Bar Charts and Beyond: Choosing Data Visualizations

switchboard
connecting resettlement experts

Table of Contents

Bar Charts and Beyond: Choosing Data Visualizations

About this Toolkit

Data visualization is the process of presenting information using visuals such as plots, charts, and diagrams. The goal of data visualization is to help your audience understand the main message of your data without introducing additional clutter. In many cases, easy-to-read forms of visualization, such as bar and pie charts, are sufficient to present your message. However, sometimes the data you want to show might be more nuanced, so more advanced visuals can help. This toolkit will introduce you to various data visualization types, starting from the simplest and progressing to more advanced options.

The toolkit includes three components: a guide, a spreadsheet, and instructions. This guide offers tips for selecting appropriate visualizations and explains more complex visuals like boxplots, spider plots, and heatmaps. It is accompanied by a spreadsheet with sample data and visualizations, meant to be used alongside the plot-by-plot instructions. The goal is for learners to use the guide, spreadsheet, and instructions to practice selecting and creating visualizations.

Ready to practice? Here's how to get the most out of this toolkit!

1. Start with this Word document ("Bar Charts and Beyond: Choosing Data Visualizations")—this **guide** will help you understand the context behind the plots. Read through the parts that are relevant to your interests.
2. Open the Excel **spreadsheet** ("Bar Charts and Beyond: Visualizing in Excel") and locate the sample data for the plot types you are interested in. Start with one or two plot types based on your comfort level, data, and available time.
3. Open the **instructions** document ("Bar Charts and Beyond: Plot-by-Plot Instructions"). Follow the instructions to replicate the example plots using the sample data in the spreadsheet.

You can refer to back to this guide as needed while working through the spreadsheet and instructions.

Getting Started: Data Visualization Considerations

Data visualization presents both a challenge and an opportunity. There are often different ways to visualize a given dataset, which provides options but also requires you to consider many factors, including:

- **Define your purpose** – As a first step, ask yourself why you are visualizing this data. Are you presenting results to a funder, writing a report, or sharing data with colleagues? Knowing the purpose of the visualization will help you decide what data to present. In addition, consider who your audience is, their stakes in your results, and how comfortable they are with interpreting data. The key is to tailor the visualization to meet the audience's needs and comfort level thus making the data easier to understand.
- **Consider your data:**
  - **How much data is there?** A **variable** is a characteristic or quantity whose value interests you, so you measure and record it. A variable's value could be a number (such as a test score), a category (such as an age group), or a word/phrase (such as a client's description of their needs). The number of variables (or data points) likely affects your visualization choice. Some visualizations become too messy or confusing when used with a large amount of data.
  - **What kind of data is it?** Consider whether your data is **numerical** (number-based) or **categorical** (word-based). A **plot** is a way to show data visually using elements such as points, lines, or bars on a graph. It helps you see relationships, trends, or patterns in the data more clearly. While some plots, like bar graphs, are best suited for numerical data, others, like pie charts, might be suitable for categorical data.
  - **How complex is your data?** – Sometimes your data tells a straightforward story, but other times, the result is more nuanced. For example, a program might show overall improvements, but some individuals may experience greater gains than others. In a situation like this, your visualization should highlight the general trend and the variations within the data.
- **Consider available space and time** – Consider the amount of additional explanation you are able to provide, whether your visualization is on a printed page, presentation, or other format. The amount of space or time to explain a visualization likely depends on its complexity. For example, a simple bar chart might not need any explanation while a heatmap will likely require some context.
- **Keep it simple.** Cognitive load theory suggests that presenting information in a simple, easily digestible format helps the brain process it faster and more accurately.:
  - Focus on clarity by eliminating unnecessary clutter, choosing readable fonts and colors, and organizing information logically. The simpler the visuals, the easier your audience can focus on the key points.
  - Use simple and accessible language by avoiding jargon or terms that might confuse your audience. Similarly, avoid overwhelming your audience with statistics they may not understand or care about.
  - Simplify your visualization. Title your plot with the key takeaway to help your audience instantly grasp your message. Additionally, use colors purposefully to highlight key data points and help your audience focus on what is most important.

