DATA VISUALIZATION IN SPORTS

How data visualization and visual storytelling is affecting the world of sports.

ICM 518
10/21/18

Brian Arfanis
Brian.arfanis@quinnipiac.edu

Intro

Data visualization is more prominent today than it has ever been before. This is due to the advancement in technology and rise in social media. According to Bagadiya (2018)[1], 100 million hours of video content are watched on Facebook daily, and around 95 million photos are uploaded each day. That is nearly 200 million combined hours and uploads spent on some type of visual. This does not even consider other sites such as Twitter, LinkedIn and Instagram. Even phones, computers, televisions and tablets are all centered around visuals. Login pages and apps are represented by visually pleasing icons that also represent the data, for example message icons include photos of chat boxes and music apps are represented by musical notes. These statistics and content highlight the fact that we are visual society.

But not only do visuals play an aesthetic role for data, they also play an effective role. White (2013)[2] of the Aberdeen Group, a market intelligence company that collects and analyzes data, published a paper recently on the effectiveness of visualization and its role in Analytics. Within the article he states, "managers in organizations using visual data discovery tools are 28% more likely than their peers that use just manages reporting and dashboard to find timely information." This is illustrated in the chart below.

Figure 1: Superior Performance of Organizations using VDD
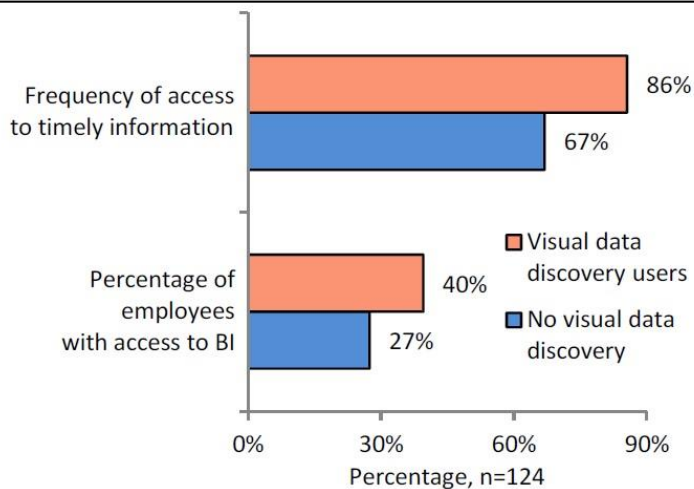
Source: Aberdeen Group, May 2013

Figure 1 White, D.W, 2013, Visualization Set your Analytics Users Free, Aberdeen Group,
https://www.tableau.com/sites/default/files/media/8604-ra-business-intelligence-analytics.pdf

This boom in visual data has changed the way businesses analyze, run and portray data and statistics. Perhaps the biggest revolution has come within sports. Very few professions and areas utilize, analyze and generate data quite like sports. Data Visualization and Analytics have taken over the way teams, coaches,

players and even fans have approached it. This paper will dive into the role data visualization and visual storytelling has in sports and how it is affecting sports overall.

Findings

Within the past 15 years, the number of jobs within sports data visualization and analytics has increased by such a significant margin that professional teams have whole departments dedicated to analytics.

The American Statistical Association(ASA)[3] hosted a webinar this past year that strictly focused on the role of statistics in sports and career opportunities within this field. Speakers from around the country came and shared their work experience and examples of projects that incorporated statistics along with visuals. One of the speakers discussed his role at a Harvard Laboratory dedicated to statistical analysis. He shared several projects that were presented at the laboratory. For example, one of the presentations focused on Reducing NFL Offensive Play-Type Predictions to a Practical Model. Students from the University of Toronto created a model that accurately predicts whether the offensive team will pass or rush on any given play. They used the below heat map to show the probability based on a certain combination of plays.
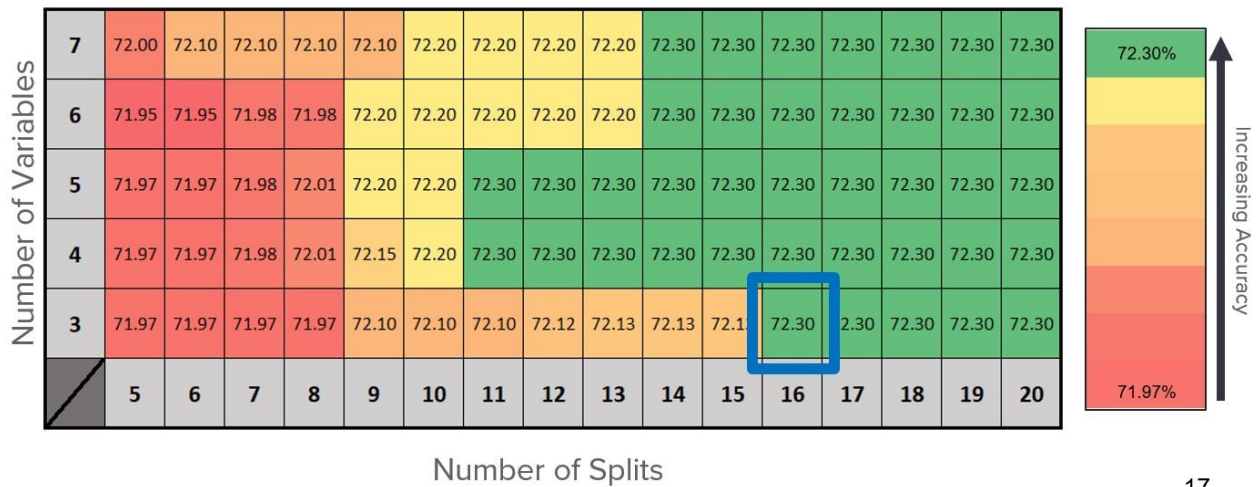

