Assessment 2: Planning and Creating Static Visualizations

Rijo Aji Kuruvilla

Master of Data Science (Professional), James Cook University (Australia)

22-A-MA5830-CNS-INT-SP83 Data Visualization

May 29,2022

VISUALIZATION - 1

WHY?

1. **Target Audience**

   The target audience are the parents of a year five student (Student ID 24).

2. **Visualization Intent**

   - To display the overall average percentage marks across all subjects for the student #24.
   - How does student #24 compare to other students of the class.

3. **Assumptions**

   ### a. Visual Assumption

   -The parents of the student are familiar with basic statistical charts and diagrams.

   ### b. Data Assumptions

   - The parents are assumed to be familiar with basic knowledge of mathematics such as minimum, maximum, total and average values.

   ### c. Interactivity Assumptions

   - The visualization should directly display the result of the concerned student.
   - Since it is a static visualization, the interactivity for this visualization can be considered as minimal.

   ### d. Time Assumption

   - Studies have shown that we have about 3–8 seconds with our audience, during which time they decide whether to continue to look at what we've put in front of them or direct their attention to something else. Therefore, it is essential to use our pre-attentive attributes wisely, even if we only get that initial 3–8 seconds, we have given our audience the idea of what we want to reveal (Knaflic, 2015, pp. 108-109). Therefore we can say that we have about 3-8 seconds to deliver our information.

4. **Desired Effect**

   To inform the parents of their child's overall performance with respect to the rest of the class.

WHAT?

Raw and New Data

Table 1 summarizes the Raw and the New data (Calculated data)

Table 1 – Meta data summary

	Variable Name	Data Type	Categories /Levels	Data Description			Intent/Desried Response	Rank	Gestalt's Principle
				Min	Max	Average			
Raw Data	Student ID	Categorical -> Nominal	30				Intent	3	
	Percentage Read Score	Numeric -> Ratio		4	82	64.53	Intent		
	Percentage Writing Score	Numeric -> Ratio		0	91	56.58	Intent		
	Percentage Spell Score	Numeric -> Ratio		17	100	69.2	Intent	4	
	Percentage Grammar and Punctuation Score	Numeric -> Ratio		17	100	63.56	Intent		
	Percentage Numeracy Score	Numeric -> Ratio		6	100	66.6	Intent		
Calculated Data	Average Percentage Marks of the Class	Numeric-> Ratio	1			64.14	Intent & Desired	2	Enclosure
	Student with the lowest class average	Numeric-> Ratio	1			15	Intent & Desired	1	Enclosure

Calculations

- Average Percentage Marks = (Percentage Read Score + Percentage Write Score + Percentage Spell Score + Percentage Grammar and Punctuation Score + Percentage Numeracy Score)/ Number of Subjects

- Student with the lowest class average = Average Percentage Marks /5

Data Limitations

Since the data set is small and has a limited number of records it makes it difficult to perform detailed analysis without violating the interactivity and time assumptions in delivering our message to the audience.

HOW?

Chart and Aesthetics

Simple Text Image

I have chosen a plain text image for displaying the visualization as I wanted to focus primarily on the overall average percentage marks of student #24 and a straightforward text in this case directly aides me in my communication by helping the audience to get a full focus on text value without any cognitive load.

For a Simple Text Image, Area will be used as a Mark and an identity channel of Colour to best interpret the image [3].

To eliminate clutter, I have used white spaces strategically by letting the text image take the centre stage and thereby capturing maximum audience attention (Knaflic, 2015, pp. 85-86)

Insert Visualization

Student #24 has an overall

15%

average marks as compared
to the overall class
percetange of 64%

FIGURE 1: Average Percentage Marks (Student #24)

Review and Improvements

FIGURE 1 directly communicates the intent and fulfills the desired effect of the visualization by providing the average percentage marks for the student #24. It also helps the audience compare the result with the class average thereby fulfilling both the visualization intent and the desired effect.

As the visualization is just a simple text, there is next to zero cognitive load on the audience and maximum information can be attained within the assumptions with respect to time (Time Assumptions).

