In this document, we will provide an explanation of the mediator response protocol, and a description of the templates used to respond to participant requests.

**Visualization templates:**

The mediator used Tableau to produce visualizations in response to participant requests. All of these responses are available in the ‘visualizations’ folder, for each participant in the study.

The most common response types were: bar charts, line charts, multi-line charts, grouped bar charts and heat-maps showing crime rate frequencies for each block within the 4 neighborhoods. These can be found in the visualizations folder.

When we first ran the study, the mediator would also provide pie charts and tree maps to participants, along with bar charts. But participants typically disliked them and would move them aside or ask for them to be removed. So, over time the mediator focused on the standard views more often.

Several times, a participant indicated a region in a heatmap, for instance concentrating on a region with high rates of theft, and would want to see visualizations associated with that region. In response, the mediator would provide a new map view with this region visually highlighted, and the requested views showing data from the selected region.

There were a few responses to participant requests that were classified as errors, because the mediator chose a poor template. These were not considered in the analysis.

**Mediator response overview:**

At the start of the study, the mediator would have a conversational interaction with the participant, in order to indicate that they could see and hear them, and to help them feel at ease, and encourage them to ‘ask anything that will help in your analysis. This can be seen in the chat box transcripts folder, one for each participant.

When a participant posed a request, the mediator would switch the status bar to ‘Processing’, and would write a comment in the chat box (such as ‘Sure, I can do that’ or ‘Yes, just a moment’), to ensure that the participant knew they were responding to their request. The mediator generally tried to provide a visualization response, even when the request was unclear, but they refrained from asking clarification questions unless absolutely necessary. If the request was completely unclear, or if they didn’t completely remember the request, they might ask the participant to repeat the request.

Participants were instructed to not pose additional requests while the mediator responded. Occasionally they forgot, and there were times where the mediator responded to these mid-response requests. The goal was to be an assistant to the participant, and the mediator sometimes responded to their request because it furthered their analysis.

The mediator would generate one or several static images of visualizations, using the Tableau export function. Using sage2, these images would be dropped at once on the display and positioned for the user by the mediator. Positioning decisions are explained below.

There are a few instances where the mediator chose to provide visualization responses in a staggered order. For instance, if there were many views that they needed to produce and the participant had been waiting, the mediator would provide some of the views, and write in the chat box ‘the remaining views are coming, just a moment.’

There are a few instances where a visualization had the wrong title, or had an error, the mediator would correct the title or error, replace the view, and explain the action in the chat box to the participant.

While the mediator created the visualizations, the in-room aid would use sage2 to move aside previous views that they believed to be unrelated to the request. This action was, at times, *not* well received by the participants, and we comment on our view positioning challenges in the paper.

There was one instance where Tableau crashed, and the mediator needed to put up the ‘error’ status, and explain the situation in the chat box. There was an instance where the microphone failed to transmit the request to the mediator at a good volume, and it needed new batteries. These disruptions were brief (less than a few minutes) and the study resumed where it left off.

**How did the mediator chose visualizations:**

Participants were told that the mediator was a human assistant, with experience creating visualizations of this data. The general approach of the mediator was to make reasonable judgements about visualization creation, focusing on the subsets of the data and data attributes the participant was interested in exploring. The intention was not to mimic the capabilities of a system, with procedural and consistent responses to a pre-defined set of request types. We did not frame or describe the study to participants in a way that would make this a ‘Wizard of Oz’ study.

The mediator acted as an assistant, and attempted to select appropriate templates for enumerated attributes and tasks (eg. line charts for temporal attributes or trend analysis questions, heat-maps for geo-spatial analysis questions, bar charts for examining frequencies or distributions).

The typical response to a targeted request was a single view, with a template selected for the single enumerated attribute, and the subset of interest used as a filter on the view.

The typical response to a comparison request for distributions or frequencies, was either a set of small multiple bar charts (one for each of the enumerated subsets of interest) or a single window with multiple sub-views. Alternately, the typical response to comparing trends was a single window with one multi-line chart.

The typical response to a browse request, was a set of views with a common subset of the data, and one visualization for each of the available attributes (7 visualizations). Some browse requests had a single subset of the data, and several specific attributes, and each one was given a separate view.

The responses to multifaceted requests were more diverse. Typically, the mediator would select some permutation of subsets and data attributes, matched to appropriate templates. The goal was to provide a set of reasonable responses which would provide alternate ways to view the data.

The typical response to a referential request would involve duplicating the template or templates, and pivoting them to the new subset or data attribute.

However, as described in the study, participants gave the mediator significant leeway in selecting sets of visualizations in response to their requests. The mediator had significant degrees of freedom in combining and subdividing enumerated attributes and subsets among available visualization templates, and at times made different choices. Here are some atypical response choices by the mediator.

At times, the mediator would provide an additional view to give context to the participant’s request. For instance, if a participant asked for a specific view about an uncommon crime type, without first seeing how often this crime type occurred compared to others, the mediator might also give an overview of the frequency of all crimes by all crime types. Or, if a participant asked to see a heat map of homicides, the mediator might provide context by giving an overview heat map showing all crimes.

At other times, the mediator might choose to respond to a comparison request involving a temporal attribute with both a multi-line chart and a set of separate line charts, because there are strengths and weaknesses for each of these visualization types. They would sometimes aggregate and divide participant requests for multiple non-overlapping subsets of the data.

The general approach was to be of assistance, and to address the question reasonably and appropriately, so that the participant could make observations and proceed with their analysis.

**Limitations in our mediator visualization selection protocol:**

We did not arrive at a system to classify participant requests until after completion of the study. So, we couldn’t fully anticipate the range of possible questions the mediator would need to respond to. And this uncertainty made it difficult to create a systematic procedure for how to produce visualizations in response to participant requests. Essentially, there is a chicken and egg problem: in order to design a clear scheme for responding to participant requests to conduct the study, we needed to know what participants would ask and have a scheme to categorize these requests, which we wouldn’t have until completing the analysis of the study.

This complexity is a major reason that we do not frame this as a Wizard-of-oz study. Participants were told that the mediator was a human assistant. The mediator’s approach to selecting visualizations was to act as an experienced assistant, and make a reasonable judgement call about which visualization or set of visualizations would address the participant’s request and allow them to continue in the analysis.

After analyzing the data, and arriving at a scheme for classifying participant requests, it may be clear in hindsight that the mediator should have responded to a particular request differently. The mediator needed to make fast judgment calls in response to varied and unexpected requests from participants, and this leads to some variability.

We address this in our analysis in several ways.

First, we did not consider how the mediator responded to participant requests in our coding scheme. We labeled the request itself, considering the references to prior visualizations on screen and the analysis context. For example, if a participant asked to see ‘Compare theft and battery, by day of the week’, a request we now understand to be a comparison request and which we may believe ought to be responded to with 2 visualizations, showing thefts in one and batteries in another. However, the mediator, making a reasonable judgment call about what might be an effective response, could give 3 visualization- a view showing all crimes, for context, and a view showing theft and battery aggregated together by day of the week, to address the uncertainty in their request, and a view showing theft and battery in a grouped bar chart, to enable comparison. By focusing on what the participant asked for, and leaving aside how the mediator responded, we bracket the question of what the optimal response to this query might be, and leave this as a discussion point for future work.

Second, we do not analyze the sequence of request types or the flow from one request type to another, because this is influenced by the response provided by the mediator. If we were studying a realized system, with pre-programmed consistent responses to participant requests, then this would be a valuable direction for an analysis, and should be considered in future work.