7-3-3 Create a Narrative Visualization

CS 498 Data Visualization

1. Introduction and Overview

In this report, I have done analysis on a dataset of exam scores. The dataset includes scores from three exams: Maths, Reading and Writing and different socio-economic background. In our analysis, I have clubbed Reading and Writing into a joint category called 'English'. However, if there is any interest in seeing the breakup, I provide that info in the tool tip. I have developed seven different scenes: First scene is introduction, with next three scenes analysing if students good in maths would be good in English also. This will help into analysing if we need to focus on Maths or English, to improve the overall score of students. I also see if this relationship varies depending on different personal, social and economic factors. In next three scenes, I analyse the impact of these factors on total scores.

The database has been taken from the data warehouse developed by Dr. Royce Kimmons and posted at http://roycekimmons.com/system/generate_data.php?dataset=exams&n=1000. It contains eight columns: 1. Gender, 2. Ethnicity, 3. Parental Education, 4. Lunch Program, 5. Test Preparation, 6. Maths Score, 7. Reading Score, and 8. Writing Score. The columns are self-explanatory. The data was relatively clean but needed some pre-processing before using it for our visualization. I used Open Refine's cluster feature on the field: Parent's education to club slightly varying parameters into one.

2. Narrative Structure

As discussed in section 1, one of the aims of the visualization is to find any pattern in the score to find different ways to improve the overall score. We have adapted the narrative style of Hybrid Structure which in the middle of narrative spectrum of reader driven layouts and reader driven layouts. In the hybrid structure also, we have taken slight inspiration from Martini glass but mostly follow the design of Interactive slide show. The first scene narrates a short story from author's point of view with slide-navigation as the sole point of interactivity. As the reader navigates to next few slides, in each slide they have the choice to get into further details by using different filters implemented using the drop-down selection feature. The scenes are ordered in specific sequence to make sure that there is continuation in story line. First three slides are narrating the point that if individuals scoring good in maths, score good in English too and also, we see if different personal/social factors show the same trend for this relationship. In last three slides, we try to analyse the relationship between total scores and training program, or parent's education or ethnicity. We also study if different genders follow the same trend. We use the consistent colour scheme in each page to maintain the continuity of story and avoid any confusion. We discuss each page in detail in next sub-sections.

3. Scenes

3.1 Scene 1: Author Driven Introductory Page

The first scene is introducing the audience with the details of the study and guide with the navigation with rest of the scenes. It uses a simple paragraph and links to rest f the pages.

3.2 Scene 2: Maths and English Comparison for Diff Genders

In this scene, we have analysed the if students scoring in Maths are equally good in English and vice-versa and if this relationship is consistent for different genders.

Visual Structure:

We first load the graph using scatter plot with Maths score in horizontal direction and English score in vertical direction. In this initial load we display both the genders at the same time. The initial scene is loaded with transition where each data point moves from origin to each destination place. It helps users in navigating the scene. Just below the graph, we have the message for the users written in bold to catch their attention and it changes with each trigger. At the bottom of graph, we have clearly visible link with "Go To" phrase working as affordance to navigate to other slides and back to main page. The current slide is also displayed clearly by showing the slide number in different colour.

Annotation:

I have used multiple forms of annotations in each scene and added some affordance as described below:

1. Details info on each data point: The tool-tip feature has been enabled to show the details of scores on each data point.
2. Balanced Score line: By using the hint from a class side, where a line was drawn to help users clearly see the country with delta population increase, I have drawn a line to display the balanced score. The line helps in observing the pattern better as it is using the 'position' attribute by separating the data into two parts.
3. Message: We also give the load message just below the graph to display the author driven message from each scene which gets refreshed with each trigger.
4. Legend: Legend is clearly mentioned to make sure that the reader can understand the graph easily.

Parameters:

There are three major parameters in this scene:

1. Maths Score: It is presented on x-axis
2. English score: It is aggregation of Reading and Writing score and presented on y-axis
3. Gender: We use this as another parameter in the drop down for the user to interact with the slide. We have sorted it lexicographically in drop down list for easy finding.

Triggers:

We have used drop-down as the trigger which has populated by the gender data. On drop-down field selection, the event is triggered to populate the graph with the data filtered only for that gender. It also triggers the Note section to reflect the new annotation.

3.3 Scene 3: Maths and English Comparison for Diff Parent's Education

In this scene, we have analysed the if students scoring in Maths are equally good in English and vice-versa and if this relationship is consistent for different level of education of parents.