17

Figure 2 Evans, S.E, 2018, Chance, Careers in Sports Analytics—Highlights of an ASA Webinar, https://chance.amstat.org/2016/11/careers-in-sports-analytics/

They later presented the below chart based on real information to consider which variables lead to a pass or to a run play. The use of colors is used to distinguish a pass play and a rush play and each box is titled to represent which factor leads to which play.
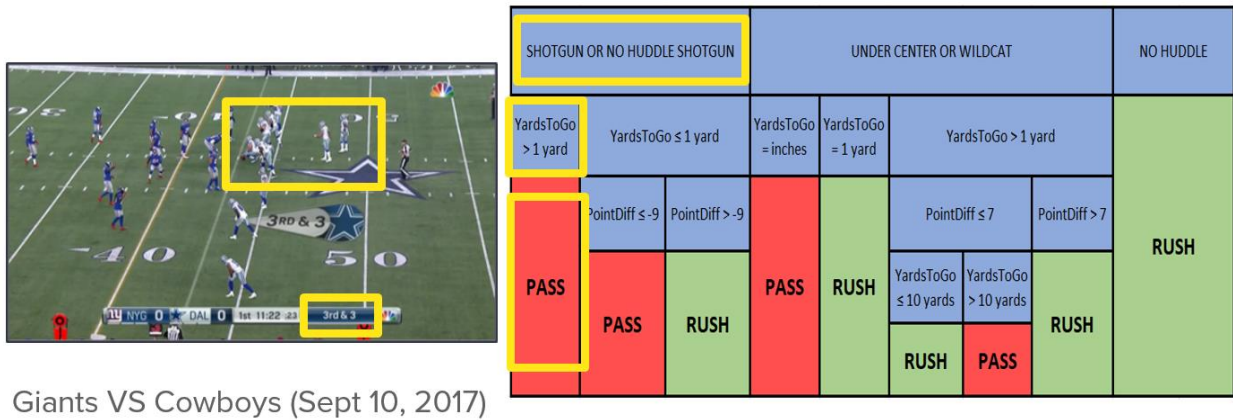

Giants VS Cowboys (Sept 10, 2017)

Figure 3 Evans, S.E, 2018, Chance, Careers in Sports Analytics—Highlights of an ASA Webinar, https://chance.amstat.org/2016/11/careers-in-sports-analytics/

To fully understand these visuals, one would have to understand the data and concepts before viewing them. However, what is important to note here is that each piece of data is complemented by a visual. This is used to better represent and understand the data. What is even more important to note is that actual NFL teams are using visuals such as the ones above to help them game plan. Dennis Lock who was one of the speakers at the ASA Webinar is entering his third season with the NFL team Miami Dolphins. He has been the head and director of the team's analytics department since 2013. He is in charge for taking large amounts of data and making sense of it for game planning, sports performance and training valuation. *Information Visualization – A Brief Introduction (2018)[4]* by the Interaction Design Foundation states it best when describing the importance of positions such as Dennis Locks. The article states "Over the last few decades computing and the internet have revolutionized our ability to create, store and retrieve information on a whim. A global economy and instant communication have created an explosion in the volumes of data to which we are exposed. Yet, the amount of data leads to a large amount of possible confusion and decision paralysis. There's more data available than we can comfortably process." The article later goes on to say, "Information visualization, the art of representing data in a way that it is easy to understand and to manipulate, can help us make sense of information and thus make it useful in our lives." Employees such as Dennis and others who presented at the ASA Webinar, are essential for businesses involved in sports data and analytics.

An NFL playbook can have over 100 plays with 5 – 6 variations of each play. That is anywhere between 500 and 600 plays. Using concepts such as data visualization is becoming essential to organize, analyze and discover patterns within such large amounts of data.

