LOEX 2019 session title: Diving Into Data Literacy Instruction
By: Andrew Battista, Katherine Boss, and Marybeth McCartin;
New York University Libraries

Lesson plan for data literacy instruction in a distributed teaching model

Location: Varies. Check the calendar invite for your session

Course: Media and Culture Analysis is a multi-section, undergraduate class (mandatory for Media, Culture, and Communication majors) in which students are required to work with data and visualize it in some way that explores an element of media and political economy. This assignment is one of four multi-media composition assignments in the class.

Goals: This interactive session, usually held in Bobst Library, will introduce students to the process of creating a data visualization. It will also give them hands on experience in working with some prepared data sets in Google Sheets. Students will gain confidence in asking questions of data and deriving new calculations from a data set. These sessions will be co-led by the Library instructor and the MCA Instructor.

Objectives:

- Evaluate the context and quality of data in order to begin developing an intellectual question
- Use data to formulate a research question related to media, access to information, and socio-political development
- Develop proficiency in Google Sheets in order to create a clear, concise data visualization
- Articulate the connections between visualizations of data and analysis of the social, political, and cultural development of a country

Materials and Handouts:

- This guide (https://guides.nyu.edu/gis/mccue14), which includes all course content and exercises, including a presentation on principles of data visualization and a .PDF of their assigned reading
- The assignment given to the MCA students
- These slides to introduce assignment

Optional Preparation:

- ❑ **Review the article "Digital Occupation" by Tawil-Souri**
- ❑ **Review the "Principles of Data Visualizations" presentation**
- ❑ **Review the slides available to introduce the assignment (optional)**
- ❑ **Practice making sample visualizations & narratives**

Class Outline at a Glance (75 mins. total):

- ❑ Introductions and exploration of sample visualizations
- ❑ Pair & Share 1
- ❑ Introduction to the dataset
- ❑ Pair & Share 2 (or looser discussion)
- ❑ Manipulation of U.N. dataset (Google Sheets)
- ❑ Small Group Exploration

Actual Class Agenda in Detail

Classroom Prep | 5 min.		
1	**Open the Libguide in the Google Chrome browser**	The libguide is here: https://guides.nyu.edu/gis/mccue14 It is important to use Google Chrome in your instruction, and stress that students should also follow along in Chrome. If they don't, they can't right-click and see the menu options.
2	**Open the class assignment and UN Millennium Development dataset**	Here is the link to their assignment and the main dataset you will be working with in the demo, the UN Millennium Development Indicators. It can be useful to be logged into your Google account via an Incognito tab and have these already open before the start of the session

Introduction and session goals | 10 min.		
1	**Introduce yourself**	Take 2-3 minutes to share who you are.
2	**Highlight the learning outcomes of the session**	The goals of the session will be to: 1: Evaluate the context and quality of data in order to begin developing an intellectual question 2: Use data to formulate a research question related to media, access to information, and socio-political development 3: Develop proficiency in Google Sheets in order to create a clear, concise data visualization
3	**Introduce the core idea of the assignment**	Mention that students are already thinking about the idea of political economy, and that the Tawil-Souri article they read and discussed in their previous class makes a bold claim: Israel increased its suppression of Palestine in 2005 and onward. Can we use data to test this claim? Note that you may want to use the slides provided to introduce this assignment.

Pair & Share: compare the example visualizations | 10 min.		
1	**Introduce example visualizations**	Introduce the students to the four example visualizations on the homepage of the libguide. It's important to explain that the visualizations are all created from the *same* dataset, "UN Millennium Development Goals Indicators," which is also a dataset they might use in their projects.

2	**Pair & share: 5 min**	Turn the students loose to compare and contrast in a small group for five minutes. You can prompt them to think about any of the following discussion questions: ● What information do these visualizations show? What information is missing or unclear? ● What questions do these maps or visualizations lead you to ask? ● What would make these maps more interesting? Better colors? More information, effects, or animations? ● What other information might make these visualizations more effective? ● What are the qualities of a good data visualization? ● How do these maps show how the same core data can be visualized differently?
3	**Discuss strengths / limitations of sample visualizations**	Return to a classwide discussion and have students report back on their findings.

