Visualization Assignment 1

Dataset: FIFA 20 complete player dataset

https://www.kaggle.com/stefanoleone992/fifa-20-complete-player-dataset?select=players_20.csv

About Data:

The datasets provided include the players data for the Career Mode from FIFA 15 to FIFA 20.This dataset contains 18,278 rows where each row represents a player's name. Each player is represented with 104 attributes. The dataset contains many null(NA) values and also contains attributes which were redundant, eg. date of birth and age. The overall dataset is good enough to compare all aspects of players.

Data preprocessing:

The dataset contains many attributes by which we can compare stats of a player across the seasons. As most player's attributes are evaluated from their real life counterparts we can also create some imaginary combination of players to check how those imaginary teams might perform in real life or how much it will cost to create such teams. For easier comparison from visualization I have taken following steps during preprocessing.

Steps of preprocessing:

1. Drop all players whose country isn't in top 7 ranking. (Ref. https://www.fifa.com/fifa-world-ranking/).
2. Drop all columns containing more than 5100 NA values.
3. Drop rows containing null values. (Because of the massive number of NA values, deleting rows first was resulting in deletion of complete dataset.)
4. Select 16 attributes which can provide more generalised insight about the players.
5. Rename some attributes.
6. Select top 700 rows which are sorted based on overall performance in descending order.

All selected attributes:

1. **Age (Numerical):**
   Age of the player till that season of FIFA.
2. **Height cm (Numerical):**
   Height of the player in centimeters.
3. **Weight kg (Numerical):**
   Weight of the player in kilogram.
4. **Nationality (Categorical):**
   Country to which the player belongs to.
5. **Overall (Numerical):**
   Over all performance of the player in all
6. **Value EUR (Numerical):**
   Net worth of the player.
7. **Wage EUR (Numerical):**
8. **Preferred foot (Categorical):**
   The left or right foot is preferred by the player.
9. **International reputation (Categorical/Numerical):**
   International reputation represented as a number. The reputation of a player is defined by level 1 to 5. Five representing the higher level of reputation.
10. **Team position (Categorical):**
    Position of the player on the game field.
    Significance of short forms: R: Right, L: Left, C: Center, M: Middle, A: Attack, W:Wing, ST: Striker, B: Back, D: Defence, F: Forward etc.
11. **Pace (Numerical):**
    Speed at which the player can run.
12. **Shooting (Numerical):**
    Shooting tells about the shooting level with the highest number as 100.
13. **Dribbling (Numerical):**
    Dribbling tells about the dribbling level with the highest number as 100.
14. **Defending (Numerical):**
    Defending tells about the defending level with the highest number as 100.
15. **Physical (Numerical):**
    Physical tells about the physical level with the highest number as 100.

16. **Skill moves (Numeric/categorical):**
    This attribute tells about the number of skills of a player. The skill count goes from 2 to 5, so it's more intuitive to represent it as a categorical variable.

Why this dataset:

I am a football lover and I have played FIFA20 or previous versions many times. The player attributes are selected very carefully to represent the real life player in most cases. The visualised dataset doesn't contain the names of players but just best on the outliers in can tell which players those points belong to for special players like Lionel Messi. The dataset also provided a wide range of attributes with information on many players which provided a lot of information for visualisation purposes.

Video link:

https://youtu.be/5uqhE8RKcbc