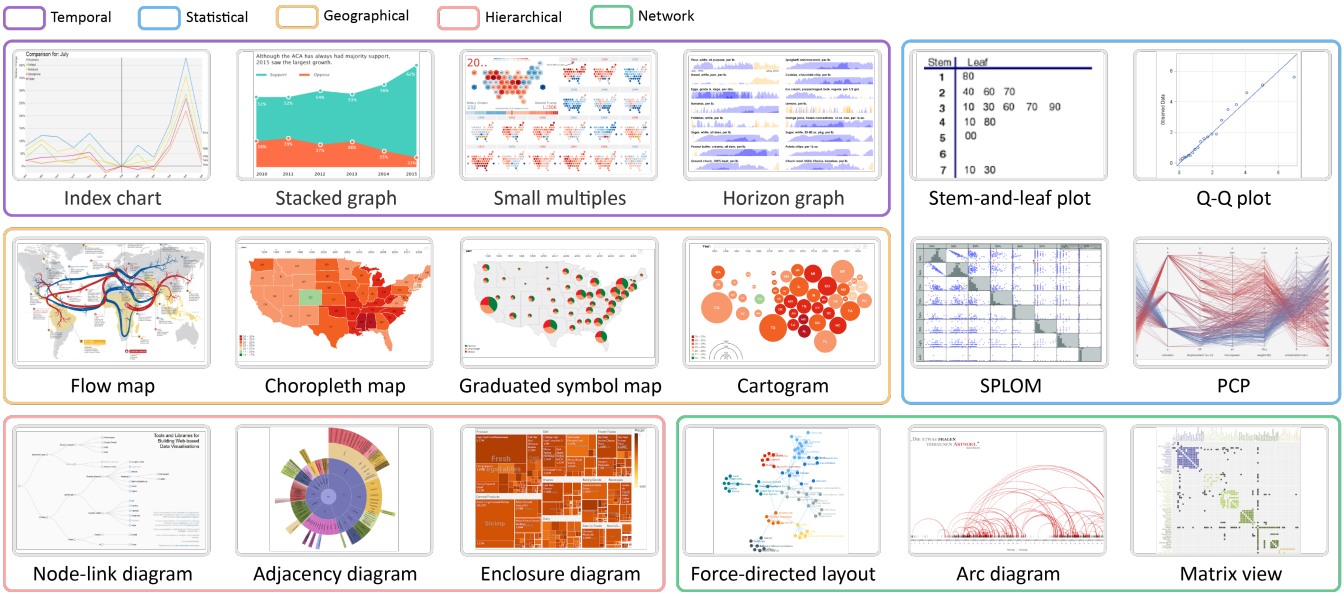

Figure 1: *The visualizations used in the preliminary study. We focused on five categories: temporal, statistical, geographical, hierarchical, and network. For each visualization type, we tried to collect maximum examples from data news and blogs. We chose data news and blogs because their visualizations are generally created from real data because we hoped to investigate the actual comprehension situations of end-users. For each sub-type of the visualizations, we finally selected 5 examples in the study. Thus,* $18 * 5 = 90$ *examples in all have been used in the study.*