Commentary: Each of these visualizations was made deliberately to contain some strong elements but also to be lacking in other areas. Students will often say that the bar chart is clear and simple, while the fourth sample visualization is the most complex and interactive (note that it is made with Tableau, which we don't cover but Data Services does support). You may want to draw out other comments and critiques. For instance, in visualization 2, you see that all countries trend upward but Finland has a precipitous spike between 2008-2012. Maybe the point of this visual is to explore why? Whatever course the discussion takes, you should use the terms and concepts of the principles of visualization slides as a nomenclature for talking about visualization, but get students to realize that each of these visualizations needs more context and clarity.

Introduction of the UN Millennium Development data set | 5 min.		
1	**Open the UN Millennium Development data set and highlight the "documentation" tab**	Impart the idea that you have to assess the data and get a sense of where it comes from, what it means, and how it's structured. This is an important part of the process of forming questions about it. Note that each dataset in our sample library has a tab called "documentation" that has information about the meaning of the variables. Remind students to focus on this and incorporate it in their presentations.
2	**Discuss why we normalize data**	How is "cell phones per 100" itself a metric to account for population variance? Why is it so important to compare data that has been normalized (comparing "apples to apples")?

Commentary: This is the perfect time to tie their current project back to the ideas of media and political economy that are outlined in the "Digital Occupation" article they read before class. Tawil-Souri argues that 2005 is when Israel disengaged militarily from the Gaza Strip, which is also the time that Palestine began to fall behind economically and socially. You can point out that the data is limited in the sense that there aren't as many data points for Palestine as there are Israel, so you'll have to account for this as you prepare to answer the question. This is an important aspect of data literacy. The question of "normalizing data" makes students realize that even "raw data" is often constructed or manipulated to show context. If raw numbers were released, obviously the countries with the highest populations would have the most cell phones. Ratios help us derive context at a glance. This is also a good time to relate the restrictions of asking questions of data to a more familiar scenario, such as a FitBit. What questions could you ask of data about the number of steps you take every day? What questions couldn't be answered by those data? What is a larger importance?

Pair & Share: asking questions of data | 5 min.		
1	**Have students look at the UN Millennium Development data set and brainstorm a question to ask of the data**	What are some questions they might ask of this data? Have each group share their most promising question. This is a good time to highlight that asking questions about data is a circular process that is wrapped up in the task of manipulating it. The first question they ask of the data might not necessarily be the most compelling, but that's ok.

Commentary: We have found it helpful to have the students test of these assessments in small group discussions. This time, just do a quick pair and share or even an informal discussion with the class. In drawing out their observations, you can ask them to conjecture on how they might overcome some of the limitations of the data they have. For example, the most glaring absence is that Palestine doesn't have data for many of the years between 2000 and 2014. This absence itself could be interpreted as a political statement. Maybe those are the years they visualize. Maybe they will need to create averages or other metrics. What about outliers that can skew the visuals? This gets them thinking about an important element to this assignment: **their own intellectual question must guide their approach to the data**. They can't just take some data that's been given to them, make a chart, and call it a day.

Manipulate UN Millennium Development Indicators dataset (Google Sheets) | 30 min.		
1	**Open the UN dataset and add it to your Drive**	Students will not be able to work with the data until they have saved a copy to their drive. Have them go to **"File" => "Make a Copy"** to add it to their drive. Tell students to name their file something distinct so they know what it is.
2	**Make a simple bar chart**	Develop a question about one of the sets that is really simple. Highlight the columns of interest (such as the names of the countries and a single year, 2014) and insert a chart. Sort the range from A-Z to give more clarity. Select the country column and a single year. In order to select two columns at once, hold down the command key. Then go to the menu and select **insert > chart**. Make a bar chart and notice how jumbled it is. The first thing that will be out of whack is that the column header is 2014, and Sheets thinks that's a number. Select the box to treat Row 1 as a header, and then this is fixed. Sorting helps identify rankings. Also, not all of the countries are shown, which you can fix by expanding the area of the chart. You can use this example to show how numbers themselves don't necessarily tell a compelling story or account for changes in time.
3	**Construct a variable that is a response to some question**	Demonstrate how to derive a new variable. Open up a new column and write a formula that illustrates an element of an intellectual question. Here is an example, but feel free to use your own: • **Which country is the most volatile over the span of time of interest to the article on digital occupation?** Note that it's better to use the education variable, because every country trends upward on access to cell phone and internet. It's better to pick an indicator that could plausibly fluctuate in either direction over time. Note that this is also a good variable to use when you have some missing years of data. To express which country is the "most volatile," open up a new column on the Education_enrollment tab and name it volatile_countries (we will say our new column is C. Then, use the formula

