IMT 562 - Assignment 3: Data Viz Makeover

Alwin Eldhose Babu

Dataset Selected: 2020 Week 44 – The Digital Gender Gap (Link to Data)

We engage with resources, services and systems powered by the Internet on a daily basis. It is hard to imagine a day without mobile phones. On the surface, the internet appears to be a universally accessible and widely used instrument that has spread and established itself globally. This, however, is not the case. The Covid-19 pandemic brought this divide to the forefront. Many services, opportunities, tools, and educational resources that were previously available in-person to offline communities, became unavailable during the pandemic. Nearly half of the world was offline in 2020, with no internet connectivity. Women in Low- and Middle-Income Countries made up the majority of this offline population (LMICs). This makeover Monday visualization was created to bring attention to the digital gender gap and promote awareness of both the progress that has been done and the wide differences in access to digital and communications technology that exist around the world.

Dataset: Source and About the Dataset:

The dataset comes from the Inclusive Internet Index, a project of the Economist Intelligence Unit. The dataset was posted as a Makeover Monday Submission in Week 44 of 2020. The dataset was straightforward with 100 rows and 4 columns, where each row corresponded to a country, and its gender gap in internet access, mobile phone access and %of households with access to internet, all expressed in percentages. Positive values in the data indicate that male access exceeds female access, while negative values indicate that female access is greater.

Original Visualization Source:

https://media.data.world/dYmiAUUSDasC0oQh9H7Q_The%20digital%20gender%20divide.png

Analysis of Original Visualization:

The original visualization showed the gender gap in internet access as a bar graph by listing all 100 countries in alphabetical order, along with the % of internet user households. The visualization then showed a map for the mobile access gender gap by color encoding the gap for each country based on its percentage of gender gap. Visualization is provided in Figure 1.