Hubspot.net publish an article titled *Data Visualization 101: How to design charts and graphs (2011)*[5]. The article lists several types of graphs and prefaces it by saying, "Your data is only as good as your ability to understand and communicate it, which is why choosing the right visualization is essential." The article shows the below two charts as being keen for data relationships.
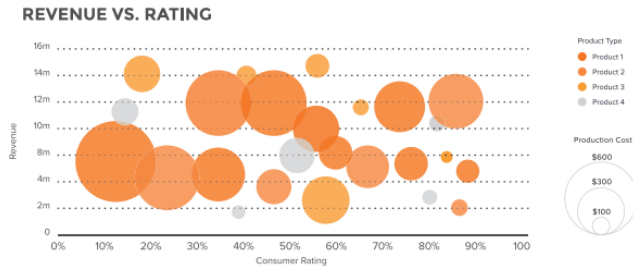
Distribution Chart

Figure 4 Data Visualization 101 Infographics: The Power of Visual Storytelling by Ross Crooks, Jason Lankow and Josh Ritchie (Wiley 2012); The Wall Street Journal Guide to Information Graphics by Dona Wong (Dow Jones & Company 2010); Visualize This by Nathan Yau (Wiley 2011) https://cdn2.hubspot.net/hub/53/file-863940581-pdf/Data_Visualization_101_How_to_Design_Charts_and_Graphs.pdf
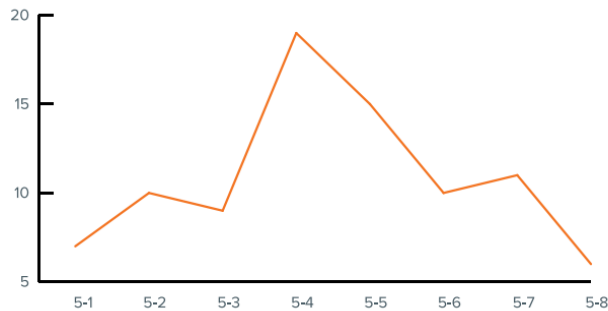
Correlation Chart

Figure 5 Data Visualization 101 Infographics: The Power of Visual Storytelling by Ross Crooks, Jason Lankow and Josh Ritchie (Wiley 2012); The Wall Street Journal Guide to Information Graphics by Dona Wong (Dow Jones & Company 2010); Visualize This by Nathan Yau(Wiley 2011) https://cdn2.hubspot.net/hub/53/file-863940581-pdf/Data_Visualization_101_How_to_Design_Charts_and_Graphs.pdf

Distribution charts show data distribution around a central value and correlation charts display two data variables that have some type of correlation with each other.

These two types of charts are very prominent in the study of baseball statistics. This statistical analysis of baseball is referred to as "sabermetrics." Baseball researcher Bill James coined the term in the 1980's and according to a paper published by the Syracuse staff (2015)[6], sabermetrics "is the empirical analysis of baseball through statistics, used to predict the performance of players, giving teams a winning edge." Thanks

to the advancement in digital technology, teams can now record players in real time which takes sabermetrics to a new level.

The below visual is similar to the distribution chart mentioned earlier. Cameras around the field digitally record the position of all players and hit balls using live camera feeds and recognition software. This calculates variables such as the difficulty of a catch, defensive range and probabilities of making a catch. The result is a distribution chart scattered with data.

FIELDf/x

FIELDf/x digitally records the position of all players and hit balls in real time using live camera feeds and object-recognition software.

Calculates the difficulty of a catch, probability of a particular player making a catch, and the true defensive range (plays that might be made).

