INFORMATION VISUALIZATION COURSEWORK

TEAM MEMBERS:

1. Prathiksha Ravoor : 2807514N
2. Akash Ananda Kumar - 2768146M
3. Adnan Qudsi - 2818495Q
4. Chiyue Cheng - 2721311C
5. Haipeng Xu - 2769175X

Q5. https://www.youtube.com/watch?v=aWTaCP5Ms4Y

QUESTIONS

Q1.

Title : Exploring the relationship between pokemon types and base stats using the pokemon dataset

Description: The dataset gives the information on various attributes which includes their names, types, abilities, moves, and base stats. The dataset contains data on nearly 800 unique pokemons and also includes 13 different variables.

The pokemon dataset can be categorized as a multivariate dataset, which includes multiple variables and features that can be used to explore the characteristics of different pokemon. The dataset has both categorical and continuous variables, such as type1, type2, total, attack, defense, special attack, and speed.

The dataset can be used to explore the relationship between pokemon types and their base stats. For example, a bar chart can be used to visualize the average base stat total for each type of pokemon. This visualization can also be useful to identify which type of pokemon is having high or low base stats. Another example is visualizing the scatter plot to explore the relationship between two different base stats, which can be Attack and Defense, with using different colors and symbols. This visualization can also be useful to identify which type of pokemon is having high or low values of each base stat.

Therefore, the pokemon dataset is a good resource for exploring the relationship between pokemon types and their base stats using different visualizations systems.

Q2.

In this pokemon dataset, users might want to perform the following actions:

1. Selection: Users might want to select a subset of the data based on any different circumstance. For example, they might want to select all the different pokemon types or only the pokemon type with particular base stat value. Selection is a useful tool for exploring and analyzing large datasets, which allows users to visualize specific subsets of data for their analysis.
2. Filtering: Users might want to filter the subset of the data based on different criteria. For example, they might want to filter the dataset to display the pokemon with a particular type of particular generation. Filtering is also a useful tool for exploring and analyzing large datasets.
3. Navigation: Users might want to navigate the dataset using different types of visualization techniques. For example, they might want to use a scatter plot to explore the relationship between different base stats and also to use a heat map to visualize the distribution of pokemon dataset with different types. Navigation tool is useful to analyze the relationship between two different patterns.
4. Reconfiguration: users might want to reconfigure the analyzed visualization in different aspects. For example, they might want to adjust the color scheme or axis labels of the graph to be in a better visualization of the dataset. This reconfiguration tool helps to customize the visualization relationships in a better view.

Q4.

Implementing the generalized selection on all three different visualization systems A, B, and C. in this pokemon dataset, the abstraction of different semantic structures levels are:

1. Entity level: This level explains each individual pokemon names, types, abilities, and base stats
2. Type level: This level explains each pokemon type which includes its names, type of pokemon generation, and its overall base stat distribution.
3. Generation level: This level explains each pokemon generation which includes the types it belongs to, and overall base stat distribution of that particular pokemon generation.

Based on this semantic structure, we can implement the traversal policy that allows the user to navigate and select the data at each given level of abstraction. For example, users can select a particular pokemon at the entity level and then select all the pokemon of specific type at the type level or specific generation at the generation level.

To implement this traversal policy in each three different systems, users can use visualization and filtering techniques as combined. For example, users can use scatterplot or heatmaps to visualize the distribution of different types of pokemon or generations and also users can select interactively different subsets of the dataset using brushing and linking techniques. And also can use filters on selections like base stat values.

Q6.