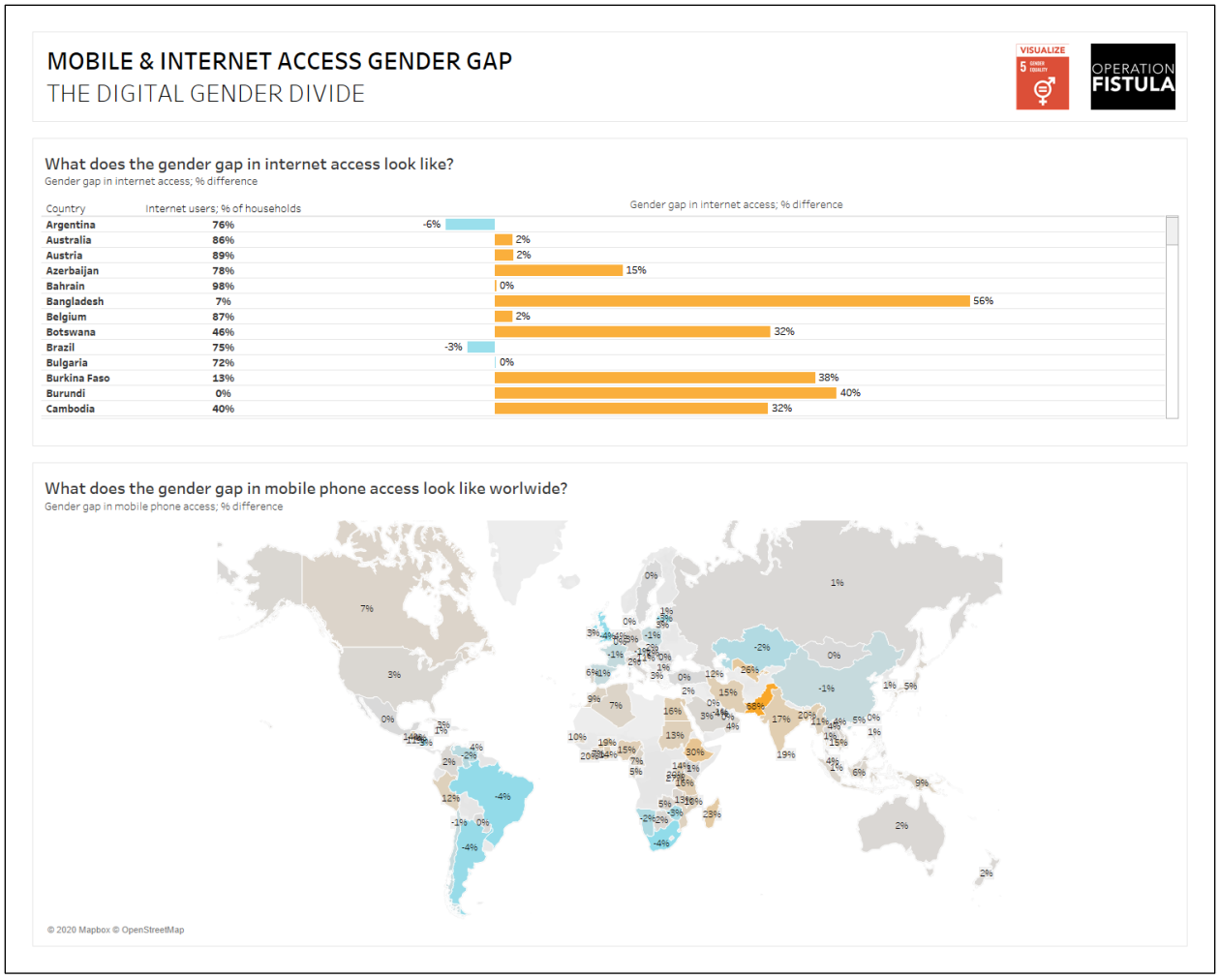

Figure 1: Original Visualization

Things that I like about the provided visualization:

1. All the data variables have been covered for analysis in the visualization.
2. All the countries have been incorporated into the visualization.
3. Color and size encoding have been performed to differentiate positive values from the negative values, both acting as pre-attentive variables.
4. Utilizing a color gradient allows for finding the countries with extreme gender gaps in mobile access.

Things that I don't like about the provided visualization:

1. There is no consideration of **'Continuity** (Gestalt Principle) in the different values that have been analyzed – internet access vs mobile access gender gaps. While one is visualized using a diverging bar graph with country listings, another is visualized using a map.
2. The map makes it very difficult to compare the values of different countries, especially if their areas are smaller (Gestalt Principe of **Figure/Ground**).
3. **Far too many colors in the map (>6). Neither a diverging nor a sequential color scale** has been used for the map. It is not clear what the countries with grey colors in the map mean. For

internet access gender gap, even though there are two colors – orange and blue – it is not clear exactly whether orange means more percentage of men or women and similarly for blue. The same values when negative have a blue color while they are grey when positive. Visualization seems to have taken a maximum value of 68% for positive as the relative comparator. There is no legend provided for the map – is it to be assumed that the map follows a similar color pattern?

4. The visualization does not allow a comparison of internet access vs mobile access – Is there a relationship between the two variables? There has been **no comparative analysis** made and it isn't possible to make one using the provided visualization.
5. There's no means to identify the top countries with internet users % of households or top countries with lowest gender gap in internet or mobile access since they are listed in alphabetical order.
6. The visualization has a low **data-ink ratio.** The listing of 100 countries (visible in original visualization) and the map takes up a lot of graphics that can be erased. Moreover, it is not easy to interpret the filled maps and the visualization is **cluttered**. What's the significance of a filled map here?

