Homework Assignment: Design and Best Practices

Part 1.1: Which Gestalt Principle could you use?

When we implement data for explanatory purposes, we often leave our presentations with our intended audience. The visualizations will include both text and visuals. This is a challenge as, many times, this creates a lot of work for our audience and increases the cognitive load. We must create visualizations that tie the text to the data. This is accomplished using the Gestalt Principles we learned in class. Looking at the following graph, which Gestalt Principles can be implemented to tie the text to the graph?

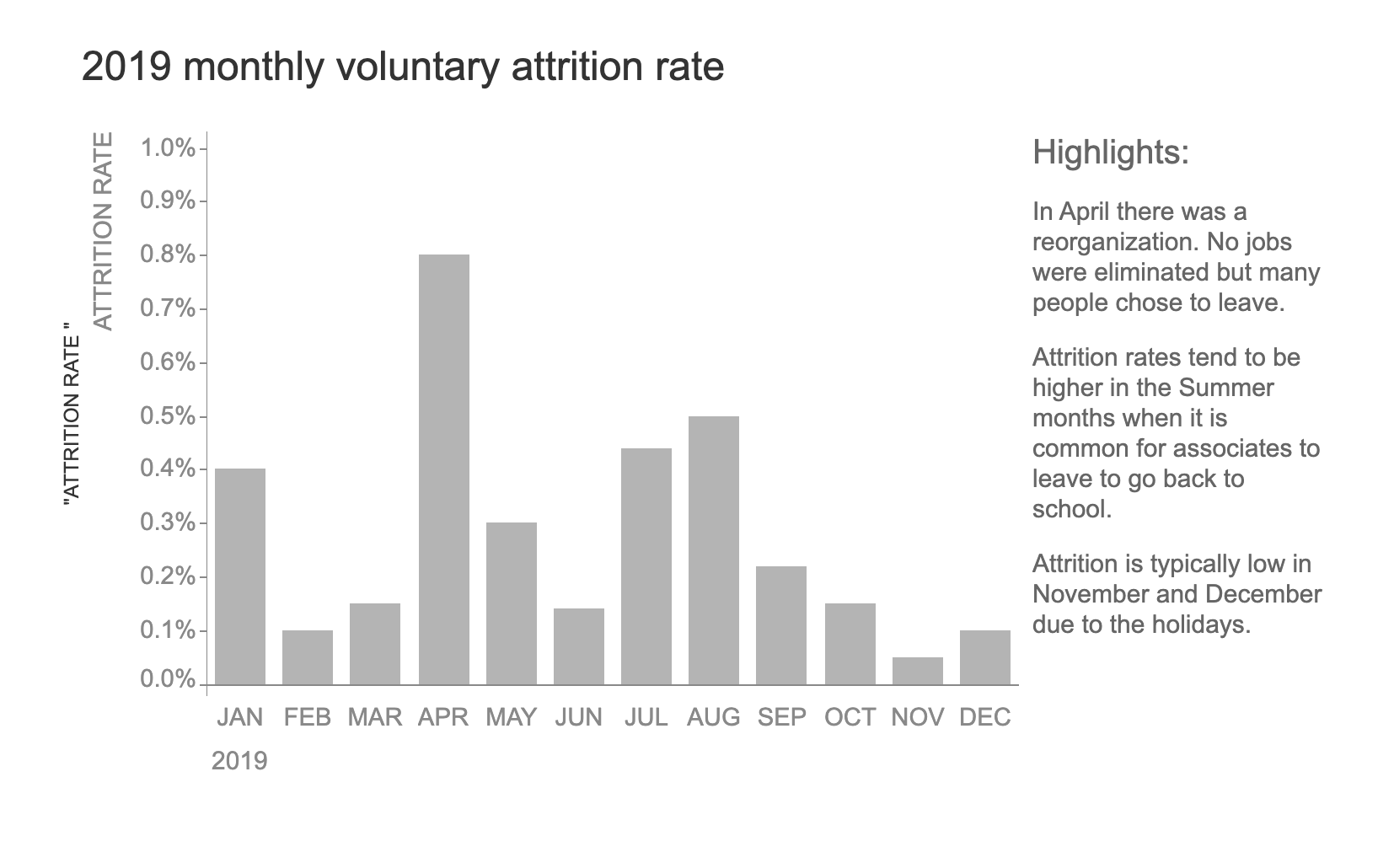


Image of a bar chart illustrating attrition rate by each month in 2019

Response

From this visualization, I can tell that Gestalt Principle of Proximity can get properly utilized for the text and the data adjacency.