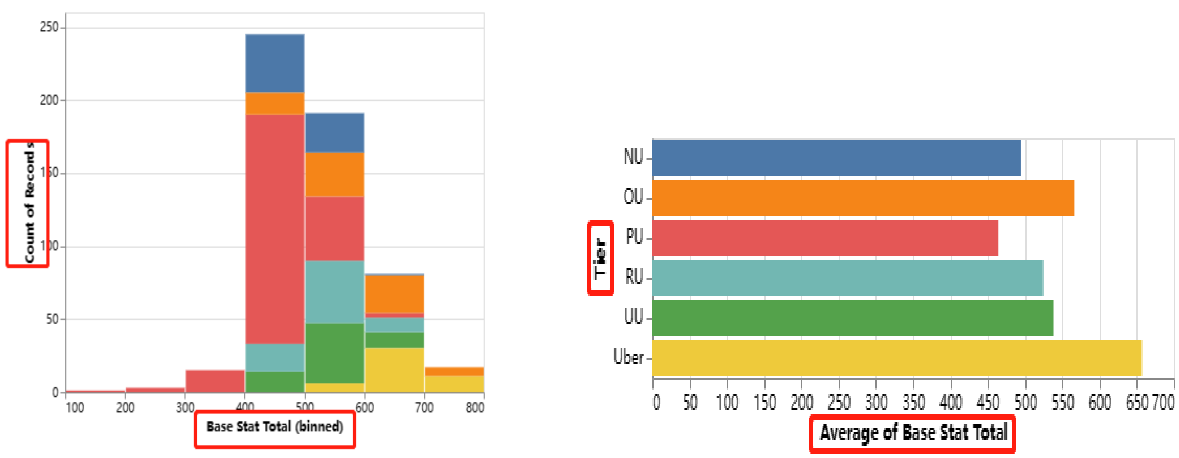
1.Axis choices:

System A1 chose "Count of Records" as the x-axis and "Base Stat Total (binned)" as the y-axis. This choice highlights the distribution of Pokemon across different ranges of base stat totals. However, it might not be the most informative axis choice for understanding the relationship between base stats and the number of moves.

Systems B and C both chose "Base stat Total" as the x-axis and "Num Moves" as the y-axis. These axes directly compare the base stats with the number of moves available to each Pokemon, making it easier to spot trends and relationships between the two variables.

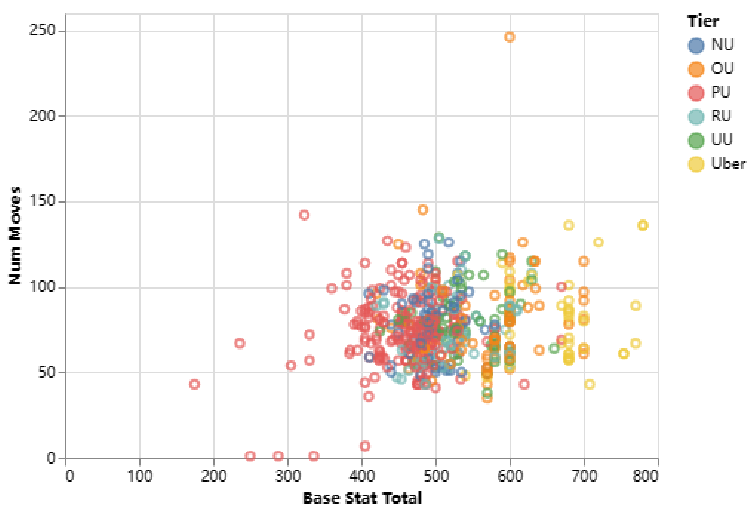
Other possible axis choices could include Pokemon types, attributes, weight, etc. Each of these options could reveal different insights into the dataset, but the choice used by systems B and C seems to be the most relevant for the given task.

The best choice might be the one used by systems B and C, as this combination more directly shows the relationship between base stats and the number of moves, making it easier to identify potential patterns or correlations.



2.Chart types:

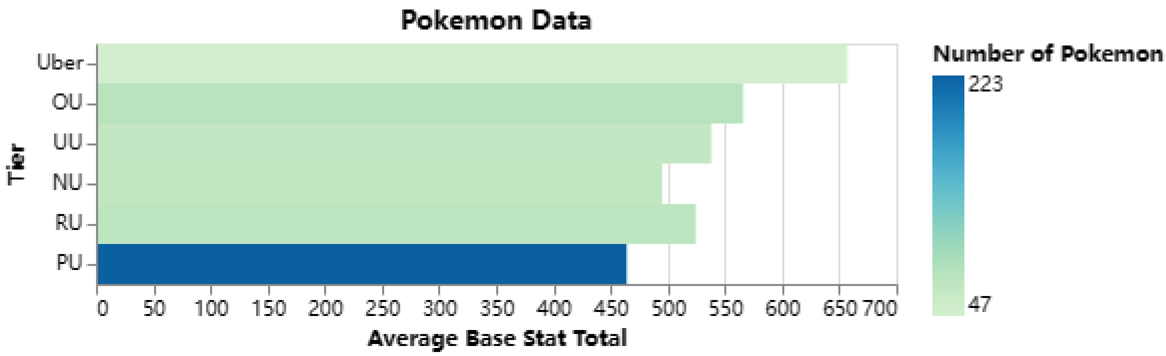
System A1 used a line chart, which can be useful for showing trends over continuous data ranges. However, this type of chart may not be the most suitable for this dataset, as the data points are discrete and do not necessarily have a continuous relationship.
System B used a scatterplot, which is an effective choice for showing the distribution and relationships between data points. This chart type makes it easy to identify clusters, outliers, and trends in the data.
System C used a bar chart, which can be useful for comparing discrete data points. However, this type of chart may not be the best choice for visualizing relationships between two continuous variables like base stat total and number of moves.
Other possible chart types include bar charts, stacked charts, etc. Each of these options has its own advantages and disadvantages, but the scatterplot (System B) seems to be the most suitable choice for this dataset.
The best choice might be the scatterplot (System B), as it more intuitively reveals the distribution and relationships between data points, making it easier to identify trends and correlations in the data.