Simple Options: Bar, Line, and Pie Charts

Bar charts, line graphs, and pie charts are a good choice for visualizing most program data. They are more accessible than complex visualizations and communicate results clearly and simply. In this section, you will learn more about the distinctions between these three visualizations and how to decide which is most appropriate for your data. You will also find examples of different ways to use bar charts.

Bar Charts

Bar charts are sometimes divided into two categories: **horizontal** bar charts and **vertical** column charts. However, in this guide, they are referred to as bar charts, with the orientation of the bars specified as necessary.

Bar charts are useful when your data is arranged in categories (e.g. Figure 1), such as the number of enrolled individuals in each program your organization runs, or the total amount of money raised in each region.

Bar charts can also be used to show changes over time, but this is best reserved for when the changes are fairly large, making it easy to distinguish the different bar lengths by eye. For example, Figure 2 uses colors and data labels to draw the viewer's eye to the most important finding, namely the drastic increase in Iraqi enrollments over time.

Figure 1. The total number of participants in the program has risen substantially between FY21 and FY24.

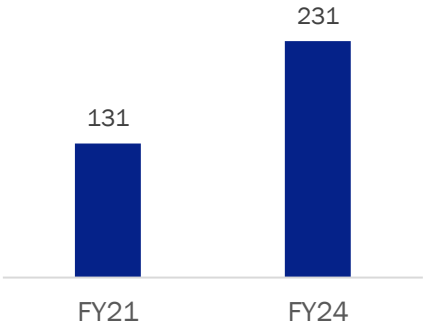
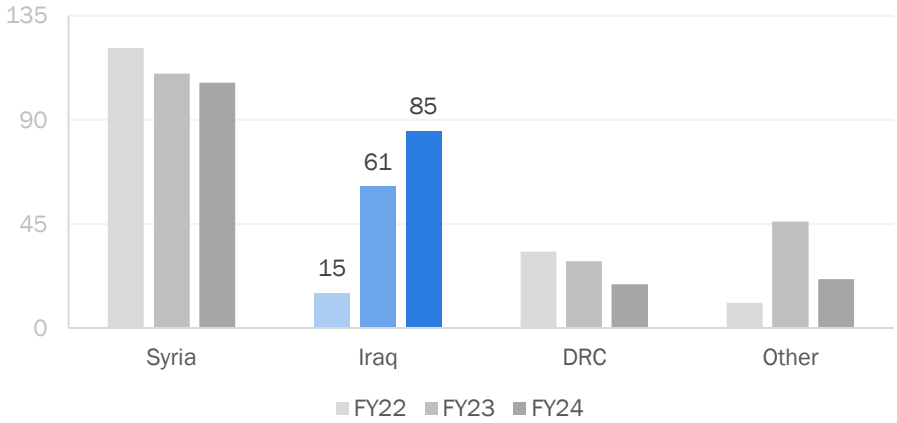
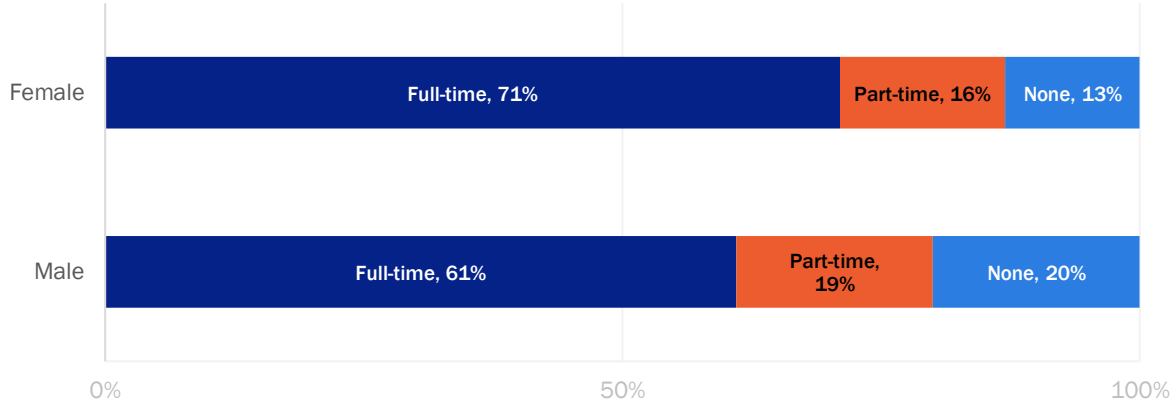


