

**Table of Contents**

1. [Fundamentals of Data Visualization 3](#_Toc169274617)

[Q1- What is the lie factor, and how does it impact Tufte’s graphical integrity rules? 3](#_Toc169274618)

[Q2- What is the lie factor in the following plot considering the values below (actual values) and that the volume of the corresponding cylinder represents values in graphics? 3](#_Toc169274619)

[Q3- What issues do we face when using unjustified 3D plots? 4](#_Toc169274620)

[Q4- Explain the concept of chart Junk and what are its advantages and disadvantages. 4](#_Toc169274621)

[Q5- What is a data-ink ratio, and should it be maximized or minimized? 5](#_Toc169274622)

[Q6- Describe what semantics can be understood from the following graphical codes. 5](#_Toc169274623)

[a- A Nested regions and partitioned regions 5](#_Toc169274624)

[b- Attached shapes 5](#_Toc169274625)

[c- Graphical objects in proximity 5](#_Toc169274626)

[d- Shapes enclosed by a contour 5](#_Toc169274627)

[Q7- Describe visual association and semantic association and how they would help in data visualization. 5](#_Toc169274628)

1. [Techniques of Data Visualization 6](#_Toc169274629)

[Q1- What are the marks channels, and several attributes encoded in the following two plots? 6](#_Toc169274630)

[Q2- Discuss the importance of using the following interaction techniques in data visualization: selection, change over time, navigation, filtering, brush and zoom, brush and link, and aggregation. 7](#_Toc169274631)

[Q3- Mention 4 interactive visualization techniques used in Google Maps. 8](#_Toc169274632)

1. [Static and Interactive Visualizations (project) 8](#_Toc169274633)

[Q1- Provide a detailed description of each of the datasets you selected, the attributes of interest, and the attributes you would like to visualize (data abstraction). 8](#_Toc169274634)

[Q2- Describe why you are looking at the data (task abstraction), and design rules for each plot you generated in R. 9](#_Toc169274635)

[Q3- Describe the encoding methods for each plot (i.e., choosing the plot based on data and tasks, marks and channels, etc.). 10](#_Toc169274636)

[Q4- Assess the methods (e.g., proper design and encoding, Tufte’s rules, Schneiderman Mantra, etc.) used to build your visualizations and how they helped (static and interactive). 25](#_Toc169274637)

[Q5- Analyze and assess the insights and findings for each plot you generated in R. 27](#_Toc169274638)

[Q6- Critically evaluate how the project visualizations impact the organization and decision-making. 27](#_Toc169274639)

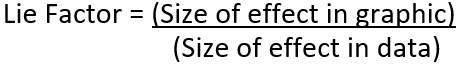
1. [Communication of Results and Findings 28](#_Toc169274640)

[Q3- Evaluate how well your project visualizations can tell the story and provide insights to the organization. 28](#_Toc169274641)

1. [References 31](#_Toc169274642)

# **Fundamentals of Data Visualization**

## Q1- What is the lie factor, and how does it impact Tufte’s graphical integrity rules?

Tufte’s graphical integrity rules consist of that the representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities measured. The lie factor is the factor that is used in visuals that may lead to misunderstanding of the true meaning behind the visual(data), which can be calculated by this formula:  


It shows how the size of the effect shown in the graphic is compared to the size of the effect in the data, so it is connected with Tufte’s graphical integrity rules because if the lie factor value after calculating is high there will be an overstating which will indicate having a misleading visual, so the Tufte’s graphical integrity rules work on reducing the lie factor in visuals. Tufte’s graphical integrity focused on minimizing the lie factor because it will produce an accurate and true representation of the data which will lead to a clear understanding of the visual to gather true and accurate insights from them.

## A cylinder and cylinder diagram Description automatically generatedQ2- What is the lie factor in the following plot considering the values below (actual values) and that the volume of the corresponding cylinder represents values in graphics?

Based on the formula for the lie factor given above:

I will start by calculating the size of the effect in the graphic by calculating the volume for each cylinder, I will be naming the large cylinder as A and the smaller cylinder as B.

**The volume of A:**

V=π\*r^2\*h

V= π \*(1.5) ^2\*5= 35.3

**The volume of B:**

V=π\*r^2\*h

V= π \*(1) ^2\*2.5= 7.8

**(V(A)-V(B))/V(B)**= (35.3-7.8)/7.8=3.5

Then, I will calculate the size of the effect in the data based on the formula above:

**Value(A)-Value(B)**=3-1=2

Finally, I will calculate the lie factor based on the formula above:

3.5/2=1.75

This indicates that the lie factor is higher than 1 which may lead to misleading in the visuals by 75%.

## Q3- What issues do we face when using unjustified 3D plots?

When we are dealing with 3D plots, we need to keep certain rules in mind and follow them, one of the most important rules is to make sure that “The Number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data”. Also, we need to not make use of unjustified 3D plots. Using 3D unjustified plots may lead to misleading the processes of understanding the plot and result in producing wrong insights into our data, adding a third dimension will make it difficult to assign distances which will result in hiding important data. As with any plots we need to color code our data, but in the 3D plots, we need to add a shadow to indicate the depth which can result in confusing the viewer and gathering false information. Finally, the interpretation of the 3D plots may require viewing the data from different angles and positions but not all interpreters may have the ability or desire to adjust the display to see all the data accurately resulting in an accurate reading of data.

