

Data Visualization Design of Sports Journalism Based on Narrative Perspective

A Case Study: Sigma Awards (2020-2021)

Pei Liu^{1*}, Qi Wang²
¹*liupeipei@mail.gdufs.edu.cn
²queeniewangkei@foxmail.com

¹School of Journalism and Communication, Guangdong University of Foreign Studies, Guangzhou, China

²School of Journalism and Communication, Guangdong University of Foreign Studies, Guangzhou, China

Abstract. With the development of big data, data visualization makes data easily understood. Data visualization has become an essential presentation of sports journalism. In this paper, we present a survey of the 18 listed projects concerning sports in the Sigma Awards. We have analyzed the types and sources of data, application areas and visualization techniques. Through content analysis, trends in the use of data and visualization techniques in sports journalism are shown: the degree of openness of data sources is high; the subject categories are relatively diverse; the visualization techniques are various.

Keywords: Visualization; Data Storytelling; Sports Journalism; Sigma Awards

1 Introduction

Under the background of big data, the forms and narratives of sports journalism have undergone profound changes. Data journalism and visualization have gradually become the mainstream of sports journalism. Data journalism can be defined as data-driven journalism, which adds value by using data-sets or databanks. In that process, the data used become part of the journalistic product and do not just serve research, and so we also talk about “data storytelling”[22]. Because the data is intensifying and various, it cannot be directly presented to readers. Hence, with the help of visualization techniques, the data and text can be transformed into a visual presentation, which makes stories more suitable for readers. Therefore, data journalism and visualization are two areas developing strongly in sports journalism [22].

This paper will take the listed projects concerning sports in Sigma Awards from 2020 to 2021 as an example. The Sigma Awards is a competition to celebrate the best data journalism from around the world. It's also here to empower, elevate and enlighten the global community of data journalists. Therefore, this paper based on in-depth analysis and content analysis of the listed project not only can enrich the development of relevant theories, but also help data visualization to play a more important role in the practice of news reporting and storytelling.

Literature review, the types, and sources of data as well as how visualization techniques assist storytelling will be respectively discussed in the following chapters.

2 Theoretical perspectives

News narrative refers to "how news is expressed" and "how to make news", which is research on the technical level of news narrative [3]. In traditional news, news narrators often use specific rhetorical methods to process news facts to tell news stories better. While the narrative of data journalism is based on the analysis and induction of data to obtain news clues. News stories will be reconstructed after that. This process achieves the transformation and innovation of discourse expression.

Sports data journalism utilizes big data and visualization techniques to elaborate sports stories, which has become a common way of sports news reporting in major media worldwide. It combines the value of intensifying sports data, excellent and rooted narrative ability, and visualization technologies, creating a new pattern of sports journalism.

Recent researches related to sports journalism in the digital environment have focused on: (1) the challenges faced by sports journalists. Horky presents a survey of the current potential of data journalism in sports. He has analyzed the challenges of assessing and publishing data in professional online sports journalists via the example of football in Germany [22]. (2) The rise of social media and its impact on sports journalism. Nölleke combines an online survey among 122 German sports journalists, an analysis of the Twitter networks of German sports journalists during the Winter Olympics 2014, and a content analysis of the most popular news items on social media. The results indicate that sports journalism and social media indeed maintain a complementary relation [8]. (3) visualization used in sports journalism. Fu summarizes the new trends of visualization in sports data journalism: data are presented in more diverse ways; data charts occupy the main body of content of data journalism [25].

Some research focus on the narrative of data journalism. For example, Sun analyses the narrative strategies of map visualization based on The Paper. she concludes that map visualization of data journalism will be more innovative, immersive, and researchable in the future [24]. However, the narrative of sports data journalism is few. Most concentrates on the news narrative of visualization applied in world competitions, like FIFA World Cup, Winter Olympics, etc.

3 Methods and materials

We searched the official website of Sigma Awards and manually selected 18 projects. The selection criteria are: (1) The projects are about sports; (2) The content of the projects must be visualized; (3) The projects can be accessible and read. Then we number the projects. See **Table 1** below.