		=MAX(D2:Z2)-MIN(D2:Z2). Then copy that formula all the way down your column and then highlight it and the countries and create a new bar graph. It's important to note that this metric doesn't necessarily denote positive change; rather, it just shows overall change from a low to high. You can arbitrarily alter the range of years (say 2013-2005) to accord more with the article and see if that changes. Talking out loud helps students think through the process.
4	**Model styling**	Show how doing things like adding labels, titles, and editing legends can help provide context for people. You can also show simple strategies, like highlighting one or two countries to make a point.
5	**Discuss different chart types (stacked bar, line, pie chart, map)**	Show how some chart types could be a better fit than others, depending on which question is being asked. Good chart types to demo include: a stacked bar chart, a line chart, a pie chart, and a map. This is a good time to mention that the library provides access to more advanced mapping software tools if they are interested, like Carto.
6	**Show how to "publish" a chart**	Click the three dots on the top right-hand corner of the chart and select "publish chart." The embed code can be pasted into a blog or website.

Commentary: This section is highlighted in red because it's the center of the instruction session. This is the part that we recommend you spend the most time practicing in advance. It's also an opportunity for you to develop your own narrative and examples in order to impart skills to students. We modeled one suggestion in the box above and will include a few other sample narrative scenarios below; however, you can illustrate other elements or make other kinds of visualizations if you so choose. Either way, talk out loud with students as you form your example. You might want to frame this process in terms of an inductive question.

In addition to posing the question of volatility, here's a couple other scenarios you can explore with the Millennium Development Indicators data:

- **Which country is the most ahead of or behind the rest of the world at a given time?** In order to do this, you need to calculate the world average. It's not a perfect metric, but it helps provide some immediate context at least. Using the internet_100 sheet as an example, start a new column called difference_from_world_average (column C). Write the formula **=D2-(AVERAGE(D$2:D$36))** Note that adding the $ symbol keeps the average relative to each range you intend. You'll see that when you make a bar graph, it's much easier to tell which countries are relatively ahead or behind in 2014.

- **Where do Israel and Palestine stand when compared to the rest of the world in terms of access to cell phones?** This is a kind of combination of the first two scenarios, but it models how highlighting two countries can make a simpler visualization and provide more clarity. One way to establish this visual and show disparity over time is compare only Israel, Palestine, and the world average. To do this, simply create a new value in **cell B37 in the cellphone_100** tab called "world average." Then, calculate that average by writing **=AVERAGE(D2:D36)** in cell D37. Note that we are skipping 2014 because there is no data for Palestine for that year. Now, drag that formula horizontally across row 37. You should have yearly averages computed. Next, use your command key and highlight row 1 (the header), row 20 (Israel), row 28 (Palestine), and row 37 (the world average). Click insert > chart, and it should automatically guess that you are doing a line graph to show the change in time. You can see visually how far behind Palestine has been.

Whatever you do, it helps to test it out examples a few times and connect it to your narrative. Eventually, you can do examples on the fly where you ask students to pose their own questions and you can follow their lead and write formulas that would respond to their questions.

Small Group Exploration | 15 min.		
1	**Explore other datasets in their groups**	Be sure to save time at the end for students to peruse the other datasets and begin to make connections on their own. They will work to explore the additional datasets linked from the libguide and discuss how they might approach the assignment. Encourage groups to use this time to strategize, plan, and ask questions.

Commentary: In the Spring 2019 semester, several of the datasets have been updated to reflect recent data additions. Also a new dataset has been added to show the price of Internet-TV-phone bundles in Europe. Definitely look at these ahead of time and feel free to connect them to your in-class examples if you want.

Wrapping up and Final Words:

Conclude by telling students that today's workshop is meant to be an introduction to the process of evaluating data and visualizing it. There are other means to get help:

- Check out Lynda.com tutorials on "Google sheets essential training" (the link is on the right hand side of the LibGuide)
- Encourage students to seek out help at Data Services if they get stuck along the way (also on the side of the guide - https://guides.nyu.edu/appointment)
- Encourage students to register for workshops facilitated by Data Services (information is also on the side of the guide)

Thank you for participating in this session. If you have any questions about the tools or data, don't hesitate to reach out to Data Services.