Part of the baseball product suite of Sportvision, a company providing television-viewing enhancements.[1]
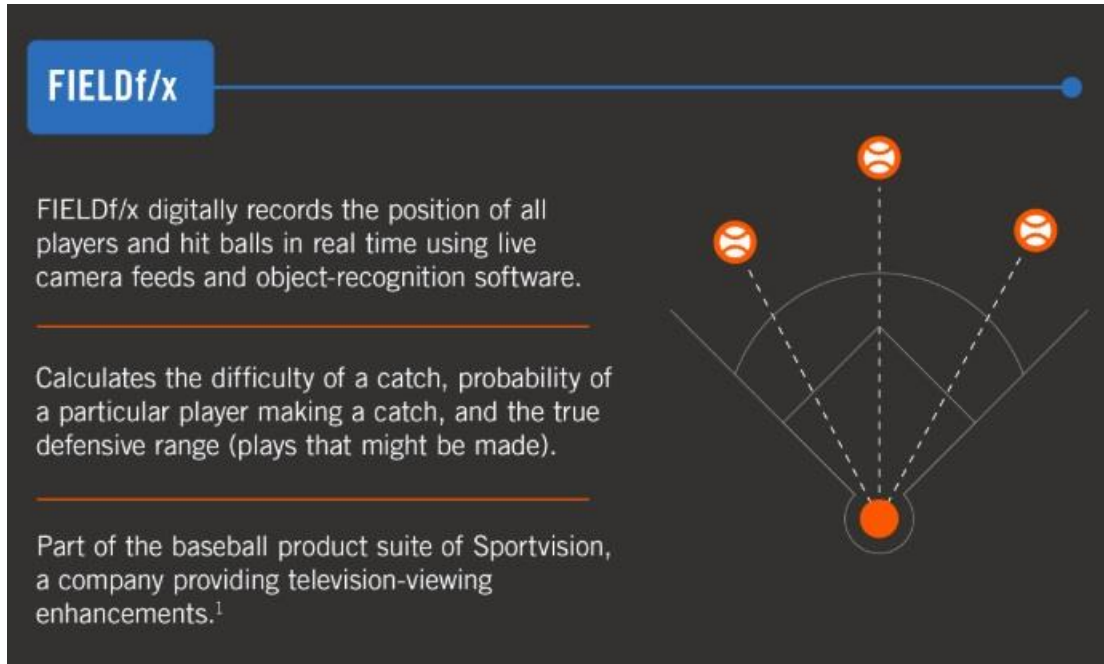
Figure 6 Syracuse Staff, 2015, Martin J. Whitman School of Management | Syracuse University Sabermetrics: Baseball Analytics and the Science of Winning [Infographic], https://onlinebusiness.syr.edu/blog/sabermetrics-baseball-analytics-the-science-of-winning/

This next visual below is similar to the correlation chart mentioned earlier. This is a result of statcast, which is an automated tool that tracks the movement of players. The below visual is an example of what statcast produces. It measures multiple variables such as velocity recorded based on release point of the pitch.
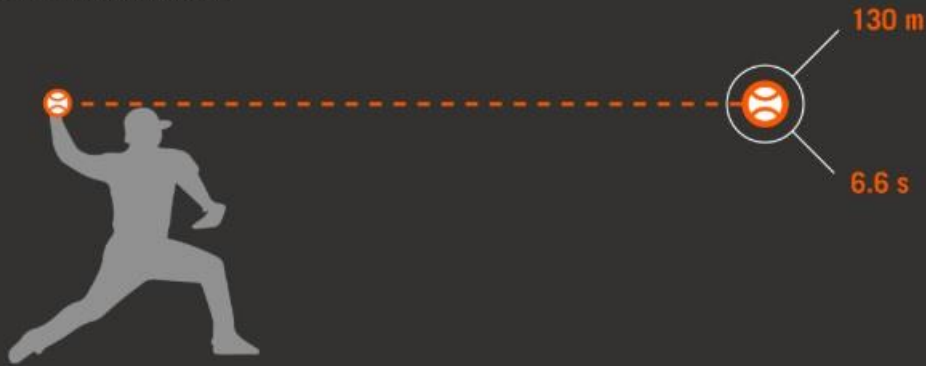
Figure 7 Syracuse Staff, 2015, Martin J. Whitman School of Management | Syracuse University Sabermetrics: Baseball Analytics and the Science of Winning [Infographic], https://onlinebusiness.syr.edu/blog/sabermetrics-baseball-analytics-the-science-of-winning/

Many believe sports analytics and visualization started with Billy Bean. Bean was the general manager for the baseball team Oakland Athletics. The book Moneyball chronicled his time as general manager and his use of sabermetrics to make player personal decisions. According to Steinberg (2015)[7], Bill Bean's use of sabermetrics "was the first known use of prioritization of statistics and data to make personnel decisions in professional sports." His use of statistics inspired sites such as FiveThirtyEight.com, which is used to provide details into much more than baseball coverage. The site includes journalists analyzing, organizing and displaying statistics for fans to gain a greater understanding of upcoming games. Sports Analytics has been turned into a science, look at the below strike zone that utilizes colors to display major league baseball player J.D. Martinez's success in different parts of the zone. Red is where he has the greatest rates of hits against a pitcher's fastball and blue is where he has the least success.
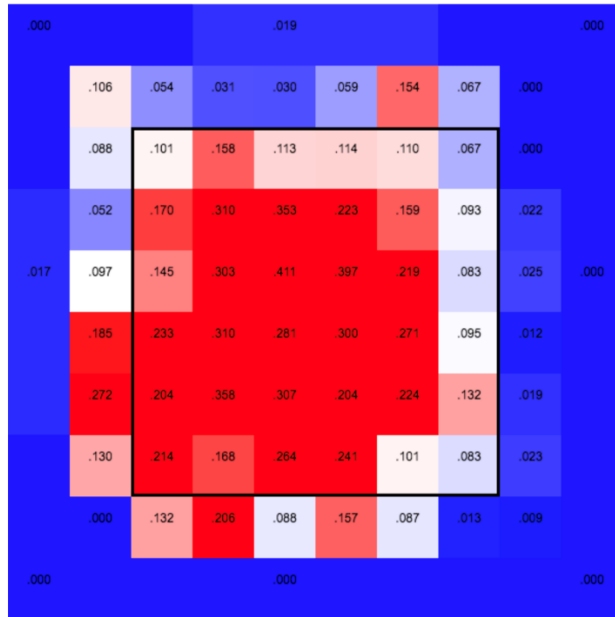
Figure 8 Travis Sawchik (2018), https://fivethirtyeight.com/features/j-d-martinez-is-always-one-step-ahead/