Table 1. Serial numbers of the projects

Number	Project title	Number	Project title
1	The Rise of Hyphenated Last	10	The State of Football Fi-

	Names in Pro Sports[12]		nance[20]
2	Spin control[14]	11	Excessive Throwing: How Many Pitches of Japanese High School Baseball Players Violate Pitch Count Guidelines[13]
3	Simple, natural, no fuss: The Food that fuels Olympic marathon champions	12	How many high school stars make it in the NBA?[11]
4	If all the Olympic champions compete together, who will be the final winner? [10]	13	Swimming abuse[4]
5	20 years Roger Federer from his first ATP Tour title[16]	14	Youth Centers Drop Women Athletes from their List of Concerns by Hager Hisham[2]
6	EURO 2020: Data Driven Sports Reporting[7]	15	Rafa Nadal's career in detail[18]
7	Hot and humid Olympic summer[21]	16	Rentals of the Spanish football stadiums[9]
8	Visualizing the FIFA Arab Cup 2021[1]	17	20 Grand Slam of Novak Djokovic[15]
9	Messi's 643 goals for the same club[17]	18	Data-driven Tokyo 2020 Olympics coverage[6]

We divided the data of the projects into three types: Image & video data, Tabular & text data as well as others. Image & video data refers to the photos and videos which contain numerous and related information from two-dimensional to three-dimensional. Tabular & text data includes structured data, unstructured data, and semi-structured data. Most of them belong to two-dimensional data. Structured data is stored in the database and can be used without cleaning and crawling. While unstructured data needs to be refined and organized. Semi-structured data is between structured and unstructured. It is stored in database but needs to be cleaned and selected with the help of Python, java, etc. Some projects do not disclose the types of data they get, so we classify them into others.

What's more, we divide the sources of data into three ways: open data, collected by themselves and special data providers.

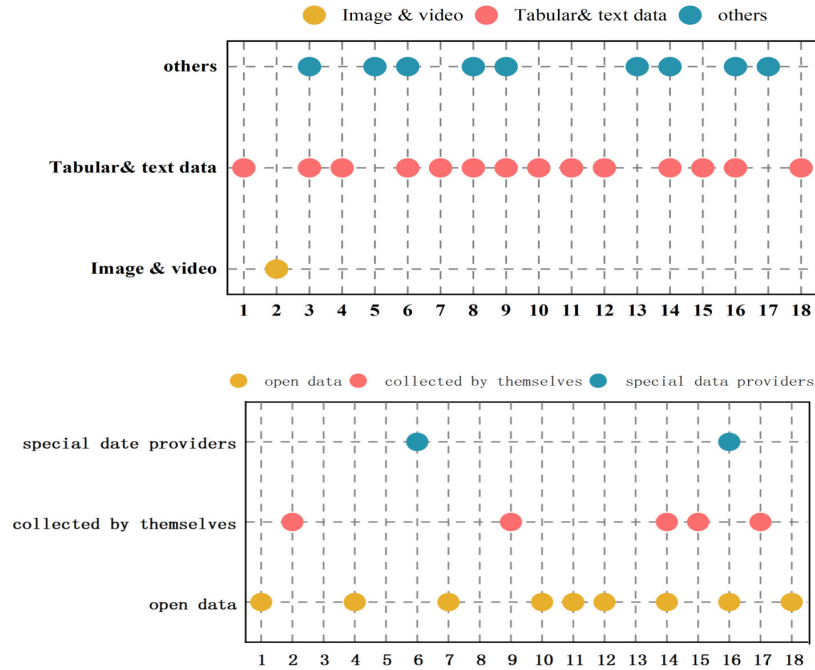


Fig. 1. The distribution of the types of data and the data sources

It can be easily seen that tabular & text data makes up nearly 77.8%. Among them, a large proportion of tabular & text data is obtained from open websites and institutions, which attributes to the openness, authority, and authenticity of the projects. We regard this source of data as open data, accounting for 50%(Fig. 1). If All the Olympic Champions Compete Together, Who Will Be the Final Winner? [10] published by The Paper is a typical example. The organization collects its data from Olympedia, IOC, and Wikipedia to present the championship results of 104 events in three major categories: track and field, swimming, and shooting at the Olympic Games from 1896 to 2016. But for different sports, the data format varies. So, they must use Python to crawl the data in each webpage of 104 sports and clean the data afterward. How Many High School Stars Make It in the NBA? is produced by the Pudding. The data they get is scraped from Basketball Reference using node.js. The front-end is built using JavaScript, D3.js, and Canvas.