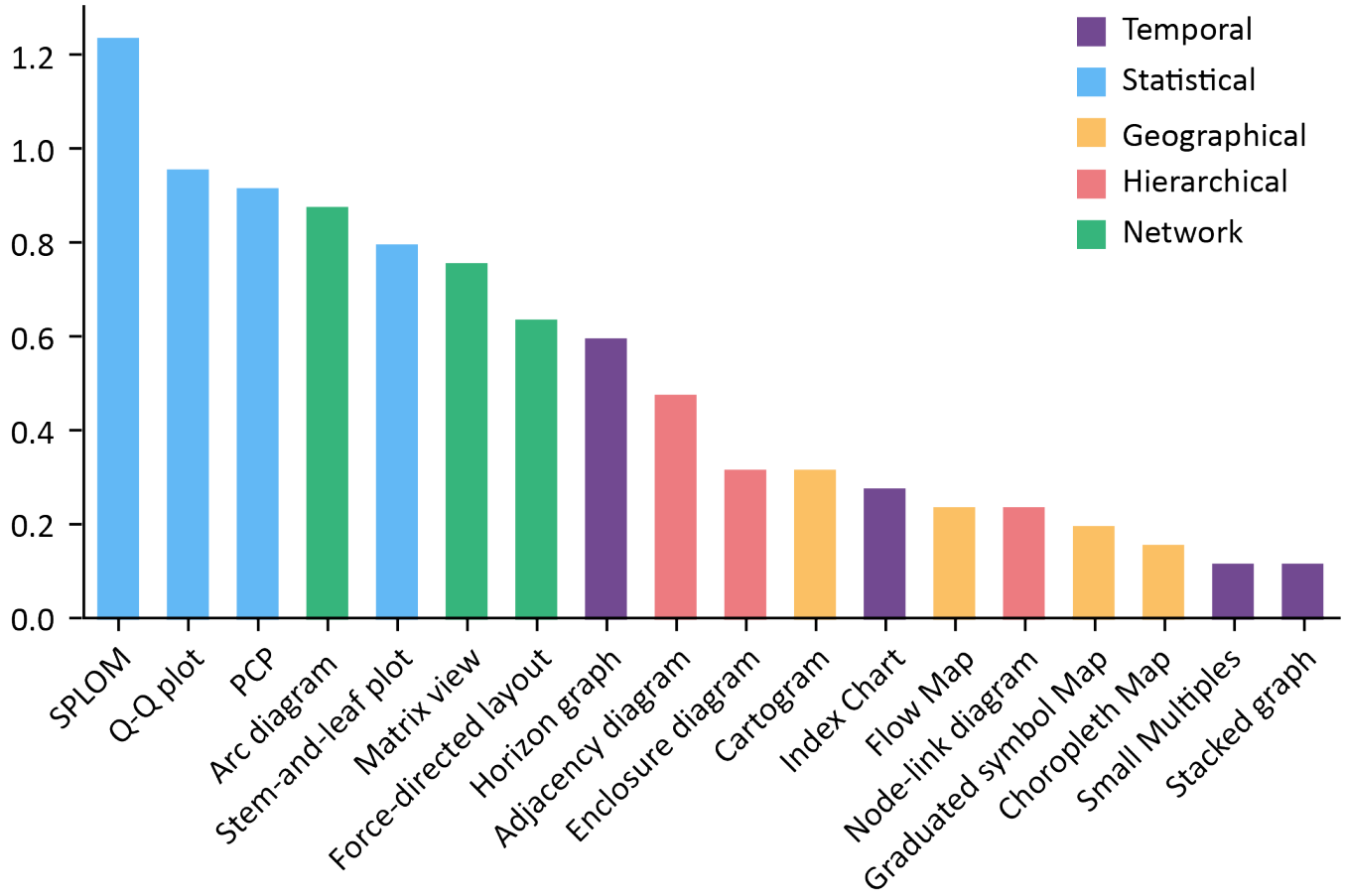

Figure 2: *The difficulty score of each sub-type of visualization. We quantify the difficulty of each visualization in the way that 0 for easy, 1 for medium, and 2 for difficult. The colors of the bars encode the high-level type of a sub-type. Statistical distribution and network are the two most difficult visualization types in our pilot study*