The importance of color in data visualization is high. That is one of the reasons why the above strike is so effective for portraying data. Dennison (2017)[8] of The Integration & Application Network published a paper, and in this paper they list color has having a major impact in communicating scientific findings. He mentions that color perception "is a powerful cognitive feature of humans. Different colors evoke different emotions and physiological responses." Dennison (2017)[8] later describes the use of red, as seen above, as something that should be used sparingly but when used right can be highly effective. Dennison (2017)[8] states "Red color, used sparingly, can serve to highlight or draw attention to certain components of a visual element. Colors can help tie different components together by grouping by color borders or having the color in a graph correspond to the color of a map location."

Thus far, the discussion has been the impact of data visualization on the front office/business side and the fan side, but how has data visualization and analytics helped the actual player? Young (2014)[9] of Eastern Washington University published a report titled *Remediation in Data Visualization: Two Examples of Learning in Real – Time Data Processing Environments.* In the report, Justin describes how data visualization is designed to teach and effect change. He cites NBA player Kevin Durant as using data visualization and statistics in his training. Durant "caters his training to numerical imbalances, memorizes situational shooting statistics." Young (2014)[9.]

Durant and other players are also being used as "subjects" by a service called SportVU. According to Young's (2014)[9] report, SportsVU is a service "that provides 25 – frames – per second optical tracking data quantifying video of player performance for visualization and analysis." The result is a visual like the one below. Camera track each player and provides raw data that is then converted to spreadsheets. At least 15 NBA teams utilize this technology to see patterns and tendencies in players. SportsVu has the potential to impact how players are coached and how the game itself is played. Never has data been presented in such a way as this.
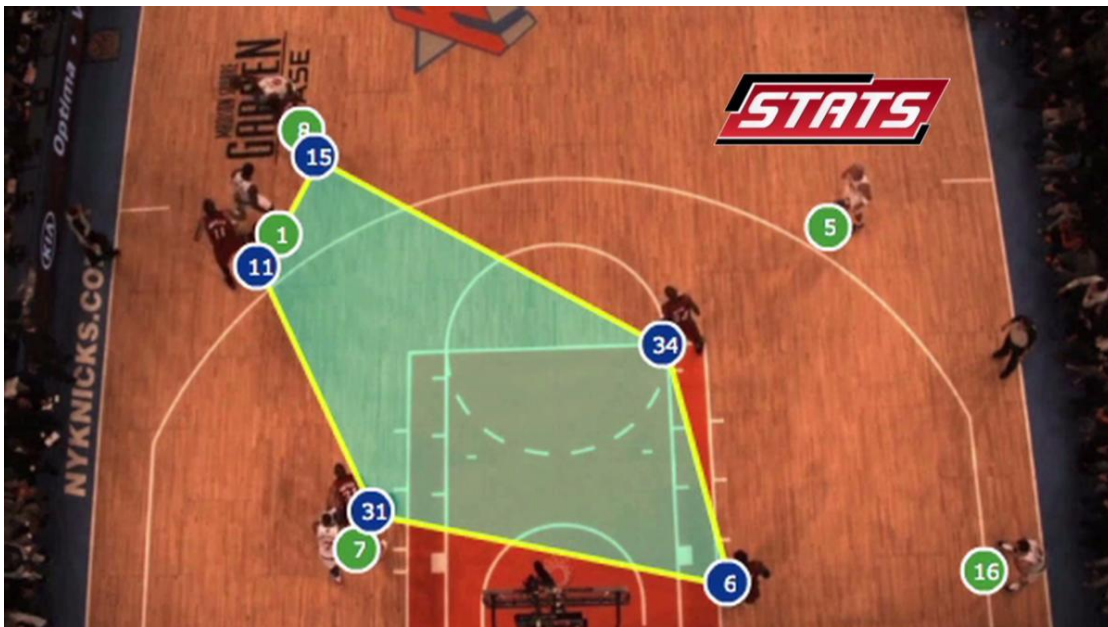
Figure 9 https://www.sporttechie.com/wp-content/uploads/2016/01/130913094124-20130913-stats-2-00000124.1200x672.jpg