With the absence of open data, the authors and organizations need to collect data by themselves, sometimes they would seek help for special data providers. The former takes up around 33.3%, the latter occupies 11.1%(Fig. 1). Swimming Abuse [4] originated from CNN BRASIL shows the report of eight former athletes who say they are touched in the private parts and harassed by the coach of swimming and the president of the Federation of Aquatic Sports of Amazonas. Youth Centers Drop Women Athletes from their List of Concerns [2] invented by ARIJ Network used the OCR application and data scraping programs from PDF files like Tabula to scrape sports activity data related to youth centers for a period of 10 years. Moreover, they manage to speak with several people related to the story and have several field

visits, like talking with affected women, experts, and government officials, which helpfully proves that women are discriminated against and neglected by not having a suitable environment for them to practice sports in youth centers.

4 Application areas

The data come from numerous sources and are mainly applied into two categories. one is professional sports information, the other is sociological, cultural, and economic issues. Each share fifty to fifty(**Fig. 2**).

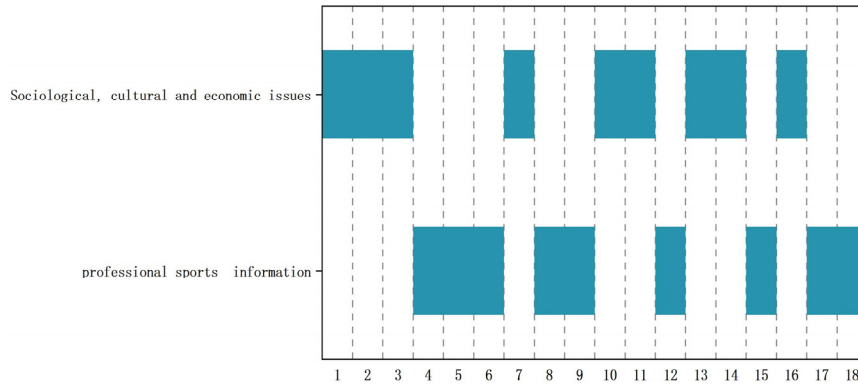


Fig. 2. The application areas of the data

4.1 Professional sports information

With the help of comprehensive data, the project could clearly provide professional sports information to readers. The project, 20 Years Roger Federer from His First ATP Tour Title [16] provides detailed information about Roger Federer's historical classification in the ATP ranking and all the Grand Slams he achieved. Spanish organization Mundo Deportivo counted and analyzed the number of goals scored, goal position, goal posture, goal time and so on in the project Messi's 643 Goals for the Same Club [17], which allows readers to fully understand Messi's competitive strength and legendary experience from a macroscopic perspective.

Comparative analysis of historical data of important sports events, weather conditions, and pre-match predictions are also common. Visualizing the FIFA Arab Cup 2021 [1] visualizes nearly 2,000 matches played between 23 Arab countries from 1957 to 2021. Once readers click any node, the time, the result of that country in a particular match, the trade value of the country's national team members and their club affiliation will be presented.

