

Recreating News Visualizations in Tableau

By Robert Kosara, Tableau Research

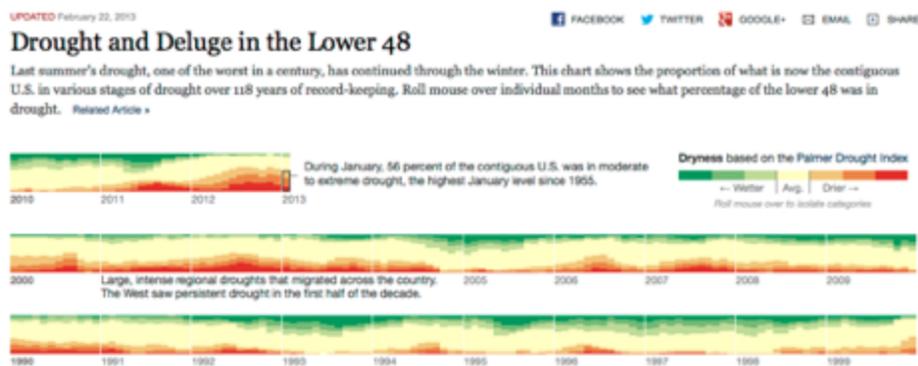
Updated by Caitlin Havlak (September 16 2016)

This guide walks you through the steps to create Tableau visualizations that mimic some particularly good news graphics. The goal is to show what can be done in Tableau, and to show/introduce some of its features in a concrete use context.

Small Multiples Area Chart of Drought Index with Decade Breaks

Drought and Deluge in the Lower 48 is a great example of a small multiples plot:

<http://nyti.ms/Obq0SH>



(1) Connect the pdsi.tsv file to Tableau. TSV (tab separated values) file is similar to a CSV (comma separated values) file except that the values are separated by tabs instead of commas.

(2) This is what the dataset looks like Excel:

key	0	1	2	3	4	5	6
189501	0	0	0	0.952	0.048	0	0
189502	0	0	0	0.923	0.077	0	0
189503	0	0	0	0.898	0.086	0.016	0
189504	0	0	0.009	0.885	0.083	0.021	0.008
189505	0	0	0.057	0.777	0.105	0.036	0.025
189506	0	0.016	0.077	0.654	0.119	0.07	0.064
189507	0.004	0.034	0.052	0.548	0.158	0.092	0.111
189508	0.008	0.038	0.072	0.529	0.184	0.087	0.081
189509	0.033	0.044	0.051	0.632	0.143	0.047	0.049
189510	0.043	0.047	0.076	0.648	0.13	0.005	0.049
189511	0.042	0.056	0.085	0.623	0.116	0.029	0.049
189512	0.024	0.053	0.112	0.631	0.084	0.05	0.046
189601	0.026	0.058	0.113	0.606	0.1	0.056	0.041
189602	0.013	0.044	0.085	0.722	0.069	0.025	0.041
189603	0.008	0.032	0.118	0.707	0.056	0.025	0.055

The key field is the date, with the first four digits being the year and the last two the month. We'll deal with that in a moment. The columns are for one of the drought severeness levels each, from 0 to 6.

- In the Data Source tab, select all of the columns except "key". In the last column (column "6"), click on the triangle and select "Pivot". This will result in two columns, one for

drought severeness level and one for amount. Rename the first column to *Level* and the second one to *Amount*. Keep the column “Key” the same.

- Make sure that each of the column data types are correct. “Key” should be a string, “Level” should be a string and “Amount” should be an number (decimal)
- Start a new worksheet. Once on the sheet, right-click in the data pane and select *Create Calculated Field...*
- Name it Date and enter this expression: DATEPARSE("yyyyMM", [Key])
 - The dateparse() function uses a pattern to match a date it can't otherwise understand. The first part, yyyy, is a four-digit year; MM stands for a numerical month (as opposed to the month name).
 - Click OK to close the formula editor
- Go back to the Data Source tab and check the format of Date to ensure that it was parsed correctly.

(3) To build the initial view, follow these steps:

- Drag *Date* to the *Columns* shelf
- Drag *Amount* to *Rows*
- Drag *Level* to *Color*
- In the dropdown at the top of the *Marks* card, select *Bar*
- Click on the little + button in the *Date* field in the *Columns* shelf so that it displays as *Quarter*
- In the *Amount* pill on the *Rows* shelf, click the little arrow and from *Quick Table Calculation*, select *Percent of Total*
- Then click that again and from *Compute Using*, select *Cell*
- This makes sure all the bar stacks are the same height. Since the numbers are cut off after three decimals, they might not always sum to 1 exactly.

(4) Now we have the initial view, but it's upside down, with the more severe drought levels (higher numbers) at the bottom.

- We can fix this by dragging the items in the colour legend so that the highest number is on top and the lowest on the bottom.
- Also, click on the *Size* field and select the “Manual” radio button. Adjust the width of the bars until there is no more visible white space between bars within a year. This creates the nice, even appearance the NY Times piece has.

(5) All the years are now laid out next to each other. But we want a more tabular layout (called a *trellis*), with ten years in each row. To do this, we need to help Tableau a little. We'll calculate two numbers for each year: the decade it is in and how far it is away from the start of the decade.

- Right-click in the data pane and create a new calculated field
- Name it *Decade* and enter the following expression:
FLOOR(DATEPART('year',[Date])/10)*10

- The DATEPART() function extracts a particular element from a date, in this case the year
- We divide that number by 10 and round the result of that down using the FLOOR() function
- For 1956, the result of that is $FLOOR(1956/10)=FLOOR(195.6)=195$
- To make the decade number nicer, we then multiply by ten again, which for the 1956 example gives us $195*10=1950$
- That is the first part of our grid (the rows), but we also need to order the years within each decade. Create a calculated field called *Decade Offset*
- Enter the expression `DATEPART('year', [Date])-[Decade]`
 - This takes the year and subtracts the decade. For the 1956 example, that would be $1956-1950=6$
- Before we can use the fields, we will need to tell Tableau that we don't want to treat these as continuous numbers. To do that, drag *Decade* and *Decade Offset* from the *Measures* area in the data pane to the *Dimensions*.
- Now drop the *Decade* on the *Rows* shelf
- Drop the *Decade Offset* on the *Columns* shelf to replace the year (either drop on the year or remove the year once the offset is there)
- Again, adjust the bar width to remove white space by clicking on the size property in the Marks card.
- This shows the 1890s at the top. To change that, mouse over the Decade axis label and click the little sort button that appears until 2010 is at the top.
- Change the coloring by clicking the triangle on the top right corner of the Level legend. Select Edit Colors.
- That is the basic layout. We can now remove the row headers by right-clicking on the little axis label saying *% of Total Amount* and deselecting *Show Header*.

A variation of this, based on different data, would be maps over time like in this example (available online at <http://nyti.ms/LwehKV>):

Drought's Footprint

More than half of the country was under moderate to extreme drought in June, the largest area of the contiguous United States affected by such dryness in nearly 60 years. Nearly 1,300 counties across 29 states have been declared federal disaster areas. Areas under moderate to extreme drought in June of each year are shown in orange below. [View Article »](#)