Figure 2. Following increased outreach to the **Iraqi community**, ESL enrollments among this group increased by 570% from FY22 to FY24.



You can also use a bar chart to show a part of the whole. This is achieved by creating a "100% stacked bar", where the entire bar represents the whole or 100%. Each component then makes up part of the bar. The horizontal format makes the text easy to read and allows you to compare more than one set of values to each other. In Figure 3, the "whole" represents all clients of that group, with the bar divided into parts to show the percentage who obtained full-time employment, part-time employment, and no employment.

Figure 3. Female clients were more likely than male clients to obtain full-time employment

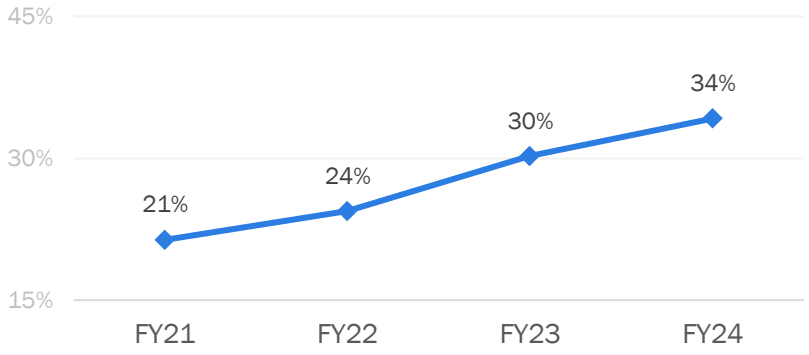


Line Graphs

Unlike bar charts, line graphs require both variables to be **numerical** or number-based. They are not suitable for plotting categories of data. Line graphs are best for showing how changing the value of one variable affects another. A common use is to show change over time. Because there is a line connecting the points, line graphs clearly show the overall trend in the data, i.e., whether a given variable is increasing or decreasing with time.

For this reason, line graphs are better than bar charts at showing modest changes over time. In resettlement programming, even small changes can have a big impact, so line graphs can be a useful tool for anyone visualizing data! For example, a line graph could be used to demonstrate the increase in arrivals your state has seen over the last few years or to show that the average time it takes for newcomers to find jobs has steadily decreased following the introduction of a new employment program.

Figure 4. The percent of participants who are women has risen steadily over the last four years.



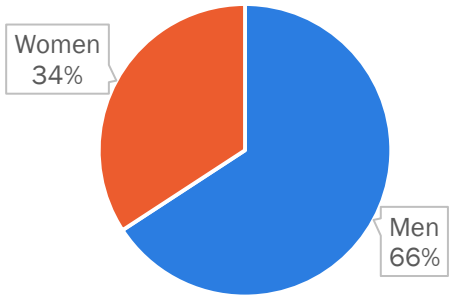
Pie Charts

Pie charts visually depict how much each "part" contributes to the whole. Common uses include showing the ethnic breakdown of a population, components of a budget, or percentage of survey respondents who chose the various options of a multiple-choice question.

Pie charts are not an appropriate choice for comparing a series of numbers that do not add up to a whole, such as the number of participants in a program across a span of years.

Pie charts are less effective when there are more than five "slices" of data (for example, when there are many ethnic groups in a population) or when several of the categories represent a small percentage of the whole. Both scenarios lead to pie charts that are challenging for readers to interpret.

Figure 5. **Female clients** still barely exceeded one-third of the participant population in FY24.



Which chart should you choose?