4.2 Sociological, cultural and economic issues

The deeper sociological, cultural and economic issues behind sporting events and phenomena also attract attention. The Rise of Hyphenated Last Names in Pro Sports [12] uses data from different sports websites to track the frequency of hyphenated last names in several different sports leagues including the WNBA, NBA, and NHL from the 1950s to the 2010s. It inspires readers to consider the sociology, family heritage, and cultural integration issues behind it. Toyo Keizai Inc. pays attention to the national high school baseball championship in Japan. They create the project Excessive Throwing: How Many Pitches of Japanese High School Baseball Players Violate Pitch Count Guidelines [13]. The most important impact of this project is that it revealed high school baseball pitchers in Japan are throwing far beyond the usual range, and most of them are due to violation of a regulation of rest days. After this project had been published, this problem drew attention to traditional media such as TV and social media such as Twitter. As a result, in November 2019, Japan High School Baseball Federation decided to make rules against over-pitching

ARIJ Network and Infotimes uses the OCR application and data scraping programs from PDF files like Tabula to scrape sports activity data related to youth centers for a time of 10 years. Their investigation reveals the failure of the Egyptian government, through the Ministry of Youth and Sports, to provide a suitable environment for Egyptian women to play sports in youth centers. This happens in various governorates of Egypt, especially in rural areas and Upper Egypt where there are rarely alternatives, and there is discriminatory against women as it deprives them of the opportunity to play sports. Youth Centers Drop Women Athletes from their List of Concerns by Hager Hisham [2] gives the story a human voice, amplifies the results and helps readers to understand what it meant for women.

5 Visualization techniques

We divide the visualization techniques into six categories: static infographics, animated infographics, interactive infographics, GIFs and animated visualizations, map and illustrated map and motion graphics and videos. The result is shown in the **Fig. 3** below.

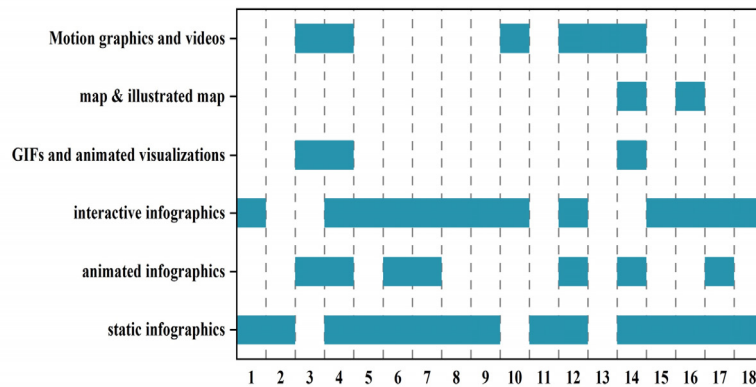


Fig. 3. The occurrences of visualization techniques

As we can see, the most used visualization technique is static infographics, sharing 83.3%. At the same time, 72.2% projects use interactive infographics. Animated infographic and motion graphics and videos take up 38.9% and 33.3% respectively. More importantly, many projects adopt more than one visualization techniques to present them, which looks more beautiful and substantial. For example, *How Many High School Stars Make it in the NBA?* [11] combines static infographics, animated infographics, interactive infographics and motion graphics and videos.

5.1 Narrative analyzation

In the dimension of time, sports data journalism can not only focus on the present but also put the present back to history narrative to realize the recontextualization of the topic [5]. One of the features of an immersive narrative is a non-linear interactive narrative. Computer media allows the creators to place all simultaneous events on a grid and provides simultaneous action in multiple paths so that interactors could browse between them. Such a multi-stage narrative and storyline structure could arouse the audience's curiosity and entice them from one scene to another [19].

5.1.1. Non-linear narrative timeline

A linear narrative timeline is mostly adopted in traditional sports journalism, which is consistent with the order of the occurrence, development and ending of sports events. While large-scale sports data journalism projects, benefit from “cybertext”, multiple subjects and events can be narrated in parallel. Hence, the news events can be separated from their traditional linear narrative timeline in the projects. The readers can understand how the different attributes of the events are represented over time through interaction. It is possible to click on a spot to get details and visually get an overview of events in linear time.

If *All the Olympic Champions Compete Together, Who Will Be the Final Winner?* [10] is published by *The Paper for 2020 Tokyo Olympics*. Before the 2020 Tokyo Olympics, they

collected the competition results of thousands of Olympic champions in 104 sports events over 120 years. Then they set the championship performance of each sport when it first entered the Olympics as a starting value and plotted the performance curves of the champions in each sport over the years to see the limits of human beings (Fig. 4). With the combination of videos, 3D animation, and data visualization, this project successfully simulated the scenes when Olympic champions competed against each other in the same field and presented the breakthroughs and limitations of human physical ability.

