MODULE ANSWER | Foundation

Use visualizations to tell a story

Differentiate between visualizations to explore the data and to communicate insights.

Quantitative or statistical attributes of datasets such as mean, median, and mode are easy to calculate and verify. However, when it comes to more sophisticated analyses, the underlying assumptions and choice of methods often plays a big role in the results.

This makes analyses on large datasets and models harder to communicate – even if these analyses might be significantly more useful to readers in their tasks than statistical attributes. Visualizations are a popular strategy to communicate complex analysis quite simply, especially if your analysis results in large data tables. But not everyone is an expert, and visualizations often require careful thought and additional treatment. Different readers will want visualizations at different fidelities. Your choice of visualization, interactive features, small changes to color palettes, naming your axes, and providing ticks at the right intervals all can significantly impact what your readers learn from the visualization.

Succinctly communicate to a variety of readers...summary visualizations are great for this.

People + AI Research

Exploratory Data Visualizations

Practitioners typically use exploratory data visualizations to delve deep into features and attributes of their datasets (exploration) or run specific analyses pertaining (execution) to their use case. These visualizations can reveal what you did and didn't know about your dataset, but often require more domain and visualization specific background knowledge to operate and interpret.

Use exploratory data visualizations to help discover insights and statistics to be communicated in a Data Card. In doing so, account for how these will need to be translated into summary visualizations so they are useful and readable in a Data Card.

Summary Data Visualizations

Summary visualizations answer specific questions by telling a data story that anyone can understand. They aggregate insights so visualizations are simple to interpret and usable by non-experts. Data Cards have to succinctly communicate to a variety of readers, and summary visualizations are great for this. Reduce and simplify your analyses into clear explanations that can be visually presented as aggregate views or trends.

Whether it's a scatterplot or an area chart, the visualization type and any additional treatment should be based on what you intend it to communicate. Use the table on purposeful visualizations to determine your plan of action.

Key Takeaways

- Visualizations are primarily of two kinds: those that summarize large datasets into digestible bites, and those for exploring and analyzing data.

- When using data tables, infographics, or data visualizations, explain what it says, how to interpret it, and offer some lifelines to help readers explore the data.

Actions For Your Team

1. **Summarize and explain.** Summarize exploratory analyses in Data Cards in appropriate explanatory visualizations, which are accompanied by captions that highlight the points you want to communicate.

2. **Provide a guided hands-on experience.** If possible, link to a demo or page where readers can also perform similar explorations, along with one or two questions for readers to investigate. This prevents unnecessary logical fallacies, like cherry-picking.

3. **Help readers recreate your visualizations.** If possible, link to code or notebooks in which readers can explore the tasks and operations performed. This encourages reproducibility and verification of your content.

4. **Annotate visualizations as necessary.** You may choose to layer in additional annotations that highlight specifics you want readers to learn from the visualization. Carefully craft a caption that signposts these insights to make them accessible, and append any other context, explanations, or supporting diagrams that will help readers interpret these results.
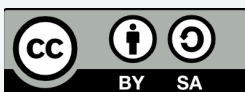
Resource

Purposeful Visualizations

Goal of Visualization	Examples of Tasks or Decisions	Visualization Tools & Libraries	Action for your team
Exploration Visualizations that help users explore their datasets and its derivatives	to make decisions about training or testing a model.	E.g. Facets, TFDV, Know Your Data	To present these visualizations in a data card, summarize them in an appropriate explanatory visualization. This should be accompanied by a caption that highlights the points you want to communicate. If possible, link to a demo or page where readers can also perform similar explorations, along with one or two questions for readers to investigate. This prevents unnecessary logical fallacies.
	validate a trained model's performance, diagnose or debug models.	E.g. What If Tool, LIT.	
Execution Visualizations that help users perform a series of tasks and operations on datasets	Adapting or transforming a dataset, or creating derivatives; or for exploring specific model behaviors on a given datapoint or subset of data points.	E.g. Visualizing the detection of anomalies, comparative distributions, trends charts; Confusion matrices, heatmaps, salience charts	Depending on what they are, these visualizations may already summarize the key takeaways you want to include in the Data Card. If not, summarize these into an appropriate explanatory visualization. Any such visualization should be accompanied by a caption that highlights the points you want to communicate. If possible, link to code or notebooks in which readers can explore the tasks and operations performed. This encourages reproducibility and verification of your content.
	Report common statistical, mathematical, or quantitative attributes of a dataset and its derivatives.	E.g. Distributions, PR, ROC curves	

	... a set of insights and paint a robust picture of a certain phenomenon present in the dataset.	E.g. [...]	These visualizations may already summarize the results and insights necessary to answer a question in the Data Card.
Explanation Visualizations that further a user's understanding of the data by demonstrating	relationships between features or the impact of a specific process to explain a certain phenomenon in the dataset.	E.g. KYD's NPMi Analysis	You may chose to layer in additional annotations that highlight specifics you want readers to learn from the visualization. Carefully craft a caption that signposts these insights to make them accessible, and append any other context, explanations, or supporting diagrams that will help readers interpret these results.

Considerations

→ What is the purpose or importance of the visualization? Does it communicate what is important about the data?

→ What is the expected impact of the visualization? Do readers have sufficient information to benefit from the visualization?

→ Does the form of visualization organize the underlying data at the appropriate fidelity? How can the visualization be misinterpreted? What details in the visualization, caption, or treatment can prevent misinterpretation?