MIS0855: Data Science
In-Class Exercise for Mar 25-27 – Creating Interactive Dashboards

Objective: Create a dashboard with interactive data filtering using Tableau

Learning Outcomes:

- Understand how to create an interactive dashboard in Tableau.
- Select and apply filters that add to the user's ability to understand the data.
- Navigate a data set using an interactive dashboard to find answers about the data.

In this exercise, you'll be working with a modified and cleaned version of a dataset from the Pew Research Project for Excellence in Journalism (see http://www.journalism.org).

The data set contains data about 20,448 news stories that ran between January 1, 2012 and May 31, 2012. In a sense, this entire data set is metadata, because it does not contain the text of the stories themselves, just key attributes of each story. These attributes include where the story appeared, when it appeared, and what it was about.

You are going to build an interactive "News Coverage Viewer" using Tableau's dashboard features. You've created dashboards before, back when you first started using Tableau. However, this time we're going to create a dashboard that dynamically changes views based on the user's selections. This makes it easier for the user to navigate the data themselves, instead of depending on a series of static views chosen by the creator of the dashboard.

Part 1: Take a look at the data set

1) Download the data set (Pew Story Data (Jan-May 2012).xlsx) and save it to your computer. Remember where you saved it!

2) Open the file in Excel and look at the "SourceCode" column (Column C). This identifies the category of publication where the story appeared and the name of the outlet.

3) Click on cell D2 (the Source Category column) and you'll see a VLOOKUP formula:

   =VLOOKUP(C2,Lookups!$A$3:$E$73,3)

   Remember, this means that it is using the value in C2 to find the correct textual description of the source category, that the lookup table is in the Lookups tab (Lookups!$A$3:$E$73), and that the third column of that lookup table contains the source category description (3).

4) Click on the Lookups tab to see the lookup table "Source Lookup."

5) Switch back to the News Coverage tab and click on E2, H2, J2, and L2. They all retrieve values from tables in the Lookup tab based on a numeric value in another column. Make sure you also look at the other tables in the Lookups tab so you're sure what's going on.

Part 2: Start Tableau and open the data file

6) Start Tableau. If you're using Windows 7, find it on the Start Menu. If you're using Windows 8, use the icon or search for "Tableau" from the Start screen.

7) Click on "Connect to data"
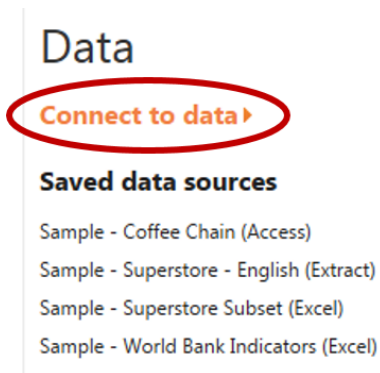
8) Click on "Microsoft Excel" under "In a file."

9) Navigate to the location where your data file is stored and select it.

10) You'll see a list of Excel worksheets at the left side of your screen. These are all the sheets contained within the workbook. Drag the "News Coverage" sheet to the workspace.

11) Wait until the data shows up and click "Go to Worksheet."
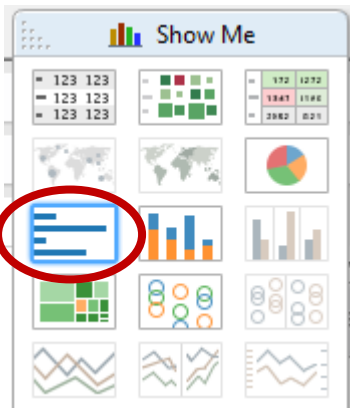
Part 3: Create the visualizations for the dashboard

The first thing we'll create is a visualization of the lead newsmakers, organized by the number of stories published about them.

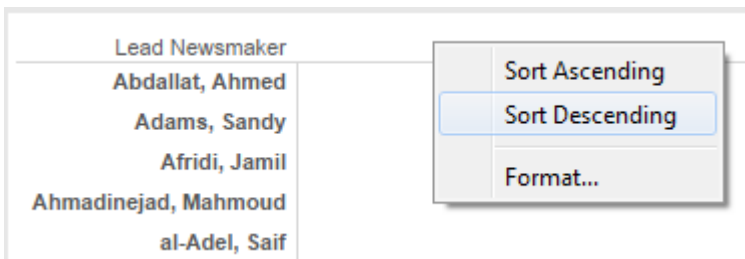1) Drag "Lead Newsmaker" from Dimensions to the Columns shelf.

