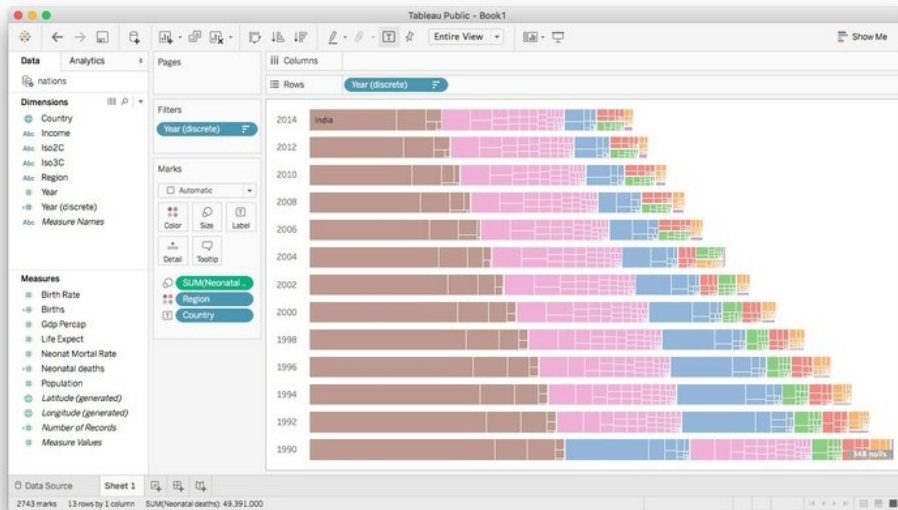
To remove some clutter from the chart, select `Format>Borders` from the top menu, and under `Sheet>Row Divider`, set `Pane` to `None`. Then close the `Format Borders` panel.

Right-click on the `Sheet 1` title for the chart and select `Hide Title`. Also right-click on `Year (discrete)` at the top left of the chart and select `Hide Field Labels for Rows`.



Then hover just above the top bar to get a double-headed drag symbol and drag upwards to reduce the white space at the top. You may also want to drag the bars a little closer to the year labels.

The labels will only appear in the larger rectangles. The chart should now look like this:



Hover over one of the rectangles, and notice the tooltip that appears. By default, all the fields we have used to make the visualization appear in the tooltip. (If you need any more, just drag those fields onto `Tooltip`.) Click on `Tooltip` and edit as follows. (Unchecking `Include command buttons` disables some interactivity, giving a plain tooltip):

A screenshot of the 'Edit Tooltip' dialog box in Tableau. The dialog has a title bar 'Edit Tooltip' and a close button. It contains a text area with the following content: 'Country: <Country>' and 'Neonatal deaths: <SUM(Neonatal deaths)>'. Below the text area are three checkboxes: 'Show tooltips' (checked), 'Include command buttons' (unchecked), and 'Allow selection by category' (checked). The 'Show tooltips' checkbox has a dropdown menu set to 'Responsive - Show tooltips instantly'. At the bottom are buttons for 'Reset', 'Preview', 'OK', and 'Cancel'.

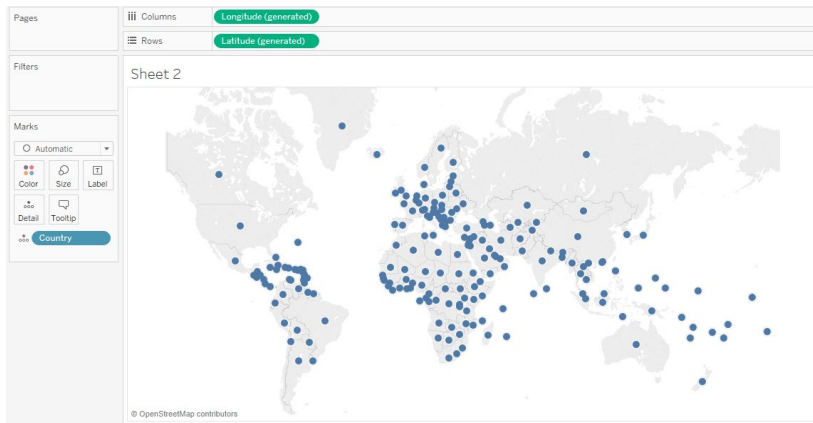
## Save to the web

Right-click on `Sheet 1` at bottom left and `Rename Sheet` to `Treemap bar chart`. Then select `File>Save to Tableau Public...` from the top menu. At the logon dialog box enter your Tableau Public account details,

give the Workbook a suitable name and click *Save*. When the save is complete, a view of the visualization on Tableau's servers will open in your default browser.