More than one type of chart may be reasonable for visualizing your data. While you might decide based on personal preferences or aesthetics, the summary below provides a good starting point for considering the strengths and weaknesses of each option.

Remember to ask for feedback on your data visualizations! Colleagues or even friends may provide valuable insights into the readability of your plots.

Chart Type	Strengths	Weaknesses
Bar chart	✓ Familiar to most people ✓ Good for comparing categories ✓ Can also show large changes over time across a few time periods	× Hard to distinguish small changes over time × Not suitable for all-numerical data × Not suited for displaying nuances in data
Line graph	✓ Familiar to most people ✓ Good for showing large and small changes over time ✓ Clearly shows the relationship between two numerical variables	× Not suitable for plotting text/categorical data × Can become cluttered if too many lines are plotted on one graph
Pie chart	✓ Familiar to most people ✓ Visually simple ✓ Suitable for showing parts of a whole	× Cannot be used for categories that are not parts of a whole × Hard to visually distinguish many categories or data with several small “slices”

More Complex Options: Box Plots, Spider Plots, Heatmaps

The next sections of this guide refer to different characteristics of datasets and plots. The glossary below offers a quick primer or refresher on these terms.

Glossary

Categorical Data – data that involves text or labels. Categorical datasets are sometimes stored as spreadsheets with dropdown menus, e.g., for sex, country of origin, native language, etc.

Disaggregation – the process of separating a dataset into smaller pieces based on one or more characteristics, such as sex, nationality, or age. This allows you to determine whether different groups experienced different outcomes.

Mean – the average value of a dataset.

Median – the middle value of a dataset.

Numerical Data – data that involves numbers. These can be whole numbers such as rankings or counts, decimal values such as dollar amounts, percentages, etc.

Outlier – a point whose value is far outside the typical range of values in a dataset, determined using a statistical formula.

Skew – how symmetrical a dataset is. A dataset that is *right skewed* will have most datapoints near a low value, with a few very high points. A dataset that is *left skewed* will have most datapoints near a high value, with a few very low points. A perfectly symmetrical dataset has no skew.

Spread – how far apart or close together numbers in a dataset are. If the data points are clustered near a particular value, this would indicate a small spread, while data points spanning a wide range of values would have a large spread.

Box Plots

Boxplots, also known as “box and whisker plots,” are used to show key statistical information about a sample. Just by looking at the plot, you can identify the **median** value. You can also see whether there are any outliers and visually assess the **spread** and **skew** of the data. Boxplots are small and visually simple, which makes them convenient for adding to a slide deck or a brief. You can add multiple boxplots to a set of axes, helping you compare pre-intervention and post-intervention results or data from multiple groups.

While they are visually simple, boxplots are not necessarily easy to understand. A boxplot may require additional explanation to help audiences grasp the concept and differentiate it from the familiar bar chart. Additionally, boxplots do not show data about individuals in a sample. You can draw conclusions about the sample as a whole but cannot speak to specific individuals’ outcomes based on the boxplot alone. A statement such as “Overall, the participants in this intervention improved their financial literacy skills” would be an acceptable conclusion to draw from a boxplot, while “Every participant improved their financial literacy skills” would not.

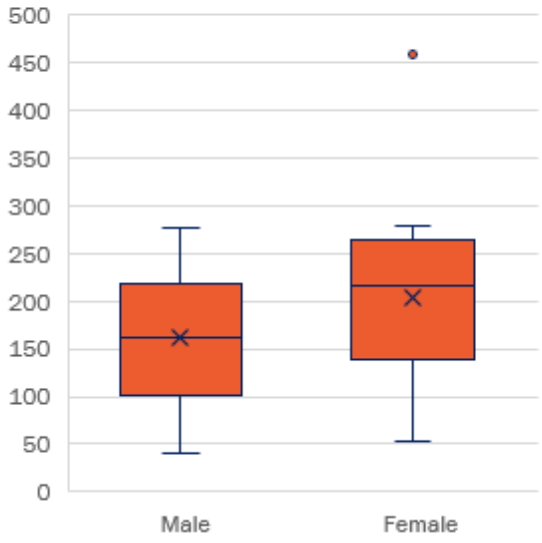
Example Scenario: Adult Literacy Intervention