Data visualization is effecting the way teams, players and fans approach sports in the front office, when constructing teams and when looking for trends in performance. It is also affecting Sports Journalism. Researchers Horky and Pelka (2017)[10] published a review that does an excellent job in discussing and analyzing the challenges and impact of publishing data in professional online sports journalism. Both Horty and Pelka (2017)[10] used a German Football article as a basis for their research and within the article, there are examples where data and visuals proved to be catching the attention of both users and media outlets. One of those examples includes Spielverlagerung, a blog founded in 2011 that is concerned primarily with football tactics. Spielverlagerung has published more than 1800 articles that include graphics and data analysis. According to Harty's and Pelka's (2017)[10] report, "Spielverlagerung's most common text form is match analysis. Here, the course of matches is replicated through strategic and tactic interactions and explained to users. Data is employed for this as support and illustration. They are not only integrated into graphic form but also into continuous text." So, both graphics and text are used to complement each other. Graphics are also adopted from other blogs and databanks in order to produce their own visuals. Spielverlagerung has caught the attention of larger media due to its appeal on social network sites such as Facebook where it has over 25,000 fans and on Twitter where it has over 17,000 followers. A very interesting quote form the report is written later that reads "There may be a trend visible here where established media align themselves with successful blogs. Since the success of Spierlverlagerung, more analysis of tactics and data journalism have appeared in traditional media's sports reporting." With data visualization and sports continuing towards a greater role in journalism, this could very well be a trend worth noticing.

With the positive there can always be a negative. Due to the high amount of statics in sports, there is opportunity to manipulate, inflate, misconstrue and mislead people.

Misleading Statistics is an all too common practice which is further exemplified by Lebied (2018)[11] of the Datapine Blog. According to Lebied (2018)[11], "Dr. Daniele Fanelli from The University of Edinburgh found that 33.7% of scientists surveyed admitted to questionable research practices, including modifying results to improve outcomes." In fact, there are so many accounts of misused and misrepresented data that the NPR (*National Public Radio, American privately and publicly funded non-profit membership media organization*) has an entire code of ethics dedicated to Visual Journalism. Some of the practices include double checking for accuracy, providing proper context, checking for any biased content and not altering any photos unless for clarity reasons. A summary of their Ethics code can be viewed here – NPR Visual Journalism (http://ethics.npr.org/tag/visual-journalism/)[12]

Having an array of statistical data can be tempting to manipulate and display in a biased/one-sided way. It is important to understand the full context of each story or graphic before using it as a source.

Findings Part 2

The computer Graphics Forum in association with The Eurographics and John Wiley published a full report this year titled *State of the Art of Sports Data Visualization (2018)[13]*, in the report they organized and reflected on recent advances and challenges in the field of sports data visualization.

The reports states that "Sports visualization offers new approaches to exploring, making sense of, and communicating sports data," and goes on to say "As visualizations can be more accessible and more meaningful than traditional statistical analysis, the number of visualizations of sports data has grown rapidly over the past decades. Print and online media companies, including The New York Times present sports data using infographics and interactive visualizations." This indicates that more large media outlets are implementing data visualization in sports related topics. The below figure shows the distribution, per year, of published research articles on sports visualization gathered by the Computer Graphics Forum Team.
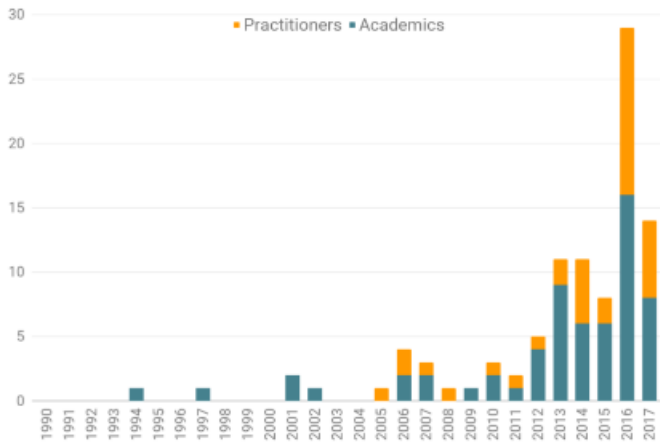
Figure 10 Perin, C.P, Vuillemot, R.V, Stolper, C.D.S, Stasko, J.T.S, Wood, J.W, Carpendale, S.C. 2018, Computer Graphics Forum 2018 The Eurographics Association and John Wiley & Sons Ltd., State of the Art of Sports Visualization. https://onlinelibrary-wiley-com

