Injury Insights: A Data Visualization Tool for Understanding and Preventing Sports Injuries

Goutham Kalikrishna Reddy Kuncham, Rahul Vaidya, and Sowbaranika Balasubramaniam

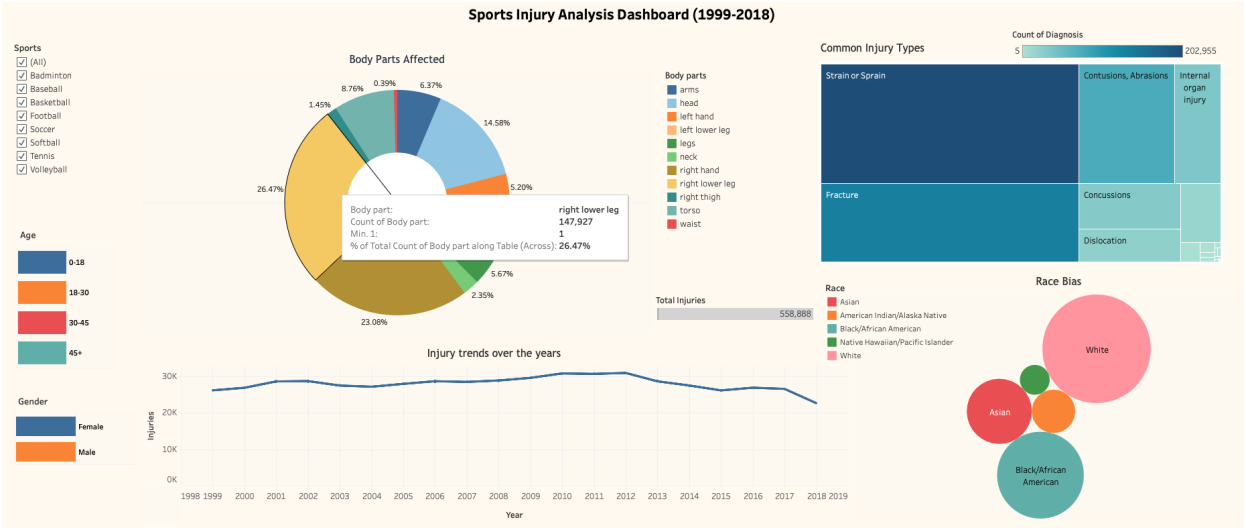


Fig. 1: Common sports injuries

Abstract— Our project's goal is to create a sports injury visualization tool that allows users to interact and visualize different sports related injuries on a human anatomy plot. We intend to use python for project development . Alpha, beta, and final iterations of the project will all include incrementally more usable and functioning features. Our project's overall goal is to give athletes, coaches, and medical professionals a useful tool to help them make better decisions about injury prevention, diagnosis, and treatment in sports.

Index Terms—Sports Injuries, Data Visualization, Human Anatomy, Data Analysis

1 Motivation

According to statistics from the Centers for Disease Control and Prevention (CDC), sports injuries significantly contribute to millions of emergency department visits each year in the United States alone, which is a big worry for athletes, coaches, and medical professionals (CDC). A complete understanding of injury incidence patterns and related risks across diverse forms is required for successful injury prevention.

Raising awareness about vulnerable areas of the human body during sport activities and recommending precautionary steps beforehand to avoid harm, constitutes the main goal of this project. It aims to equip athletes, trainers, and health care specialists alike with knowledge necessary to make crucial decisions when it comes down to preventing injuries from ever happening in the first place can promote one's quality of life markedly better. Adopting these practical approaches can bring along drastic improvements for an athlete in terms of their health conditions which could concurrently limit both likelihoods and severities pertaining to various types of sport-related traumas.

2 Approach

In order to visualize the injury patterns for different types of sports, the first step is to collect data from publicly available dataset. Once the data is collected, the next step is to preprocess the pertinent information and patterns of injury occurrence are analyzed. Based on the frequency and locations of injury occurrence, an interactive visualization will be designed using Python libraries to create 2D plots that resemble human anatomy. Users will be able to browse injury data, filter by sport, gender, and age-group, and see injury patterns over time with the dashboard that is provided once the visualizations have been created.

The approach to solve this problem requires data analyzer and a data visualization expert. Our team will work together and make sure that the data is precisely analyzed, and the visualization are intuitive as well as interactive.

