Part 3: The Basics of Data Analysis and Visualization & Historic Examples of Data Visualization

The goal of these slides is to help participants understand the basics and history of data analysis and visualization.

This document is designed to walk you through the content and to collect notes that might be helpful as you go.

Slideshow Contents

- Focus/Team meeting
- Group Check-in
- The Basics of Data Analysis and Visualization
    - By Caleb Deitch and Confidence Oguebu, *WTT?* TA's!
- [Halfway point]
- One-on-one meetings
- Historic Examples of Data Visualization
    - By Greg Zapata, BARI Education Program Manager
- Wrap-up

Team Meeting

Change as you'd like!

- Icebreaker
    - Any volunteers?
- Group check-in
- Volunteer opportunity
- Next week
- Today's agenda
- Check in with the surveys

Group Check-in

- How are we feeling about the projects?
- What is something that you feel good about?
- What is something you have questions about?
- What do you want to have accomplished by the end of the week?
- Any other questions?
    - Finish drafting them and have us look them over.
    - Write 2-5 sentences about your bill (important for survey)

The Basics of Data Analysis and Visualization

By Caleb Deitch and Confidence Oguebu, WTT? TA's!

Data Visualization and Storytelling

- **Data visualization** is the graphical representation of information and data. By using visual elements like charts, graphs, and maps.

- **Storytelling** through data implies understanding the data, doing exploratory data analysis in order to decide what story to tell, Identifying powerful and meaningful stories and Creating visualizations telling those stories.

- The elements of good data storytelling are represented on the image.

A Good Data Visual Should Have…

- **Clear Purpose:** Clearly define the purpose of the visualization. What question is it trying to answer? What story is it trying to tell?

- **Accurate Representation:** Ensure that the data is represented accurately. Avoid misleading scales, and other manipulations. Use proportions and scales that accurately reflect the data values.

- **Appropriate Chart Type:** Choose the right type of chart for the data.

- **Clear Labels:** Use Clear and Concise Labels, Titles and Subtitles.

- **Effective Use of Color:** Use contrasting colors to distinguish different data series or categories. Maintain color consistency.

Displaying Data

- These types of graphs work for categorical data or numerical data that has been groups into bins
    - **Bar Graph:** displays categorical data with rectangular bars, where the length of each bar represents the value of the category.
    - **Histogram**: shows the frequency distribution of a continuous dataset by grouping data into bins
    - **Pie Chart**: represents data as slices of a circular pie, where the size of each slice is proportional to the percentage of the whole.
- These types of graphs should only be used with numerical data
    - **Scatter Plot**: uses dots to represent the values of two different variables
    - **Box Plot:** displays the distribution of a dataset, often highlighting outliers.
    - **Line Graph**: connects individual data points with straight lines to show trends or changes over time
- What do you think?e.g., “Describe your three biggest concerns regarding AI.”
    - The chart below represents a good visual and three visuals of poor quality. Any thoughts? Share them!

EXPORTING DATA FROM A QUALTRICS SURVEY

Tip: If you have a filter applied to data, you will export only the filtered data.

- Step 1: Go to the Data section of the Data & Analysis tab.
- Step 2: Click Export & Import.
- Step 3: Select Export Data.
- Step 4: Select your desired file format- CSV, TSV, Excel, XML, SPSS, Google Drive and User-submitted files.
- Step 5: Select Download all fields if you would like to export all of the data you collected.
- Step 6: Decide whether you'd like to download data in a quantifiable, numeric format (Use numeric values) or with the answer choices as written (Use choice text).
- Step 7: Click Download.

Data Visualization In Google Sheet

- Step 1: Highlight the cells you want to visualize by clicking on the column letter(s).

- Step 2: Click on Insert and select Chart.

- Step 3: A default chart will pop up alongside the chart editor pane.

- Step 4: Use the chart editor pane to select the desired/suitable chart for your visual, choose its aggregation, and fine tune the chart using labels and legends to effectively represent the data.