In content, they firstly set suspense in the title to arouse the curiosity of the readers. Next, the theme is introduced by the video of the same event across different times and spaces. The sound of the game is added at the same time, bringing the audience into the scene and revealing the answer in person. Then they set the championship performance of each sport when it first entered the Olympics as a starting value and plotted the performance curves of the champions in each sport to compare and visualize the championship results of 104 sports over 120 years to see the limits of human beings. Y-axis symbolizes the results of different sports events. The height of the position reflects the speed of improvement. The passage of time over the years is not a simple accumulation of particles, but a gradual disclosure of human limits to the audience. Past events cast shadows on the future, restricting the scope of the following events, enabling the audience to perceive the narrative lines and experience suspense, and fully highlighting the value of data.

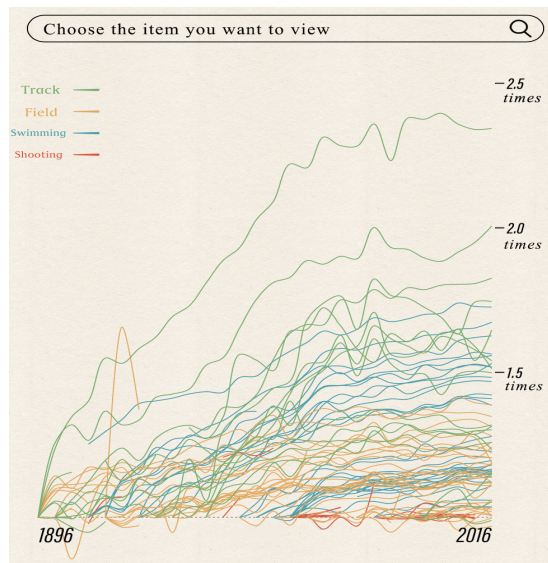


Fig. 4. If All the Olympic Champions Compete Together, Who Will Be the Final Winner?

5.1.2 Linear and periodic narrative timeline

Linear time assumes a starting point and defines a linear period of data elements from the past to the future. Many natural processes have cyclical laws, such as the seasonal cycle [23]. There are many excellent sports journalism projects in Sigma Awards that adopt various types of linear narrative timelines. For example, circular diagrams, spiral charts and so on. These types of charts could better present the cyclical characteristics of specific sports data.

To celebrate the 20 years of Roger Federer's first victory, Spanish media organization, Mundo Deportivo made this infographic in paper format and in digital, detailing the historical classification in the ATP ranking and all the Grand Slam achieved. The main picture of the work is the sports moments of Roger Federer. With the scrolling technique, all the Grand Slams are shown separately. Because the four tournaments were held in the same year, the circular chart is the better choice, which can represent the cycle time structure of Roger Federer's entire career. When the readers scroll the mouse, The red, blue, orange, green, blue, and gray circular time information charts are successively presented in the visual center of the web page to show the achievements of the four Grand Slam tournaments. In addition, all kinds of supplementary charts are arranged in the blank to introduce sorts of important information. The design is simple but full of visual impact.

5.1.3 Chronicle narrative timeline

A chronicle narrative timeline conforms to the human habit of time recognition. Much media attention in the run-up to the games was focused on how Japan would navigate the ongoing COVID-19 pandemic through the event. Reuters stepped back from that coverage and looked at another potential health problem at the Olympics—the sweltering heat. So, they launched the project Hot and Humid Olympic Summer [21]. This data-driven, visually led piece looked to highlight how heat as well as the outsized humidity of Tokyo over this period could put athletes at an elevated risk for heat illness, an issue further complicated by the ease of confusion for first responders with coronavirus patients due to the similarity of symptoms. This story sparked widespread debate about how climate change could affect sports events like the Olympics.