The visualization itself doesn't require any improvement as it ticks all boxes in successfully delivering the message to the audience.

VISUALIZATION - 2

WHY?

1. **Target Audience**
   The target audience for the second visualization is the teacher for the Writing Subject.

2. **Visualization Intent**
   To display the students with the below average percentage score in the Writing subject.

3. **Assumptions**

   a. Visual Assumptions
      - The teacher is familiar with identifying basic statistical charts.

   b. Data Assumptions
      - The teacher is assumed to be familiar with basic knowledge of mathematics such as minimum, maximum, total and average values.

   c. Interactivity Assumptions
      - The visualization should directly communicate the result and the interactivity is next to minimal as it is a static visualization.

   d. Time Assumption
      - Time here is assumed to between 3-10 seconds for the audience to gasp their attention.

4. **Desired Effect**

   To highlight the students who have a below average grade in the Write score.

WHAT?

Raw and New Data

Table 2 summarizes the Raw and the New data (Calculated data)

Table 2 – Meta data summary

	Variable Name	Data Type	Categories /Levels	Data Description			Intent/Desried Response	Rank	Gestalt's Principle
				Min	Max	Average			
Raw Data	Student ID	Categorical -> Nominal	30	1	1	1	Intent	3	
	Percentage Writing Score	Numeric -> Ratio		0	91		Intent	4	
Calculated Data	Average Percentage Score for Writing	Numeric -> Ratio				56.8	Intent & Desired	2	Enclosure
	Below Mean Writing Score	Categorical -> Ordinal	1				Intent & Desired	1	Similarity

Calculations

Average Percentage Write Score = (Total Sum of Grades in Writing)/Number of Students

Data Limitations

As we have a small data set, it is difficult to explore more complicated graphs and charts such as the box plot for standard deviation and variance analysis without increasing the cognitive load for the user. The disadvantage of small datasets is that tends to lower the precision in your results which risks in generating comparisons between the result among various groups as biased.

HOW?

Chart and Aesthetics

Bar Chart

A bar chart helps in plotting numeric values for levels of a categorical feature. Bar charts will be easy on the eyes to read. I have chosen a bar chart to display the grades of the student that are below the average marks for the write subject.

The teachers can compare the end points of the bars, so it is easy to see quickly which category is the biggest, which is the smallest, and the incremental difference between categories (Knaflic, 2015, p. 50).

As I have used a bar chart to display the visualization, the following will be the mark and channel for the bar chart:

Mark – Line [3].
Channel – Vertical Spatial Position for the Quantitative Attribute (Below Average Write Marks) [3].
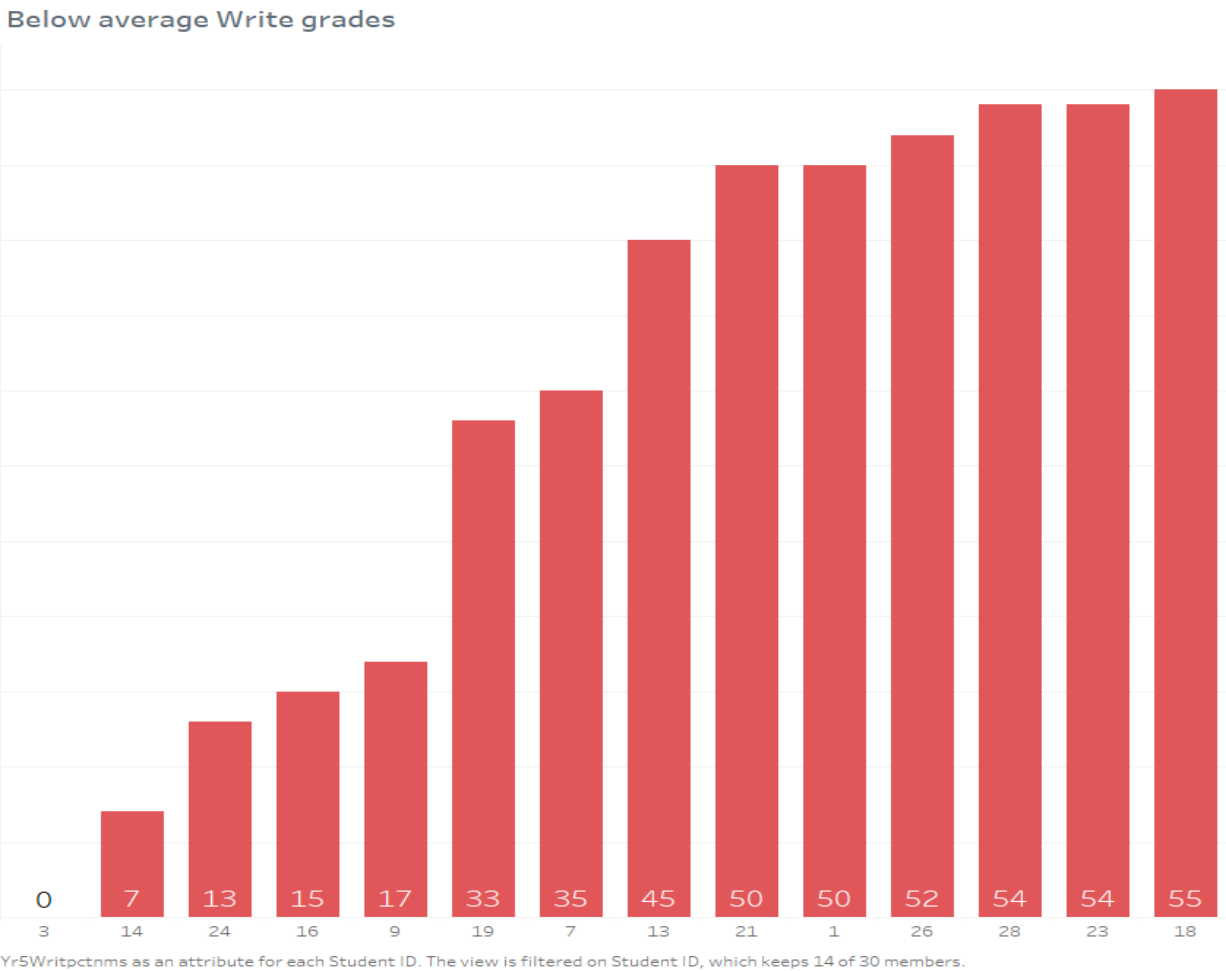
Student ID being a categorical value will be best represented by an identity channel. Spatial position can be considered as the most expressive and effective channel for categorical data [3].

The below average write grades is a derived attribute which was filtered from the average grade of write grades and it directly communicates the desired response to the audience. The Gestalt's Principle of Similarity and Enclosure [2] best describes this bar chart and is included with the visual channel of colour (hue) [3].

The clutter in this visualization has been reduced by placing the scores inside of the bars for the audience to directly interpret the score without much strain.

Below average Write grades

	0	7	13	15	17	33	35	45	50	50	52	54	54	55
	3	14	24	16	9	19	7	13	21	1	26	28	23	18

Yr5Writpctnms as an attribute for each Student ID. The view is filtered on Student ID, which keeps 14 of 30 members.

FIGURE 2: Visualization Two

Review and Improvements

The visualization in FIGURE 2 helps the audience to identify the students who fall below the average marks for the Writing subject.

As listed in the assumptions the teacher is expected to know the concepts of minimum, maximum and average and the visualization helps the audience get the overview of the students who are at the lower end of the class within the time frame mentioned in the assumptions.

Since we have a limited number of data, more complicated statistical plots such as the box plot would not be feasible to create and if used, there might be high probability that cognitive load on the audience might increase. Hence, the bar chart seems more appropriate medium to deliver the information taking care of all assumptions and ensuring that there is not much cognitive load on the audience.

Visualization - 3

WHY?

1. **Target Audience** – The target audience for this visualization is the teacher for the Numeracy subject.

2. **Visualization Intent**
   – To display the performance of the students in the numeracy subject.
   – To identify the students who have scored above average and below average in Numeracy.