Imagine that your organization has implemented an adult literacy project. You measure participants' scores before and after the intervention, using a test scored out of 500. You want to understand whether the intervention will be appropriate for all clients, so you decide to look at the men's and women's scores before implementing. You could certainly accomplish this with a bar chart, but you may wish to understand the **spread** of the data as well to be sure you understand your sample. A boxplot will give you this extra information. Below is an example boxplot, created from simulated (fake) data.

Let's walk through what this plot shows you, piece by piece:

- **The "box"** component shows the range of values of the middle 50% of the data. In this case, that means that half of the male prospective participants are in the score range of 102–219, and half of the female participants are in the range of 139–266. If you would like to show those exact values, you may add data labels to the plot.
- **The "whiskers"** show the upper and lower 25% of the data. Here, we can see that the lower 25% of the male sample has a score range of 41–101 while the female sample is from 54–138. Likewise, the upper 25% for the male sample ranges from 220–278, and 267–280 for the female sample.
- **The median** of the data is represented by a horizontal line in the box section. In this plot, the male median score is 163, while the female median is 216, indicating a much stronger median literacy score among the female sample.
- **The mean** of the data is represented by an X in the box section. This part of the plot is optional; you can choose to include or omit the mean. In this example, the mean of the male subset happens to be the same as the median (163), while the female mean is slightly lower than the median with a score of 204.
- **Outliers** are plotted in a boxplot as separate points so that they do not skew the results too much. Here, we can see that one woman with an unusually high English literacy level (a score of 460) enrolled, and this individual's result is plotted as its own point far above the rest of the boxplot.

Figure 6. Male and female participants achieved a similar range of scores on the pretest, but female participants tended toward higher scores on average.



What conclusions could you draw from this data visualization? As noted above, the median score for the female sample was substantially higher than for the male sample. You could get this information from a bar graph, though. What does the boxplot add?

A boxplot tells you about the data's **distribution**. In this case, it shows us that the female sample is skewed—its data points are more clustered around higher values. You can see this by comparing the length of the lower whisker and the higher whisker; the higher whisker is much shorter, indicating that the top 25% of the sample is distributed among a narrow range of scores. Some possible interpretations of this plot could be:

- "Though the median literacy test score among female participants was substantially higher than that of male participants, the range of the two samples was similar. This indicates that both groups include significant variation in score."
- "While the range of the male and female scores were similar, the female scores had a significantly higher median value and were clustered around higher values, suggesting that there is a group of women in the sample with a relatively high English literacy level."

In essence, the boxplot here gives us insight into the baseline levels for each group, which will help us to assess the intervention's effectiveness later. We can compare the posttest scores to understand how each group improved.

The **statistical significance** of your results, often expressed through a "p-value," is a way of indicating how confident you can be in your result. It is more commonly used in fields such as health care in which it's important to understand how significant the differences between two groups are. For example, statistical significance is crucial information for interpreting the results of a clinical trial, since it can help readers understand whether a treatment was effective. This concept is used less frequently in resettlement work but may be useful for those working with data related to the effectiveness of an intervention. Consider whether this is information your audience can use and understand.

Statistical significance can be challenging to incorporate into a visual. A simple way to do this is to treat it like a footnote: include an asterisk next to the relevant value(s), and then add a note at the bottom of the plot or in the caption with the p-value(s).