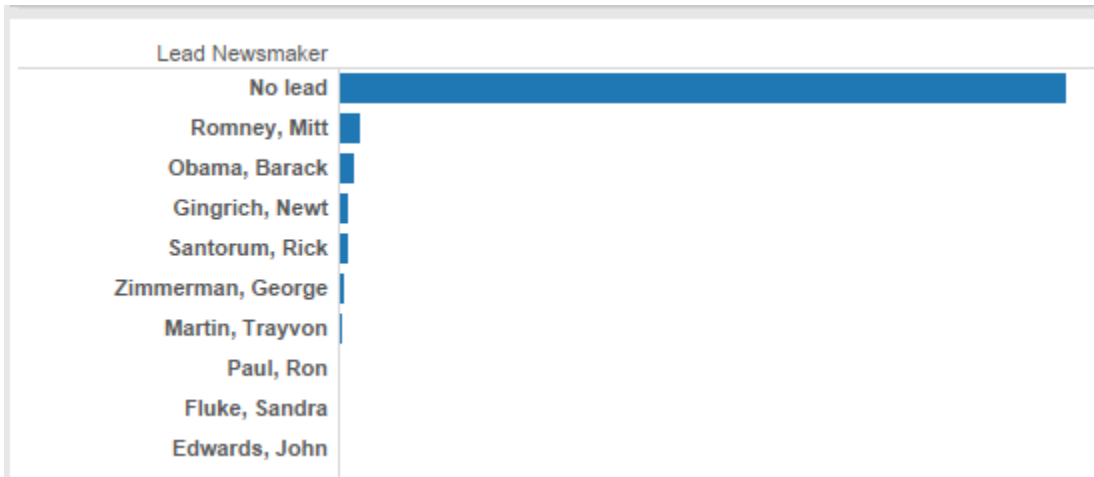2) Drag "Number of Records" from Measures to the Text icon in the Marks box.
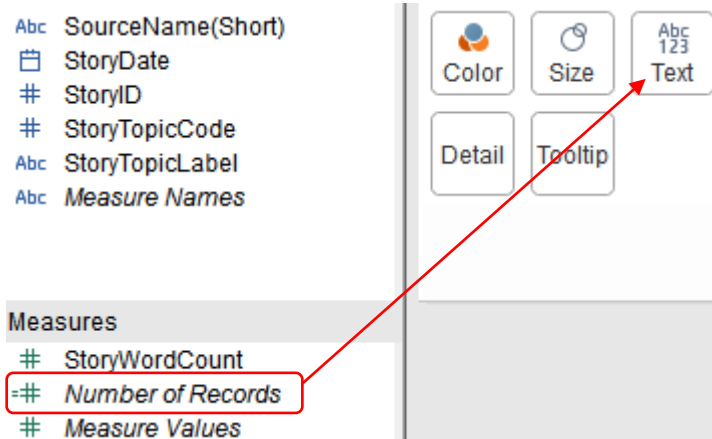
3) Under "Show Me," click on the horizontal bar chart:

4) Order the results from highest to lowest by right-clicking at the top of the chart, to the right of the "Lead Newsmaker" column heading and select "Sort Descending."

5) You'll notice that it looks like there wasn't a lot of activity:

However, that's because most of the stories had no lead. Hover over the bar next to "No lead" and you'll see that it accounts 17,552 records out of 20,448. Tableau has to make the other bars very small in order to be able to fit the "No lead" bar within the display.