Design and Visualization Improvements:

On analyzing the file with the details of the dataset, I understood that the positive value in the data indicates that male access exceeds female access, while negative values indicate that female access is greater. This information could have been implemented in the visualization The dataset contains information about 100 countries. After identifying these factors, I decided to create an improved visualization with the following considerations:

1. Analyze if there is any relationship between internet access and mobile phone access gender gap, possibly using a different type of chart.
2. Allow the usage of a filter to deep dive into a particular country.
3. **Use colors (<6) and a sequential color gradient** for comparing ratio measures
4. Allow for analysis of the top and bottom countries in the dataset against each of the parameters, using a different type of chart.
5. Use **Gestalt's principles** into consideration while creating the visualization, especially principles of similarity, continuity, proximity, connectedness, and symmetry, wherever possible.
6. Allow for answering the questions: Which are the best/worst countries in terms of gender gaps in internet access, mobile phone access and % of users.

Makeover Visualization and Design Choices:

Taking the improvements into consideration, while I considered making slope graphs and box plots, I think the best and most appealing way that the data could be understood is using a scatterplot in this case. Hence, I decided to go with a scatterplot of the Gender Gap in Mobile phone access vs Internet Access to see the relationship between them among different countries and how they stack up overall.

Additionally, I decided to create bar graphs with parameter filters to show the top and bottom 40 countries against each of the parameters according to the interest of the user. I chose the number 40 as it would cover 80 out of 100 countries, that is 80% of the data.