The growth has been exponential within the last few years. Not only has the number of visuals in sports increased, but also the quality and detail of visuals has increased. Look at the below scorecard located on the left of the San Francisco Giants baseball team, and then take a look at the chart on the right that displays detailed information based on a recent championship soccer match. Due to technology of today, graphics such as the right one can be automated using programs and trackers. Programing jobs and graphical design jobs are now common for sports teams.
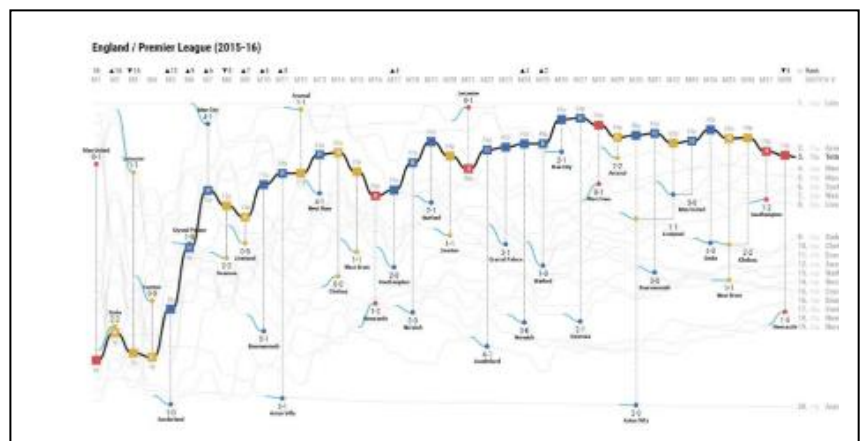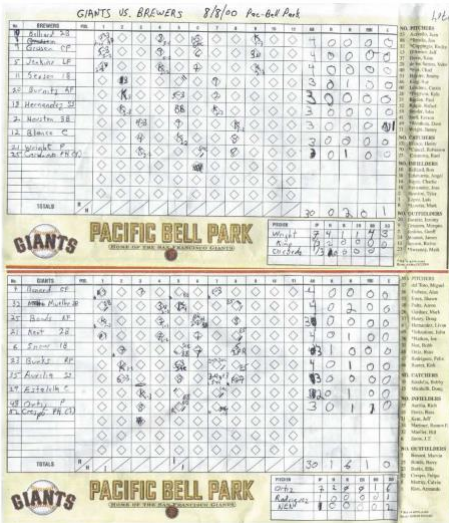
Figure 11 Perin, C.P, Vuillemot, R.V, Stolper, C.D.S, Stasko, J.T.S, Wood, J.W, Carpendale, S.C. 2018, Computer Graphics Forum 2018 The Eurographics Association and John Wiley & Sons Ltd., State of the Art of Sports Visualization. https://onlinelibrary-wiley-com

Conclusion

Based on all these findings, the usage of data visualization in sports is trending upwards. It is affecting sports in many ways including: high rates of analytical analysis, the usage of graphs and other visuals to make player decisions, player tracking tools to create visuals that help players/coaches and general managers and data visualization is creating new opportunities in sports journalism. Entire college departments are being dedicated to sports data and nearly every NFL, NBA and MLB team has some sort of data analytics and graphics position.

A graphic/picture brings new meaning and emotions to viewers. It is one thing to read about something, it is another to see it. DeMeré (2016)[14] highlights this point by providing a real-life example in her post on Hubspot. See below for a visual about the number of tweets sent in a second.
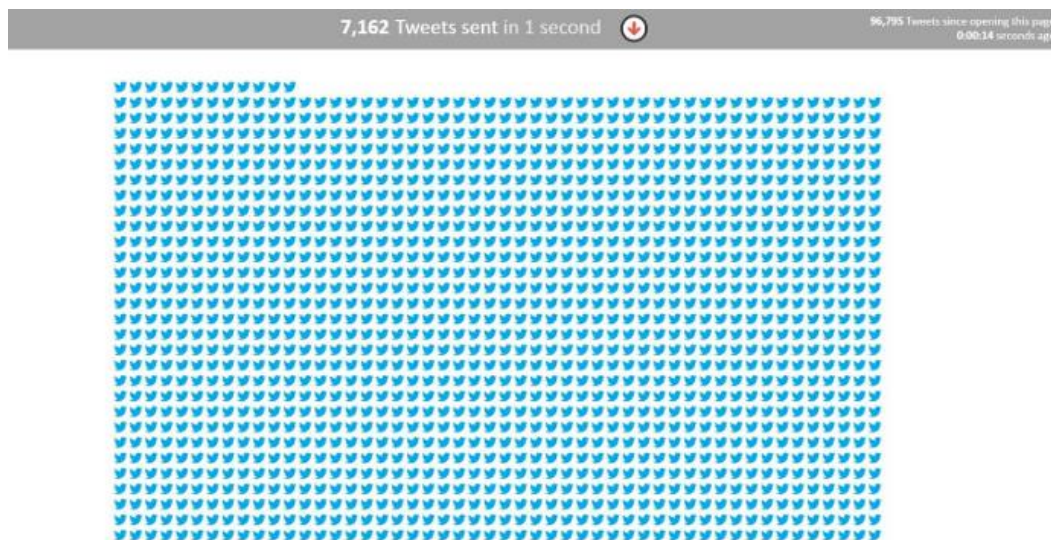