## Q4- Explain the concept of chart Junk and what are its advantages and disadvantages.

Chart Junk is a concept of representing normal visuals in an eye-catching way, which will attract the attention of the viewers but does not work on improving the viewer's understanding of the insights because it focuses on the way of presenting the plots, and the same insights will be drawn from the two different plots (Chart Junk and the Normal Chart). One advantage of using a Chart Junk is that it will make the chart more memorable for the viewer, so it will stick in their mind and have the ability to retain the information longer. It will also boost energy and viewer engagement in the room while introducing visuals that will maintain focus and help in understanding the concept of the presented data because Chart Junks may introduce a way of storytelling our data. On the other hand, there are some disadvantages regarding Chart Junks as it will be time-consuming to waste much time designing the Junk charts which will sometimes result in the viewer’s deviation from the main concept of the chart while losing professionalism while presenting and earning the chart unseriousness, also, not capturing the key ideas and trends in the data will result in misinterpretation of the data and producing wrong insights.

## Q5- What is a data-ink ratio, and should it be maximized or minimized?

A data-ink ratio is a phrase that Tufte creates, and it is the indication of the ink used to show data.

By following the formula:

A black text on a white background

Description automatically generated

This maximizes the data-ink ratio by minimizing the data-ink in the graph like the 3D elements, using too many colors and the non-data elements, and ensuring the most ink used is for the representation of the actual data. The maximized data-ink ratio works to deliver an accurate and clear representation of data.

## Q6- Describe what semantics can be understood from the following graphical codes.

### A Nested regions and partitioned regions

Nested regions and partitioned regions follow the hierarchical concept of pattern semantics.

The nested regions may indicate the relationship between different elements such as a visual representing a country divided into states, and every state is divided into countries.

The partitioned regions may indicate the division into distinct parts such as dividing the population into distinct groups such as the year income gender, etc.

### Attached shapes

The attached shapes follow the parts of the conceptual structure concept of pattern semantics, which indicates a connection between shapes, such as how the tasks are connected in a flowchart through arrows to show the structure concept.

### Graphical objects in proximity

The graphical objects in proximity follow the (similar concepts, related information) concept of pattern semantics, which indicates that the objects placed near each other have a connection because usually, the proximity indicates relevancy, such as the website home page every tab on it are placed closer to each other indicating their functional relationship.

### Shapes enclosed by a contour

Shapes enclosed by a contour follow the (object, idea, entity, node) concept of pattern semantics, which indicates the identification of discrete elements within the visual context, which will produce a clearer understanding of information.

## Q7- Describe visual association and semantic association and how they would help in data visualization.

Both the Semantic and Visual associations contribute to the clarity and efficacy of information delivery, but they do so in different ways. The Semantic association, for example, focuses on interpreting data based on meaning and context by understanding data based on the underlying contextual meanings between the data, which will ensure enhancing comprehension by presenting the data in a meaning-full manner and facilitate the viewer's ability to derive meaning-full insights. Visual association, on the other hand, focuses on interpreting data based on the relationship between visual elements based on their proximity or similarity. Presenting data in a visual format will assist the viewer in quickly understanding the relationship between data as well as in identifying patterns and trends in the data by visually grouping related data points.

Both provide storytelling to the concept of data through the use of visual elements in visual association and the meaningful data produced by semantic association, and both can enhance the story of our data. They also enable viewers to dig deeper, explore, and analyze data by providing useful visuals and meaningful information to do so.

# **Techniques of Data Visualization**

## Q1- What are the marks channels, and several attributes encoded in the following two plots?

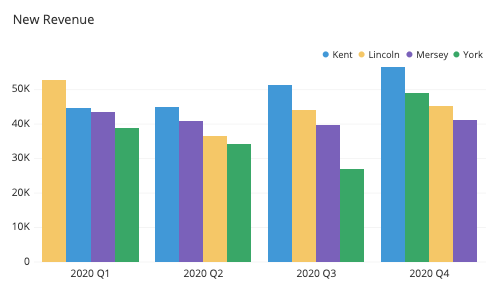
1. A graph of a graph showing the number of sales

   Description automatically generated with medium confidence

Marks: points.

Channels: Position (Both), color, size(area).

The number of attributes encoded: 4 attributes.

B)

Marks: lines, area.

Channels: Position (Both), color, size (Both).

The number of attributes encoded: 3 attributes.

## Q2- Discuss the importance of using the following interaction techniques in data visualization: selection, change over time, navigation, filtering, brush and zoom, brush and link, and aggregation.

**Selection:** Provides the viewer with the ability to concentrate on a particular data point inside the visualization, allowing them to conclude by contrasting and comparing various parts.

**Change over time:** By showing data across time for temporal data analysis, this feature enables viewers to comprehend previous data and make predictions based on it.

**Navigation:** By zooming or panning over various sections, the viewer may quickly examine massive data sets and dive down into specific areas of interest.

**Filtering:** By allowing the viewer to dynamically modify the data shown based on particular criteria, filtering helps the viewer to concentrate on subsets of data pertinent to their analytical objectives and improves visual clarity.

**Brush and Zoom:** allow the viewers to interactively investigate features within the visualization. They may use the brush tool to choose a particular region of interest, and then zoom in to inspect it more closely for more convenient data exploration.

