Inaccurate and Misinformative Data Visualisations in Sports-Field

Sneha S Prabhu

Student, Computer Science and Business Systems, Dayananda Sagar College of Engineering

Abstract: *Data is vital in modern sports, it is the lifeblood because it determines everything from player performance analysis to game strategy. Data is helping in optimizing player rotations, enhancing fan engagement, it basically is reshaping the sports field, making it more precise and strategic. This prominence of data in the sports field has lead to dishonest data visualisations to captivate the audience. Unveiling the impressive yet fraudulent data visualisations in sports field that distort statistics, alter narratives, and mislead the audience. By inspecting sports-related visualizations, the study points out common techniques used to manipulate figures, twist numbers and rewrite tails. This work contributes to understanding the techniques used to create manipulative data visualisations across all fields, not just confined to the sports field.*

Keywords: Data visualization, Misinformative data, Sports

1. Introduction

"Sports " plays a prominent role in fostering the physical health and mental well-being of an individual. It provides a platform to develop discipline. It also nurtures leadership, time management and strategic thinking. It also imparts a sense of responsibility. Due to these benefits it has gained huge value commercially. The enthusiasm for sports has prevailed from a long time. With this increasing popularity it has proved to be a source of revenue. The league- based events have contributed to the growth of the economy. Firms and brands around the world invest huge amount of money in high performing athletes and top teams. Global sponsorship for sports players has increased rapidly over the last years.

Cricket is a popular sport in our country and is deeply embedded in the national culture. It boosts revenue to the economy in various ways like ticket sales, merchandise, sponsorship deals and broadcasting rights. Substantial revenue is generated through cricket matches especially international and premier league games. It has also paved a path in giving global recognition to the top performing players of the nation.

Successful cricketers serve as role models and motivate young individuals. It remains as an integral part of the country.

Data is crucial to this industry, affecting several aspects like player performance to fan engagement. Data analytics keeps a track every aspect of an athlete's performance. Coaches analyse data to prepare game strategies while it also enhances fan engagement. Data visualization is a prominent method that effectively communicates information to those who lack expertise. It helps in making informed decisions, targets marketing efforts and secures sponsorship deals. Visualisation has made complex data accessible and actionable. The influence and impact it has, has lead to misinformative data visualisations. Often cherry- picking of data and poor narratives have been used to fool audience and increase the popularity of sports. This paper aims to recognize these techniques used in creating misinformative data visualisations.

Mainly focusing on how poor Data Visualisation techniques are used in Cricket to mislead the viewers.

The research has done the following:
- Collects Cricket Data Visualisation from online sources.
- Categorizes these data visualisations using misinformative visualisation techniques.

2. Methodology

The main purpose of finding misinformative data visualisations in the field of sports is to prevent misleading conclusions. It is important that we avoid the spread of misinformation because the media may lose its credibility. Also, accurate performance metrics is required for real-time insights.

This paper mainly focuses on the misinformative data that is being spread about Indian Premier League (IPL) i.e, which is a franchise based game popular in India. IPL is watched by more than 35 crores of people across India. It is expected to generate approximately Rs 48,390 crores between 2023 to 2027. The Fantasy cricket players rely heavily on data visualisations to select teams, keep track of the top performing players and make informed decisions. But, often this tool is misused to give faulty insights and misleading audience to maintain fan engagement and the popularity of the league. Therefore, it is very important that we uncover the truth these poor data visuals that are misleading fans.

Figure 1: The logo of IPL

The prerequisites required for the methodology includes

Cricket Media Landscape. Research method includes Data collection, Data Tagging, Data Analysis, Categorizing and Evaluation.

1) Cricket Media Landscape

Cricket data is generated from various sources such as,

a) Broadcasting Networks

Draws conclusion for future misinformative data visualisations. Major Sports Television channels like ESPN , Star Sports and others broadcast live matches and analysis.

b) Online Platforms

There are streaming services like Hotstar, Jio Cinema, ESPN+ and also there are official websites of leagues and cricket boards.

c) Social Media

Various platforms like Twitter, Facebook, Youtube and Instagram supply user-generated content.

d) Mobile Apps

There are various apps like Cricbuzz, ESPN Cricinfo and BBC Sport which provides in-depth and exclusive content also provides comprehensive coverage and live updates.