## Make a map to use as a color legend

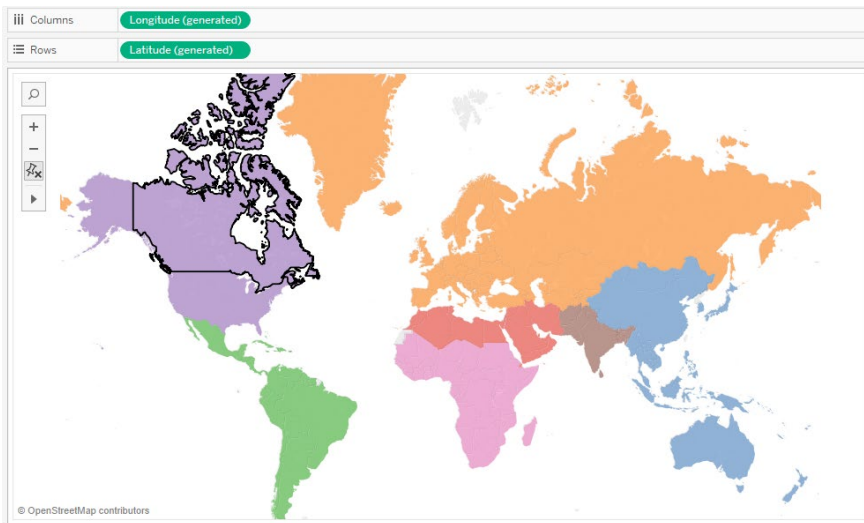
Select *Worksheet>New Worksheet* from the top menu of Tableau Desktop, and double-click on *Country*. Tableau recognizes the names of countries and states/provinces; for the U.S., it also recognizes counties. Its default map-making behavior is to put a circle at the geographic center, or centroid, of each area, which can be scaled and colored to reflect values from the data:



However, we need each country to be filled with color by region. Using *Show Me*, switch to the *filled maps* option, and each nation should fill with color. Drag *Region* to *Color* and see how the same color scheme we used previously carries over to the map. Click on *Color*, set the transparency to 75% to match the bubble chart and remove the borders in the same *Color* setting. Also click on *Tooltip* and uncheck *Show tooltip* so that no tooltip appears on the legend.

We will use this map as a color legend, so its separate color legend is unnecessary. Click the color legend's title bar and select *Hide Card* to remove it from the visualization. Also remove the *Sheet 2* title as before.

Center the map in the view by clicking on it, holding and panning, just as you would on Google Maps. It should now look something like this:



Rename the worksheet `Map` `legend` and save to the web again.

Make a line chart showing neonatal mortality rate by country over time

To address our second question, and explore the neonatal death rate over time by country, we can use a line chart.

First, rename `Neonat Mortal` as `Neonatal death rate (per 1,000 births)`. Then, open a new worksheet, drag this variable to `Rows` and `Year` to `Columns`. The chart should now look like this:

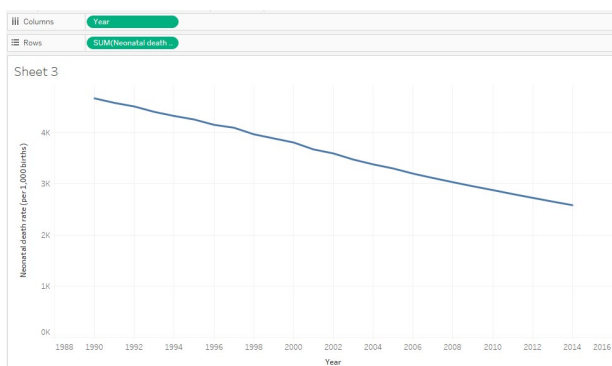
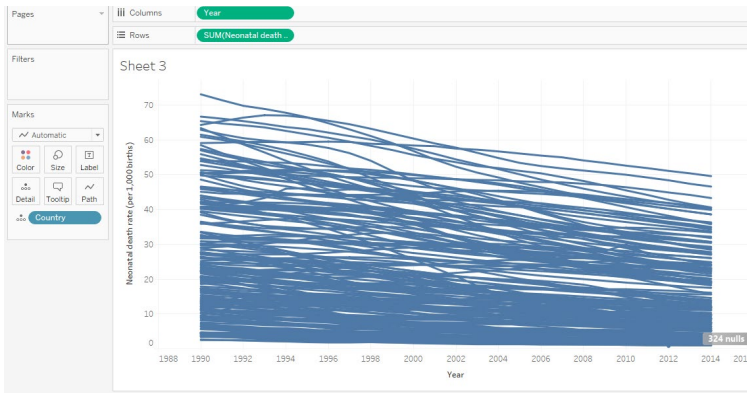
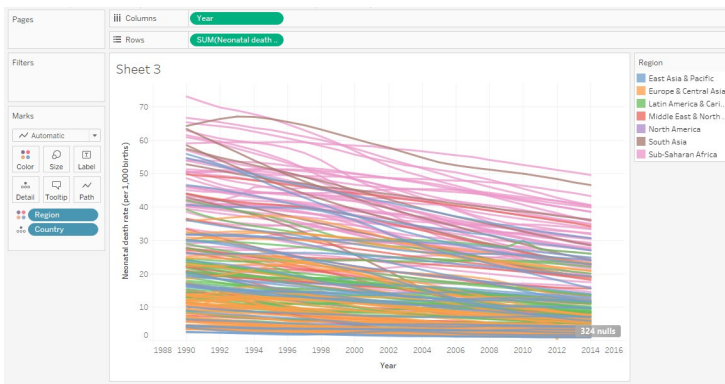


Tableau has aggregated the data by adding up the rates for each country in every year, which makes no sense here. So drag `Country` to `Detail` in the `Marks` shelf to draw one line per country:



Drag region to **Color** and set the transparency to 75%.



Now right-click on the X axis, select **Edit Axis**, edit the dialog box as follows and click **OK**:

Edit Axis [Year]

General

Tick Marks

Range

☐ Automatic
 ☐ Uniform axis range for all rows or columns
 ☐ Independent axis ranges for each row or column
 ☒ Fixed

Fixed start

Fixed end

1990

2,016

Scale

☐ Reversed
 ☐ Logarithmic
 

☒ Positive
 ☐ Symmetric

Axis Titles

Title

Year

Subtitle

☒ Automatic

Reset

Right-click on the X axis again, select **Format**, change **Alignment** to **Up** and use the dropdown menu set the **Font** to bold. Also remove the Sheet 3 title.