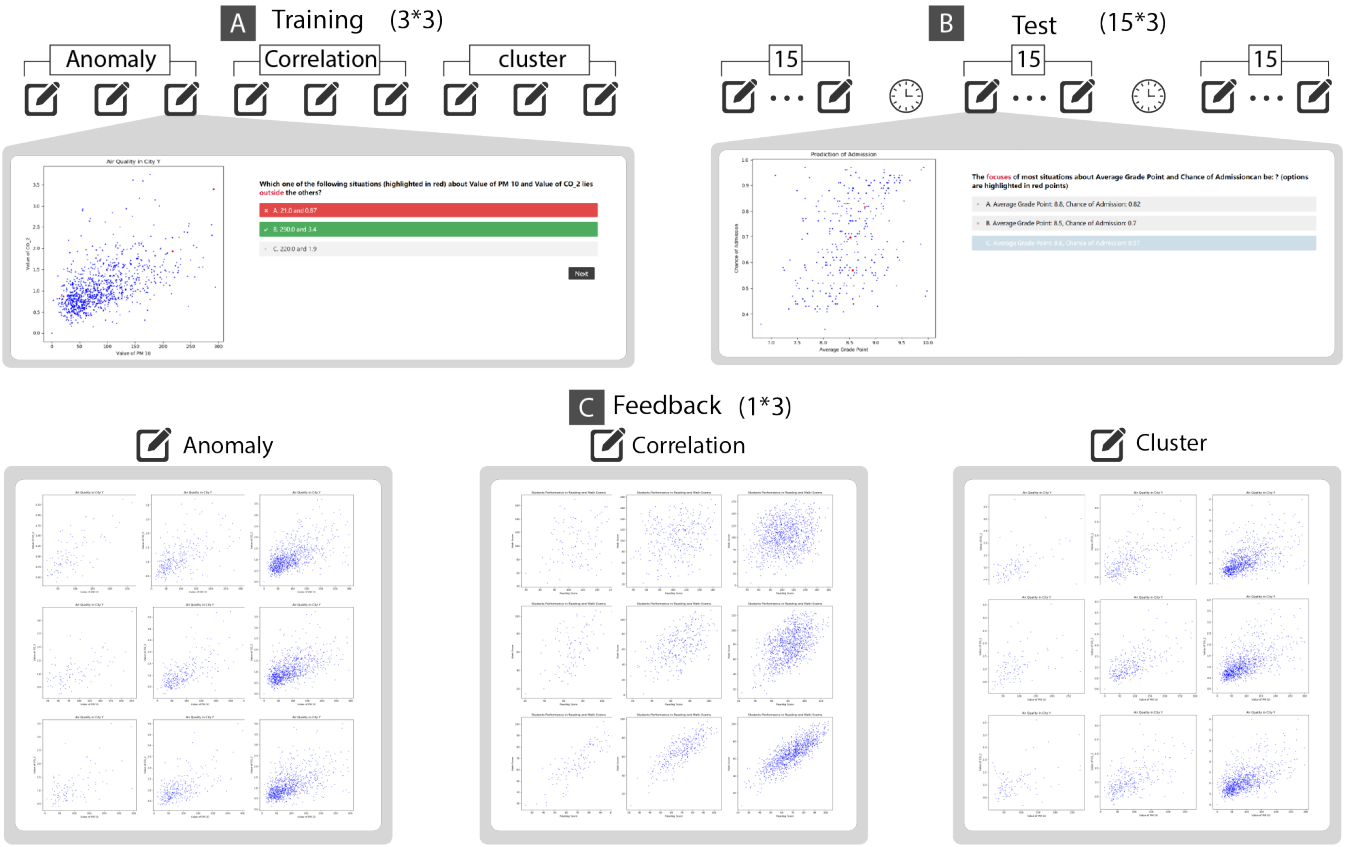

Figure 3: *The three sections in the experiment. The training section (A) contained 9 questions (3 (question) × 3 (task)) for a participant to get familiar with the tasks and system interactions. The answers were told to the participants. The test section (B) contained 45 questions 3 (condition) × 15 (dataset). No answers were provided, and participants required to finish the questions as soon as possible. A break was set after every 15 questions. The feedback section (C) contained three questions about the most difficult condition of the scatterplot in each kind of task.*