These are some of the ways in which viewers receive the enormous data about cricket. There are 10 franchises that participate in the league. The major franchises with huge fanbases are Royal Challengers Bangalore (RCB) , Chennai Super Kings(CSK) and Mumbai Indians(MI). These franchises are enormously popular. Thus, people are creating misleading data visuals just to conserve the admirations and demand it gets in the country.
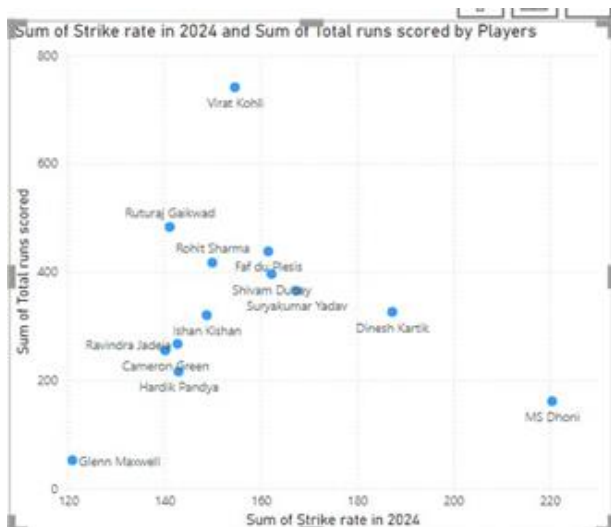

Figure 2: Scatter plot of the top performing players in 2024

3. Research Techniques

This repository mainly aims to identify the misleading data visualizations in IPL . It serves as a collection point, sourced

from various sports-related subreddits.,
https://github.com/snehaaprabhu/Research/tree/ba2ae5954bc945df1c9ae39a75b807df61e918fa

3.1 Data Collection

The data visualisation samples were collected from various sources like Cricbuzz, IPL Official Page, Kaggle datasets and from Reddit.

Data is collected from cricket-related subreddits.
- /r/Cricket
- /r/IPL
- /r/Indian Premier League
- /r/ICC

And ten of these subreddits focus on a specific franchise:
- /r/Royal Challengers Bangalore
- /r/Mumbai Indians
- /r/Chennai Super Kings
- /r/Lucknow Super Giants
- /r/Sunrisers Hyderabad
- /r/Gujarat Titans
- /r/Rajasthan Royals
- /r/Delhi Capitals
- /r/Kings XI Punjab
- /r/Kolkata Knight Riders

I selected three team subreddits based on the popularity and the most activity based fanbase. For example, I assumed Royal Challengers Bangalore, Mumbai Indians, Chennai Super Kings are the most popular teams in in IPL.

This collection process was automated through (PRAW) Python Reddit API Wrapper. PRAW is a library that allows to collect recent posts with the Reddit API using different sort orders. In total, I collected 521 images from these subreddits.

3.2 Data Tagging

I manually filtered the collected data and tagged the images after the process of collection of data. First, I discarded all those images that didn't appear to be data visualisations. After this, there were 62 images that identified as data visualisations. These were imported to TagSpaces, a cross-platform that allows to apply tags and finding files on cloud as well as local devices. In total, 5 unique tags were were applied to the corpus.

3.3 Data Analysis

I manually analysed each tagged data visualisation to classify them based on the taxonomy. The classification of the misinformative data included Cherry picking data, Truncated Y-axis, Omitting of data, Misleading pie charts, Omitting of the labels for x-axes and y-axes and also using the wrong chart type.

These fragmented narratives destroy the biggest components of a successful insight. These lead to poor decision-making.
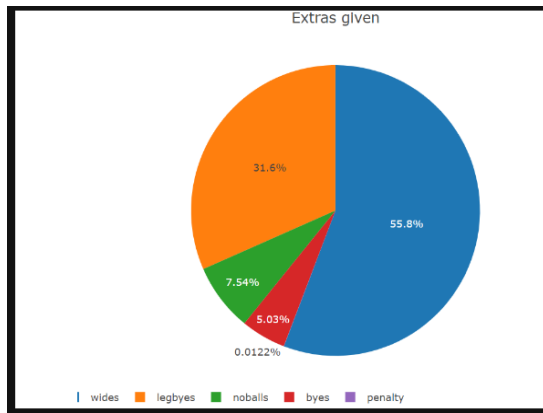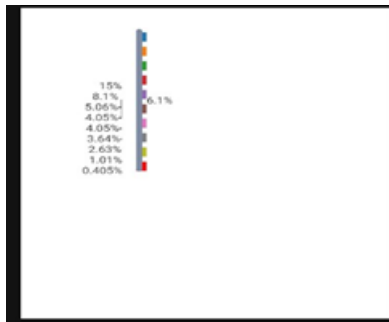
Figure 3: Pie Chart with no Heading

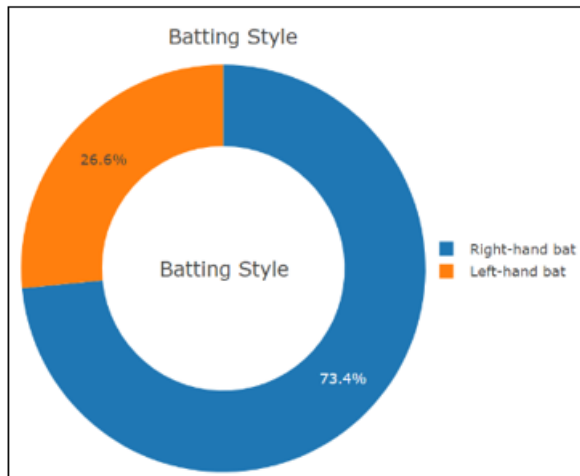**Figure 4:** Bar Chart where no axes are defined

Figure 5: Donut chart where the player's name is missing.