Visual Structure:

We first load the graph using scatter plot with Maths score in horizontal direction and English score in vertical direction. In this initial load we display all level of parent's education at the same time. The initial scene is loaded with transition where each data point moves from origin to each destination place. It helps users in navigating the scene. Just below the graph, we have the message for the users written in bold to catch user's attention and it changes with each trigger. At the bottom of graph, we have clearly visible link to navigate to other slides and back to main page. The current slide is also displayed clearly by showing the slide number in different colour.

Annotation:

I have used multiple forms of annotations in each scene and added some affordance as described below:

1. Details info on each data point: The tool-tip feature has been enabled to show the details of scores on each data point.
2. Balanced Score line: I have drawn a line to display the balanced score. The line helps in observing the pattern better as it is using the 'position' attribute by separating the data into two parts.
3. Message: We also give the load message just below the graph to display the author driven message from each scene which gets refreshed with each trigger.
4. Legend: Legend is clearly mentioned to make sure that the reader can understand the graph easily.

Parameters:

There are three major parameters in this scene:

1. Maths Score: It is presented on x-axis
2. English score: It is aggregation of Reading and Writing score and presented on y-axis
3. Parent's Education: We use this as another parameter in the drop down for the user to interact with the slide. We have sorted it lexicographically in drop down list for easy finding.

Triggers:

We have used drop-down as the trigger which has populated by the parent's education data. On drop-down field selection, the event is triggered to populate the graph with the data filtered only for that parent's education set. It also triggers the Note section to reflect the new annotation.

3.4 Scene 4: Maths and English Comparison for Diff Ethnicity

In this scene, we have analysed the if students scoring in Maths are equally good in English and vice-versa and if this relationship is consistent for different ethnicity.

Visual Structure:

We first load the graph using scatter plot with Maths score in horizontal direction and English score in vertical direction. In this initial load we display all the ethnicities at the same time. The initial scene is loaded with transition where each data point moves from origin to each destination place. It helps users in navigating the scene. Just below the graph, we have the message for the users written in bold to catch user's attention and it changes with each trigger. At the bottom of graph, we have clearly

visible link to navigate to other slides and back to main page. The current slide is also displayed clearly by showing the slide number in different colour.

Annotation:

I have used multiple forms of annotations in each scene and added some affordance as described below:

1. Details info on each data point: The tool-tip feature has been enabled to show the details of scores on each data point.
2. Balanced Score line: By using the hint from a class side, where a line was drawn to help users clearly see the country with delta population increase, I have drawn a line to display the balanced score. The line helps in observing the pattern better as it is using the 'position' attribute by separating the data into two parts.
3. Message: We also give the load message just below the graph to display the author driven message from each scene which gets refreshed with each trigger.
4. Legend: Legend is clearly mentioned to make sure that the reader can understand the graph easily.

Parameters:

There are three major parameters in this scene:

1. Maths Score: It is presented on x-axis
2. English score: It is aggregation of Reading and Writing score and presented on y-axis
3. Ethnicity: We use this as another parameter in the drop down for the user to interact with the slide. We have sorted it lexicographically in drop down list for easy finding.

Triggers:

We have used drop-down as the trigger which has populated by the ethnicity data. On drop-down field selection, the event is triggered to populate the graph with the data filtered only for that ethnicity. It also triggers the Note section to reflect the new annotation.

3.5 Scene 5: Impact of Parent's Education on Total Score

In this scene, we have analysed the impact of parent's education on total score. We also see if this relation is same for different genders too.

Visual Structure:

We first load the graph using scatter plot with parental education in horizontal direction and total score in vertical direction. In this initial load we display all genders at the same time. The initial scene is loaded with transition where each data point moves from origin to each destination place. It helps users in navigating the scene. Just below the graph, we have the message for the users written in bold to catch user's attention and it changes with each trigger. At the bottom of graph, we have clearly visible link to navigate to other slides and back to main page. The current slide is also displayed clearly by showing the slide number in different colour.

Annotation:

I have used multiple forms of annotations in each scene and added some affordance as described below:

1. Details info on each data point: The tool-tip feature has been enabled to show the details of scores on each data point.
2. Middle Score line: I have drawn a line to display the middle score. The line helps in observing the pattern better as it is using the 'position' attribute by separating the data into two parts.
3. Message: We also give the load message just below the graph to display the author driven message from each scene which gets refreshed with each trigger.
4. Legend: Legend is clearly mentioned to make sure that the reader can understand the graph easily.