3.Legend representation:

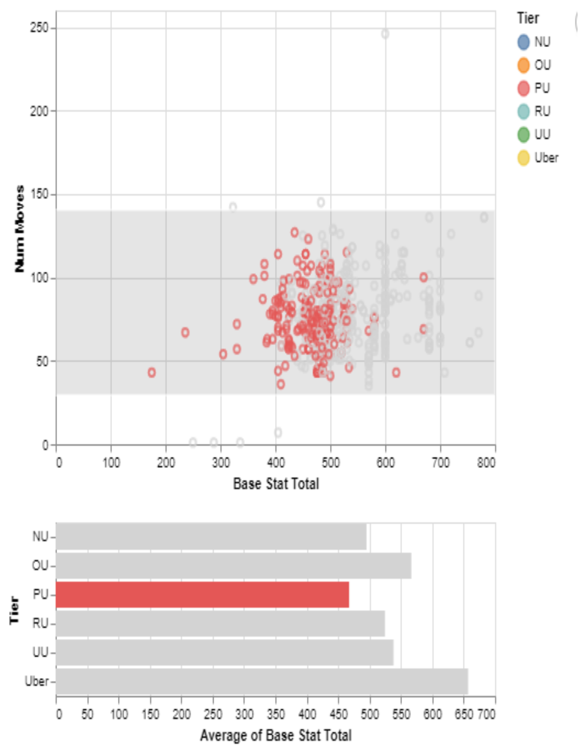
System A1 used differently colored horizontal lines to represent the legend. This choice may be clear and easy to understand but could become visually cluttered if there are many categories or lines.
System B used differently colored circles to represent the legend. This choice is visually simple and clean, making it easy to identify which data points belong to which category. However, using circles for a legend can be less informative if the size of the circle also conveys information.
System C used a gradient to represent the legend. This choice is visually appealing and can be useful for showing continuous data ranges. However, it may be less informative if the data is discrete and has clear categories.

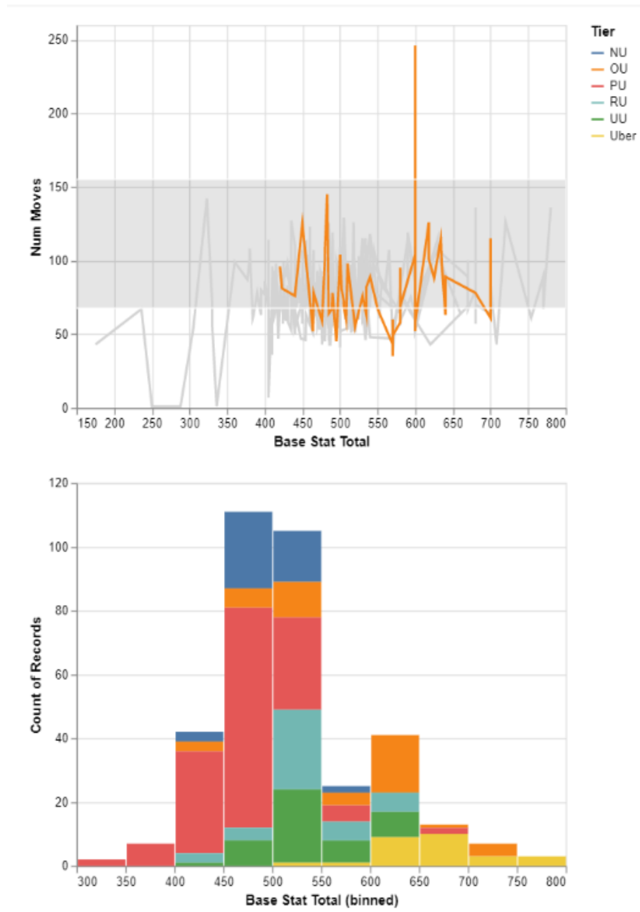
Other possible legend representations include squares, shapes, etc. Each of these options has its own strengths and weaknesses, but the gradient used by System C seems to be the most visually appealing and informative choice.
The best choice might be the gradient used by System C, as it better represents the continuity of the data and is visually appealing without being overly cluttered.