The highlights/annotations are not on the relevant data point they are describing and that is an added burden on the viewer. This is how Proximity would be better utilized:

1. Instead of a separate highlights box, the April reorganization note might sit right next to — or be connected with a dotted line to — the highlighted spike in April's bar. This would provide an instant visual relationship between the data point and its story.
2. The remark on higher attrition in summer months could be placed right above or below the cluster of bars on July-August of higher value to group those two by proximity.
3. The note about low attrition during the holiday season could be placed next to the Nov-Dec bars, allowing the connection between the data and explanation to be clear on first glance.

So what would we do by taking advantage of the Principle of Proximity?

* Minimize the eye travel between the data and the explanation
* Form natural visual groups that allow readers to quickly ascertain which explanations are related to which data points
* Reducing cognitive load: by not forcing them to connect separated but nonetheless related bits of information

Other Gestalt Principles such as Similarity (e.g. using the same colors or shapes to group related elements) or Connection (e.g. connecting text with data points using lines or arrows to explicitly link them) could also be applied, but Proximity would work very well in this case, allowing to easily convey intuitive spatial relations between the data and its context.

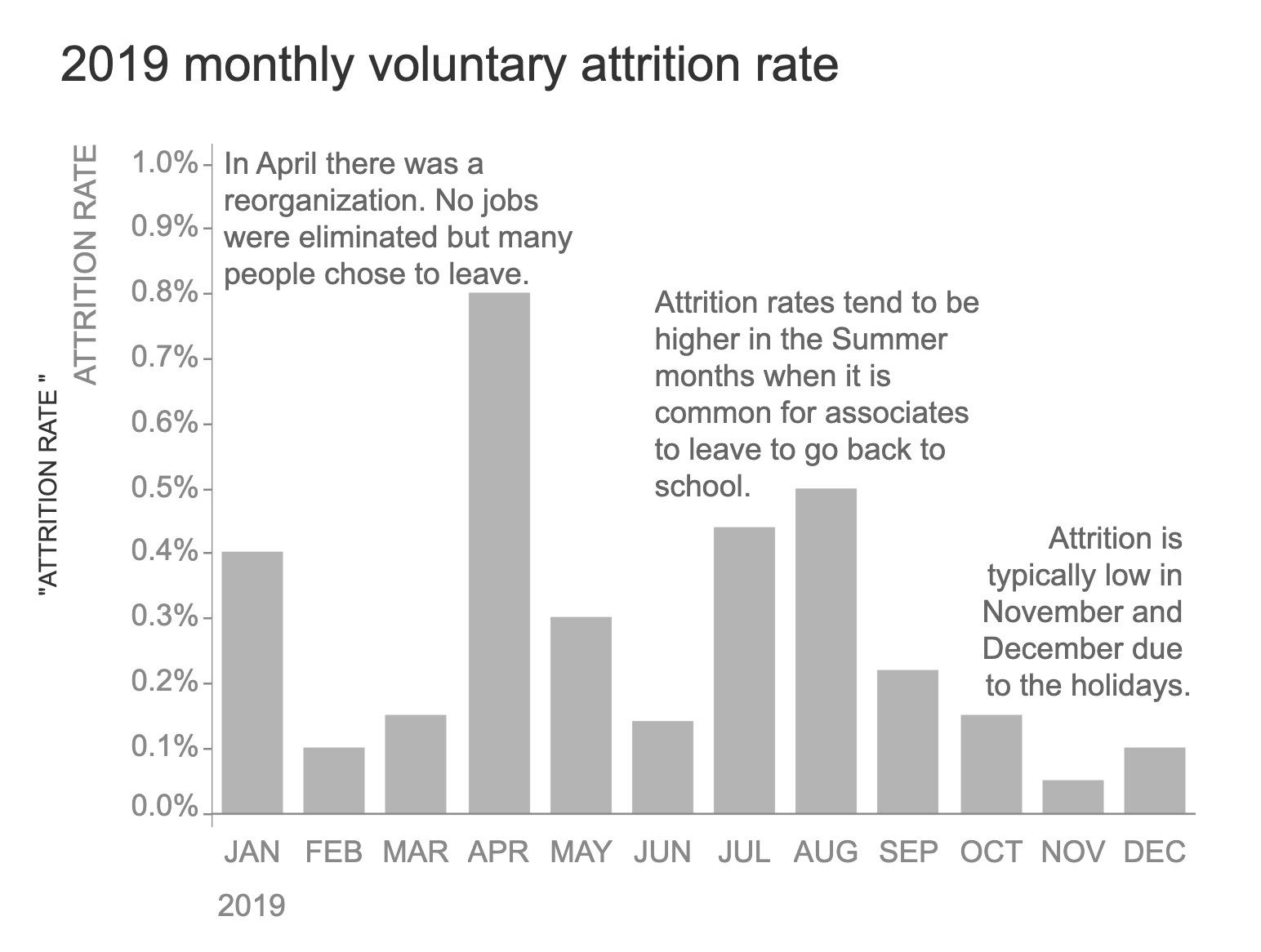
Part 1.2:

Match the following Gestalt Principles AND explain the differences between the four graphs with the graphs below. Additionally, indicate which graph you believe would create the least cognitive load for the audience.

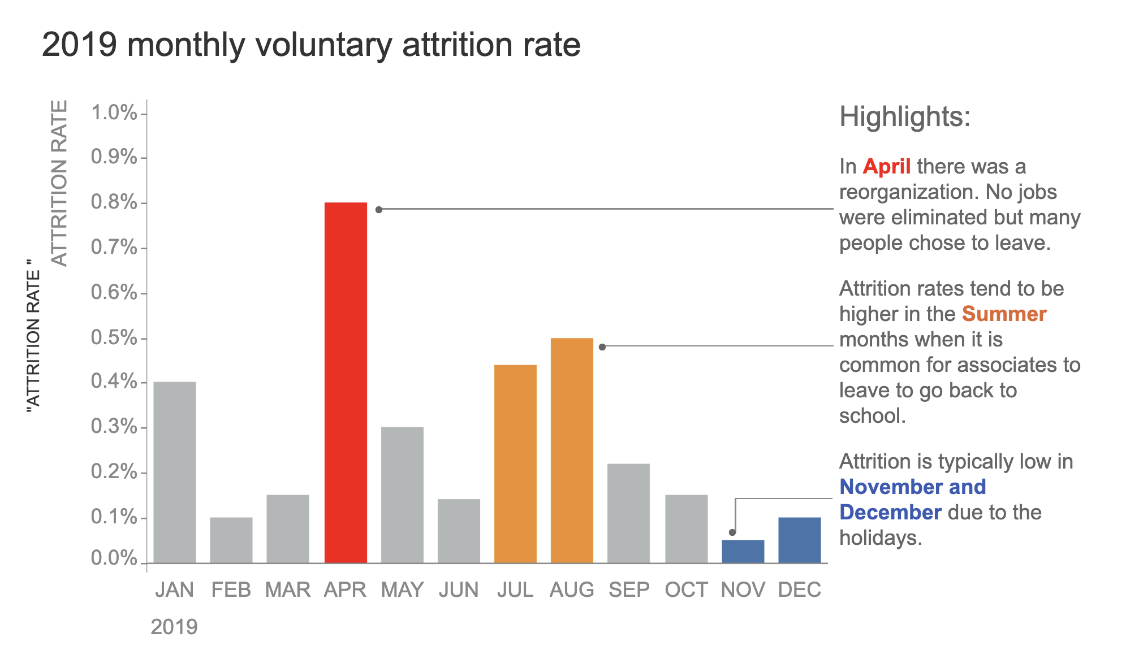
1. Proximity
2. Similarity
3. Enclosure plus similarity
4. Connection plus similarity