Multiple Choice/ Matrix

Split multi-value fields into columns changes how your data is exported (one column for each statement in matrix)

- Selected:
    - each choice will include its own column in the dataset. On each participant's row, there will be a "1" in the columns of the choices they selected.

- Deselected
- question will have only one column in your dataset, and the values of the selected choices will be comma-separated beneath

[Halfway point]

Nesting Functions in Google Sheets

- IF(logical _ expression, value _ if _ true, value _ if _ false)

Bar Chart from choose multiple

- =COUNTIF(range, condition) will find sum of records with condition true

- Use "* *" to search for any case of the string (be careful)

Histogram/ Pie Chart

- Similar to Bar Chart

- Use PercentIF() to get percentile

- Make sure categories are distinct

Rank Order

- The downloaded dataset for a rank order question includes a column for every item being ranked. In each column, you'll see the ranking each participant awarded that particular item.

- **** Mean, Median, Mode are important to look at for these******

Text Entry

- Each text entry box in your survey will be shown as a column in your downloaded data, with a row for each response

Let's Dive In!

- Represent the total wealth by region using a bar chart. Pick any colour of your choice and use appropriate titles/legend where necessary.

- Try doing same using a pie chart. Which of the two visuals do you think best represents the data and why?

- Now let's do something fun, using the countries and their Wealth ($B) represent this data using a line graph, a scatter plot, a bar graph, a pie chart and finally a geo map! Tell me what you think.

[Halfway Point]

One-on-one meetings

Historic Examples of Data Visualization

- So far you talked about…
    - Types of data
    - How we collect it
    - How we clean it
    - How we visualize it
    - Definitions of data visualization
    - Methodologies:
        - Charts
        - Graphs
- Now let's see them in action, with two notable examples from the past and present.

- Edward Rolf Tufte
    - American statistician and professor emeritus (fancy way of saying "retired) at Yale University
        - Political Science
        - Statistics
        - Computer science

- From his website: "He is noted for his writings on **information design** and as a pioneer in the field of **data visualization**."

- Data Visualization
    - The practice of presenting data
    - Usually it's raw data
    - Many ways to organize it
        - Charts
        - Graphs
        - Matrices
        - Diagrams
        - Plots
        - Dashboards
    - But it all has ONE GOAL… but what?
    - **VISUALLY ENGAGING**
        - Why?
            - Example: 1861 illustration by Charles Joseph Minard. Depicts Napoleon's march to and retreat from Russia (1812-1813)
            - Larger version on slide 30:
                - The Minard diagram shows the losses suffered by Napoleon's army in the 1812–1813 period.
                - Six variables are plotted: the size of the army, its location on a two-dimensional surface (x and y), time, direction of movement, and temperature.
                - This multivariate display on a two-dimensional surface tells a story that can be grasped immediately while identifying the source data to build credibility.
                - Edward Tufte wrote in 1983 that: "It may well be the best statistical graphic ever drawn."
                - This brings us to INFORMATION DESIGN

Information Design:

- Wikipedia: “The practice of presenting information in a way that fosters an efficient and effective understanding of the information.”
- Many uses and applications across many careers
- All careers use data in some way
- Communicating that data is a key to success
- Information Design is often taught in graphic design programs

Edward Tufte is perhaps THE big name

- His most famous book is *The Visual Display of Quantitative Information*
    - The book has become a bible of the trade
- Focuses on theory and practice of designing charts, tables, etc.
- Part history book/part manual for data display
- Topics include:
    - Data-ink ratio
    - Aesthetics
    - Do’s and don'ts of graphic design
    - Avoiding distracting clutter (“*Chartjunk*,” as Tufte calls it)

- Tufte's next major book: Visual Explanations (1997)
    - Dan was kind enough to lend us his copy (*FIND A COPY)*

- Focuses on (from the publisher)
    - pictures of verbs
    - the representation of mechanism and motion
    - process and dynamics
    - causes and effects
    - explanation and narrative
- Essentially a companion piece for Quantitative Information