This report uses D3 technology to build interactive calendar views. The background is a static calendar view spread over the entire page, covering all the time nodes from 1964 at the top to 2021 at the bottom. The X-axis symbolizes the month, Y-axis represents the year. Each full row represents annual temperatures across the year. Each small block shows the highest temperatures recorded each day in Tokyo, Japan. The red blocks show days where the temperature was higher than 30°C (86°F). Rather than just a brightly colored static chart, when the readers scroll, specific date segments related to weather stories over the years are highlighted, emphasizing the key events, distilling out the data that the readers need, and bringing the readers' attention to the core issues behind the vast amount of weather data (Fig. 5).

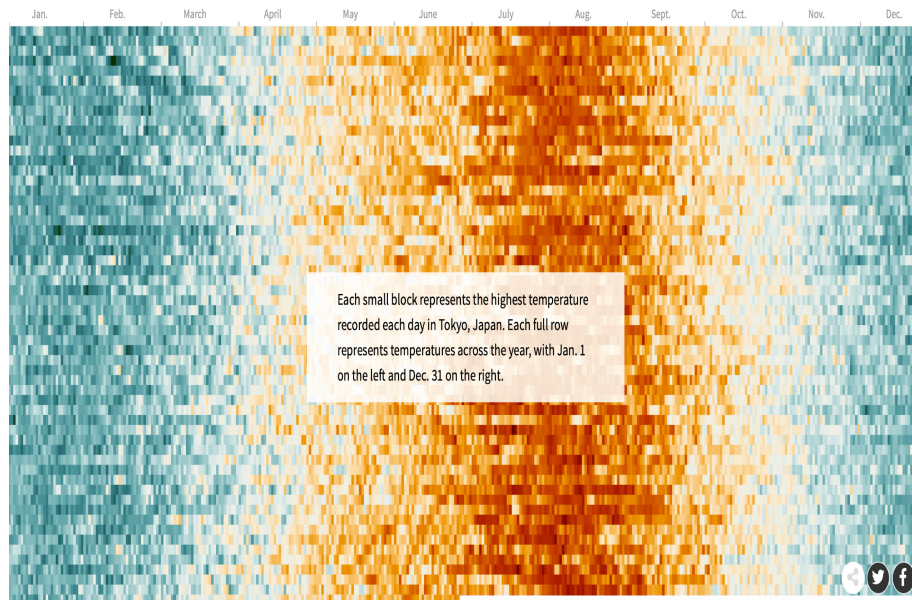


Fig. 5. Hot and Humid Olympic Summer

6 Conclusion

By in-depth analysis and content analysis of 18 listed projects concerning sports in Sigma Awards, we conclude the visualization features in sports data journalism: (1) data sources are diverse and publicly available, which lay foundation for the objectivity and transparency of data journalism; (2) the sociological, cultural, and economic issues under the sports events and phenomena are given attention; (3) the visualization techniques are various, which help readers better understand the news stories.

For narrative perspective, non-linear narrative timeline, linear and periodic narrative timeline and chronicle narrative timeline are shown in the projects. Adopting different narrative timelines to illustrate news stories plays important role in improving the level of news reporting and providing readers better reading experience.

Acknowledgment

We are grateful to all the projects listed in Sigma Awards permitted us to read. Hence, we also want to express our appreciation to the organization of Sigma Awards to hold this worldwide data journalism competition and to the academics and practitioners who promote data journalism and news narratology in their practices.