3 Authors Roles and Responsibilities

1. Sowbaranika Balasubramaniam (balasubramaniam.29) - will be primarily working on analyzing the data and working along with Rahul in finalizing the design.
2. Rahul Vaidya (vaidya.84) - will be mainly working on preprocessing the data and working along with Sowbaranika in finalizing the design.
3. Goutham Kalikrishna Reddy Kuncham (kuncham.2) - will be implementing the final data visualization.

4 Dataset

The dataset for this project is obtained on the U.S. National Electronic Injury Surveillance System (NEISS) Data on injuries, 1999-2018. It consists of 7,352,939 rows and 31 columns.

5 Initial Questions

1. What are some of the most common injuries in the dataset?
2. Which sport causes the most/least number of injuries?
3. In what ways have injury rates changed over the years?
4. Are certain age groups or genders more prone to certain types of injuries?
5. What are the most common body parts affected by injuries?

6 Deliverables

1. Alpha stage design and capabilities:
   - Data preprocessing and exploratory data analysis, highlighting data trends across various sporting activities.
   - Initial visualizations in the form of human anatomy plot for one sport and gender.
   - Initial understanding of the dataset in terms of injury occurrences by sport, gender, age.
2. Beta stage design and capabilities:
   - Data analysis - exploring trends in data and identifying interesting patterns/trends across different sports, based on age group and gender.
   - Visualizations of injuries on the human anatomy plot by gender, sport, and age group.
   - Ability to choose a particular sport and filter out by gender and age from age group from the drop-down list.
3. Final stage and capabilities:
   - Interactive dashboard that allows users to interact with the plot and filter the data based on type of sport, gender, age group, and year.
   - Recommendations and strategies to prevent injuries.
   - A final project report that summarizes key findings/contributions.

7 Evaluating the Usefulness

1. User Feedback and Comparative Analysis: We will send a feedback form to the users to evaluate the usefulness and effectiveness of the project
2. Comparative analysis: We intend to evaluate the project by performing a comparative analysis of similar projects.

8 Planned Extensions

1. Trend Analysis: This would involve identifying trends in injury patterns over time, such as seasonal patterns, and analyzing the factors that contribute to these trends.
2. Advanced Visualizations: This extension would involve building more advanced visualizations, such as 3D models of the human body, to provide more detailed insights into injury patterns and location.

9 Annotated Bibliography

1. Paper [1]: The authors of the paper emphasize the importance of visualizing complex and large data sets and the significance of selecting appropriate visualization tools. Some of the techniques for visual data mining, such as scatter plots, heat maps, and tree maps are briefly explained. The paper includes several case studies that show how information visualization and visual data mining can be used in drug design, such as identifying potential drug targets and analyzing molecular structures. Overall, the paper provides a useful overview of the application of information visualization and visual data mining in drug design, highlighting the potential benefits of these techniques for advancing drug discovery and development. Similarly, our project provides an interactive design to understand the major areas of injury caused by different sports based on age and gender, which can be used for injury prevention and care in the future.
2. Paper [2]: The paper studied the incidence, severity, aetiology, and prevention of sports injuries by reviewing existing literature. The authors discovered that sports injuries were widespread and varied across activities and age groups. They also discovered that some sports had higher rates of severe injuries than others. They also reviewed several preventative techniques, such as adequate warm-up and conditioning routines, safety equipment, and rule amendments. The authors concluded that a holistic approach incorporating education, training, and injury prevention techniques can lower the frequency and severity of sports injuries.
3. Paper [3]: In the blog, the author performs exploratory data analysis and emphasizes the need for tailoring the presentation of injury data to the intended audience, highlighting the need for different approaches when communicating with athletes, coaches, or policymakers. It also summarizes a simple presentation of the sports injury data based on age, severity during training as well as competition. Similarly, our project provides similar data analysis but our final design involves visualization over a human anatomy plot to better understand the results.