The chart should now look like this:

Format Year

A

Fields

Axis

Pane

Default

Font

Tableau Boo..

Shading

Scale

Ticks

Numbers

123456

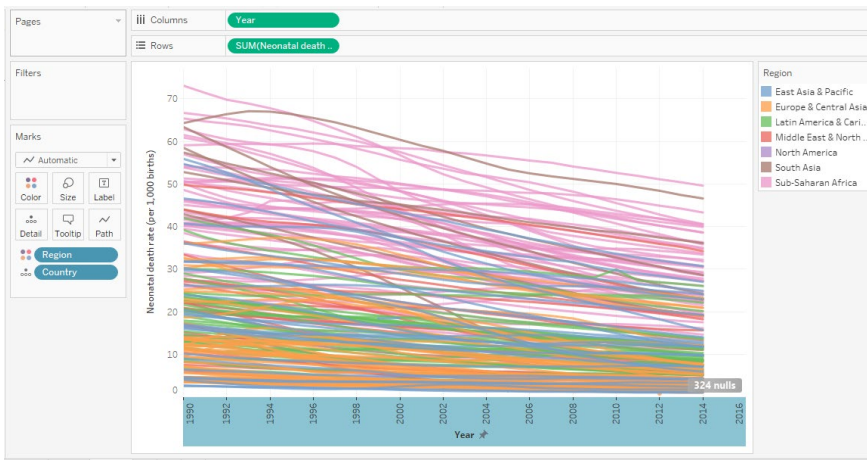
Alignment

Up

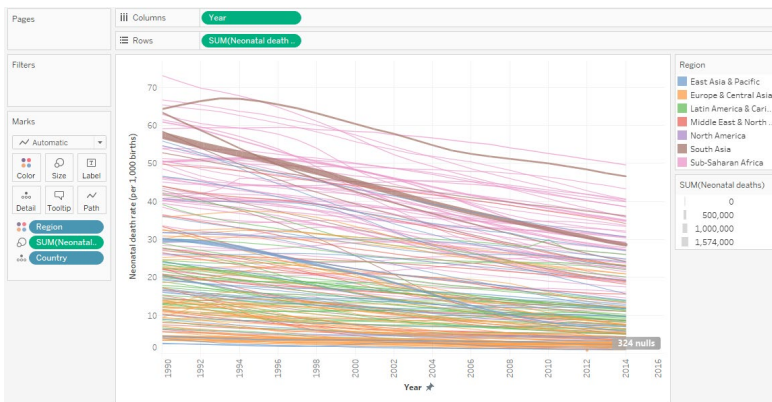
Title

Font

Tableau Me..



We can also highlight the countries with the highest total number of neonatal deaths by dragging `Neonatal deaths` to `Size`. The chart should now look like this:

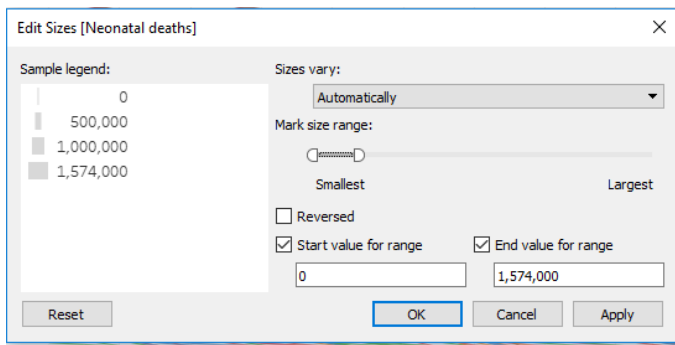


This line chart shows that the trend in most countries has been to reduce neonatal deaths, while some countries have had more complex trajectories. But to make comparisons between individual countries, it will be necessary to add controls to filter the chart.

Tableau's default behavior when data is filtered is to redraw charts to reflect the values in the filtered data. So if we want the Y axis and the line thicknesses to stay the same when the chart is filtered, we need to freeze them.

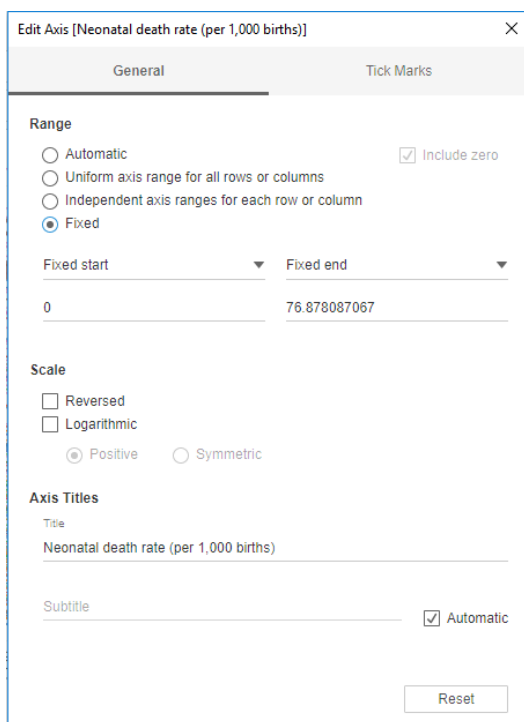
To freeze the line thicknesses, hover over the title bar for the line thickness legend, select `Edit Sizes...` and fill in the dialog box as follows:





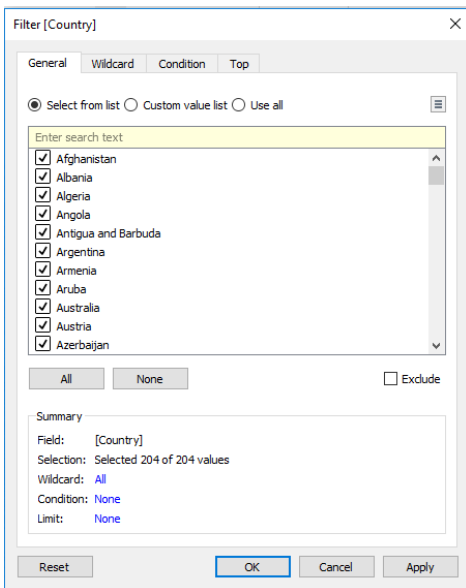
Now remove this legend from the visualization, together with the color legend. We can later add an annotation to our dashboard to explain the line thickness.

To freeze the Y axis, right-click on it, select `Edit Axis...`, make it `Fixed` and click `OK`:

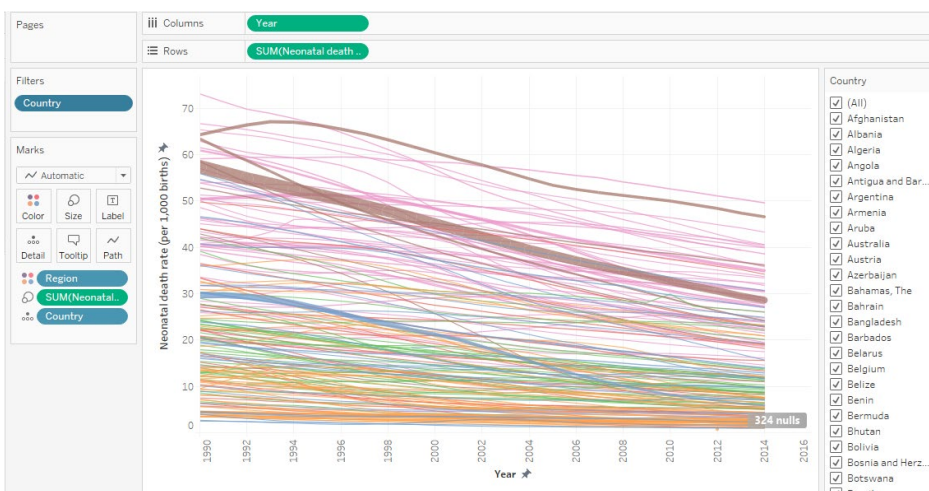


Right-click on the Y axis again, select `Format...` and increase the font size to `10pt` to make it easier to read.

Now drag `Country` to `Filters`, make sure `All` are checked, and at the dialog box, click `OK`:



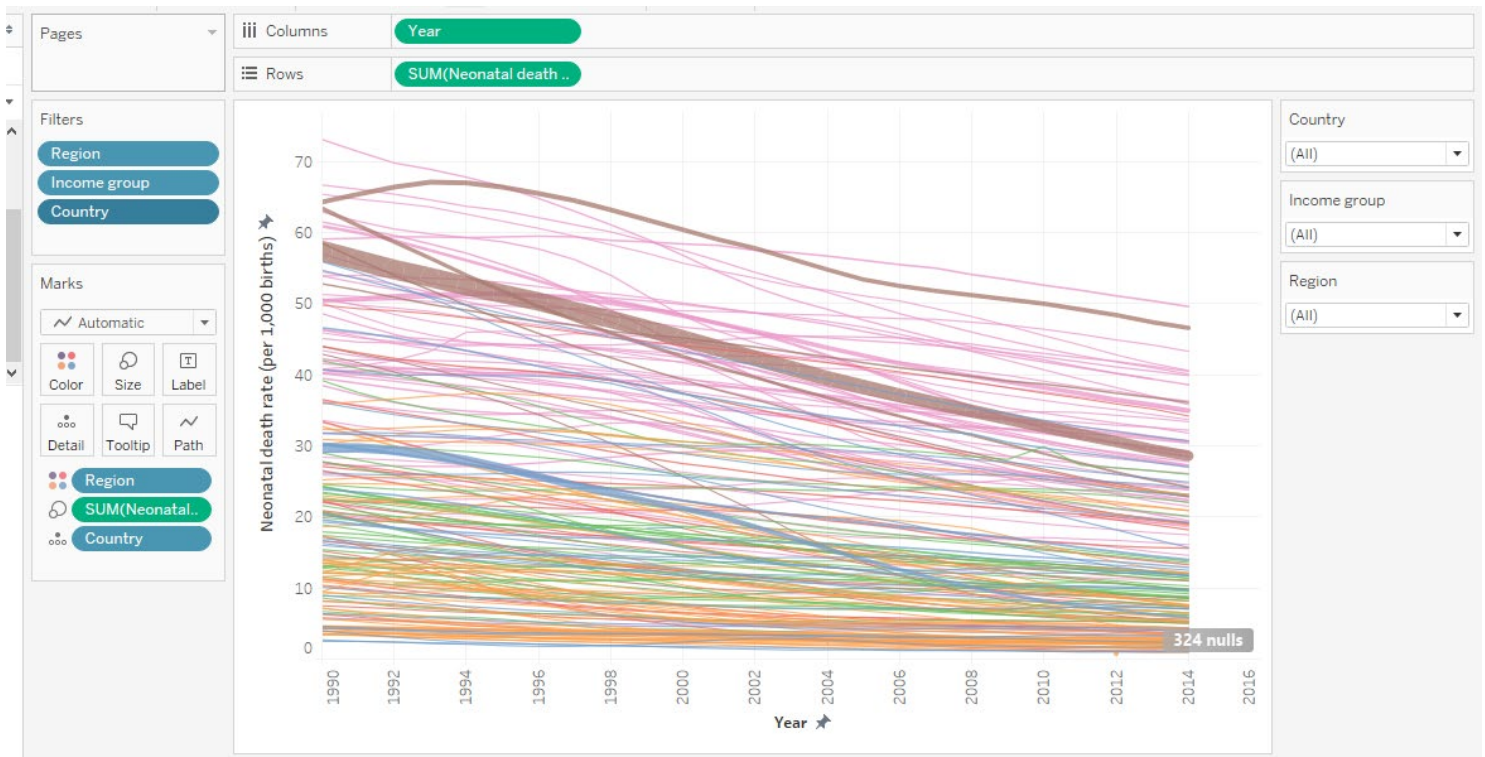
Now we need to add a filter control to select countries to compare. On `Country` in the `Filters` shelf, select `Show Filter`. A default filter control, with a checkbox for each nation, will appear to the right of the chart:



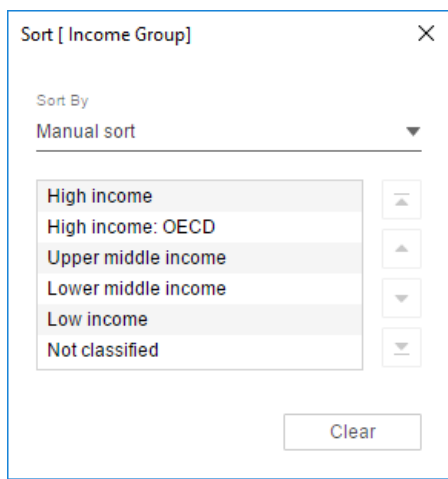
This isn't the best filter control for this visualization. To change it, click on the title bar for the filter, note the range of filter controls available, and select `Multiple Values (Custom List)`. This allows users to select individual countries by starting to type their names.

Take some time to explore how this filter works.

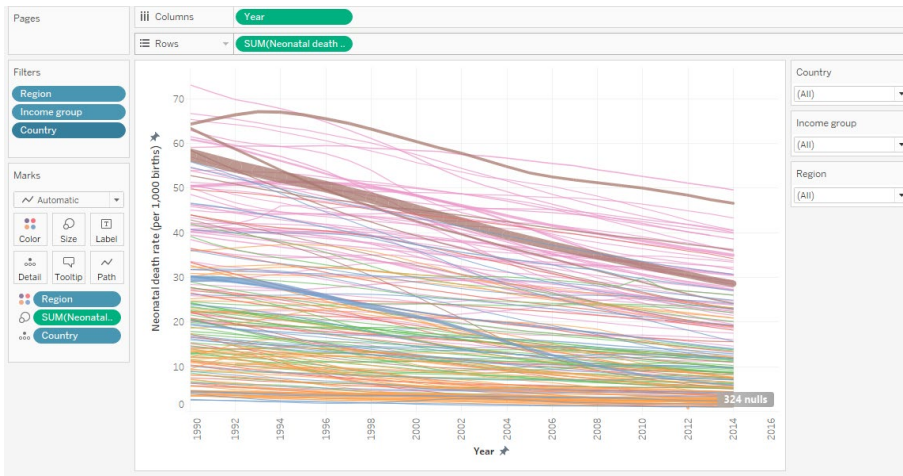
Rename `Income` to `Income group`. Then add `Region` and `Income group` to `Filters`, making sure that `All` options are checked for each. Select `Show Filter` for both of these filters, and select `Single Value Dropdown` for the control. Reset both of these filters to `All`, and the chart should now look like this:



Notice that the `Income group` filter lists the options in alphabetical order, rather than income order, which would make more sense. To fix this, right-click on `Income group` in the data panel and select `Default Properties>Sort`. At the dialog box below, select `Manual sort`, edit the order as follows and click `OK`:



The chart should now look like this:

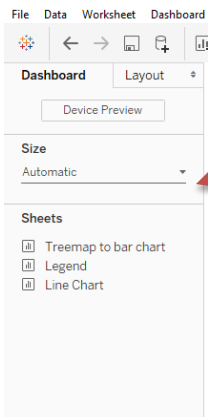


Finally, click on **Tooltip** and edit as follows:

Rename the sheet **Line chart** and save to the web.

## Make a dashboard combining all charts

From the top menu, select **Dashboard > New Dashboard**. Set its **Size** to **Automatic**, so that the dashboard will fill to the size of any screen on which it is displayed:



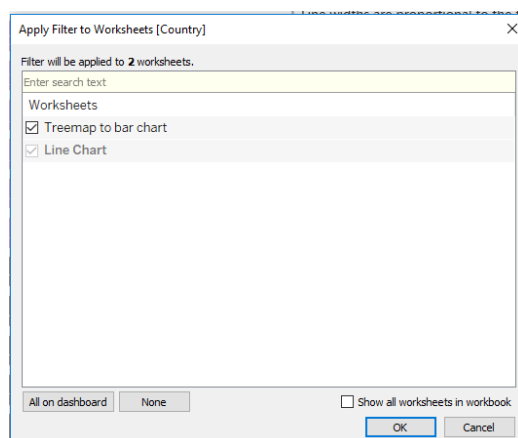
To make a dashboard, drag charts, and other elements from the left-hand panel to the dashboard area. Notice that Tableau allows you to add items including: horizontal and vertical containers, text boxes, images (useful for adding a publication's logo), embedded web pages and blank space. These can be added **Tiled**, which means they cannot overlap, or **Floating**, which allows one element to be placed over another.