**Brush and Link:** This tool enables viewers to coordinate their exploration across many connected visualizations for improved correlation and comparative analysis. It lets viewers choose data points in one visualization and similar data points in other linked visualizations.

**Aggregation:** Summarizing the data and presenting it at various levels enables viewers to aggregate the data to higher levels.

## Q3- Mention 4 interactive visualization techniques used in [Google Maps](https://www.google.com/maps/@31.9628747,35.9032147,12.1z?entry=ttu).

**Select:** Using Google Maps, viewers may pick particular map elements, such as marking locations they want to get more information about to decide which way to go.

**Navigate:** Navigating through maps to locate directions is possible with Google Maps. Viewers may move around the map, and zoom in, or out to select the best routes.

**Filtering:** Google Maps gives viewers the ability to alter the data shown on the map to enhance their research and locate pertinent information more quickly.

**Brush and Zoom:** Viewers of Google Maps may choose a region on the map to enlarge. Zoom levels on the map may be dynamically changed by viewers to concentrate on and investigate particular locations in more detail.

# **Static and Interactive Visualizations (project)**

## Q1- Provide a detailed description of each of the datasets you selected, the attributes of interest, and the attributes you would like to visualize (data abstraction).

**Name of the Data Set:** Beach Volleyball Data set.

**Description:**

The dataset contains information on match-level results, player details, and match-level statistics for beach volleyball matches. Most of the data is from the international [FIVB tournaments](https://en.wikipedia.org/wiki/FIVB_Beach_Volleyball_World_Championships) but about 1/3 is from the US-centric [AVP](https://en.wikipedia.org/wiki/Association_of_Volleyball_Professionals). Each row represents a single match played by two teams, with each team comprising two players, the data as I mentioned contains information for both the winning and the losing team about the tournament details(tournament, year, date), Match details(gender, match\_num, bracket, round), player details(name, birthdate, age, height, country), match statistics (attacks, kills, errors, hit percentage, aces, serve errors, blocks, digs). The purpose of the data set is to analyze beach volleyball matches, understand players' performance, and identify trends and patterns over time to guide insights to shape strategies and assist in the beach volleyball decision-making process.

My dataset has 65 columns, but I ran my project on 16 columns. The 16 columns used are as follows:

Attributes of interest:

**Year:** Year of the tournament.

**Score:** Match score separated by a dash and matches separated by a comma, e.g. 21 points to 12 points is 21-12.

**Gender:** Gender of team.

**w\_p1\_tot\_attacks:** Winner player 1 number of attacks (attacking swings over the net).

**w\_rank:** Winner team rank.

**w\_p1\_age:** Winner player 1 age.

**w\_p1\_hgt:** Winner player 1 height in inches.

**w\_p1\_tot\_kills:** Winner player 1 number of kills (point ending attacks).

**w\_p1\_tot\_blocks:** Winner player 1 total blocks - point ending blocks.

**w\_p1\_tot\_aces:** Winner player 1 total ace - point ending serves.

**w\_player1:** Winner player 1 Name.

**w\_p1\_country:** Winner player country.

**l\_p1\_country:** Losing player 1 country.

**w\_p1\_tot\_digs:** Winner player 1 total digs - successful defense of an attack.

**l\_p1\_tot\_digs:** Losing player 1 total digs - successful defense of an attack.

Visualize attributes:

* Player Age Distribution.
* Height Distribution of Players.
* Top Players by Total Blocks.
* Top Players by Total Kills.
* Total Matches Played by Year.
* Performance Over Time for Top Players (Total Aces).
* Average Total blocks vs. Player Height
* Average Total Kills vs. Player Height
* Winning Percentage by Country.
* Correlation Heatmap of Performance Metrics.
* Player Height vs. Winning Percentage.
* Number of Matches by Year and Gender.
* Player Rank vs. Age.
* Player Rank vs. Total Attacks.
* Player Statistics Over Time.
* Top Players by Total Digs.

## Q2- Describe why you are looking at the data (task abstraction), and design rules for each plot you generated in R.

The goal of using this specific data is to analyze beach volleyball matches, assess player performance, and find trends and patterns over time to lead insights into shaping plans and assisting in beach volleyball decision-making.

Task Abstraction:

**Understanding Player Demographics:**

Visualize age and height distributions to understand the demographic profile of players.

**Evaluating Player Performance:**

Identify top players based on performance metrics such as kills, blocks, and aces.

Compare player performance across different metrics and over time.

**Analyzing Match Trends:**

Visualize trends in the number of matches played over the years.

Analyze winning percentages by country to identify dominant countries.

**Exploring Performance Relationships:**

Investigate the relationship between player height and performance metrics.

Examine correlations between different performance metrics.

**These tasks and designing rules are answered below for each plot.**

## Q3- Describe the encoding methods for each plot (i.e., choosing the plot based on data and tasks, marks and channels, etc.).

Task Abstraction:

**Understanding Player Demographics:**

Visualize age distributions to understand the demographic profile of players (Analyze).

* **Player Age Distribution:**

A graph of a distribution

Description automatically generated

Designing Rules:

* Choose appropriate bin widths for clarity.
* Providing context: titles, axis labels, and breakpoints on the x-axis for clarity.

Marks:

Lines, area

Channels:

Both positions (horizontal represents the age distribution, and vertical represents the frequency), size (length, area)

A number of variables are encoded:

1(age)

Insight:

The plot demonstrates that the peak of the age distribution is around mid to late 30 which shows that most beach volleyball players are around these ages. Also, there is a noticeable drop-off in the number of players as the age increases beyond 35.