4. Results

4.1 Impact

From the 521images corpus, it was narrowed down to 62 IPL data visualisations that can be considered inaccurate and misinformative.

These were some of the examples where I found that there were inaccurate data visualisations that were available to public.

The Examples are given in the 2.3 Data Analysis section of the paper.

The taxonomy found are as follows:

Table 1: The taxonomy on which the data visuals were divided.

Type of Chart	Type of Inaccuracy	Count
Pie Chart	Missing Titles	19
Bar Chart	Missing Titles	10
Bar Chart	Truncated Axes	3
Pie Chart	Inappropriate Chart Type	2
Bar Chart	Incomplete Chart	4
Pie chart	Incomplete Chart	8
Donut Chart	Missing Titles	4
Table	Cluttering	1
Scatter plot	Inappropriate Chart type	2
Bar Charts	Missing Labels	9

There were majorly 6 taxonomies based on which I classified the inaccurate data visualisations that were collected.

The categories that were considered were Missing Titles, Truncated Axes, Inappropriate Chart Types, Incomplete Charts, Cluttering and Missing Labels. These kinds of inaccurate data when consumed by the public and the coaches end up making poor decisions. Since IPL, has been a major source of revenue in the country, it is important that people should not be fed with inaccurate data.

3.2 Evaluation

The data collected was relatively small in this work, as it was done over a period of one week. But I believe its results are quite solid. The strong point being that there aren't any works done on this specific field and on this specific topic making it unique and one of the first in its fields.

This work, however, could improve as:
- The data was collected only from Subreddits from the Reddit API, Kaggle datasets and Cricbuzz official, the corpus would be more complete if the data would have been collected from a wider range of variety like Social Media Platforms such as Twitter, Instagram and other websites.
- Also, the data was only collected for a period of one week. It would be more efficient if it was collected for over an entire year.

5. Conclusion

4.1 Impact

This work on the intersection of Cricket (Sports media) and misinformative/inaccurate data visualisations is a rather new topic which has not been explored much. It uncovers how the powerful tool such as data visualisation has been misused to alter people's perspectives. If the taxonomies are followed appropriately and we get rid of all the inaccurate and misinformative data, it would lead to improved decision-making by the coaches and the viewers.

4.2 Future Work

There are various potential approaches to build on the results of this work. First, would be collecting a more complete corpus may reveal more opportunities to contribute to misinformative and inaccurate taxonomies. Second, would be to collect data over a longer period of time to uncover more accurate results. Third, would be create a more accurate and automated technique to filter out the collected data instead of manually filtering and tagging them.

References

[1] Drew Scott, "Misinformative Data Visualisations in the Sports Media Domain", April 03,2004.

[2] Harri Jalonen, Sasu Touminen, Arto Ryoma, Jaakko Haltia, Juho Nenonen, Anna Kuikka, "How does Value Creation Manifest Itself in the Nexus of Sport and Business ?: A systematic literature review", 2018.

[3] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini and Huamin Qu, "Misinformed by Visualization: What Do We Learn From Misinformative Visualizations?", Volume 41, 2022.

[4] M. Lisnic, C. Polychronis, A. Lex, and M. Kogan. Misleading Beyond Visual Tricks: How People Actually Lie with Charts. preprint, Open Science Framework, Oct. 2022.

[5] CHEN M., EDWARDS D. J.: "isms" in visualization. In Foundations of Data Visualization. Springer, 2020

[6] KLASS G.: Just plain data analysis: Common statistical fallacies in analyses of social indicator data. Statlit. org (2008).

[7] SZAFIR D. A.: The good, the bad, and the biased: Five ways visualizations can mislead (and how to fix them). Interactions 25, 4 (2018), 26–33.

[8] BRESCIANI S., EPPLER M. J.: The pitfalls of visual representations: A review and classification of common errors made while designing and interpreting visualizations. Sage Open 5, 4 (2015), 2158244015611451.

[9] LINKS:
[1]https://www.kaggle.com/code/rishpande/indian-premier-league-ipl-data-visualization