4.Data interaction features:

System C does not have brushing and linking functionality, which limits the interactivity and exploration potential for users. However, this system does provide detailed information when the user hovers over a data point, offering a more informative user experience in this aspect.
Other possible interaction features could include zooming, panning, or filtering capabilities. Each of these options would provide users with additional ways to explore the data and focus on areas of interest.
The best choice might be a combination of System B's brushing and linking with System C's hover information display. This would offer users the ability to select specific regions of the data, while still providing detailed information about individual data points when needed.

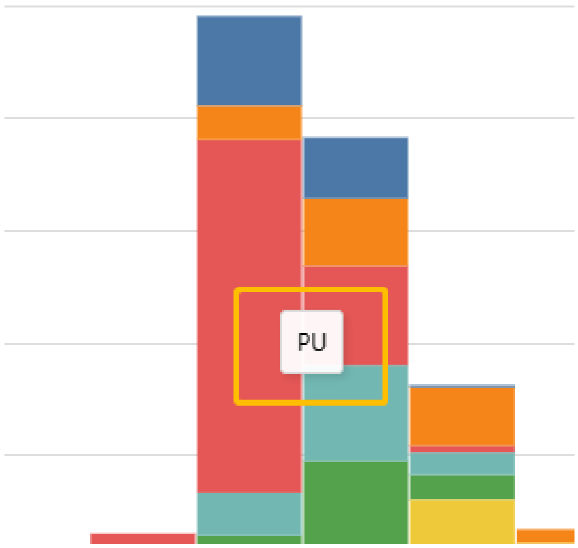


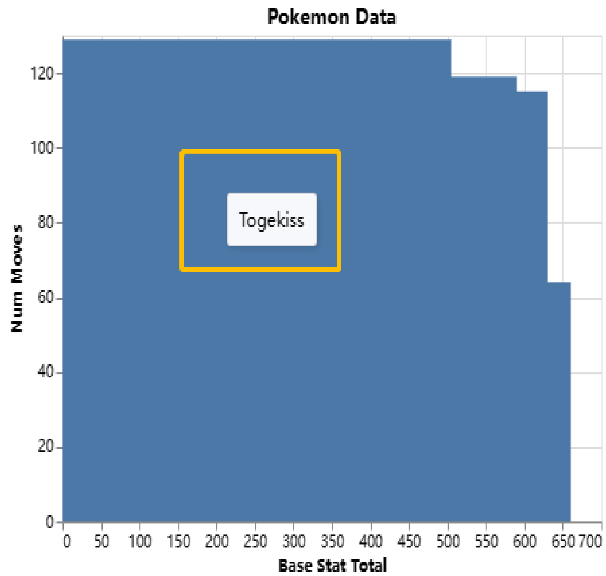


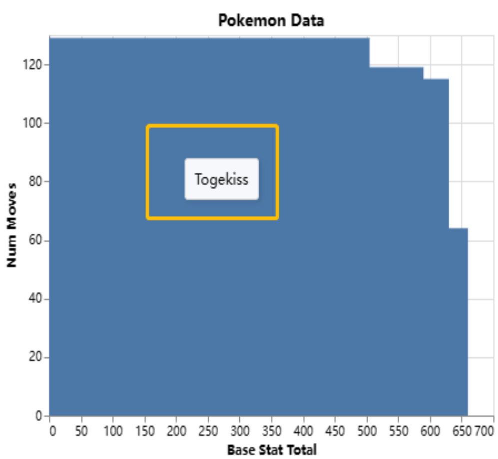
5.Color schemes:

System A1 used six different colors for the line segments, making it easy to distinguish between the different data categories. However, using too many colors could potentially become visually overwhelming and confusing.
System B used six distinct colors for the circles in the scatterplot, allowing users to easily identify the categories. However, using many colors in a scatterplot may lead to visual clutter and difficulty in recognizing patterns.
System C used a gradient color scheme, ranging from light to dark shades. This approach provides a visually appealing and continuous representation of the data, which might be more suitable for continuous data ranges.
Other possible color schemes could include monochromatic, diverging, or qualitative palettes. Each of these options has its own advantages and disadvantages, but the gradient color scheme used by System C seems to be the most visually appealing choice for this dataset.