Formed Decision:

The stakeholders should focus on younger players, specifically targeting the age group of early 20s to mid-20s in their training programs as they are the most likely to participate in upcoming competitions.

For the players in their late 20s and early 30s, the stakeholders should think of creating a support program to extend the player's careers.

Task Abstraction:

**Understanding Player Demographics:**

Visualize height distributions to understand the demographic profile of players(Analyze).

* **Height Distribution of Players:**

A graph of a graph

Description automatically generated with medium confidence

Designing Rules:

* Ensure consistent bin widths.
* Providing context: titles, axis labels, and breakpoints on the x-axis for clarity.

Marks:

Lines, area

Channels:

Both positions (horizontal represents the height distribution, and vertical represents the frequency), size (length, area)

A number of variables are encoded:

1(height)

Insight:

The plot demonstrates that the height distribution for players is that most players are between 70 to 80 inches tall, with a small skew toward the taller players.

Formed Decision:

The stakeholders should focus on putting strategies for recruitment targeting tall athletes as it appears that height is a significant factor in the player field.

In the training, do a specialized session for taller players to focus on blocking and attacking skills.

Task Abstraction:

**Exploring Performance Relationships:**

Examine correlations between different performance metrics(Analyze).

* **Correlation Heatmap of Performance Metrics:**

A diagram of performance metrics

Description automatically generated

Designing Rules:

* Maximize Data Ink Ratio: Displays the relationship between player's statistics with minimal distractions.
* Ensure a clear color gradient.

Marks:

area

Channels:

Both positions (horizontal represents the performance metrics, and vertical represents the performance metrics)

color (gradient for correlation values)

A number of variables are encoded:

8 which are the performance metrics

Insight:

It is observed from the plot that the correlations between certain performance metrics are as high as the attacks and kills, which shows that the players who attempt more attacks attend to receive more kills, and so on for the other metrics who are observed to have a high correlation.

Formed Decision:

One of the important decisions regarding these insights is to use these correlations to tailor a specialized training program to focus on metrics that impact the match outcomes.

Task Abstraction:

**Exploring Performance Relationships:**

Analyze winning percentages by height to understand their relation and identify trends(Analyze).

* **Player Height vs. Winning Percentage:**

A graph of a number of dots

Description automatically generated with medium confidence

Designing Rules:

* Ensure clear axis labels.
* Providing context: titles, axis labels, and breakpoints on the x-axis for clarity.
* Serve a clear purpose: Demonstrates the relationship between a player's height and winning performance.

Marks:

points.

Channels:

Both positions (horizontal represents the height distribution, and vertical represents the winning percentage)

A number of variables are encoded:

1which is the height

Insight:

As shown from the plot there is a high correlation between the winning percentage and the height of taller athletes, which indicates that having tall players will tend to have more winning match outcomes.

Formed Decision:

The decision based on these insights is to have training programs to maximize the potential of shorter players to focus on their speed and strategic playing. Also, to put the priority on taller athletes in the scouting time and the priority on developing the skills of these taller athletes to build a competitive team.

Task Abstraction:

**Evaluating Player Performance:**

Identify top players based on performance metrics such as kills, blocks, and aces.

Compare player performance across different metrics and over time.

* **Player Statistics Over Time:**

A graph with a line

Description automatically generated

Designing Rules:

* Allow user selection of players and statistics.
* Users manipulate visual elements directly.
* Provide context: titles and topic labels for clear understanding.

Marks:

lines

Channels:

Both positions (horizontal represents the year, and vertical represents the specified statistical value)

A number of variables are encoded:

2 which is the total block and year

Insight:

The results demonstrate the peak performance periods in a standout year for certain players. Also, shows the shift and change in key performance metrics with some players always improving while others show changeable performances over the years.

Formed Decision:

The decisions must focus on tracking the performance metrics for players over time to identify the potential areas for improvements. Also, put long-term plans for player developments on the weak metrics and adjust based on the changeable performance trends.

Task Abstraction:

**Evaluating Player Performance:**

Identify top players based on the total dig performance metric.

Compare player performance across different metrics and over time. (Analyze).

* **Top Players by Total Digs:**

A graph with lines and numbers

Description automatically generated

Designing Rules:

* Allow user selection of top players.
* Users manipulate visual elements directly
* Provide context: titles and topic labels for clear understanding.

Marks:

lines

Channels:

Both positions (horizontal represents the year, and vertical represents the total digs)

Color represents different players.

Size(length)

A number of variables are encoded:

2 which is the total block and year

Insight:

The plot shows that certain players always appear at the top of the total gigs which indicates strong defense skills. Also, the players having top dig performance tend to have an important impact on the team's defensive abilities.

Formed Decision:

The decisions should focus on the top dig players by using them as benchmarks to train other players as well as develop a specialized training program for teaching the correct defensive tactics to improve the overall team defense.

Task Abstraction:

**Analyzing Match Trends:**

Visualize trends in the number of tournaments played over the years comparing between genders. (Analyze).

* **Number of Matches by Tournament Year and Gender:**

A graph with a line and a blue and green line

Description automatically generated

Designing Rules:

* Displays the total number of matches played each year with minimal distractions.
* Provide context: Users dynamically query and see real-time results.
* Clear vision and understanding: Ensure that key points and conclusions are clear and graphically represented.