6) We're interested in newsmakers, so we're going to remove the stories with no lead from this visualization. Make sure you're still hovering over the bar and click the Exclude button.
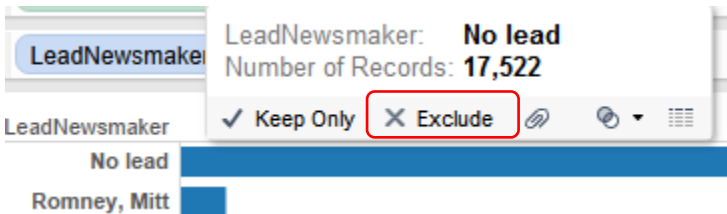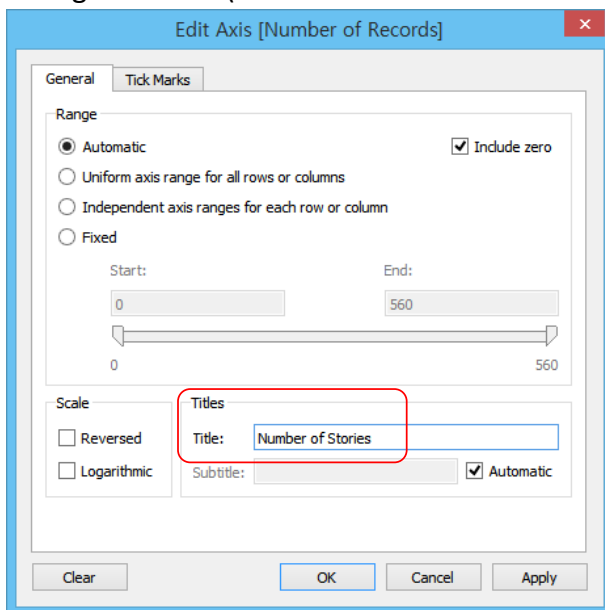
This creates a quite a different view, and one that's easier to work with:
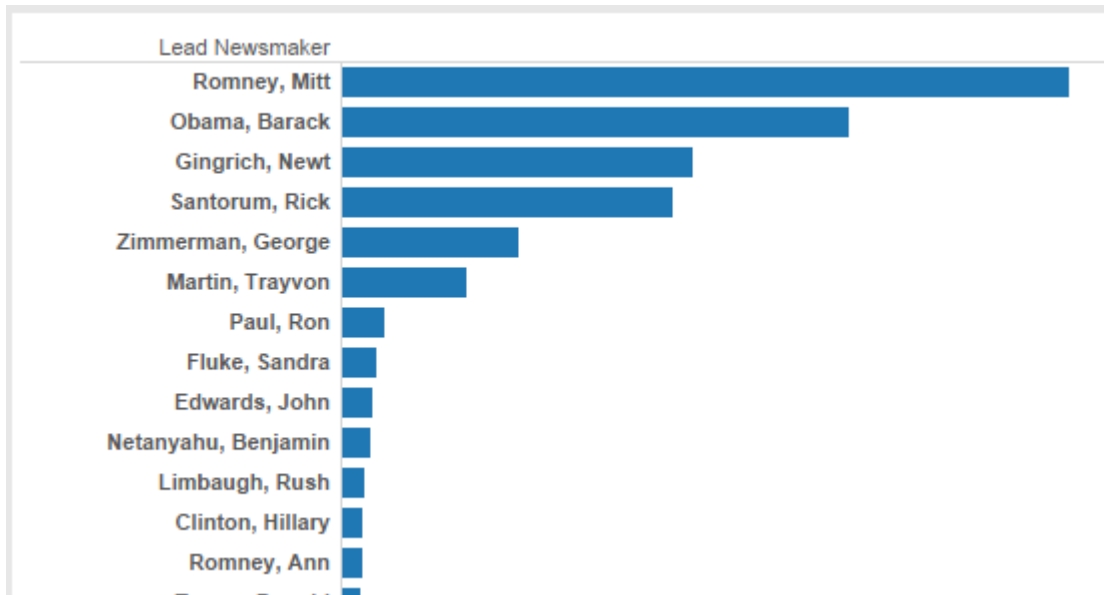
7) The x-axis is labeled "Number of Records" – let's make the axis title more helpful. Double-click on the words "Number of Records" at the bottom of the bar chart.
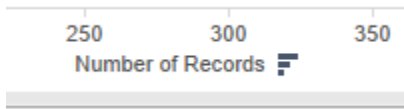

8) Change the Title (under Titles in the Edit Axis dialog) to "Number of Stories". Click OK.

9) We want our dashboard to be able to filter by the story's publication date and its geographic focus. So drag "Story Date" from Dimensions to the Filters box underneath "Lead Newsmaker" (the Filters box is to the left of the bar chart).
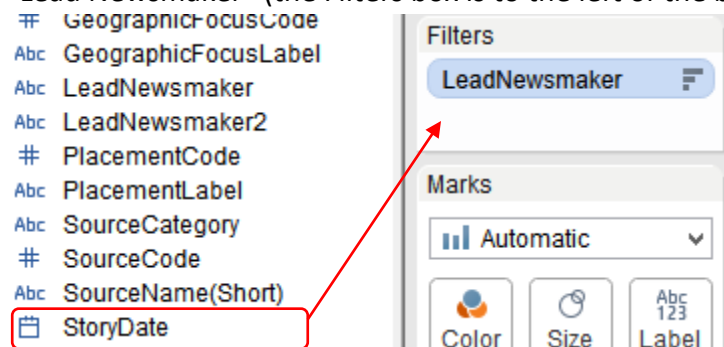
10) A "Filter Field" dialog will come up. Select "Range of Dates" and click Next.