The best choice might be the gradient color scheme used by System C, as it offers a visually appealing and continuous representation of the data, making it easier to identify trends and patterns.







6.Coordinate properties:

System A1 used a different set of coordinate properties compared to Systems B and C. The unique properties in System A1 might provide a different perspective on the data, but could also make it harder to compare the visualizations directly.
Systems B and C used the same coordinate properties, allowing for easier comparison between the two systems. Consistency in coordinate properties can help users better understand the relationships and patterns in the data across different visualizations.
Other possible coordinate property choices could involve transformations, such as logarithmic or power scales. These options could reveal different insights into the data, but might also introduce additional complexity and confusion for users.

The best choice might be the consistent coordinate properties used by Systems B and C, as this approach makes it easier to compare the visualizations and better understand the relationships and patterns in the data.

Q7. User Evaluation Comparison:

1. Evaluation Methodology:

To evaluate the systems, we conducted a user questionnaire with seven participants. The questionnaire was divided into three sections:

The first section asked the users to rate the layout and content of each system on a five-point Likert scale.

The second section asked the users what improvements they would like to see in each system.

The third section asked the users to answer questions comparing the systems, such as which system is easiest to operate, which system can help analyze data more quickly, which system provides more intuitive data visualization and their appropriateness for helping users make decisions..

The purpose of the questionnaire was to collect qualitative data on user perceptions of the three systems, identify any strengths or weaknesses of each system, and understand how the systems can be improved.

2. Data Collection:

We collected data from seven users who were asked to evaluate the three systems based on their experience and expertise as trainers in the Pokémon universe. We selected seven participants as a manageable sample size, given the limited scope of the project.

The users were provided with a brief description of the three systems and given access to each system for a limited time to explore its features. Then, they were asked to complete the questionnaire to provide their feedback on the systems.

3. Data Analysis:

To analyze the data, we calculated the percentage of participants who agreed or disagreed with each statement for each system and the mean scores for each Likert scale question and tabulated the responses to the open-ended questions.

We also compared the results of the questionnaire to identify any trends or patterns across the systems. We then compared the results across the three systems to identify which aspects of each system were revealed to be the "best." For example, we looked at the responses to the question "which system is easiest to operate?" and compared the scores to determine which system had the highest score. We also looked at the responses to the question "which system can help analyze data more quickly?" to determine which system had the highest score.

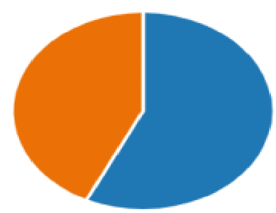
4. Results:

Based on the results of the questionnaire, we found that System 2 received the highest ratings in terms of layout and content, with six participants strongly agreeing that the system was logical and aesthetically pleasing. System 1 and System 3 also received high ratings, with four and five participants strongly agreeing, respectively, but with some suggestions for improvement.

1. I believe that **the layout and content** of System 1 is logical and aesthetically pleasing.

更多详细信息

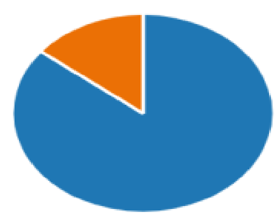
Strongly agree	4
Agree	3
Neutral	0
disagree	0
strongly disagree	0



2. I believe that **the layout and content** of System 2 is logical and aesthetically pleasing.

更多详细信息

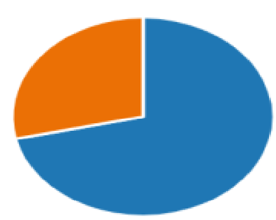
Strongly agree	6
Agree	1
Neutral	0
Disagree	0
Strongly disagree	0



