**Coding process**

**First coding pass:**

We started our coding scheme by creating an excel spreadsheet, which is in this repository. We coded from collected data all created visualization during each session. For every visualization, we included information about how this visualization was created and its relations to other visualization in the analysis session. The following explains columns in the spreadsheet.

**Basic information**

Column A: Session number

Column B: Description of the event of creating a visualization

**Visualization details**

Column E: number of the created visualization

Column E: type of the visualization

Column G: dataset from where the visualization data comes

Column H: used attributes to filter/create visualization

**Relation to other views & tasks**

Column J: data scope of the visualization either a new subset or derived from another one

Column K: if new subset, relation to other subsets

Column L: the task that the visualization was derived for

Column M: relation to other views if any

Column N: multi level task, when a task followed by another task with more visualization (i.e. browse a and b, then correlate a, b, and c)

Column O: what changed when the visualization was derived form an earlier one

**Second coding pass:**

Then, using the data from the spreadsheet, we created search tress for all sessions. All final search trees are in this repository under the folder “search trees from the study”. Every node in the search tree represents a visualization created during the analysis session. Nodes are plotted in chronological order of creation.

During this phase, we added links between all nodes that share something (i.e. derived from, same data subset, used in same task, etc.)

**Finalizing coding scheme:**

Then, we grouped these relationships between nodes into a theme of two sets of codes.

First, we observed that some nodes in the search tress are those visualizations when participants filter a new subset of data. Those visualizations have no ancestors but they have descendants. So, the first set of codes represent the relationship of the new subset to other subsets in the session, and represent the initiation of analysis paths.

When a visualization is created with a new subset of data, we observed that this has one of the following options:

* This subset of data shares attributes with another subset but they are from different datasets 🡪 they are ***parallel***
* This subset of data is from the same dataset as another subset but from a different dimension space 🡪 they are ***subsequent***
* This subset of data is from the same dataset as another subset, but it is a drill down in the dimension space of that subset 🡪 it is ***focused*** dimension

Second, for every other visualization, we identified its relationships to other visualizations from the analysis task perspective, i.e. what analysis task was performed on this visualization, and with other visualization if any. Descendant visualizations compose the evolution of analysis paths. We observed that visualizations were created to do one of the following tasks:

* ***Compare***: between visualizations that share an attribute  and visual representation.
* ***Correlate***: exploration task to infer the de- pendency of one attribute on another mostly between visualizations  of the same representation but different at tributes to infer possible relationships.
* ***Validate***: between visualizations of similar attributes  and representation but different dimension  space of data points to validate observations made on an earlier visualization.
* ***Browse***: browsing interesting attributes that may involve interesting trends.