11) At the next screen click OK.

12) Now drag "Geographic Focus Label" from Dimensions to the Filters box, underneath "Story Date."

13) Another "Filter" dialog will appear. Click the "All" button and then "OK." Your Filters box should now look like this:
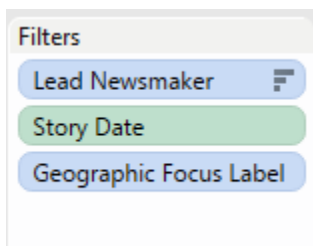
14) You'll notice that the bar chart has not changed even though you've introduced filters. That's ok. You haven't filtered anything yet – you're just giving the worksheet (and your dashboard, eventually) to use those filters.

15) Change the name of the worksheet from "Sheet1" to "Newsmakers".

Now we want to create a visualization of the number of stories published in each type of source (i.e., newspapers, online, radio, and television).
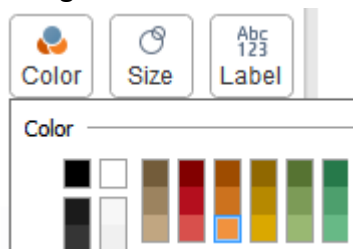
16) Create a new worksheet by clicking on the icon or selecting New Worksheet from the Worksheet menu.

17) Drag the "Source Category" Dimension to the Columns shelf.

18) Drag the "Number of Records" Measure to the Rows shelf.

19) You'll see a very basic bar chart.

20) Let's add some color. Click on the Color button under Marks and change the color to orange.
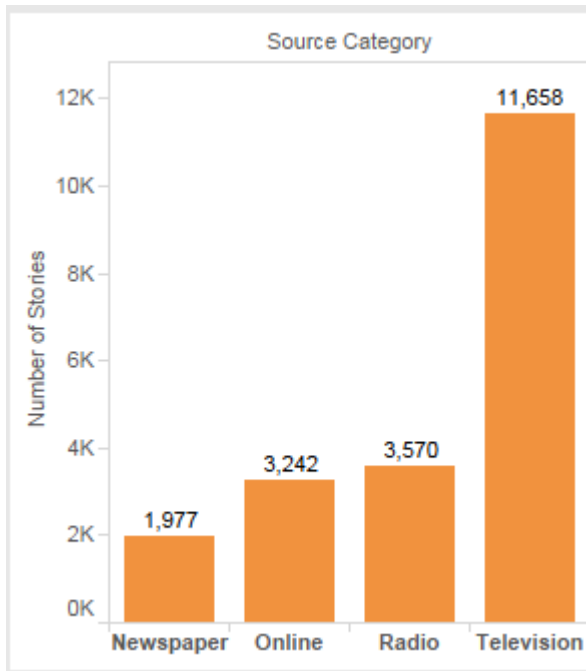

21) Click on the Label button under Marks and select "Show Mark Labels." You'll now see exact values above each bar.

22) Change the y-axis label from "Number of Records" to "Number of Stories," just like you did for the previous bar chart.

23) Your bar chart should now look like this:

24) Change the name of the worksheet from "Sheet2" to "Stories by Source".

Part 4: Put together the dashboard

1) Create a new dashboard by clicking on the [icon] icon or selecting New Dashboard from the Dashboard menu.

2) Drag the Newsmakers worksheet onto the Dashboard canvas (the blank space on the right side of the window).

3) Drag the Stories by Source worksheet onto the Dashboard canvas, placing it in the lower half of the window. The Newsmakers bar chart will adjust to make room for the other chart.