Parameters:

There are three major parameters in this scene:

1. Parents Education: It is presented on x-axis
2. Total score: It is aggregation of Reading and Writing score and presented on y-axis
3. Gender: We use this as another parameter in the drop down for the user to interact with the slide. We have sorted it lexicographically in drop down list for easy finding.

Triggers:

We have used drop-down as the trigger which has populated by the gender data. On drop-down field selection, the event is triggered to populate the graph with the data filtered only for that gender. It also triggers the Note section to reflect the new annotation.

3.6 Scene 6: Total Score of Different Ethnicity

In this scene, we have analysed if all ethnicities score similar or different. The ethnicity name has been made abstract to make it general data. We also see if this relation is same for different genders too.

Visual Structure:

We first load the graph using scatter plot with parental education in horizontal direction and total score in vertical direction. In this initial load we display all genders at the same time. The initial scene is loaded with transition where each data point moves from origin to each destination place. It helps users in navigating the scene. Just below the graph, we have the message for the users written in bold to catch user's attention and it changes with each trigger. At the bottom of graph, we have clearly visible link to navigate to other slides and back to main page. The current slide is also displayed clearly by showing the slide number in different colour.

Annotation:

I have used multiple forms of annotations in each scene and added some affordance as described below:

1. Details info on each data point: The tool-tip feature has been enabled to show the details of scores on each data point.
2. Middle Score line: I have drawn a line to display the middle score. The line helps in observing the pattern better as it is using the 'position' attribute by separating the data into two parts.
3. Message: We also give the load message just below the graph to display the author driven message from each scene which gets refreshed with each trigger.

4. Legend: Legend is clearly mentioned to make sure that the reader can understand the graph easily.

Parameters:

There are three major parameters in this scene:

1. Ethinicity: It is presented on x-axis
2. Total score: It is aggregation of Reading and Writing score and presented on y-axis
3. Gender: We use this as another parameter in the drop down for the user to interact with the slide. We have sorted it lexicographically in drop down list for easy finding.

Triggers:

We have used drop-down as the trigger which has populated by the gender data. On drop-down field selection, the event is triggered to populate the graph with the data filtered only for that gender. It also triggers the Note section to reflect the new annotation.

3.7 Scene 7: Impact of Training on Total Score

In this scene, we have analysed if training program improves the score. The We also see if this relation is same for different genders too.

Visual Structure:

We first load the graph using scatter plot with training status in horizontal direction and total score in vertical direction. In this initial load we display all genders at the same time. The initial scene is loaded with transition where each data point moves from origin to each destination place. It helps users in navigating the scene. Just below the graph, we have the message for the users written in bold to catch user's attention and it changes with each trigger. At the bottom of graph, we have clearly visible link to navigate to other slides and back to main page. The current slide is also displayed clearly by showing the slide number in different colour.

Annotation:

I have used multiple forms of annotations in each scene and added some affordance as described below:

1. Details info on each data point: The tool-tip feature has been enabled to show the details of scores on each data point.
2. Middle Score line: I have drawn a line to display the middle score. The line helps in observing the pattern better as it is using the 'position' attribute by separating the data into two parts.
3. Message: We also give the load message just below the graph to display the author driven message from each scene which gets refreshed with each trigger.
4. Legend: Legend is clearly mentioned to make sure that the reader can understand the graph easily.

Parameters:

There are three major parameters in this scene:

1. Training: It is presented on x-axis

2. Total score: It is aggregation of Reading and Writing score and presented on y-axis
3. Gender: We use this as another parameter in the drop down for the user to interact with the slide. We have sorted it lexicographically in drop down list for easy finding.

Triggers:

We have used drop-down as the trigger which has populated by the gender data. On drop-down field selection, the event is triggered to populate the graph with the data filtered only for that gender. It also triggers the Note section to reflect the new annotation.

4. Conclusions

In this project, we dealt with almost all the concepts of data visualization which was explained throughout the course. The free form of exercise forced us to analyse the data and finding ways to visualize it. In section 1, we have given an overview of complete data set and the message we are trying to convey through the data visualization. In section 2, we describe the narrative structure used to convey the story. In each sub-sections of section 3, we describe each scene with the visual structure used in it, and different attributes such as Annotation, Parameter, and Trigger.

Finally, I take this opportunity to thank the professor and instructors/TAs for so much help throughout the course.