Response

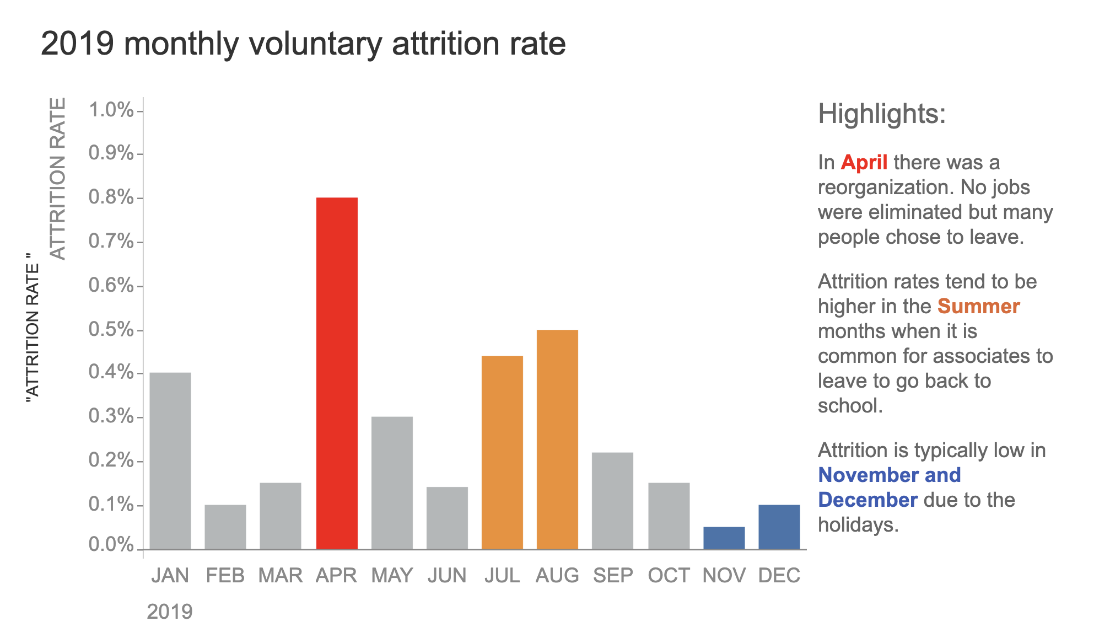
Bar chart showing separation counts per month of year for 2019 Another line says "No jobs were cut; many people decided to leave.") Attrition rates were also high concurrently in summer months. Nov and Dec are usually not high attrition months due to the holidays.