4) Add the filters controls to the Dashboard by right-clicking in the grey bar at the top of the Newsmakers bar graph (it's way at the top). Select the Quick Filters menu and then Story Date.
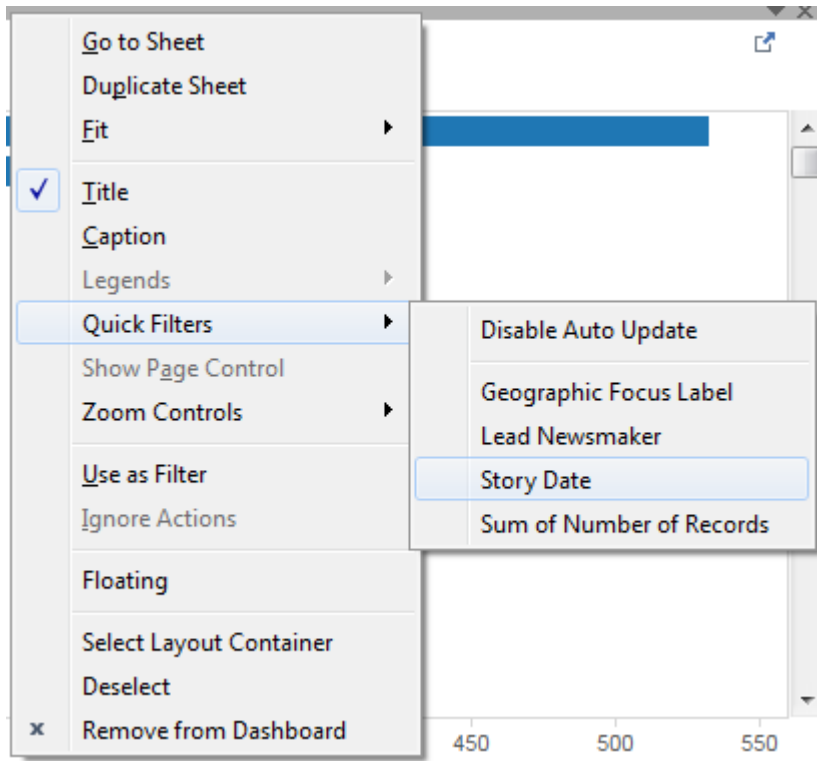
5) Now add the Geographic Focus Label Quick Filter the same way. Your Dashboard will look something like this:
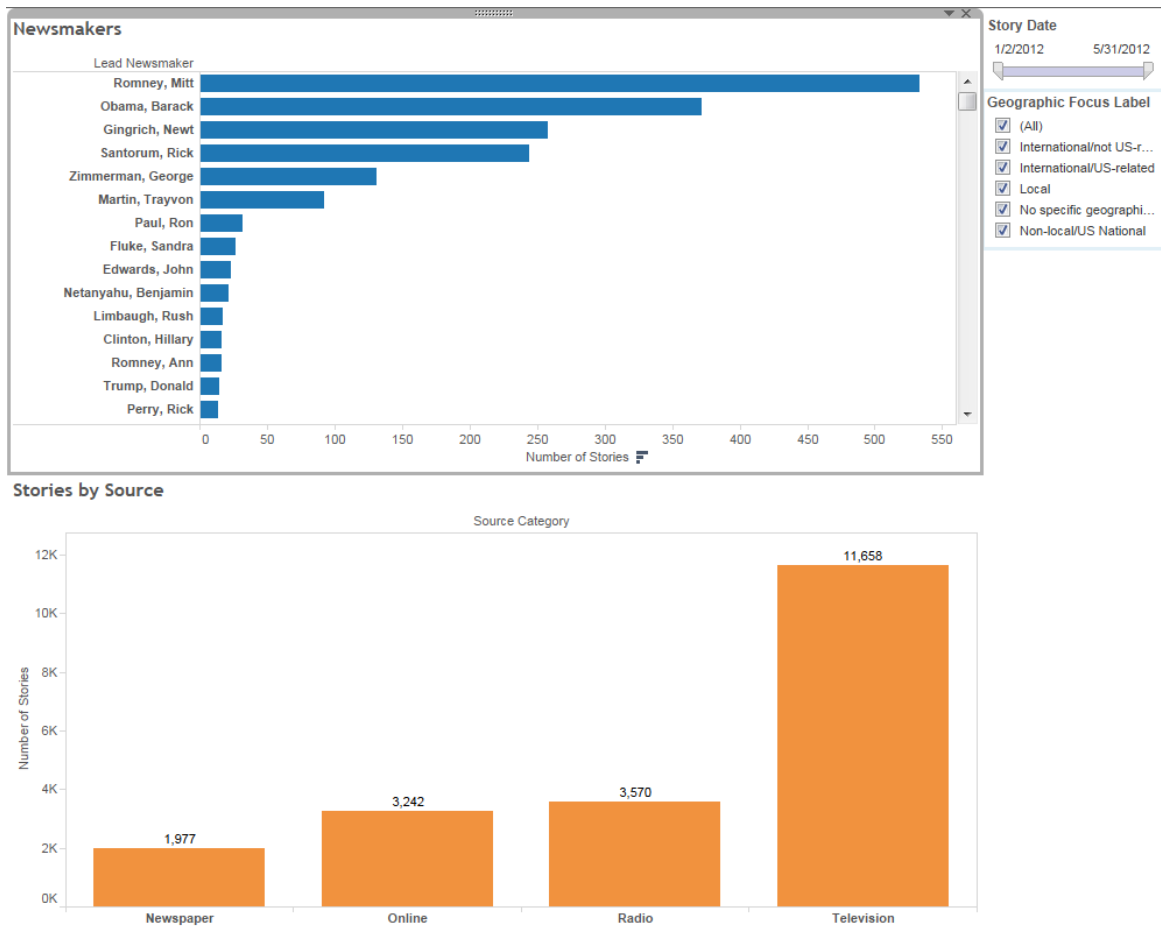