3. I believe that **the layout and content** of System 3 is logical and aesthetically pleasing.

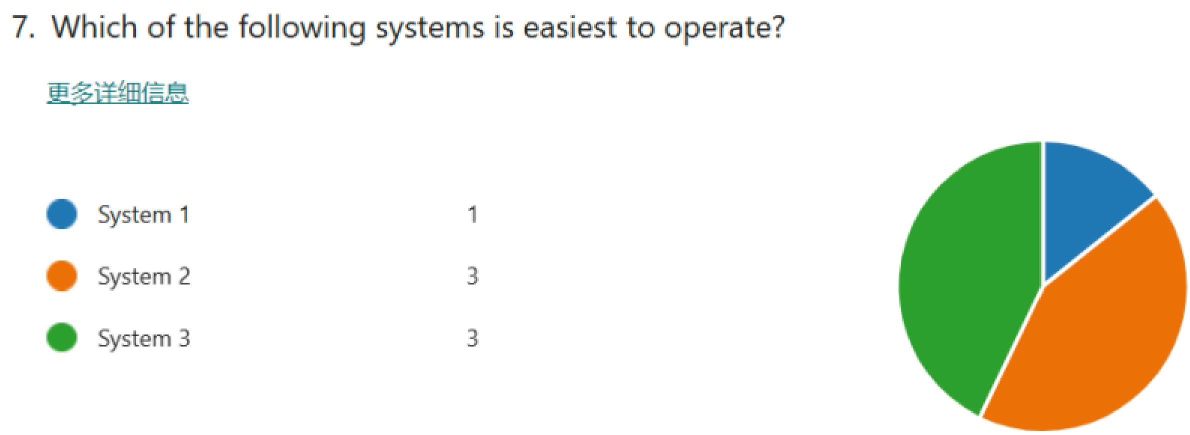
更多详细信息

Strongly agree	5
Agree	2
Neutral	0
Disagree	0
Strongly disagree	0

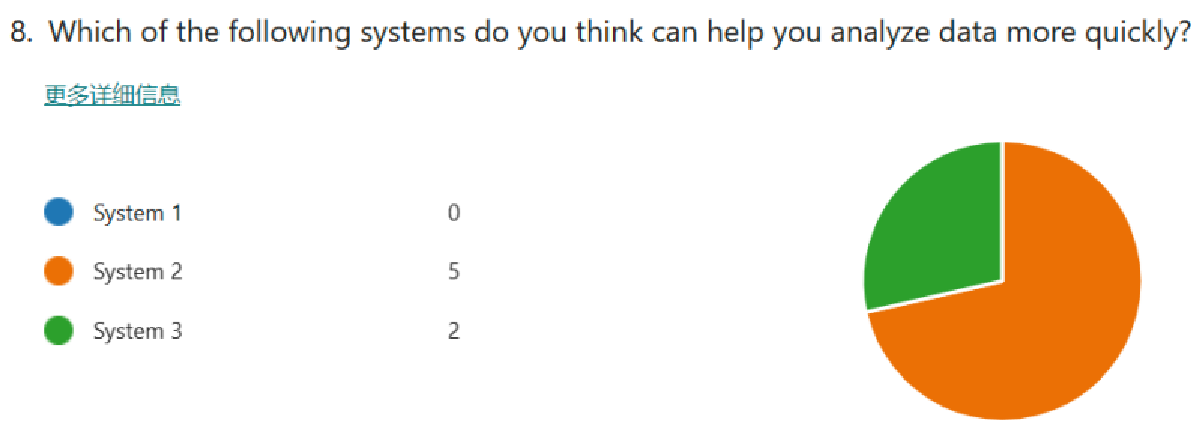


Regarding improvements, participants suggested adding more gaps to the bar chart in System 1 and presenting more multi-dimensional data through pipes of size, shape, etc. in System 2. Participants did not suggest any specific improvements for System 3.

In terms of ease of operation, System 2 and System 3 received the same score, with three participants each selecting these systems. Both being considered easier to operate than System 1. System 1 received the lowest score, with only one participant selecting it.



In terms of the ability to analyze data quickly, System 2 was rated the highest, with five out of seven participants selecting it as the best option. System 3 was the second most highly rated, with two out of seven participants selecting it as the best option. No participants selected System 1 as the best option for analyzing data quickly.



Regarding data visualization, System 3 received the highest score, with four participants selecting it. System 2 received three selections, and System 1 received zero.

