IFT 598: Data Visualization & Reporting for IT

(Taha H. Ahmed, Evan Sequeira, Neeraj Patel, Saaransh Rana)

**Project - Phase I: Planning**

Dr. Asmaa Elbadrawy

November 24, 2021

**Section 1: Used Visualization Tools**

We are primarily using Tableau as the visualization tool of choice. We chose it over other tools like Power BI, for the following reasons:

* We were taught Tableau as a part of our class and have implemented all content in it.
* It has an easy-to-use interface, with simple drag and drop features and customization options, and also cleans data accordingly.
* Since the database we are working with is not that big, Tableau is suitable for visualizing the dataset.
* It has the option of reading many different file formats such as Excel, text files, Microsoft Access, PDFs et cetera.
* It is cross-platform. Hence, we can share and access the same file across different Operating Systems.

**Section 2: Explanation of Required Data Pre-processing, if any**

The data pre-processing done is as follows:

* The ‘Loaned From’ column had mostly null values. The entire column was dropped from the dataset.
* The columns ‘Value’, ‘Wage’, ‘Release Clause’ were initially in the format ‘€110.5M’, ‘€565K’ and were recognized as strings. These were stripped so that only the numbers remained and were then recognized as such.
* Similarly, the ‘Height’ column was in the format, ‘5’7’, ‘5,8’ and was recognized as strings. The apostrophe ‘’’ was replaced by a period ‘.’ for the numbers to be read as float values.
* Similarly, the ‘Weight’ column was in the format, ‘158lbs’, ‘172lbs’ and was recognized as strings. These were stripped so that only the numbers remained and were then recognized as such.
* The columns ‘LS’, ‘ST’, ‘RS’, ‘LW’, ‘LF’, ‘CF’, ‘LF’, ‘RW’, ‘LAM’, ‘CAM’, ‘RAM’, ‘LM’, ‘LCM’, ‘CM’, ‘RCM’, ‘RM’, ‘LWB’, ‘LDM’, ‘CDM’, ‘RDM’, ‘RWB’, ‘LB’, ‘LCB’, ‘CB’, ‘RCB’, ‘RB’ all were in the format ‘89+2’, ‘93+2’ to show potential growth of the player’s stats. For the purpose of this database, these added numbers were stripped so that only the base numbers (89, 93) remained.

**Section 3: List of Final Sets of Questions**

1. Does a correlation exist between a player’s age and reaction time?
2. What is each player’s rating in each of the physical statistics (shooting, agility, composure, et cetera)?
3. For every player, is there a relation between their strength, shot power and finishing, given their age?
4. Is a player able to score more freekicks based on his accuracy and the curve he can apply on the ball?
5. Which players are paid the most/least?
6. What is the age distribution of the players?
7. What is the player distribution over the world?
8. Who are the Top N players for each position on the field?
9. Which club has the best overall rating, based on the players?
10. Do players from certain nations have better ratings than others?

**Section 4: Dashboard Plots**

For each of the questions,

1. Does a correlation exist between a player's age and reaction time?
   1. Explain what it shows and how that relates to the set of questions.

The graph plots reactions of the players vs their ages and shows a trend line to observe correlation. Hovering over each point shows each player’s name.

* 1. List the set of used pre-attentive attributes and colors.

Pre-attentive attributes: Shape, Color (Hue)

Colors: Blue.

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.

Chart, scatter chart

Description automatically generated

1. What is each player’s rating in each of the physical statistics (shooting, agility, composure, et cetera)?
   1. Explain what it shows and how that relates to the set of questions.

It shows play related attributes (Acceleration, Aggression, Agility, Balance, Ball Control, Composure, Crossing, Curve, Dribbling, FK Accuracy, Finishing, Heading Accuracy, Interceptions, Jumping, Long Passing, Long Shots, Marking, Overall, Penalties, Positioning, Potential, Reactions, Short Passing, Shot Power, Skill Moves, Sliding Tackle, Special, Sprint Speed, Stamina, Standing Tackle, Strength, Vision, Volleys, Weak Foot).

for a player of the user’s choice. The user can select which player and which attributes to be displayed.

* 1. List the set of used pre-attentive attributes and colors.

Pre-attentive Attributes: Length, Color (Hue).

Colors: Tableau Palette (multiple colors).

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.

Graphical user interface, application

Description automatically generated

1. For every player, is there a relation between their strength, shot power and finishing, given their age?
   1. Explain what it shows and how that relates to the set of questions.