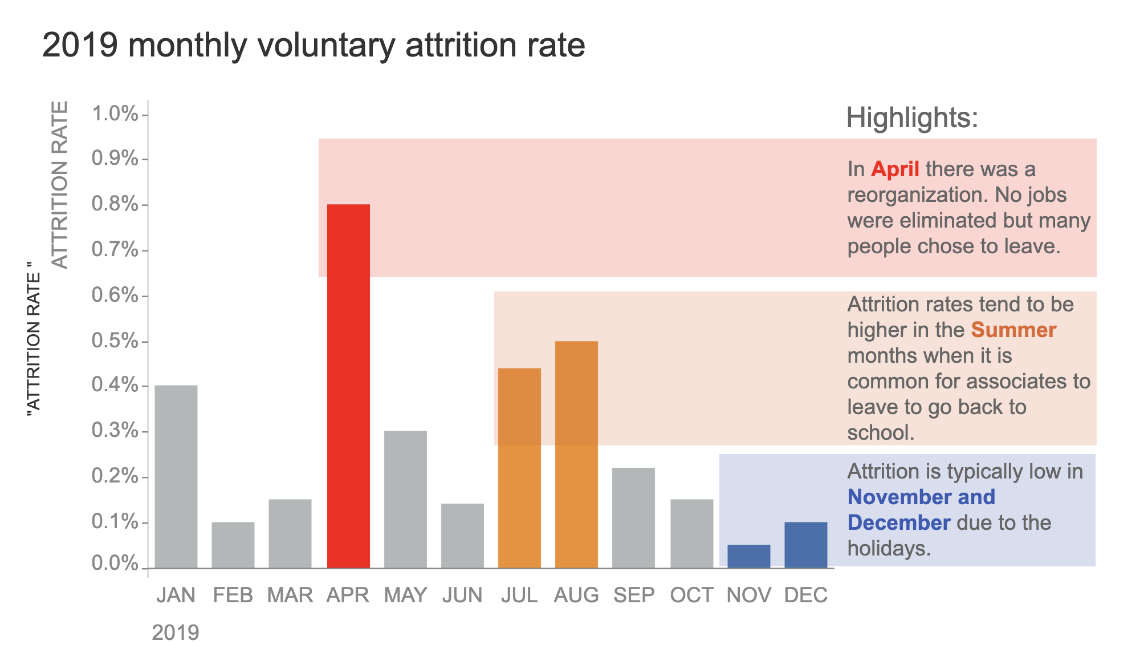
Neutral gray bars: this is a baseline visualization with simplistic proximity. The bars are in chronological order, and the annotations sit next to the corresponding points in the data. This is the Proximity principle (#1), where we group elements that are closer together.



Colored lines with arrows In this view, similarity (the same color) is used to cluster data points related to each other (April spike in red, summer months in orange, holiday period in blue), with arrows representing annotations. It is Connection + Similarity (#4) because it uses both connecting lines and color grouping to depict relationships.



You can see that this is also related to which lever was used (#2 -- Similarity) through levers using colors alone to technically group pertaining data points (red for April spike, orange for summer months, blue for holiday months), yet so asymmetrically with the un-connecting lines. Color coding enables the audience to related elements more at a glance.



This adds to the combination of Enclosure and similarity (#3) by both color coding the bars and shading the background areas that contain related elements. Different colored areas on the background produce separated clusterings while preserving the color relationship between the bars and their associated annotations.

Regarding cognitive load,: Image C would create the least cognitive load for the audience because:

1. Color-coded system nicely applied, but not in a noisy way
2. It keeps it simple while providing the most important patterns
3. It also does not contain other graphical features, such as arrows or shading of backgrounds, which may only be distracting
4. We also annotate with the same color everywhere in the chart where the bars show the color coding.
5. this lets the viewer absorb the 3 primary shapes (reorg spike, summer, holiday) without additional visual processing

Each of the other versions has its disadvantages as well:

* Image A takes more cognitive load to make connections between related components
* The arrows from Image B create visual noise not needed
* Background shading in Image D for the purpose of minor grouping, but an additional visual layer to process

All four options successfully communicate the information, but the third (C) image best balances visually emphasizing key patterns with a clean, minimalist presentation.

Part 2: Color Selection

*Write a paragraph describing at least three colors explaining how the colors can affect our visualizations.*

Color is an essential element in data visualization but a double-edged sword – misused; it can mislead or be interpreted with an unintended emotional response. Being a warm color, Red works effectively to stimulate a sense of urgency, danger, or passion in Western culture, while Chinese culture links ideas of luck to the color Red. Uses Of Red Red is one of the best colors for emphasizing the data points or warning signals — like the attrition spike visualization above and should be used judiciously for proper emphasis. Blue generally gives off feelings of trust, stability, and professionalism in the Western world; however, in some Middle Eastern cultures, it is also associated with protection, spirituality, and/or energy. This calming quality helps show the baseline data (or positive data, as seen in the holiday period highlight). Orange, a warm color between red and yellow in the visible spectrum, transports warmness, high spirits, and vitality in most cultures. Still, it signifies loss or grief in some Middle Eastern countries. In the attrition visualization, orange brings attention to the summer months' pattern without creating an immediate alarm analogous to red, but still, the usage of color remains captivating. Colors play the central role in visualizations, but beyond their cultural and emotional meaning, they must also be chosen, considering that the color palette is compatible with colorblind viewers and that they must create a visual hierarchy that guides the eye to the correct order of details. As seen in visualizations of monthly attrition rates, the combination of warm colors (red, orange) against cool colors (blue) made for an efficient contrast yet sustained harmony.