Spider Plots (Radar Graphs)

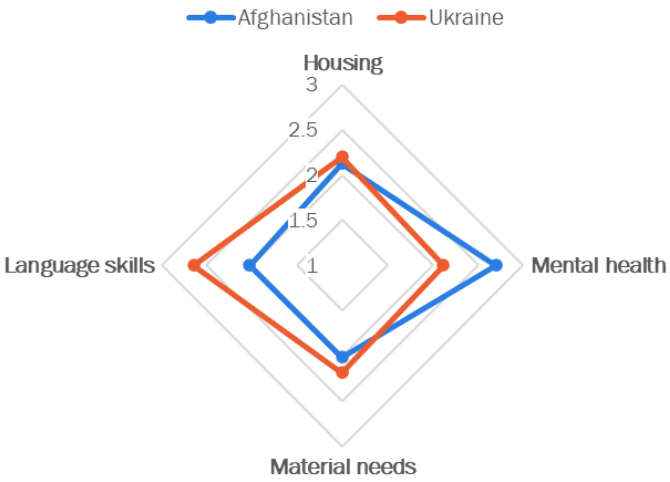
Spider plots, also known as radar graphs, are useful for showing the values of several variables at once. Often these variables are related to each other in some way and are thus arranged around a circle. For example, they may all represent components of a broader concept, such as wellness or financial stability. Spider plots can also be used to show the change in these variables from pre-intervention to post-intervention. If you want to compare data from a few groups (e.g., people of different sexes or ethnicities), you can place them on the same spider plot for easier visual comparison.

On the other hand, spider plots can sometimes be visually confusing. This is especially true if there are many variables, which can cause the plot to look cluttered. It is also difficult to show results from many groups or datasets in a compact way, so spider plots may not be the ideal visualization for a larger volume of data if space constraints exist.

Example Scenario: Needs Assessment

Before planning your next project, you want to better understand the needs of your community. In particular, you want to determine whether your clients from Afghanistan and Ukraine have different needs. Following a needs assessment survey that asked clients to rate the severity of different potential needs from 1 (mildest) to 3 (most severe), you produce the plot below (generated here from fictitious sample data).

Figure 7. Mental health is the highest-rated problem among **Afghans** compared to language skills among **Ukrainians**.



How can you read this plot and communicate it to others? First, let's discuss the structure:

- On a spider plot, the **variables** are positioned evenly around the outer edge. The position along the edge does not signify anything about the variable's value (e.g., housing is not more important simply because it is at the top here).
- Rather than axes running vertically and horizontally, the spider plot has a **radial axis**. The value of 0 is at the middle, with values increasing toward the edges, making it easy to visually compare values of a given variable.

What conclusions could you draw from this graph? Here are some examples:

- "Afghans rate mental health as their most severe problem, while Ukrainians emphasized language skills as a major issue."
- "Both groups rated each of the four problems with an average severity of 1.5–2.7, indicating that they fall in the moderate to severe range."

Heatmaps

Heatmaps are arranged in a matrix, with one variable running vertically and another running horizontally. Values of an "outcome" or "dependent" variable are displayed in each cell. Heatmaps are useful for visualizing the effects of multiple variables on another outcome variable. If you have two factors that may be potentially affecting your outcome variable, a heatmap can help you assess these impacts more easily than a table or bar chart. Heatmaps are also a compact way to convey a lot of information without appearing cluttered. The matrix structure keeps data orderly.

To visualize your data in a heatmap, the two variables that create the matrix must be categorical. You can put your numerical variable into this form by "binning it," i.e., by creating groupings of values. One example of this would be income data. To use people's incomes as one of the variables in the matrix for a heatmap, you would first have to bin the incomes. For example, you may define ranges that correspond to "low," "medium," and "high" income.

Besides the fact that this requires an extra step, it can be limiting if you want to identify a clear numerical correlation between variables. With a heatmap, you will not be able to see the individual values of each variable

(i.e., the value of the variable for each individual client in the heatmap). Finally, heatmaps take up more space than other types of visualizations, especially if you have several variable categories.

Example Scenario: Adult Literacy Intervention

Returning to the adult literacy example, let's imagine you have additional demographic data for the sample. You may know participants' country of origin and age, for example, and you want to understand whether some groups are more literate than others. You decide to use a scale in which "below 1" indicates very basic English literacy, while a 5 indicates mastery of complex text.

You produce this heatmap of participants' age groups and countries of origin:

Figure 8. The English literacy scores of Venezuelans aged 30-39 and Eritreans aged 50-64 lag behind those of other groups.