3. **Assumptions**

   a. Visual Assumptions
      - The teacher is familiar with identifying basic statistical charts.

   b. Data Assumptions
      - The teacher is assumed to be familiar with basic knowledge of mathematics such as minimum, maximum, total and average values.

   c. Interactivity and Time Assumptions
      - The same interactivity and time assumptions used in Visualization two has been used in this visualization.

4. **Desired Effect**

   To be able to differentiate between the students who stand above average and those below the average score in year five.

WHAT?

Raw Data and New (Calculated Data)

Table 3: Meta Data summary

	Variable Name	Data Type	Categories/ Levels	Data Description			Intent/Desried Response	Rank	Gestalt's Principle
				Min	Max	Average			
Raw Data	Student ID	Categorical -> Nominal	30	1	1	1	Intent	5	
	Percentage Numeracy Score	Numeric -> Ratio		6	100	66.6	Intent	4	
Calculated Data	Total Sum of Grades in Numeracy	Numeric-> Ratio				1998	Intent	3	Continuity
	Average Percentage Score for Numeracy	Numeric-> Ratio				66.6	Intent/Desried Response	2	Enclosure
	Above/Below Mean	Categorical-> Ordinal	2				Intent/Desried Response	1	Similarity

Calculations

- Total Sum of Grades = Sum of Grades for All Students

- Average Percentage Numeracy Score = (Total Sum of Grades in Numeracy)/Number of Students

Data Limitation

Data Limitation for this visualization will be same as that of Visualization 2.
As we have a small data set, it is difficult to explore more complicated graphs and charts such as the box plot for std deviation and variance analysis without increasing the cognitive load for the user.

HOW?

Chart and Aesthetics

Bar Chart

For this visualization, I have used a bar chart since I wanted to display a comparative analysis between students who were above and below average. The visualization at a first glance itself displays a clear separation between the students on the high end of the axis as well on the lower end of the axis.

As I have used a bar chart to display the visualization, the following will be the mark and channel for the chart:

Mark – Line [3].
Channel – Vertical Spatial Position for the Quantitative Attribute (Below Average Write Marks) [3].

The total sum of grades for Numeracy being a quantitative attribute must be displayed using a magnitude channel [3]. The mark used here is a line and the attribute is encoded using the vertical spatial position.

The Average Numeracy score which is a calculated data is displayed using a dotted line across the bar chart helps differentiate between the two categories. The Gestalts principle of Enclosure which specifies that objects which are physically enclosed together are perceived to be a part of a group helps in this case [2].

The Above/Below average is a categorical-> ordinal attribute and is therefore encoded with an identity channel of Colour (Hue) [3].

For reducing the clutter, the score values are placed inside each bar. The Gestalt principle of Similarity ensures that even though there are many bars, the audience can easily understand the difference between the two contrasting groups [2].