6) Double-click on the title "Geographic Focus Label" to change it to "Geographic Focus".

7) Now try moving the Story Date slider and clicking on the categories in the Geographic Focus control. You'll see that the Newsmakers bar chart adjusts to the restrictions you place, but the Stories by Source bar chart doesn't change at all! Let's fix that.

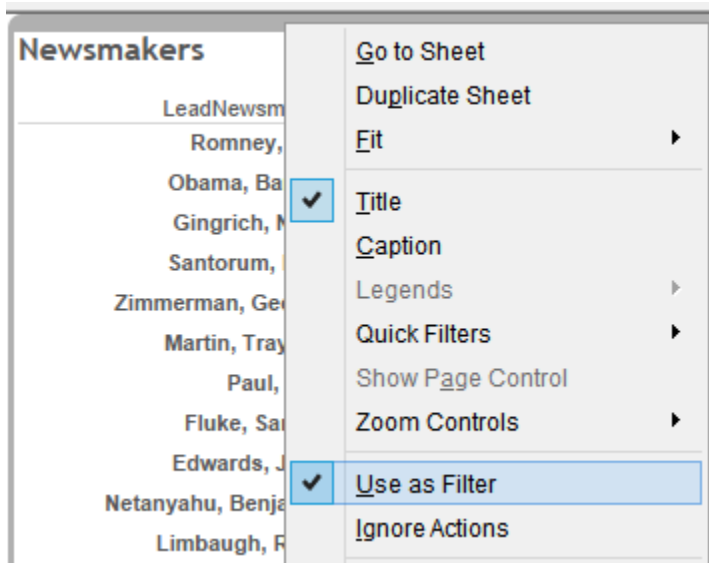8) Right-click inside the Story Date slider control and select Apply to Worksheets/Selected Worksheets:
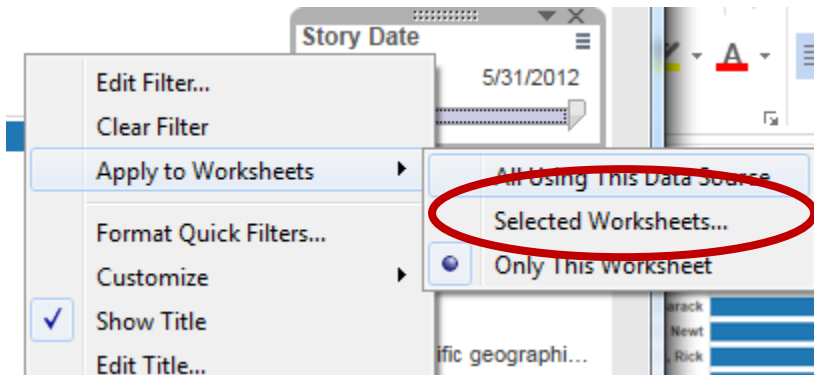
9) In the Apply Filter to Worksheets dialog, make sure Newsmakers and Stories by Source are both checked. Then click "OK."
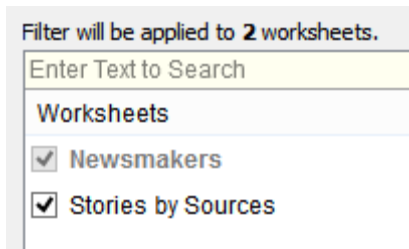
10) Do the same thing for the Geographic Focus control.