PART 3: Properties of Data

*Classify each variable (or attributes/columns) of the 2024\_track and\_field\_dataset\_messy in Canvas (under the Dataset module) based on the following types of data discussed in the lectures:*

* *N: Nominal*
* *O: Ordinal*
* *Q (Interval): Quantitative (Interval)*
* *Q (Ratio): Quantitative (Ratio)*

Track and Field Dataset Variable Classification

|  |  |  |
| --- | --- | --- |
| **Variable/Column** | **Classification** | **Justification** |
| Athlete ID | N | Unique identifier that represents each athlete. Numbers are used as labels without quantitative meaning. |
| Athlete Last Name | N | Categorical data representing names with no inherent order or numerical value. |
| Weight(K) | Q (Ratio) | Has a true zero point (0 kg means no mass) and ratios are meaningful (80kg is twice as heavy as 40kg). |
| Age | Q (Ratio) | Has a true zero point (0 years means no age) and ratios are meaningful (30 years is twice as old as 15 years). |
| Gender | N | Categorical data representing gender identity with no inherent order. |
| Event Name | N | Categorical data representing different types of events with no inherent order. |
| Event Date | O, Q (Interval) | Can be both: Ordinal as dates can be ordered, and Interval as differences between dates are meaningful but there's no true zero point. |
| Event Location | N | Categorical data representing different venues with no inherent order. |
| Event Country | N | Categorical data representing different countries with no inherent order. |
| Event Weather | O | Represents weather conditions that can be ordered (1-5 scale), but intervals between values may not be equal. |
| Event Results | Q (Ratio) | Has a true zero point (0 means no performance) and ratios are meaningful. |
| Event Rank | O | Represents finishing positions (1st, 2nd, 3rd, etc.) with clear ordering but not necessarily equal intervals. |
| Event Record | N | Binary categorical data (TRUE/FALSE) indicating whether a record was set. |
| event\_surface | N | Categorical data representing different types of surfaces with no inherent order. |
| Notes | N | Text descriptions with no inherent order or numerical value. |
| Event\_Temp in C | Q (Interval) | Temperature in Celsius has meaningful intervals but no true zero (0°C is not absence of temperature). |
| Event\_Temp in K | Q (Ratio) | Temperature in Kelvin has both meaningful intervals and a true zero point (0K is absolute zero). |

Part 4: Create a New Chart in Tableau

1. Chart 1: Average Performance by Gender and Surface Type (Bar Chart) A graph of blue rectangular bars

   Description automatically generated with medium confidence

Why I chose this chart:

* Bar charts are excellent for comparing categorical data
* It allows us to see the relationship between multiple variables (Gender, Surface Type, and Performance)
* The vertical bars make it easy to compare performance levels across different categories
* Shows clear distinctions between different surface types for each gender

Business Insights:

* We can identify which surface types produce the best performances for each gender
* This information can be valuable for:
  + Training program optimization based on surface type
  + Facility planning and investment decisions
  + Competition scheduling and venue selection
  + Athlete preparation strategies for different surfaces

1. Chart 2: Performance Distribution by Age Group (Bar Chart)

A screenshot of a graph

Description automatically generated

Why I chose this chart:

* Shows the relationship between age and athletic performance
* Allows us to see performance trends across different age groups
* The distribution helps identify peak performance age ranges
* Provides a clear visual representation of age-related performance patterns

Business Insights:

* Helps identify optimal age ranges for peak performance
* Can be used for:
  + Talent development and recruitment strategies
  + Training program customization by age group
  + Competition category planning
  + Long-term athlete development planning
  + Resource allocation for different age groups

Reference:

Nussbaumer Knaflic, C. (2015). *Storytelling with data* (C. N. Knaflic, Ed.). John Wiley & Sons.