10 Initial Sketch

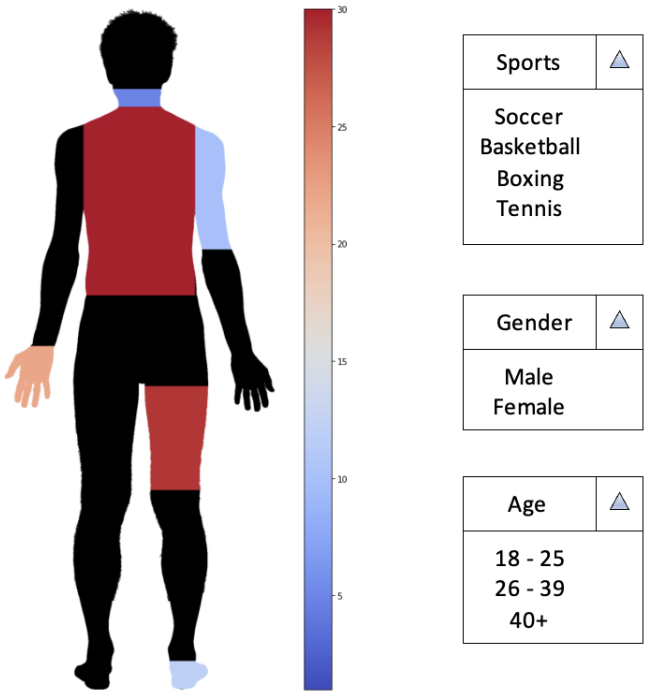


Fig. 2: Human Anatomy plot highlighting sports injuries

11 IMPLEMENTATIONS

11.1 Data Preprocessing:

For our sports injuries data visualization project, we performed several data preprocessing and transformation steps to clean and organize the data. Firstly, we loaded the original data from NEISS, and then removed any extra columns that were not relevant to our analysis. Next, we filtered the sports column to focus only on the sports of interest, while also updating the column value by removing unwanted characters. We then retrieved the unique values in the sports name column and dropped any sports fields that weren't necessary for our analysis. Additionally, we dropped any body parts rows that weren't relevant. To better categorize age data, we created labels for age categories, and then we assigned age values to these categories. By performing these steps, we were able to clean and transform the data to better analyze and visualize sports injuries data.

Column Name	Used for
Treatment Date	Used for Trend analysis (Line Chart)
Age (in years)	Used for filter (Dashboard)
Sex	Used for filter (Dashboard)
Body Part	To visualize body parts injured (Dount chart)
Sports	Used for filter (Dashboard)
Race	Used for Raise bias (Bubble Chart)
Diagnosis	To find common injury diagnosis (Tree Map)
Year	Used for Trend analysis (Line Chart)
Month	Used for Trend analysis (Line Chart)

11.2 Visualizations:

Our project's implementation involves the use of a variety of tools and methodologies to produce an in-depth and informative set of visualizations that can help people who work in the sports industry prevent and manage injuries. We performed data cleaning and processing tasks using Python, and then we used Tableau to generate a variety of visualizations that provide an interactive and in-depth view of the data. We created a wide range of charts, including a donut chart to depict injured body sections, a tree map to illustrate the most prevalent sports injuries, a bubble chart to illustrate racial bias, and a line chart to depict the evolution of injuries over time and a human body heatmap to provide a more comprehensive and interesting view of injury locations on the body. To convey the insights from the data effectively, we employed various encoding techniques such as color, size, shape, and position in our visualizations. We also created a dashboard that integrated all the visualizations, except for the human body heatmap, and incorporated filters for users to interact with the data. This interactive feature provided stakeholders with a more meaningful way to explore injury patterns based on factors such as sport, gender, and age, allowing them to extract more specific insights from the data.

11.2.1 Donut Chart

We used a donut chart [Fig.4] to represent the distribution of injured body parts based on factors such as sport, gender, and age. The chart provided an intuitive visualization of the proportion of injuries by body part for different subgroups of the population. The donut chart was particularly useful in highlighting the most common body parts affected by injuries for different groups, such as male or female athletes, or for different age groups. By using a donut chart instead of a pie chart, we were able to make the visualizations more effective by including an inner circle that displayed additional information, such as the percentage of injuries that were severe or moderate. The donut chart helped to provide a more comprehensive view of the data and facilitated a better understanding of injury patterns across different subgroups of the population.