Figure 12 DeMeré, N.E.D, 2016, The Power of Visual Storytelling: 15 Stunning Examples to Inspire You, https://blog.hubspot.com/marketing/visual-storytelling-examples

 "The visual of 7,162 birds is far more impactful than the header saying "7,162 Tweets sent in 1 second." It's a different kind of visual storytelling, but compelling nevertheless." - DeMeré (2016)[14]

Sports contains a vast array of data and with the new tools and technology available, data visualization and visual storytelling will continue to evolve.

Sources

Modules:

[14] *(Module 1) DeMeré, N.E.D, 2016, The Power of Visual Storytelling: 15 Stunning Examples to Inspire You,*
https://blog.hubspot.com/marketing/visual-storytelling-examples

[5] *(Module 3) Data Visualization 101 Infographics: The Power of Visual Storytelling by Ross Crooks, Jason Lankow and Josh Ritchie (Wiley 2012); The Wall Street Journal Guide to Information Graphics by Dona Wong (Dow Jones & Company 2010); Visualize This by Nathan Yau (Wiley 2011)*
https://cdn2.hubspot.net/hub/53/file-863940581-
pdf/Data_Visualization_101_How_to_Design_Charts_and_Graphs.pdf

[4] *(Module 3) 2018, Interaction Design Foundation, Information Visualization - A Brief Introduction*
https://www.interaction-design.org/literature/article/information-visualization-a-brief-introduction

[8] *(Module 6) Dennison, B.D, 2017, Integration & Application Network, Practical visual literacy for science communication*, http://ian.umces.edu/blog/2017/03/28/practical-visual-literacy-for-science-communication/

[11] *(Module 7) Lebied, M.L, 2018, Misleading Statistics Examples – Discover The Potential For Misuse of Statistics & Data In The Digital Age, The Datapine Blog,*https://www.datapine.com/blog/misleading-statistics-and-data/

[12] *(Module 4) Visual Journalism, Ethics Handbook*
http://ethics.npr.org/tag/visual-journalism/

Internet:

[1] *Bagadiya, J.B., 2018, 171 Amazing Social Media Statistics You Should Know in 2018, SocialPilot.co,*
https://www.socialpilot.co/blog/social-media-statistics

[2] *White, D.W, 2013, Visualization Set your Analytics Users Free, Aberdeen Group,*
https://www.tableau.com/sites/default/files/media/8604-ra-business-intelligence-analytics.pdf

[3] *Evans, S.E, 2018, Chance, Careers in Sports Analytics—Highlights of an ASA Webinar,*
https://chance.amstat.org/2016/11/careers-in-sports-analytics/

[6] *Syracuse Staff, 2015, Martin J. Whitman School of Management | Syracuse University Sabermetrics: Baseball Analytics and the Science of Winning [Infographic],* https://onlinebusiness.syr.edu/blog/sabermetrics-baseball-analytics-the-science-of-winning/

[7] *Steinberg, L.S, 2015, Forbes Media, CHANGING THE GAME: The Rise of Sports Analytics,* https://www.forbes.com/sites/leighsteinberg/2015/08/18/changing-the-game-the-rise-of-sports-analytics/#5a209af4c1fd

Peer reviews:

[13] *Perin, C.P, Vuillemot, R.V, Stolper, C.D.S, Stasko, J.T.S, Wood, J.W, Carpendale, S.C. 2018, Computer Graphics Forum 2018 The Eurographics Association and John Wiley & Sons Ltd., State of the Art of Sports Visualization.* https://onlinelibrary-wiley-com.libraryproxy.quinnipiac.edu/doi/full/10.1111/cgf.13447

[10] *Thomas Horky & Philipp Pelka (2017) Data Visualisation in Sports Journalism, Digital Journalism, 5:5, 587-606, DOI: 10.1080/21670811.2016.1254053,* https://www-tandfonline-com.libraryproxy.quinnipiac.edu/doi/pdf/10.1080/21670811.2016.1254053?needAccess=true

[9] *Young, J.Y, Potter, C.P, 2014, Communication Design Quarterly 2.2, Remediation in Data Visualization: Two Examples of Learning in Real-Time Data Processing Environments,* http://delivery.acm.org.libraryproxy.quinnipiac.edu/10.1145/2600000/2597474/p29-young.pdf?ip=162.221.8.54&id=2597474&acc=ACTIVE%20SERVICE&key=A79D83B43E50B5B8%2E4DBAA2C86BE5786A%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1539469901_3245d9632c79b2cec2c32553f9e22af2