Drag **Treemap bar chart** from the panel at left to the main panel. The default title, from the worksheet name, isn't very informative, so right-click on that, select **Edit Title ...** and change to **Total deaths**.

Now add **Line Chart** to the right of the dashboard (the gray area will show where it will appear) and edit its title to **Death rates**. Also add a note to explain that *"line widths are proportional to the total number of deaths."* The dashboard should now look like this:



Notice that the **Country**, **Region** and **Income group** filters control only the line chart. To make them control the treemaps, too, click on each filter, open up the dropdown menu from the downward-pointing triangle, and select **Apply to Worksheets>Selected Worksheets...** and fill in the dialog box as follows:



The filters will now control both charts.

Add `Map legend` for a color legend at bottom right. (You will probably need to drag the window for the last filter down to push it into position.) Hide the legend's title then right-click on the map and select `Hide View Toolbar` to remove the map controls.



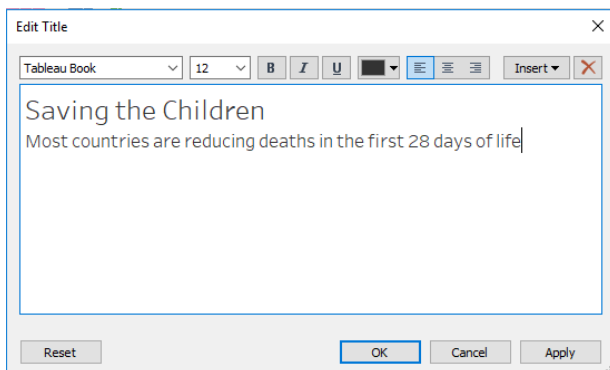
We can also allow the highlighting of a country on one chart to be carried across the entire dashboard. Select `Dashboard>Actions...` from the top menu, and at the first dialog box select `Add action>Highlight`. Filling the second dialog box as follows will cause each country to be highlighted across the dashboard when it is clicked on just one of the charts:

The 'Add Highlight Action' dialog box is shown. It has a 'Name' field set to 'Highlight1'. Under 'Source Sheets', 'Dashboard 1' is selected, and 'Legend', 'Line Chart', and 'Treemap to bar chart' are checked. The 'Run action on:' buttons are 'Hover', 'Select', and 'Menu'. Under 'Target Sheets', 'Dashboard 1' is selected, and 'Legend', 'Line Chart', and 'Treemap to bar chart' are checked. Under 'Target Highlighting', 'Selected Fields' is selected, and a list of fields is shown: 'Country' (checked), 'Region' (unchecked), 'Year' (unchecked), and 'Year (discrete)' (unchecked). 'OK' and 'Cancel' buttons are at the bottom.

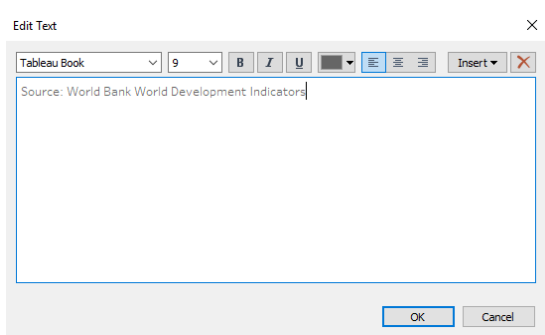


Click **OK** on both dialog boxes to apply this action.

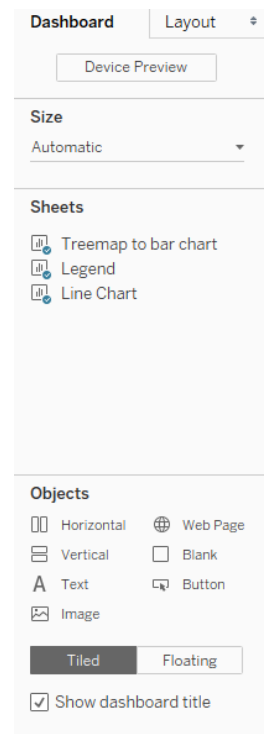
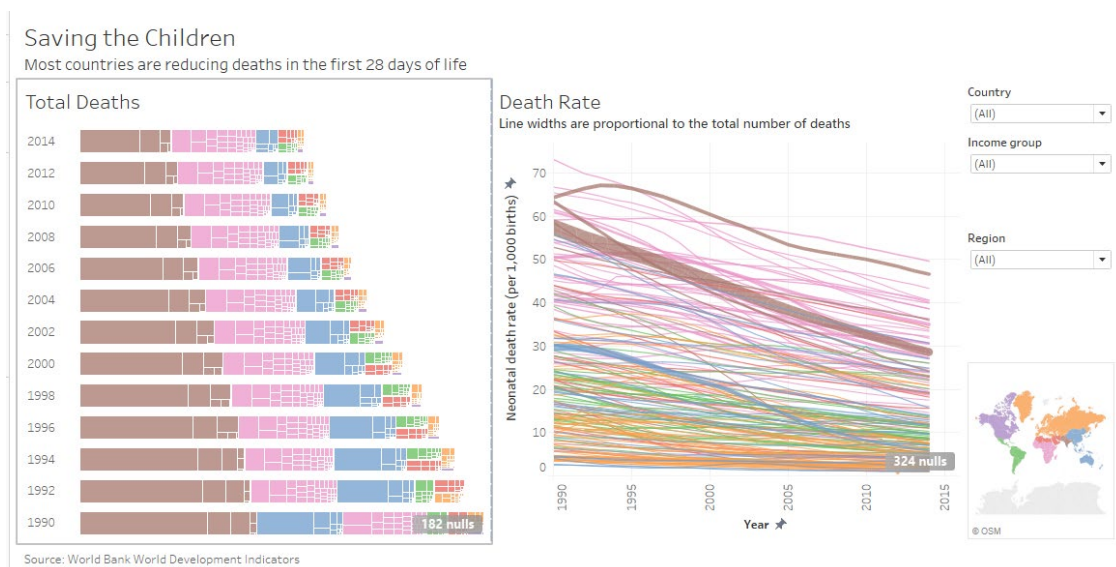
Select **Dashboard>Show Title** from the top menu. Right-click on it, select **Edit Title...** and change from the default to something more informative:



Now drag a **Text** box to the bottom of the dashboard and add a footnote giving source information: (Source: World Bank World Development Indicators)



The dashboard should now look like this:



Save this dashboard one more time to your Tableau Public account.

Now save to the web once more. Once the dashboard is online, use the `Share` link at the bottom to obtain an embed code, which can be inserted into the HTML of any web page.

## Further reading/viewing

[Tableau Public training videos](#)

[Gallery of Tableau Public visualizations](#): Again, you can download the workbooks to see how they were put together.

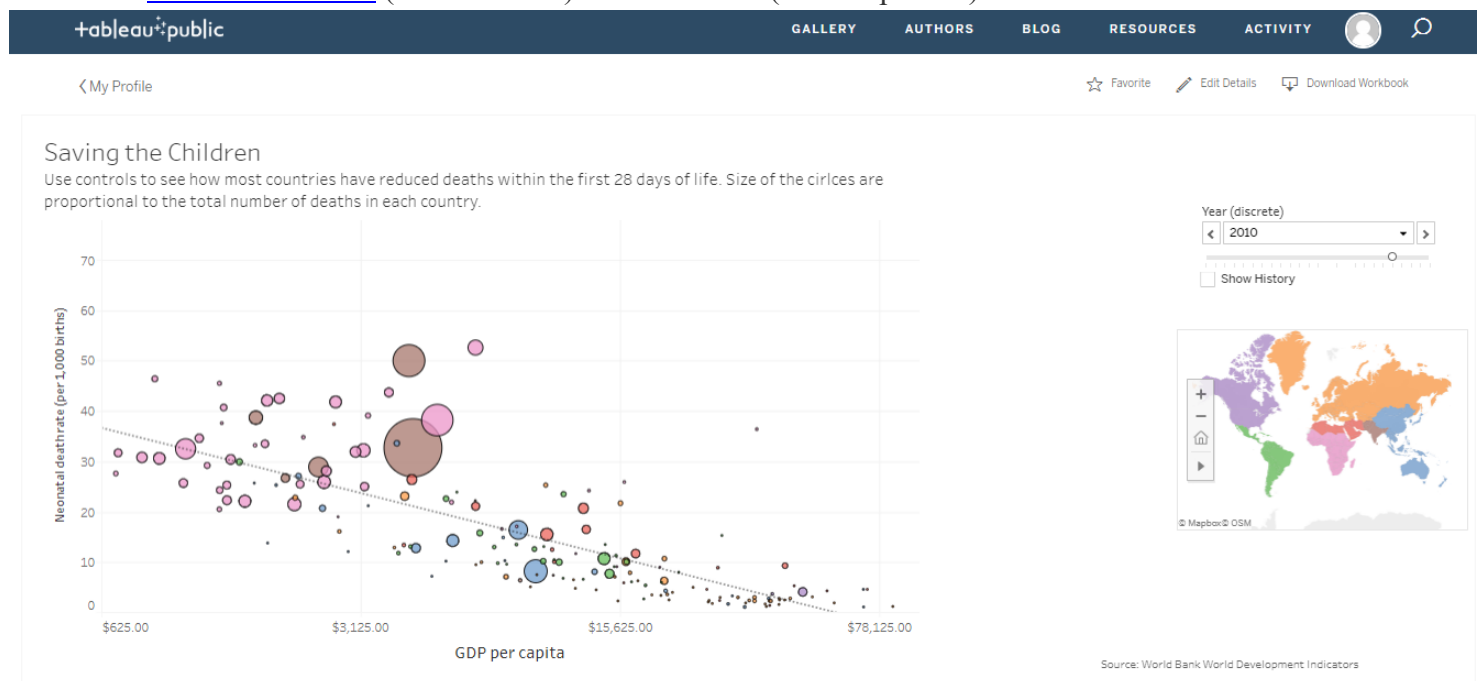
[Tableau Public Knowledge Base](#): Useful resource with the answers to many queries about how to use the software.

## Week 12 Tableau Public HW Assignment

1. Follow the directions to complete the “Saving the Children” Dashboard. Save your completed dashboard and **submit your link in the Week 12 assignment dropbox by Tuesday, \_\_\_\_\_**.
2. If you would like additional practice working in Tableau, I have provided two extra visualizations (with hints) you can try to complete for extra credit. **Each of these is worth up to 1.5 points extra credit.** Have fun!