11.2.2 Treemap

Treemap chart [Fig. 5] was utilized in our sports injuries data visualization project to display the most common diagnosis across different sports. We used Tableau to create this visualization, which provided a clear and concise view of the most common types of injuries in a hierarchical view. The treemap was designed in such a way that the size of the rectangles represented the frequency of the injuries, and the color-coding was used to differentiate between different types of injuries. To make the treemap chart more informative and effective, we used different encoding techniques such as color and size. For example, we used color to indicate different types of injuries, and size to represent the frequency of injuries

11.2.3 Bubble Chart

We utilized a bubble chart [Fig. 6] in our research project to visualize race bias in sports injuries. The chart displayed the number of injuries over time, with the bubbles representing the different races of the athletes. The size of the bubbles indicated the number of injuries, and the color was used to distinguish between different races.

This visualization allowed us to compare injury trends across different races and identify any disparities or trends that may be present. By using a bubble chart, we were able to convey the magnitude of injuries suffered by athletes of different races and identify any patterns that may exist in the data.

11.2.4 Human Body Heatmap

We used a human body heatmap as a novel visualization technique to better understand the locations and severity of sports injuries. The heatmap allowed us to visualize injury patterns across different body parts and to identify the severity of those injuries. The heatmap was color-coded to indicate the level of injury severity, with red indicating the most severe injuries.

11.3 Filters

In order to make the visualizations more interactive and user-friendly, we incorporated three different filters in our dashboard. These filters were designed to allow the users to dynamically visualize the data and explore the injury patterns based on various factors such as sports, age, and gender.

The first filter, "Sports"[Fig.8.a], allowed the users to select specific sports and view the injury patterns specific to that sport. This feature was particularly useful for coaches and athletes, as it helped them identify the most common types of injuries in their respective sports and take necessary precautions to prevent them.

The second filter, "Age" [Fig. 8.b], allowed users to filter the data based on different age groups. This feature helped to identify the injury patterns in different age groups and determine if there were any age-related factors that were contributing to the injuries.

The third filter, "Gender" [Fig. 9.c], allowed users to view injury patterns specific to males and females. This feature was important for medical professionals, as it helped them identify any gender-related factors that may be contributing to the injuries. By incorporating these filters, our dashboard became a comprehensive tool for exploring the injury patterns and trends

11.4 Results

The pie chart visualization showed that the most commonly injured body parts were the knee, ankle, and head/face. These injuries occurred across a wide range of sports and affected both male and female athletes. The treemap visualization provided insights into the most common types of injuries, with the top three being sprains/strains, contusions, and fractures. The bubble chart allowed us to identify race bias in sports injuries, where certain races were more prone to injuries than others.

The line chart visualization revealed an overall trend of increasing sports injuries over time, particularly for injuries such as concussions and knee injuries. The human body heatmap provided a comprehensive view of injury locations on the body, showing that the most common locations for injuries were the knee, ankle, and head/face.

The dashboard, which included all visualizations except for the human body heatmap, allowed users to dynamically interact with the data based on factors such as sport, age, and gender. Through this dashboard, users could identify injury patterns across different sports and demographics, and potentially use this information to inform injury prevention strategies.