Marks:

lines

Channels:

Both positions (horizontal represents the tournament years, and vertical represents the number of matches)

Colors represent different genders

A number of variables are encoded:

2 which is the number of matches and tournament year

Insight:

The data shows that the number of matches played each year has a changeable rate of increase which indicates the growing popularity and participation in beach volleyball. Also, both the male and female matches show similar trends over the years.

Formed Decision:

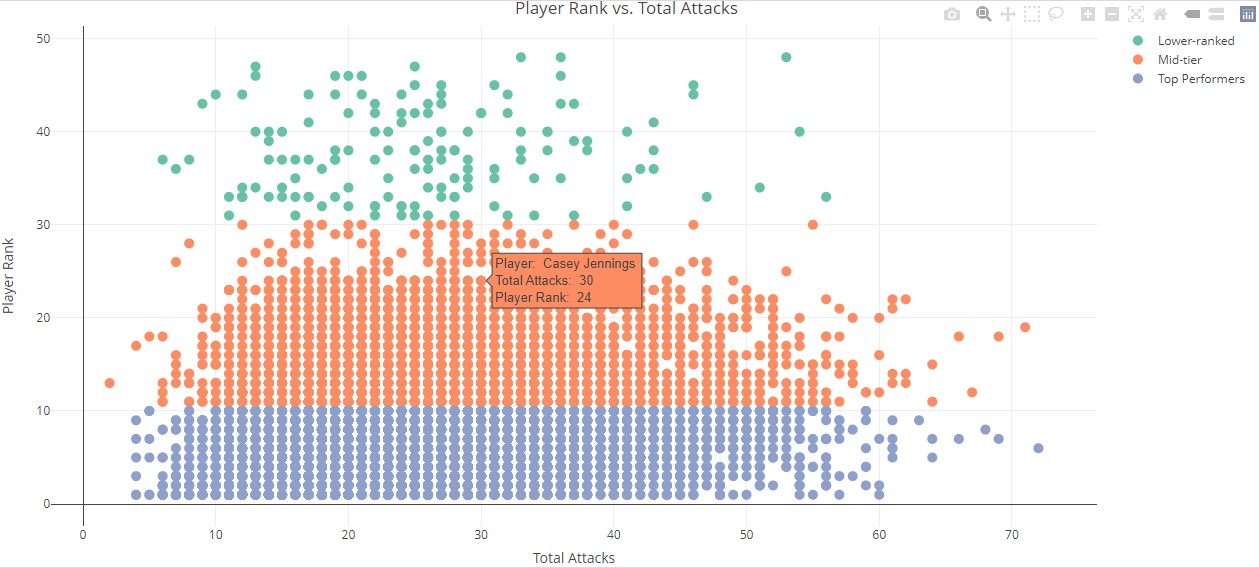
The decisions should focus on continuously promoting and investing in beach volleyball to obtain further increased participation. Also, ensure that these investments are equally distributed between female and male matches to maintain balanced achievements.

Task Abstraction:

**Evaluating Player Performance:**

Identify the rank of players based on performance metrics such as attacks. (Analyze).

* **Player Rank vs. Total Attacks:**



Designing Rules:

* Highlights the relationship between player ranking and total attacks.
* Provide context: Users manipulate visual elements to directly reveal data at multiple levels of detail
* Encourage comparisons: Facilitates comparisons between player ranks and total attacks.
* Allows viewers to hover over points to get specific details about the player.

Marks:

points

Channels:

Both positions (horizontal represents the total attacks, and vertical represents the player rank)

Colors represent different performance categories (top performers, mid-tier, lower-ranked)

A number of variables are encoded:

1 which is the total attacks.

Insight:

These insights show that the higher-ranked players mostly have a higher number of total attacks which suggests that the offensive play is important to obtaining higher rankings. Also, that can lead to the insight that the most successful players are more aggressive in their playing style.

Formed Decision:

The decisions should focus on encouraging an offensive mindset in the training sessions which focus on the attacking strategies. Also, focus on the higher-ranked players and alter play styles to improve overall team performance.

Task Abstraction:

**Evaluating Player Performance:**

Identify the rank of players based compared by age. (Analyze).

* **Player Rank vs. Age:**

A screen shot of a computer screen

Description automatically generated

Designing Rules:

* Displays the relationship between a player's rank and age with minimal distractions.
* Provide context: Users manipulate visual elements to directly reveal data at multiple levels of detail
* Allows viewers to hover over points to get specific details about the player.

Marks:

point.

Channels:

Both positions (horizontal represents the player's age, and vertical represents the player's rank)

Colors represent different performance categories (top performers, mid-tier, lower-ranked)

A number of variables are encoded:

1 which is the age.

Insight:

The data shows that age can be an important key in performance and ranking which shows that the younger players tend to be the higher-ranked players. Also, shows that there are trends where peak performance is associated with younger age groups, with some exceptions.

Formed Decision:

The decisions should focus on identifying young talents to build a strong group of highly-ranked players for future scouting. Also, developing age-specific training programs that meet the physical and skill development needs of young athletes.

Task Abstraction:

**Exploring Performance Relationships:**

Investigate the relationship between player height and kills as performance metrics. (Analyze).

