**Data Viz Workshop Script**

Color:

* Hue example: graph showing countries in relation to something, each country is it’s own category, so can be a different color
* Chroma/Value example: age ranges
* Saturation - strength of color (weakened by adding gray)
* Value - how light or dark the color is
* Context: can be a helpful or harmful tool for your visualization - helpful if the colors and their context are related to your intended message. Harmful if the colors have a context unrelated to your message because it can distract from the main point

Accessibility:

* Color vision deficiency is fairly prevalent - 1 in 12 men and 1 in 200 women have some form of color vision deficiency, since it is a spectrum. Considering some of the more common forms when selecting the colors of your visualization make the visual more universally accessible.

Gestalt:

* How people instinctively view images
* Multiple types of gestalt principles:
* Proximity - close location of elements indicates a group
* Similarity - how similar or dissimilar elements are relates to grouping
* Enclosure - elements contained in a border or space are grouped together
* Connection - elements connected, even when located further apart, are grouped
* Continuity - how our  minds automatically follow the path of an element
* Symmetry - how we group symmetrical elements despite amount of space between
* Figure & ground - how positive and negative space are viewed
* Closure - how we automatically complete an element or image in our mind
* Common fate - similar orientation or intersection of elements are sorted into groups
* We mention this because:
* Helpful to know how people automatically perceive elements of your design - you may intend for a certain element to be highlighted but others may perceive the information differently
* Important to keep in mind when designing a visualization - you can use these principles to highlight or draw the eye in a certain way.

Spotting misleading visualizations - transition into talking about bias in visuals and how visuals can be purposely or inadvertently made to be misleading

Bias in visuals:

* Visualizations can be inherently biased because bias can occur at each stage of a visualizations creation, whether unconsciously or consciously:
* data /research stage
* Bias can occur in the planning, data collection, analysis, and publication phases of research, Especially through selecting or encouraging one outcome or answer over others
* Researcher trying to prove their hypothesis, etc.
* Creator bias
* Format selected and tools used
* Tools might shoe-horn in research to a set template
* Publication
* Given the nature of visuals to highlight specific aspects of data, one of the most frequent biases occurs with self-confirmation
* Bias is also present in viewing and interpreting the visualization
* Confirmation bias
* Representational bias
* Thought of in two ways: Focused on how the information is presented and who is or isn’t included. The way information is presented encourages thinking and reasoning in certain ways at the expense of others.
* “When people rely on representativeness to make judgments, they are likely to judge wrongly because the fact that something is more representative does not actually make it more likely.[4] The representativeness heuristic is simply described as assessing similarity of objects and organizing them based around the category prototype (e.g., like goes with like, and causes and effects should resemble each other).[2] This heuristic is used because it is an easy computation.[4] The problem is that people overestimate its ability to accurately predict the likelihood of an event.”
* Familiarity bias
* Overconfidence bias
* See book on Cognitive Bias in Visualization: <https://aedeegee.github.io/bookchapter18.pdf>

To help identify biases in visualizations and when viewing visualizations, there are a few strategies you can use. One strategy is the DIG method.

DIG Method

* Analyze
* Who, what, when, and where do you see **represented** in the image?​
* What **textual information** is provides (title, caption, date, etc.)​
* How does the image make you **feel**?
* Interpret
* Who **created** the image? Who owns or **published** the image?​
* What is the **message**? Who is the intended **audience**?​
* How does **context** (social, cultural, political, etc.) inform the image?
* Evaluate
* How might your reaction **impact** how you view the image?​
* Is the image **reliable**and **accurate**?​
* Has the image been **misrepresented** or **manipulated**?
* Comprehend
* Base your judgement on your evaluations and all the information available.​
* Is the purpose of the image to **persuade, inform, entertain or sell**? Why do you think so?

DIG Method Example:

* Analyzing:
* Showing United States, voter distribution, political affiliation
* Text is very bare - only title and data source showing
* A lot of red -
* Interpreting:
* Source: two sources listed at bottom, but publisher not shown
* Message: voter distribution geographically shown
* Additional context: red associated with Republican, Blue with Democrat
* Evaluating:
* Is it accurate? Is this misleading?
* Comprehending:
* Is this visual meant to persuade, inform, entertain, or sell? What is our final judgement of the visual?

Now let’s take it a step further - how would you change this image to make it better?

**SLIDE #24**

Thank you, Shelby.  OK, we will proceed and this time we are going to focus on different forms of visualizations.