Each player’s strength, shot power and finishing are plotted, and a trend line is used to observe correlation. Name and Age is seen when hovering over each datapoint.

* 1. List the set of used pre-attentive attributes and colors.

Pre-attentive attributes: Shape, Color (Hue), Spatial Clustering.

Colors: Orange-Blue Diverging Color Palette.

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.

Chart, scatter chart

Description automatically generated

1. Is a player able to score more freekicks based on his accuracy and the curve he can apply on the ball?
   1. Explain what it shows and how that relates to the set of questions.

Curve vs Freekick Accuracy is plotted, with finishing as a color. A trend line is added to show correlation between these attributes.

* 1. List the set of used pre-attentive attributes and colors.

Pre-attentive attributes: Spatial grouping, Color (Hue), Shape.

Colors: Red-Green Diverging Color Palette

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.

Chart, scatter chart

Description automatically generated

1. Which players are paid the most/least?
   1. Explain what it shows and how that relates to the set of questions.

The graph shows name vs wage of each player, allowing the user to see which players are paid the most/least.

* 1. List the set of used pre-attentive attributes and colors.

Pre-attentive Attributes: Length, Color (Hue)

Colors: Blue

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.

Chart, histogram

Description automatically generated

1. What is the age distribution of the players?
   1. Explain what it shows and how that relates to the set of questions.

This graph shows the age distribution of the players, it can help us understand the average age of players, youngest/oldest players, et cetera. This can also be narrowed down to age distribution of players in a particular team.

* 1. List the set of used pre-attentive attributes and colors.

Color of the line graph (Hue).

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.

A picture containing text, whiteboard

Description automatically generated

1. What is the player distribution over the world?
   1. Explain what it shows and how that relates to the set of questions.  
      This graph shows the distribution of the players over various countries.
   2. List the set of used pre-attentive attributes and colors.

Pre-attentive attributes: Spatial positioning, Color of the choropleth chart (Intensity).

Colors: Blue Color Palette

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.  
     Map

     Description automatically generated

1. Who are the top N players for each position on the field?
   1. Explain what it shows and how that relates to the set of questions.  
      This graph can show us the names of the players and their ratings for different positions on the field. This can help us gain valuable information about how a player performs in different positions on the field.
   2. List the set of used pre-attentive attributes and colors.

Pre-attentive attributes: Length of the bar chart, Color of the bar chart (Hue).

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.

Chart, bar chart

Description automatically generated

1. Which club has the best overall rating, based on the players’ overall ratings?
   1. Explain what it shows and how that relates to the set of questions.  
      This plot shows us the best football club based on the sum of the players’ overall performance ratings.
   2. List the set of used pre-attentive attributes and colors.

Length of the bar chart.

Color of the bar chart (Hue).

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.

Chart

Description automatically generated

1. Do players from certain nations have better ratings than others?
   1. Explain what it shows and how that relates to the set of questions.  
      This plot shows the players average overall rating based on their nationality. This helps us understand if a players from a certain country are better than those from others.
   2. List the set of used pre-attentive attributes and colors.

Pre-attentive attributes: Spatial Clustering, Color (Hue).

* 1. Include a rough, hand-drawn or computer-drawn, figure of the plot.

Chart, treemap chart

Description automatically generated

**Section 5: Dashboard Interactivity**

The dashboard has a few planned interactive features.

Measure Values Filter:

* List what they will be used for

Used to select which of the player’s attributes to be visualized.

* Which plots are connected to each one

The plot for ‘What is each player’s rating in each of the physical statistics (shooting, agility, composure, et cetera)’.

* The value range for each control and whether or not it is loaded from a certain attribute in the data.

Value ranges: 0 – 100.

Name Filter:

* List what they will be used for

1. Used to select the player for whom attributes will be visualized.
2. The plot for ‘Player Distribution over Country’

* Which plots are connected to each one

The plot for ‘What is each player’s rating in each of the physical statistics (shooting, agility, composure, et cetera)’.

* The value range for each control and whether or not it is loaded from a certain attribute in the data.

List of all the names of the players. (Single select list). Loaded from Names.

Top N Filter:

* List what they will be used for

Used to select Top N players to be displayed in the visualization.

* Which plots are connected to each one

The plot for ‘Top N Best Player for each position on the field’

* The value range for each control and whether or not it is loaded from a certain attribute in the data.

Value range : 1 - 10,000