* **Average Total Kills vs. Player Height:**

A graph with blue lines and numbers

Description automatically generated

Designing Rules:

* Maximize Data Ink Ratio: Represents average performance metrics versus operator height.
* Providing context: titles, axis labels, and breakpoints on the x-axis for clarity.
* Serve a clear purpose: Demonstrates the relationship between a player's height and performance metrics.

Marks:

lines

Channels:

Both positions (horizontal represents the player's height, and vertical represents average total kills)

A number of variables are encoded:

1 which is the total kills and height.

Insight:

The data shows that the taller players resort to having a higher average number of total kills, emphasizing the advantage of height in offensive operations.

Formed Decision:

Decisions should focus on utilizing height advantages in the offensive strategies. Also, develop a training session that will focus on enhancing the attack efficiency for the taller players to increase their impact and performance in the matches.

Task Abstraction:

**Exploring Performance Relationships:**

Investigate the relationship between player height and blocks as performance metric. (Analyze).

* **Average Total Blocks vs. Player Height:**

A graph showing a number of blocks

Description automatically generated

Designing Rules:

* Maximize Data Ink Ratio: Represents average performance metrics versus operator height.
* Providing context: titles, axis labels, and breakpoints on the x-axis for clarity.
* Serve a clear purpose: Demonstrates the relationship between a player's height and performance metrics.

Marks:

lines

Channels:

Both positions (horizontal represents the player's height, and vertical represents average blocks)

A number of variables are encoded:

1 which is the total blocks and height.

Insight:

The data shows a strong relation between the player's height and the average total block, also the taller the player the more they are effective in blocking.

Formed Decision:

The decision should focus on prioritizing the development of the blocking skills of the taller players to improve the defending teams. Also, implementing training sessions that exploit height advantage for blocking purposes.

Task Abstraction:

**Evaluating Player Performance:**

Compare player performance across aces performance metrics over time. (Analyze).

* **Top 2 Players Performance Over Time (Total Aces):**

A graph with red and blue lines

Description automatically generated

Designing Rules:

* Maximize Data Ink: Highlights performance trends of top players over time with minimal distractions.
* Provide context: titles, axis labels, and legend to identify the player.
* Integration with descriptions: The chart is closely integrated with statistical descriptions for complete understanding.

Marks:

lines

Channels:

Both positions (horizontal represents the years, and vertical represents total aces)

Color represents different players.

Size (length)

A number of variables are encoded:

1 which is the total aces and year.

Insight:

The data shows that these top 2 players show peak performances with some periods also it shows the variations in total aces. Also, the analysis of their performance trends can provide insights into elements contributing to their success.

Formed Decision:

The decisions must use the performance data of the top players to identify key factors and contribute to their success. Also, embrace the best training sessions and practicing and successful strategies from top players in training programs for other athletes.

Task Abstraction:

**Analyzing Match Trends:**

Analyze winning percentages by country to identify dominant countries. (Analyze).

* **Top 10 Countries by Winning Percentage:**

A graph with blue and white bars

Description automatically generated with medium confidence

Designing Rules:

* Maximize Data Ink Percentage: Represents winning percentages across countries.
* Providing context: titles, axis labels, and categorized data for clarity.
* Clear vision and understanding: Ensure that key points and conclusions are clear and graphically represented.

Marks:

lines and area

Channels:

Both positions (horizontal represents the winning percentage and vertical represents countries)

Size (length & area)

A number of variables are encoded:

1 which is the country.

Insight:

The data highlight the predominance of a few countries in international competitions. Also, Some countries consistently show higher winning percentages, indicating strong beach volleyball programs and player development systems.

Formed Decision:

The decisions should focus on training sessions to identify the best strategies for top countries, also having collaborations with successful countries to improve the beach volleyball programs

Task Abstraction:

**Analyzing Match Trends:**

Visualize trends in the number of matches played over the years. (Analyze).

* **Total Matches Played by Year:**

A graph of a number of green bars

Description automatically generated with medium confidence

Designing Rules:

* Provide context: titles and labels for clarity.
* Ensure clear year labels.
* Clear vision and understanding: Ensure that key points and conclusions are clear and graphically represented.

Marks:

lines and area

Channels:

Both positions (horizontal represents the years, and vertical represents total matches played)

Size(length & area)

A number of variables are encoded:

1 which is the year and number of matches.

Insight:

The data shows that there is a general increase in the total number of matches played each year which reflects the growing popularity of this sport also shows that in some years like 2011 and 2012, the number of matches is low which is caused by certain factors.

Formed Decision:

The decisions should focus on the continuous promotion of this sport to gain popularity also keeping in mind having support systems to accommodate the increased number of matches and participants.

Task Abstraction:

**Evaluating Player Performance:**

Identify the top 10 players based on kills as a performance metric. (Analyze).

* **Top 10 Players by Total Kills:**

A graph of a number of players

Description automatically generated

Designing Rules:

* Maximize Data Ink Ratio: Highlights top players based on total kills with minimal distractions.
* Provide context: titles and topic labels for clear understanding.
* Serve a clear purpose: Clearly shows the top players based on total kills.

Marks:

lines and area

Channels:

Both positions (horizontal represents total kills, and vertical represents the players)

Size (length & area)

A number of variables are encoded:

2 which is the total kills and player.

Insight:

The data shows that the top 10 players are consistently dominating in terms of offensive performance which is impacting the team's success. Also, the players show a great performance of constancy in their ability to score points through attacks.