## Optional Additional Tableau Practice #1

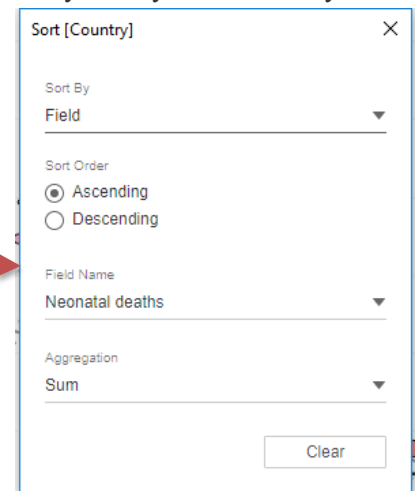
Create the [second dashboard](#) (shown below) from the data. (this is optional)





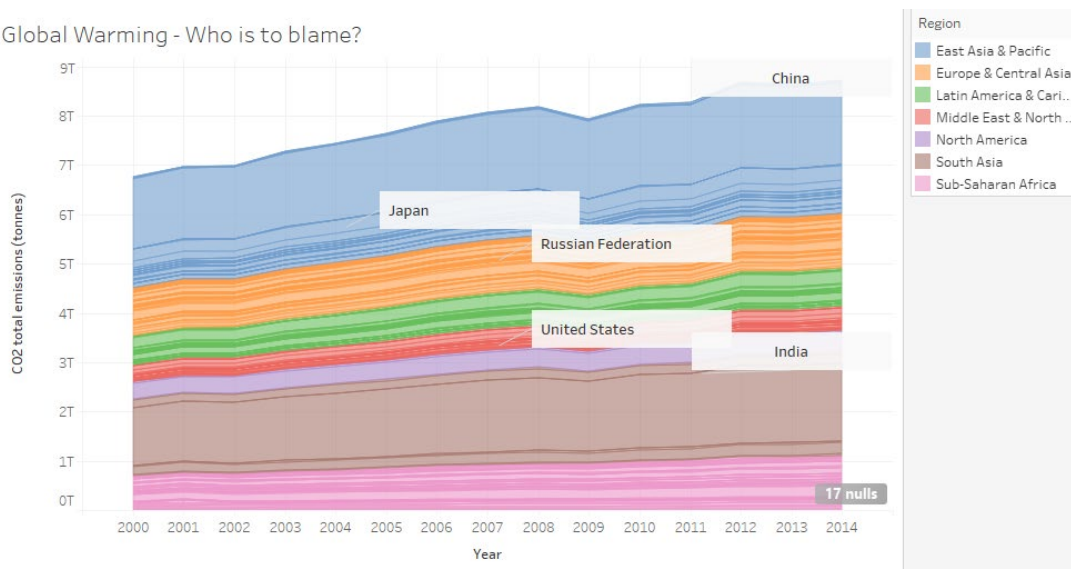
Here are some hints:

1. Drop Year into the Pages shelf to create the control to cycle through the years.
2. You will need to change the Marks to solid circles and scale them by the total number of neonatal deaths. You will create a scatterplot of GDP Per Capita with Neonatal Death Rate (use the original column of information provided in the `nations` dataset. Add `regions` to Color and Neonatal Death Rate to the size. Having done so, you will also need to increase the size of all circles so countries with small numbers of neonatal deaths are visible. Good news: Tableau's default behavior is to size circles correctly by area, so they will be the correct sizes, relative to one another.
3. You will need to switch to a Logarithmic X axis and alter/fix its range. You will also have to change ticks to power 5.
4. Format GDP per capita in dollars by clicking on it in the Data panel and selecting Default Properties>Number Format>Currency (Custom).
5. Create a single trend line for each year's data, so that the line shifts with the circles from year to year. Do this by dragging Trend line into the chart area from the Analytics panel. You will then need to select Analysis>Trend Lines>Edit Trend Lines... and adjust the options to give a single line with the correct behavior.
6. Getting the smaller circles rendered on top of the larger ones, so their tooltips can be accessed, is tricky. To solve this, open the dropdown menu for Country in the Marks shelf, select Sort and fill in the dialog box as follows.  Now drag Country so it appears at the top of the list of fields in the Marks shelf.
7. Save this dashboard under your tableau public profile and share the link with me in your assignment dropbox for week 12.



## Other Optional Extra Assignment #2

Global Warming - Who is to blame?



1. Download data in Excel format on carbon dioxide emissions in metric tons per capita, from [here](#). If you don't have Open Refine, skip steps 2 and 3 and use this data set: [co2\\_global\\_emissions\\_clean](#).
2. Remove the first set of rows that do not have data BEFORE you bring it in to Open Refine
3. Process this data in Open Refine using the same procedure from the class exercise above for dates between 2000 to 2014, and export as a CSV file in your week 8 folder.
4. Join to the exported data to the file `nations.csv` in Tableau as above, **then create the simple dashboard shown below**, recording total carbon dioxide emissions by region and nation, from 2000 to 2014. Hint: You will need to create a calculated variable for `total emissions`, using the per capita emissions data multiplied by the population data from `nations.csv`.
5. Add labels using “annotate” just for the biggest CO2 producers: China, Japan, Russian Federation, United States, and India.
6. Save this dashboard on your tableau public profile and share the link with me in assignment dropbox week 8.