Insert Visualization
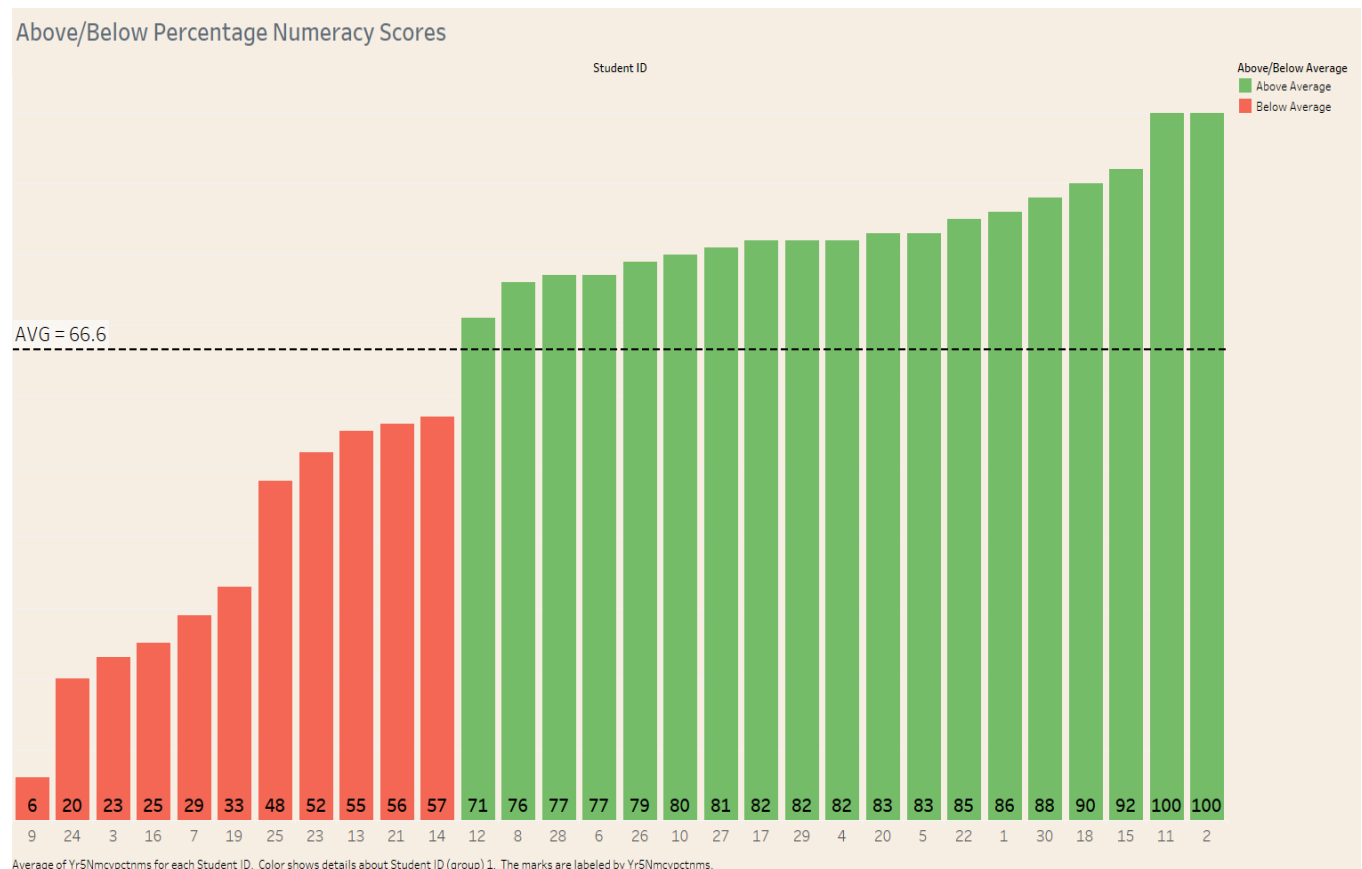
FIGURE 3: Comparison of Above/Below Percentage Numeracy Score

Review & Improvements

The visualization in FIGURE 3 directly delivers the required desired effect by segregating the students who are above/below the average percentage Numeracy grades.

The audience can easily distinguish between the students who have score excellent grades and can also focus on the students who are on the lower end of the grading scale. The visualization is highly informative, the title itself displays clearly what the chart shows and therefore not only fulfills the time constraints (Time assumptions) but also connects with the desired meaning of the visualization.

Like the visualization two, box plot and variance analysis would provide a much-detailed information but considering the time and the lack of records, the chances of the desired effect not being fulfilled may increase. Therefore, the bar chart is the most appropriate medium to convey the desired information to the audience without adding additional cognitive load while maintaining the time assumptions.

REFERENCES

[1].   Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals* (pp. 35–59) [Review of *Storytelling with Data: A Data Visualization Guide for Business Professionals*]. John Wiley & Sons.

[2].   Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals* (pp. 71-98) [Review of *Storytelling with Data: A Data Visualization Guide for Business Professionals*]. John Wiley & Sons.

[3].   Munzner, T. (2015). *Visualization Analysis and Design* (pp. 94–114) [Review of *Visualization Analysis and Design*]. CRC Press LLC.

[4].   Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals* (pp. 99-126) [Review of *Storytelling with Data: A Data Visualization Guide for Business Professionals*]. John Wiley & Sons.