9. Which of the following system data visualization is more intuitive?

更多详细信息

System 1	0
System 2	3
System 3	4

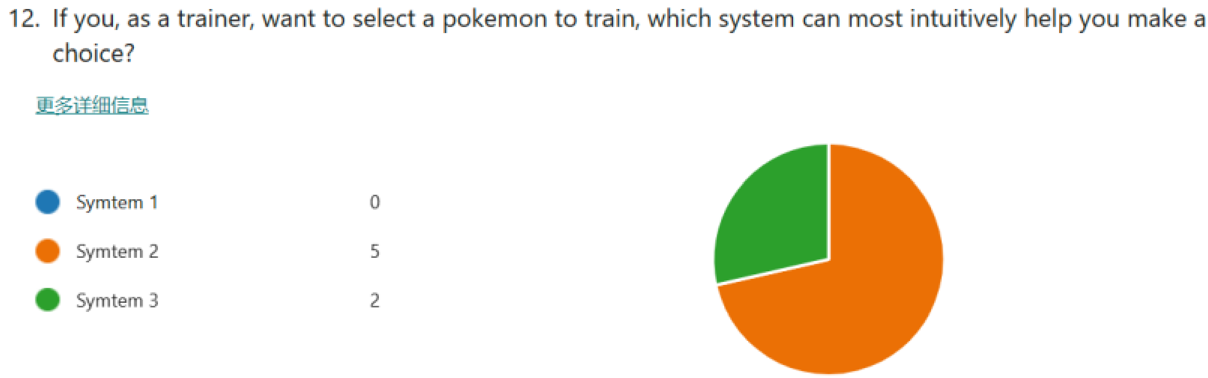
Regarding the horizontal and vertical coordinate spacing, System 2 received the highest score, with five participants selecting it. System 3 and System 1 received two and zero selections, respectively.

11. Which of the following three systems has the most appropriate horizontal and vertical coordinate spacing?

更多详细信息

Symtem 1 (heatmap and stacke...	0
Symtem 2 (scatterplot)	5
Symtem 3 (linechart bargraph)	2

When asked which system was most appropriate for helping users make decisions selecting a Pokémon to train, System 2 received the highest score, with five participants selecting it. System 3 and System 1 each received two and zero selections, respectively. Both System 2 and System 3 being considered more appropriate than System 1.



The results of the questionnaire suggest that System 2 performs the best in terms of layout and content, ease of operation, and data analysis. However, System 3 performs the best in terms of data visualization. System 1 received lower scores across all categories. Overall, our results indicate that System 2 was the most highly rated system, scoring the highest percentage of "strongly agree" responses across most of the evaluation criteria. In contrast, System 1 received the lowest percentage of "strongly agree" responses across most of the evaluation criteria, although it still received generally positive feedback. System 3 was rated in the middle, with generally positive feedback, but also with a few areas for improvement.

Q8. Future Work:

Based on our user evaluation, we would recommend the following changes to improve each system:

System 1:

Improve image quality: Some users reported that the image used in the system was not clear enough. Improving the quality of the image used could help make the system more visually appealing and easier to understand.

Add gaps to bar chart: One user suggested that gaps could be added to the bar chart as appropriate. This could help to visually separate the bars and make it easier to interpret the data.

Improve chart layout: While most users found the layout of the system to be logical and aesthetically pleasing, some minor adjustments to the chart layout could help to make the data easier to interpret.

System 2:

Add more dimensions to data visualization: One user suggested that more multi-dimensional data could be presented through pipes of size, shape, etc. Adding more dimensions to the data visualization could help to provide more insight into the relationships between different data points.

Increase customization options: While most users found the system easy to operate, adding more customization options could make the system more versatile and better suited to individual user needs.

Add tooltips: Adding tooltips to the scatterplot could help users to understand the data more easily by providing additional context and information.

System 3:

Increase customization options: Similar to System 2, adding more customization options could make the system more versatile and better suited to individual user needs.

Simplify chart layout: Some users found the layout of the chart to be somewhat cluttered, so simplifying the chart layout could help to make the data easier to interpret.

Refining the line chart and add more interactive features: Adding more interactive features could help to make the system more engaging and intuitive to use.

Overall, improving the quality of the data visualization, providing more customization options, and simplifying the chart layout could help to make each of these systems more effective for users. Additionally, adding more interactive features and improving the quality of the images used could help to increase user engagement and make the systems more visually appealing.