Formed Decision:

The decision is to use the top performers' playing styles and techniques to train and develop the other players’ offensive skills.

Task Abstraction:

**Evaluating Player Performance:**

Identify the top 10 players based on blocks as a performance. (Analyze).

* **Players by Total Blocks:**

A graph of a number of players

Description automatically generated

Designing Rules:

* Maximize Data Ink Ratio: Compares top players based on total blocks.
* Providing context: Headings and labels for clarity and context.
* Integration with descriptions: The chart is closely integrated with statistical descriptions for complete understanding.

Marks:

lines and area

Channels:

Both positions (horizontal represents the total blocks, and vertical represents the players)

Size (length & area)

A number of variables are encoded:

2 which is the total blocks and player.

Insight:

The data shows that the top players are very important for their team's defensive strategies, and it shows that they are always performing well in blocking the opponent’s attacks.

Formed Decision:

The decision will include using the defining skills of the top blocking performers in building team strategies. As well as develop training sessions for the other players focusing on their blocking skills to enhance the overall team defense.

## Q4- Assess the methods (e.g., proper design and encoding, Tufte’s rules, Schneiderman Mantra, etc.) used to build your visualizations and how they helped (static and interactive).

**Static Visualizations: Tufte’s Rules & It’s importance:**

My statice visuals follow Tufte’s Rules to ensure efficiency, clarity, graphical integrity, and precision, here is how my visuals follow this rule:

* Clarity & preciseness:

Tufte's principles guarantee that static visualizations convey data clearly and exactly, without needless distractions. This is important to stakeholders who need to rapidly understand key insights without going through irrelevant content.

* Maximized Data-Ink Ratio:

Visuals become more efficient and compelling when they focus on key facts reducing unnecessary ink. This assists in highlighting the most relevant trends and patterns, making it easier for stakeholders to understand the data. For example, the Player Height vs Win Percentage chart employs basic points and axes, enabling the data to be the primary focus without any additional designs.

* Avoidance of Chart junk:

The removal of distracting components such as excessive gridlines and 3D effects ensures that graphics are clean and professional. This increases the trustworthiness of the facts and helps keep the focus on what is important. The Performance Measurements Heatmap captures this with a simple style and fresh color grading, making the correlations between measurements easy to identify without visual clutter.

* Integrity of Text and Graphics:

Integrating Text and labels increases visual understanding. This facilitates the identification and comprehension of key data points. For example, the "Top Players by Total Blocks" table features simple player name labels, making it simple to recognize top performers.

**Interactive Visualizations: Schneiderman Mantra & It’s importance:**

* Details-on-Demand:

Allowing users to click or hover over data points to uncover additional information delivers a greater degree of insight while keeping the primary view tidy. This is essential for stakeholders who want thorough information to make informed judgments.  
For example, in the Top Players by Total Digs visualization, selecting the player's data displays comprehensive performance information.

* Dynamic Querying:

Enabling dynamic querying and real-time results allows for a more responsive and interesting data exploration experience. Stakeholders may test various theories and see the findings quickly, allowing them to make more dynamic decisions. For instance, users may dynamically query and view real-time results in a visual player rating vs total attack.

* Direct Manipulation:

Dragging sliders or clicking checkboxes to filter data are examples of interactive components that make data exploration more realistic. This simple adjustment makes the data analysis process more understandable and user-friendly to stakeholders.  
For example, "Top Players by Total Digs" lets you switch between players and select specific players to try and compare their performance.

In Summary, by following Tufte's rules for static visualizations, the project guarantees that the data is displayed, precisely, and efficiently. This clarity and concentration are critical for stakeholders who must immediately grasp essential findings. Schneiderman's belief, on the other hand, when applied to interactive visuals, produces a dynamic, user-friendly experience that allows stakeholders to thoroughly look over the data, focus on specific areas of interest, and receive deep insights into demand. These tactics work together to make visuals informative and interesting, giving stakeholders complete and useful info.

## Q5- Analyze and assess the insights and findings for each plot you generated in R.

**This question is answered above for each plot there is an insight and informed decision.**

## Q6- Critically evaluate how the project visualizations impact the organization and decision-making.

**Decision Making**

The visualizations I generated in this project are impacting the decision-making within the beach volleyball organizations, by presenting my data clearly and understandably as well as in a visually appealing manner, these plots will enable the stakeholder to quickly key trends by merely grasping the complex information, as you can see below is a more detailed way of proving how my visualizations can impact decision making:

* Strategic planning:

Strategic planning can benefit from a visualization for example displaying the winning percentage by country will help in identifying countries with high winning percentages which can enable strategic collaborations or the adoption of training methods from those countries with high winning percentages. Also, the trends over time as the number of matches played over the years will provide insights into the popularity of this sport which will lead to future investments.

* Talent scouting & recruitment:

In the busy time of talent scouting and recruitment my visualization will be available for stakeholders to help them in decisions such as the correlation heatmaps and the player ranks, so these visualizations will help them understand the key performance metrics that correlate with a higher rank and choosing these players that show the availability of these metrics, using these metrics will identify promising talent so scouts and recruiters can focus on recruiting players who fit the profile of successful athletes.

* Resource Allocation:

My visualizations can track the top players’ performance throughout time to help the stakeholders make informed decisions about where to invest the resources for player support and development. By understanding which of the players are performing well these resources can be directly go toward them to support these p; players to enhance their potential.