References

- [1] AI Jazeera, Visualising the FIFA Arab Cup 2021, 2021, Accessed September 14, 2022. <https://sigmaawards.org/visualising-the-fifa-arab-cup-2021/>
- [2] ARIJ Network, Infotimes, Youth Centers Drop Women Athletes from their List of Concerns by Hager Hisham, 2020, Accessed September 14, 2022. <https://sigmaawards.org/youth-centers-drop-women-athletes-from-their-list-of-concerns-by-hager-hisham/>
- [3] C. He, News Narratology, Hunan: Hunan Yuelu Shushe, 2016.
- [4] CNN BRASIL, Swimming Abuse, 2020, Accessed September 14, 2022. <https://sigmaawards.org/swimming-abuses/>
- [5] C. Zhang, Four Dimensions of the Complex Narrative of Data Journalism, TV Research, 2018 (02):38-40.
- [6] Folha de São Paulo, Data-driven Tokyo 2020 Olympics Coverage, 2021, Accessed September 14, 2022. <https://sigmaawards.org/data-driven-tokyo-2020-olympics-coverage/>
- [7] De Tijd, EURO 2020: Data Driven Sports Reporting, 2021, Accessed September 14, 2022. <https://sigmaawards.org/euro-2020-data-driven-sports-reporting/>
- [8] D. Nölleke, C. G. Grimmer, T. Horvák. 2017. News Sources and Follow-up Communication, Journalism Practice, 11:4, 509-526.
- [9] El Confidencial, Rentals of the Spanish Football Stadiums, 2019, Accessed September 14, 2022. <https://sigmaawards.org/rentals-of-the-spanish-football-stadiums/>
- [10] The Paper, If all the Olympic champions compete together, who will be the final winner?, 2021, Accessed September 14, 2022. <https://sigmaawards.org/if-all-the-olympic-champions-compete-together-who-will-be-the-final-winner/>
- [11] The Pudding, How Many High School Stars Make It in the NBA?, 2019, Accessed September 14, 2022. <https://sigmaawards.org/how-many-high-school-stars-make-it-in-the-nba/>
- [12] The Pudding, The Rise of Hyphenated Last Names in Pro Sports, 2019, Accessed September 14, 2022. <https://sigmaawards.org/the-rise-of-hyphenated-last-names-in-pro-sports/>
- [13] Toyo Keizai Inc., High School Baseball Players Violate Pitch Count Guidelines, 2019, Accessed September 14, 2022. <https://sigmaawards.org/excessive-throwing-how-many-pitches-of-japanese-high-school-baseball-players-violate-pitch-count-guidelines/>
- [14] MIA “Rossiya segodnya”, Spin Control, 2020, Accessed September 14, 2022. <https://sigmaawards.org/spin-control/>
- [15] Mundo Deportivo, 20 Grand Slam of Novak Djokovic, 2021, Accessed September 14, 2022. <https://sigmaawards.org/20-grand-slam-of-novak-djokovic/>
- [16] Mundo Deportivo, 20 Years Roger Federer from His First ATP Tour Title, 2021, Accessed September 14, 2022. <https://sigmaawards.org/20-years-roger-federer-from-his-first-atp-tour-title/>
- [17] Mundo Deportivo, Messi’s 643 Goals for the Same Club, 2020, Accessed September 14, 2022. <https://sigmaawards.org/messis-643-goals-for-the-same-club/>
- [18] Mundo Deportivo, Rafa Nadal’s Career in Detail, 2019, Accessed September 14, 2022. <https://sigmaawards.org/rafa-nadals-career-in-detail/>
- [19] M. Yu, L. Peng, Storytelling with Street View: An Immersive narrative account of the Ferguson case, Press Circles, 2015 (01):68-73.
- [20] Reach Data Unit, The State of Football Finance, 2020, Accessed September 14, 2022. <https://sigmaawards.org/the-state-of-football-finance/>
- [21] Reuters Graphics. Hot and humid Olympic summer, 2021, Accessed September 14, 2022. <https://sigmaawards.org/hot-and-humid-olympic-summer/>

- [22] T. Horky, P. Pleka. 2016. Data Visualization in Sports Journalism, *Digital Journalism*, 5:5, 587-606.
- [23] W. Chen, *Data visualization*, Beijing: Publishing House of Electronics Industry, 2013.
- [24] Y. Sun, Narrative Strategies of Map visualization: An analysis based on The Paper. *Youth Journalists*, 2022 (01) :60-63.
- [25] X. Fu, D.Zhang, The New Trend of Sports data visualization: An analysis of sports news practices in 2016, *News and Writing*, 2017 (09) :96-98.