11) Right-click inside the top bar of the Newsmakers bar chart and select Use as Filter.

12) Change the name of the tab from "Dashboard 1" to "News Coverage Viewer".

13) When you click the controls you'll see the visualizations change:

- When you move the slider, you'll see the counts adjust to reflect that date range.
- When you select or deselect the sources, you'll see the counts adjust to reflect only those sources.
- When you select a "newsmaker," you'll see the counts adjust to reflect only that person. Keep in mind you can also control-click to select multiple people.

14) So now you've created a "News Coverage Viewer" dashboard, where you can now see the level of coverage someone receives during a specified period of time.

   Test it out by trying the following:

- Select a date range of 5/1/2012 to 5/14/2012 using the slider.
- Uncheck everything under Geographic Focus except "Non-local/US National"
- Click the bar next to "George Clooney."

   You'll see the following Stories by Source chart:
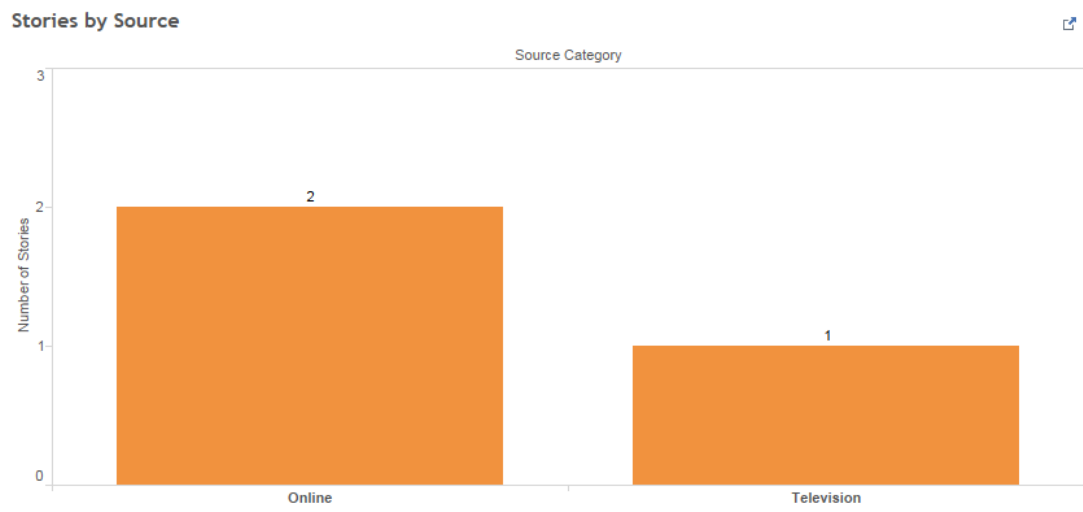
   This tells you that from May 1 through May 14, there were two online stores and one television story where George Clooney was identified as the primary subject.

15) Now reset your filters by expanding the date range back to January 1 to May 31 and check the "All" box under Geographic Focus.

Part 5: Now try it on your own!

Change the Dashboard to reflect major story topics instead of newsmakers – keep the filters the same.

To do this you should:

1) Create a new worksheet, called "Story Topics," that lists the Story Topic labels in a bar chart, arranged in descending order. Feel free to play around with fonts and colors to get it the way you want it.

2) Create a Date filter (range of dates) and Source Category (all) filter for that worksheet.

3) Create a Dashboard with Story Topics at the top, Stories by Source at the bottom, and the filters Story Date and Source Category on the right side.

4) Set the filters so that they affect both worksheets in the Dashboard, and make sure you set Story Topics to be a filter for the rest of the Dashboard (i.e., turn on "Use as Filter").

5) Call your Dashboard "Story Topic Viewer."

6) Test it out by choosing the date range of March 15 to April 11, 2012 and make sure Newspaper, Radio, and Television are checked under Source Categories. Then answer these questions:

Q1	How many stories about Crime appeared on the radio during this time?	
Q2	How many stories about Health/Medicine appeared on television during this time?	
Q3	How many stories about Government appeared in newspapers during that time?	
Q4	Which story topic is covered most frequently by television?	
Q5	Which story topic is covered most frequently by radio?	