**Stakeholders and Their Use of the Project**

The main stakeholders in the beach volleyball area that can benefit from my visualization are the coaches and trainers, team managers, scouts and recruiters, sports analysts and sponsors, and investors. How each can be benefited are as follows:

* Scouts & Recruiters:

They can benefit by using the performance metrics and demographic data to identify promising talents and will focus on recruiting players who can fit the profiles of successful players which will ensure a competitive team roster.

* Coaches & Trainers:

They can develop tailored training sessions and allocate key areas of improvement for the players depending on the performance metrics, which will help in the overall performance and effective training sessions.

* Team Managers:

They can build strategic decisions on the player selection and the match preparations depending on the performance metric for the team players to ensure that the team is well prepared and equipped for the matches and competitions.

* Sports Analysts:

They can analyze the trends and patterns to provide deeper insights into the matches and the data Drevin obtained can provide strategic advice that can influence the game strategies and team dynamics.

* Sponsors & Investors:

They can understand the popularity and growth of this sport to make informed investments to make a well-studied decision as well as identify high-performing players and teams for the sponsorship opportunities, ensuring that they are investing and sponsoring teams and players as well as gaining a huge impact.

# **Communication of Results and Findings**

## Q3- Evaluate how well your project visualizations can tell the story and provide insights to the organization.

**Storytelling and Insight**

The visuals in this project are quite successful in telling the story of beach volleyball dynamics, player performance, and match patterns. This is how they contribute:

* Clarity and Comprehension:

My visualizations make the complex data into a simpler form which makes It more understandable and accessible to all the stakeholders especially if they differ in their statistical experience and their position in the organization. Also, the use of clear titles and legends will make it easier when interpreting the insights providing on-the-spot insights.

* Significance and Targeting:

Each one of the plots focuses on a specific area of the sport like the player demographics, performance metrics as well as match information. This targeting way will provide insights that are directly significant to the stakeholder's needs.

* Actionable Insights:

The plots provide actionable insights into the highlighted key elements such as the dominance of specific countries or the effect of the player height on the performance which can directly inform the decision-making process.

* Historical and Predictive Analysis:

My plots provide both predictive insights and historical analysis, for example, by analyzing the performance over time the stakeholders can predict future performance and make dependable decisions as well as showing the trends in match numbers throughout the years will help in predicting the growth and plan of the sport for the future depending on that.

**Why My Project Is the Right Project?**

My project is the right project for the organizations in beach volleyball for some reasons which are listed below:

* Data-Driven Decision-Making:

My project focuses on offering a thorough, data-driven approach to understanding and developing beach volleyball, replacing intuition and guesswork with hard statistics and visual proof. As well as covering all facets of the sport, from player demographics and performance statistics to match trends. This extensive coverage provides a holistic perspective, which is critical for making educated decisions.

Customizable and Interactive:

My project has the special element of the interactivity of visuals to focus on a specific player, metrics, and periods of time, which gives the ability to meet the specific needs of the users and stakeholders and allows them to explore the data in depth, which make my project a dynamic project to use.

* Strategic Insights:

The project produces strategic insights that have an important impact on recruiting, training, resource allocation, and overall team success. This makes it a vital resource for coaches, managers, and analysts.

* Improving Competitive Edge:

The organization may gain a competitive advantage by using the information from these visualizations. This results in enhanced player performance and team success on a global scale.

To summarize, this project not only gives a thorough understanding of the dynamics of beach volleyball, but it also provides stakeholders with the tools and insights they need to make educated decisions, resulting in the growth and success of beach volleyball.

# **References**

* Ziemkiewicz, C., Ottley, A., Crouser, R.J., Yauilla, A.R., Su, S.L., Ribarsky, W. and Chang, R., 2012. How visualization layout relates to locus of control and other personality factors. *IEEE transactions on visualization and computer graphics*, *19*(7), pp.1109-1121.
* Widder, M. and Gorsky, P., 2013. How students use a software application for visualizing 3D geometric objects to solve problems. *Journal of Computers in Mathematics and Science Teaching*, *32*(1), pp.89-120.
* Schwabish, J., 2021. The practice of visual data communication: what works. *Psychological Science in the Public Interest*, *22*(3), pp.97-109.
* Pattanaik, S.N. and Wiegand, R.P., 2021. Data visualization. *Handbook of Human Factors and Ergonomics*, pp.893-946.
* Sagerer, G. and Niemann, H., 2013. *Semantic networks for understanding scenes*. Springer Science & Business Media.
* Fabrikant, S.I. and Buttenfield, B.P., 2001. Formalizing semantic spaces for information access. *Annals of the Association of American Geographers*, *91*(2), pp.263-280.
* Schloss, K.B., Leggon, Z. and Lessard, L., 2020. Semantic discriminability for visual communication. *Ieee transactions on visualization and computer graphics*, *27*(2), pp.1022-1031.
* Tominski, C., 2022. *Interaction for visualization*. Springer Nature.
* Elmqvist, N. and Fekete, J.D., 2009. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE transactions on visualization and computer graphics*, *16*(3), pp.439-454.
* Fua, Y.H., Ward, M.O. and Rundensteiner, E.A., 2000. Structure-based brushes: A mechanism for navigating hierarchically organized data and information spaces. *IEEE Transactions on visualization and computer graphics*, *6*(2), pp.150-159.