**SLIDE #25**

There are many different types of visualizations based on the data used and its function.  Each type of data needs its own form.  For instance, is the data quantitative or categorical?  Starting with this information can help you determine the direction to go on in creating a visualization to convey your message or storytelling.

**SLIDE #26**

Here is a sample of the different types of charts based on classification: Data comparison and Data Reduction.

**Data Comparison:** Here, you want to compare data – as suggested – you can decide on charts that either compare (bar, bullet, dot plot); composition (pie), or distribute (histogram)

**Data Reduction:**Data reduction is where – again, as suggested – reducing the data by eliminating data that is invalid or showing data in summary form or in aggregate levels.

**SLIDE #27**

**Line Graph:** Line graphs can reveal trends and changes over time.  Line graphs works when the data is continuous versus if the data starts and stops.  Use solid lines and, again, consistency counts so the audience can focus on the data.

**SLIDE #28**

**Pie Chart:** Pie charts are used if you are representing parts that make up of a whole.  For instance, using a static number (i.e. 100%) and dividing it up into multiple categories that will add up to the static number.  Make sure to differentiate between slices that represent the categories.

**SLIDE #29**

**Bar Graph:** It is best to use the bar graph if you are displaying negative numbers; comparing 10 or more items; working with lengthier labels.  Use consistent colors so the audience can focus on the data.  Bar graphs are good to use to keep track of data over time as well as to compare different sets of data.

**SLIDE #30**

**Histogram:** Histograms show a distribution of data over a period of time.  Histograms are useful to see where the data is concentrated as well as determine probability.  Referring back to what Shelby described, more content doesn’t always equal better context, if you are comparing 3 or more variables in data sets, it is best not use a histogram.  Make sure the bars consistent in size and not too wide or too narrow as well as consistent in colors and labels.  Use equal round numbers.

Does anyone have any questions?

**SLIDE #31**

It is common that people confuse bar graphs from histograms.  They are similar when you see them.  There are differences.  First, bar graphs have spacing in between the bars whereas there is no spacing in an histogram.  Second, bar graphs shows categories on one of its axis (i.e. countries) whereas histograms show number ranges.  Another key difference is that bar graphs are used to compare variables – as noted earlier that they are under the category of Data comparable.  Histograms show the distribution of variables.  Earlier, I mentioned that some chart types are quantitative while others are categorical.  Bar graphs are categorical and histograms are quantitative.

**SLIDE #32**

Maps can show distribution and concentration of data in geographical area(s).  Maps are good when you want to translate your data visually but also providing aesthetics that engage your audience.  Map visualizations are clearer and are intuitive and allows people to search further for information that they need to make decisions. Maps also allow people to compare activity across locations. There are many different types of maps: heatmaps, statistical maps, 2D maps, 3D maps, static maps, but the most common are bubble maps and dot maps.  Bubble maps displays location and population.  Dot maps displays location. 

**SLIDE #33**

This chart shown on this slide has a lot of information on it, but it is another good source to refer too.  This shows different chart types in which they are grouped by their function such as comparison, distribution, composition, and relationship.  This PowerPoint presentation will be made available for you to refer too as part of the LibGuide, so you can have more time to view this image.

**SLIDE #34**

There are tools that can help guide you in selecting a chart type to fits the data you have. These are are two examples of tools that are available: Storytelling with Data Chart Guide and Data Viz Catalogue.  These are also on the LibGuide and available to you to explore.

**SLIDES #35-35 – ACTIVITY**

**Slide #35:**

We will now begin our second activity.  Before we start, let’s review the activity’s instructions.

**Slide #36: (BRING UP THE GOOGLE FORM – “Selecting Visualization Types”).**

We will assign approximately 3 students to each breakroom room.  For this activity, each group will work on this activity using a Google Form we have for this activity. The link is on this slide as well as in the chat box.  Amongst your group, please assign an individual to fill out the Google Form.  On the form, you will be presented with five scenarios.  We just discussed the common visualization types – bar chart, line graph, histogram, map, and pie chart.  For each scenario, you will choose a visualization type that fits with the data presented.  If your group needs to review the visualization types that we discussed, the link to the LibGuide is in the chat box for your reference.  The link is also on the top of the Google Form you are working on. You will have 10 minutes to complete the assignment.  Once done, we will all return to the main room to discuss.

**Slide #37:**

Welcome back everyone!  How did the activity go overall for the groups?  Let’s take this time to review each scenario.