	18-22	23-29	30-39	40-49	50-64	65-75	76+
Afghanistan	2.78	3.36	2.67	3.14	2.69	2.75	2.60
Ukraine	1.80	3.00	2.10	2.14	3.00	2.53	2.17
Myanmar	2.50	1.75	2.67	2.00	2.00	2.44	2.58
Venezuela	2.46	2.00	1.25	2.40	2.00	2.36	1.64
Eritrea	2.46	2.64	3.00	2.58	1.50	2.45	2.70
Sudan	2.70	2.77	3.67	2.11	3.11	2.75	2.60
Syria	3.75	2.53	2.64	2.14	2.90	2.40	2.71

In this case, darker orange colors indicate lower English literacy levels. Here is how to read this heatmap:

- Each cell represents a pairing of origin country and age group. For example, the bottom right cell indicates Syrians aged 76 and older.
- The value in the cell tells you the average English literacy level for people who meet those characteristics. For example, Syrians aged 76 and older averaged a literacy level of 2.71 (out of 5).
- The color scale helps you do a quick visual inspection of the data. In this case, the darker orange cells immediately jump out, showing which groups have particularly low English literacy levels.

Some potential messages from this heatmap could be:

- "Younger Syrian newcomers appear to have much stronger English literacy levels than most other groups. Their literacy scores exceed those of young adults from other countries, as well as older Syrians."
- "Venezuelans aged 30–39 and Eritreans aged 50–64 appear to have particularly low English literacy levels. These groups may benefit from customized lesson plans or additional instruction."

Which chart should you choose?

As with the bar, line, and pie charts, more than one type of chart may be reasonable for visualizing your data. While you might decide based on personal preferences or aesthetics, the summary below provides a good starting point for considering the strengths and weaknesses of boxplots, spider plots, and heatmaps.

Remember to ask for feedback on your data visualizations! Consider testing your visualizations on people who have a similar background and data literacy level to your intended audience, when possible.

Chart Type	Strengths	Weaknesses
Boxplot	✓ Shows key statistical values of a sample ✓ Visually displays outliers, spread, and skew ✓ Useful for comparing groups or pre-/post-intervention data	× Less familiar to most people, potentially requiring more explanation × While useful for showing changes in a population, does not include information about individual changes
Spider plot	✓ Good for showing the values of multiple related variables on one plot ✓ Can be used to compare groups or pre-/post-intervention data	× Less familiar to most people, potentially requiring more explanation × Can be visually confusing if there are too many groups or variables
Heatmap	✓ Good for showing how three variables are related ✓ Clear matrix structure keeps the plot visually clean	× Data must be organized into categories (e.g., a numerical variable such as age must be binned into age groups) × Can become large/unwieldy

Conclusion

Data visualization is a powerful communication tool. It can help you convey key messages to funders, clients, staff, and other stakeholders, in addition to helping you understand your own data. With so many options (including many not covered here), selecting the right visualization for your data can be a challenge. However, using this toolkit, you can choose the visualization that is best for your data and audience. With the accompanying spreadsheet and instructions, you can practice recreating the sample plots before trying out your new skills on your own data!

Recommended Resources

- Archived Webinar: Using Data to Tell the Stories of Our Immigration Programs
- Blog Post: Data Communication for Resettlement Organizations
- Blog Post: Using Data to Tell Stories Without Putting Your Audience to Sleep

About Switchboard

Switchboard is a one-stop resource hub for refugee service providers in the United States. With the support of the Office of Refugee Resettlement (ORR), we offer tools and materials, learning opportunities, research, and technical assistance on resettlement-related topics. From employment, education, and health to monitoring and evaluation, Switchboard's focus areas reflect real-world needs.

The IRC received competitive funding through the U.S. Department of Health and Human Services, Administration for Children and Families, Grant #90RB0053. The project is 100% financed by federal funds. The contents of this document are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. Department of Health and Human Services, Administration for Children and Families.

switchboard
connecting resettlement experts