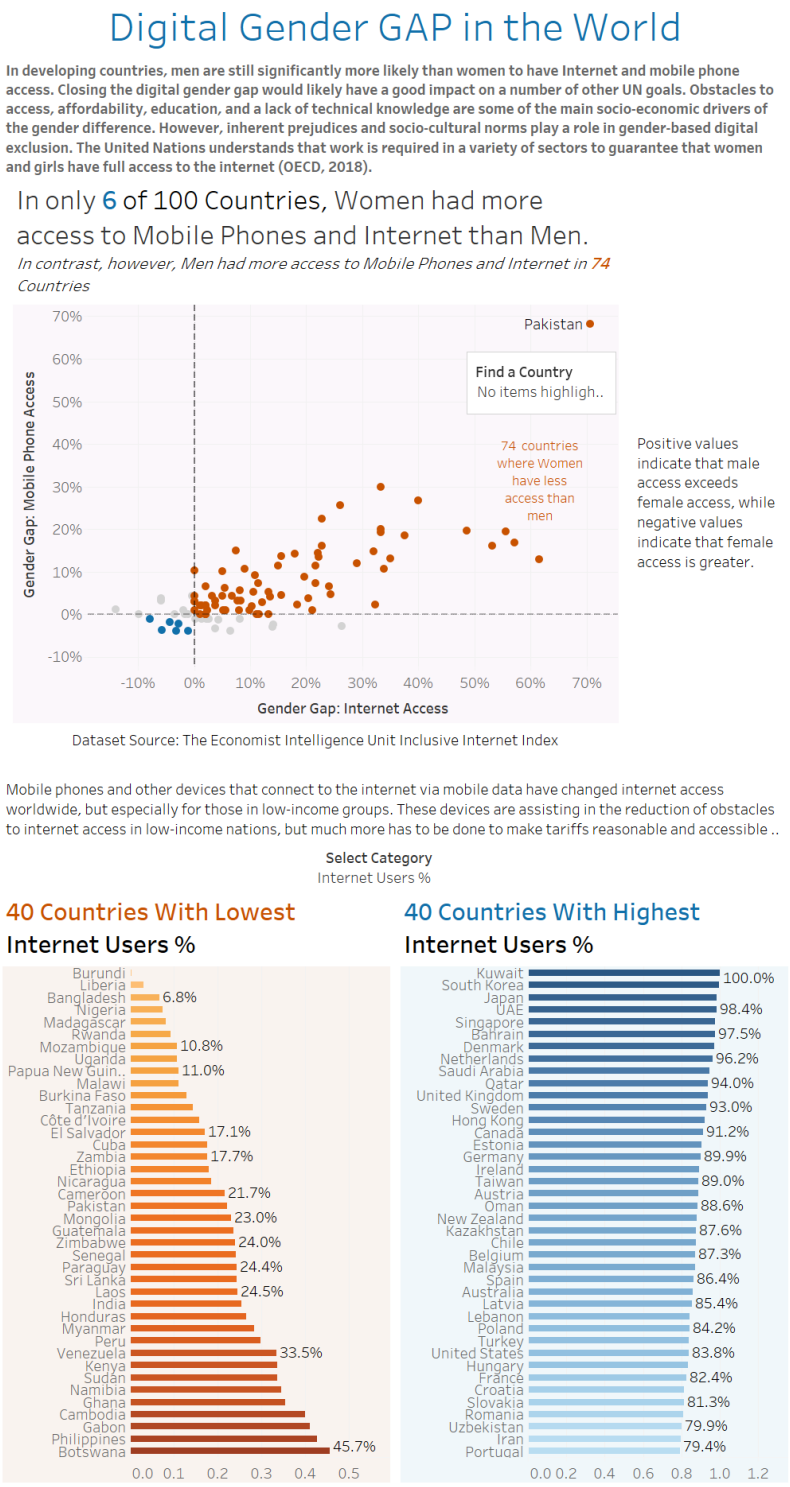

Figure 2: Makeover Visualization

Design Choices

My idea was to create a clean and decluttered visualization with the best data-ink ratio possible as well as all the details that are needed.

Scatterplot:

1. Divided the data into 4 quadrants to visualize the countries where both gender gap parameters are negative/positive
2. Created a highlighter for country to identify the value of a particular country of interest
3. Used 2 colors to differentiate the best and the worst, using a **Colorblind Tableau Template**. Red was chosen to be the negative value of interest and blue as the positive value of interest, as they are those of interest. The other countries with either one parameter were marked in grey.
4. Created a measured field to differentiate the 4 quadrants in the scatterplot.
5. Added a text for identifying the number of countries where improvement is required
6. Called out the country with both parameters lowest, and outlier – 'Pakistan'
7. Added title, caption and footnotes to detail the findings.

Bar Graphs:

1. Created two parallel graphs showing top & bottom 40 countries respectively.
2. Created a parameter filter as a drop-down to select the parameter of choice
3. Used the same blue and red colors as **sequential color gradients** to analyze the countries
4. Divided the data into 4 quadrants to visualize the countries where both gender gap parameters are negative/positive
5. Created a highlighter for country to identify the value of a particular country of interest
6. Used 2 colors to differentiate the best and the worst, using a **Colorblind Tableau Template**. Red was chosen to be the negative value of interest and blue as the positive value of interest, as they are those of interest. The other countries with either one parameter were marked in grey.
7. Created a measured field to differentiate the 4 quadrants in the scatterplot.
8. Added a text for identifying the number of countries where improvement is required
9. Called out the country with both parameters lowest, and outlier – 'Pakistan'
10. Added title, caption and footnotes to detail the findings.

Effects of Makeover Visualization and Analysis:

The created visualization addressed all the 6 improvements that I was looking to integrate. It also allowed for analysis of effects of the variables and we could gather that 74 countries have % of male access to mobile phones as well as internets as higher than female. Only 6 countries had better female access than males, which were China, Argentina, United Kingdom, Kazakhstan, Venezuela and Argentina. This shows that there is much to be improved upon and the countries need to work on the same. According to the Inclusive Internet Index, "56% of female respondents use a mobile phone to complete financial duties, compared to 48% of men," among those with access to a mobile phone. These gadgets are important agents for gender parity, economic participation, and financial independence and control, as seen by the 8% positive differential. Avoiding a map allowed for direct analysis of countries.

I also created a dashboard with interactivity additionally, allowing to dive into the top/bottom 40 countries only, if required by clicking a button. Both the dashboards have been published onto Tableau Public site within the same story.

Tableau Public Link : Digital access to the internet and mobile phones compared across men and women | Tableau Public