For Scenario #1, we are focusing on the average residential data from the American Water Works Association. What were your answers?  You can type in your group’s responses in the chat box.  (DEPENDING ON THE ANSWER).  The correct answer is a pie chart.  A couple of clues in the scenario description was the data was represented by the percent of TOTAL water usage and the categories presented: shower, toilet, faucet, washing machine, leaks, and other.

**Slide #38:**

For Scenario #2, we are looking at COVID-19 data based on Race/Ethnicity in Los Angeles County.  What chart type did the groups choose?  You can type in your group’s responses in the chat box. (DEPENDING ON THE ANSWER).  The correct answer is a bar graph.  A bar graph makes sense since you are using categories (Race/Ethnicity) and the data consists of variables that you want to compare.  A bar graph is good for this purpose.

**Slide #39:**

Before we move on to the third scenario, is there any questions?

For Scenario #3, we are looking at data from the World Atlas of Language Structures database.  What answers did you select?  You can type in your group’s responses in the chat box.  (DEPENDING ON THE ANSWER).  The correct answer is a map chart.  A map chat would be a good fit for this kind of data since it you can visualize see where the living languages currently are in the world and get a sense of concentration.

**Slide #40**

For Scenario #4, we are looking at data from the Los Angeles County on COVID-19 cases by age group.  What answers did you select?  You can type in your group’s responses in the chat box.  (DEPENDING ON THE ANSWER). The correct answer is a histogram.  A histogram is a good fit for this kind of data since you can see the range or progression through the age groups.  You can also see where the COVID-19 cases are concentrated.

**Slide #41**

For Scenario #5, for something on a lighter side, we are looking at a marketing data from American Pet Products on the number of households that own a pet and by pet type.  What answers did you select?  You can type in your group’s responses in the chat box.  (DEPENDING ON THE ANSWER).  The correct answer is a bar graph.  A bar graph is good for this type of data because you can compare the data easier and use a handful of categories.  Also, referring back to the differences between bar graphs and histograms, for this bar graph, we’re using categories (pet types) whereas in Scenario #4 for this histogram, we used numbers (age groups) in one of the axis.

**SLIDE #42**

There are a lot of tools available that you can use to create visualizations for your data.  Some of the tools you need to pay for and some are for free. Here on this slide are some of the main tools used by research here:

**Tableau** is a visual analytics platform.

<https://www.tableau.com/>

**Excel**is part of the Office 365 package. As students you should have access to this package through the university’s ITS department.  Once you input your data in Excel, you can transform your data into a visualization format working under the “Insert” tab where “Charts” is located.  There are videos on YouTube as well as LinkedIn/Lynda.com that provide tutorials in how to use Excel to create visualizations for your data.

<https://its.lmu.edu/whatwedo/computingsoftware/at-homesoftwareforstudents/>

**R**is also another statistical computing open software to use in creating data visualizations.

<https://www.r-project.org/>

**SLIDE #43**

During the course of this workshop, we hope that you learned the following:

Principles of visualizations such as colors and the Gestalt Principles.

Assessing Visualizations using the DIG Method

Types of Visualizations: Selecting a chart or graph type that fits the needs of the data and can accurately tell a story.

Remember, on chart type may not be the right fit for all of your data needs.  Be cognizant of selecting the right chart type to clearly visualize your data accurately and not confuse your audience.

**SLIDE #44**

For resources to take with you or that you can access anytime, we have a LibGuide that provides information on data visualizations including tools – as we mentioned earlier – that you can check out and learn more.  This PowerPoint presentation will also be available on the LibGuide for your reference.  The link to the LibGuide is in the chat box.

**SLIDE #45**

We do appreciate your feedback on today’s workshop.  Your feedback will help us improve on our workshops but also gives us a better idea of what you learned and take away from these workshops.  Please take a moment to fill out our evaluation form.  The link is here on this slide and in the chat box.

**SLIDE #46**

On behalf of my colleagues, reference librarians Shelby Hallman and Nataly Blas (who is serving as our Zoom Navigator for this session), it has been a pleasure providing instruction on data visualization to all of you.  Do you have any questions for us?  If you do not have any questions right now, but have questions later on whether about this workshop or relating to your research assignments, please don’t forget that we have our online chat that you see here on this slide (it is on the library’s website).  Once the library building opens in full capacity, we will also be available in person to assist the LMU community.  You can also email any of the librarians and staff including myself, Shelby and Nataly.  Again, thank you for your participation in today’s workshop.