Fig. 3: Line chart visualizing trend of injuries for years 1998-2019

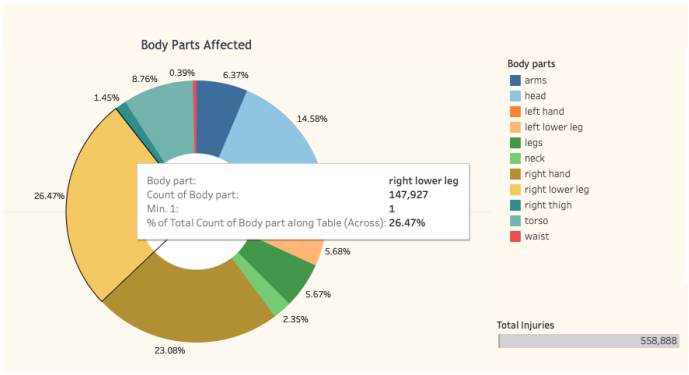


Fig. 4: Donut chart highlighting body parts affected

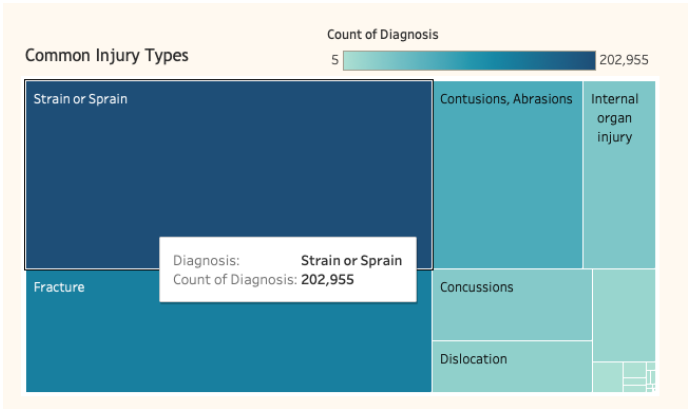


Fig. 5: Treemap highlighting most common injury diagnosis

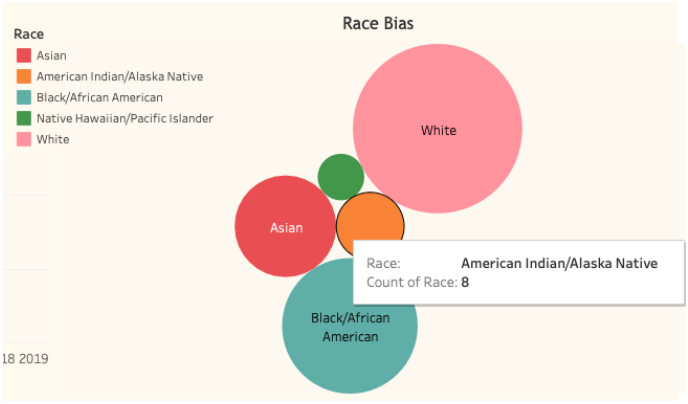


Fig. 6: Bubble chart visualizing Race Bias

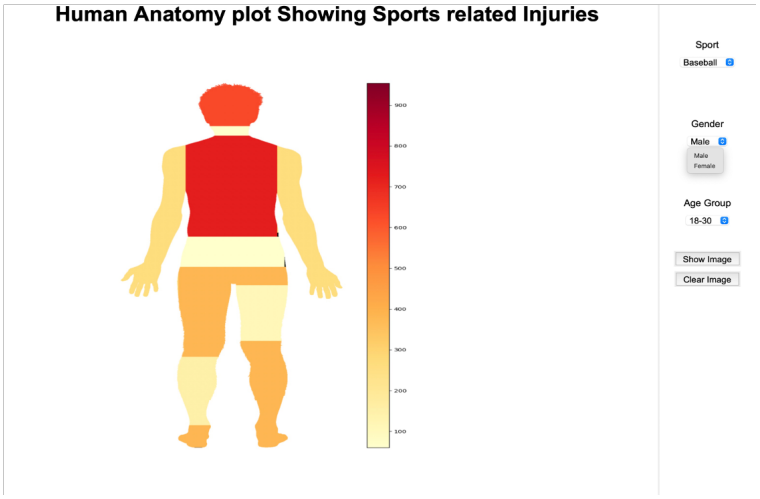
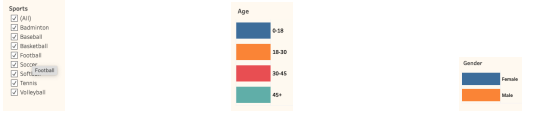


Fig. 7: Human Body Heatmap



(a) Sports filter (b) Age filter (c) Gender Filter

Fig. 8: Filters

It is important to note that these visualizations were created using publicly available data, and caution should be exercised when drawing conclusions based solely on these visualizations. Future studies should aim to incorporate additional data sources and include input from athletes and medical professionals to provide a more comprehensive analysis of sports injuries.

11.5 Future Work

Future work could explore the integration of machine learning algorithms into the data visualization system to provide more accurate injury risk assessments and personalized injury prevention recommendations. Additionally, the system could be expanded to include more advanced features, such as real-time injury monitoring and the integration of wearable devices for collecting sports injuries data.

REFERENCES

[1] B. Brewer and T. Van Brewer, "Information visualization, visual data mining, and its application to drug design." J Funct Morphol Kinesiol, December 2020. 2

[2] W. van Mechelen, H. Hlobil, and H. Kemper, "Incidence, severity, aetiology and prevention of sports injuries," *Sports Medicine*, vol. 14, 1992. 2

[3] P. Edouard, "Keep it simple: how to best present the extent of a sports injury problem." https://blogs.bmj.com/bjsm/2019/12/19/keep-it-simple-how-to-best-present-the-extent-of-a-sports-injury-problem/, 2019. Accessed: April 20, 2023. 2