- Pass the book around and ask yourselves…
    - What survey data is most important?
    - What story does it tell? About communities, trends, etc.
    - How would you WANT to visualize it?
- Regroup in 15 minutes

- *Visual Display* is a great book for many reasons
- One of the biggest: storytelling
- Applies to one of the most-used methods of data communication today: the slideshow
- Nancy Duarte (slide:ology) has 5 rules:
    - Tell the truth,
    - Get to the point,
    - Pick the right tool for the job,
    - Highlight what is important,
    - Of course, keep it simple.
- Duarte and Tufte are right… but someone came first
    - Go to picture of WEB DuBois on next slide
-

- W.E.B. Du Bois (1868-1963) achieved many things as a civil rights activist:
    - First African American to earn a doctorate (Harvard, 1895)
    - The Souls of Black Folk (1903) was a landmark book that talked about Black society from a Black perspective

- Less known is his contribution to data science
- 1900: A cultural conference called "The Paris Exhibition, 1900" is held
    - The goal: show how people around the world live.
    - DuBois, a history professor at Atlanta University, helps develop an exhibit depicting the post-Civil War living conditions of African Americans

- Exhibit showcased progress and living conditions of Black people in 1900
- Exhibit included over sixty studies, graphics and photos taken by Du Bois and his team!

- These studies and graphics are in the Library of Congress
    - Have uploaded to the LoC website.
- They've also been published in the collection W.E.B. Du Bois's Data Portraits: Visualizing Black America
- These studies and graphics are in the Library of Congress
- Have also been uploaded to the LoC website.
    - https://www.loc.gov/pictures/search/?sp=2&co=anedub&st=grid
- Feel free to open
    - **CW**: contains terms and words that are not acceptable today!
- Let's see some examples!
    - Go through slides 43-46 and explain each
    - Play brief documentary (https://www.youtubeeducation.com/watch?v=gr9l7GX9Qsk)

- One more example: Maps!
    - Data Visualization: not just about graphs and charts
    - Maps are also a method of visualizing data
    - Can see data within data
        - Redlining:“the process of denying services to residents of specific neighborhoods or communities (usually neighborhoods with non-white people) either by direct methods or by raising prices.”

- The “Mapping Inequality” Project
    - A collaboration between Virginia Tech, University of Maryland, University of Richmond and Johns Hopkins University
    - Part of the American Panorama project that tracks american history through maps
- Feel free to open

- CW: contains 1930s terms and words that are not acceptable today!
    - https://dsl.richmond.edu/panorama/#maps

- Redlining Background:
    - In the 1930s, as part of the New Deal, the government creates the Home Owners Loan Corporation (HOLC).
    - One of the causes of the Great Depression: collapse of the housing market
    - HOLC sends agents to cities and towns to determine which areas would "safe" for banks to issue mortgages.
    - The housing racism of 1937 lays the groundwork for the racism of 2023
        - Let's look at an example (Slides 52-54)

- This tool is important because it allows you to see…
    - How communities were divided on maps
    - How the people in charge thought (word choice, etc.)
    - What the government's priorities were
        - and think about whether or not they've changed

- Data visualization can sometimes say more than a person means to…

- Visualizations like Du Bois's are just one way we can measure and present findings.

- Programs like Excel, Tableau and others have made visualization even easier.

- (Slides 56 -58) This is a modeled heat index from 2017, a measure of heat and humidity. Mapped by The Trust for Public Land Notice where the hottest spots are… I'm going to pull up a redlined Boston map. What do you think we would see if we compared this map to the redlining map?

- More to these visualizations than heat and property values
- Imagine if we could quantify the BIG ISSUES of this internship
- Where would access to them pop up?

- Last thoughts…
    - Maps, charts, bars, graphs… what do they all have in common?
    - All ways to illustrate the same thing
    - Application depends on circumstance

Wrap-up:

- Next Time
    - Rough draft of report and slides
    - We'll also do some practice sessions

[End of Content]