

Enjoy this site?  Gift the author a WordPress.com plan.

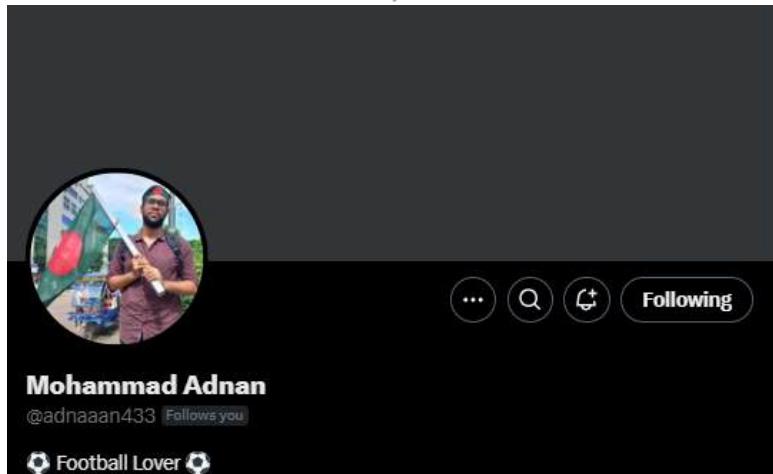
[Gift](#)

GETTING STARTED WITH FOOTBALL ANALYTICS – MOHAMMAD ADNAN

Hi all,

I sit down with Mohammad Adnan, who shares some great resources as he grows in his own football analytics journey.

Find him on **X**, and **Github**.



CJ: For those that are unaware, could you tell us a little about your football analytics journey to date?

Was it a combination of enjoying football, and also just wanting to learn to code?

MA: My journey into football analytics began during my Mechanical Engineering studies. While pursuing my degree, I took a minor course in Python programming, which sparked my interest in coding. As a lifelong football enthusiast, I naturally gravitated towards applying my newfound programming skills to the sport I loved.

I started by following the excellent tutorials provided by **@mckayjohns** on YouTube and leveraging the powerful ‘mpl soccer’ library. These resources made learning football data analysis both accessible and enjoyable. I dedicated significant time to replicating existing projects, experimenting with different approaches, and gradually developing my own unique insights.

McKay Johns

@McKayJohns • 19.4K subscribers • 89 videos

Hi everyone and welcome to my channel! ...[more](#)

mckay-s-site.thinkific.com/courses/football-analytics-course and 2 more links

[Subscribe](#)

[Home](#) [Videos](#) [Shorts](#) [Live](#) [Playlists](#) [Community](#) 



How I Would Learn Sports Analytics (If I Could Start Over)

42,566 views • 10 months ago

Get Free Sports Data by Building Your Own Web Scraping Pipeline: <https://mckay-s-site.thinkific.com/>

use code YOUTUBE at checkout for 25% off :)

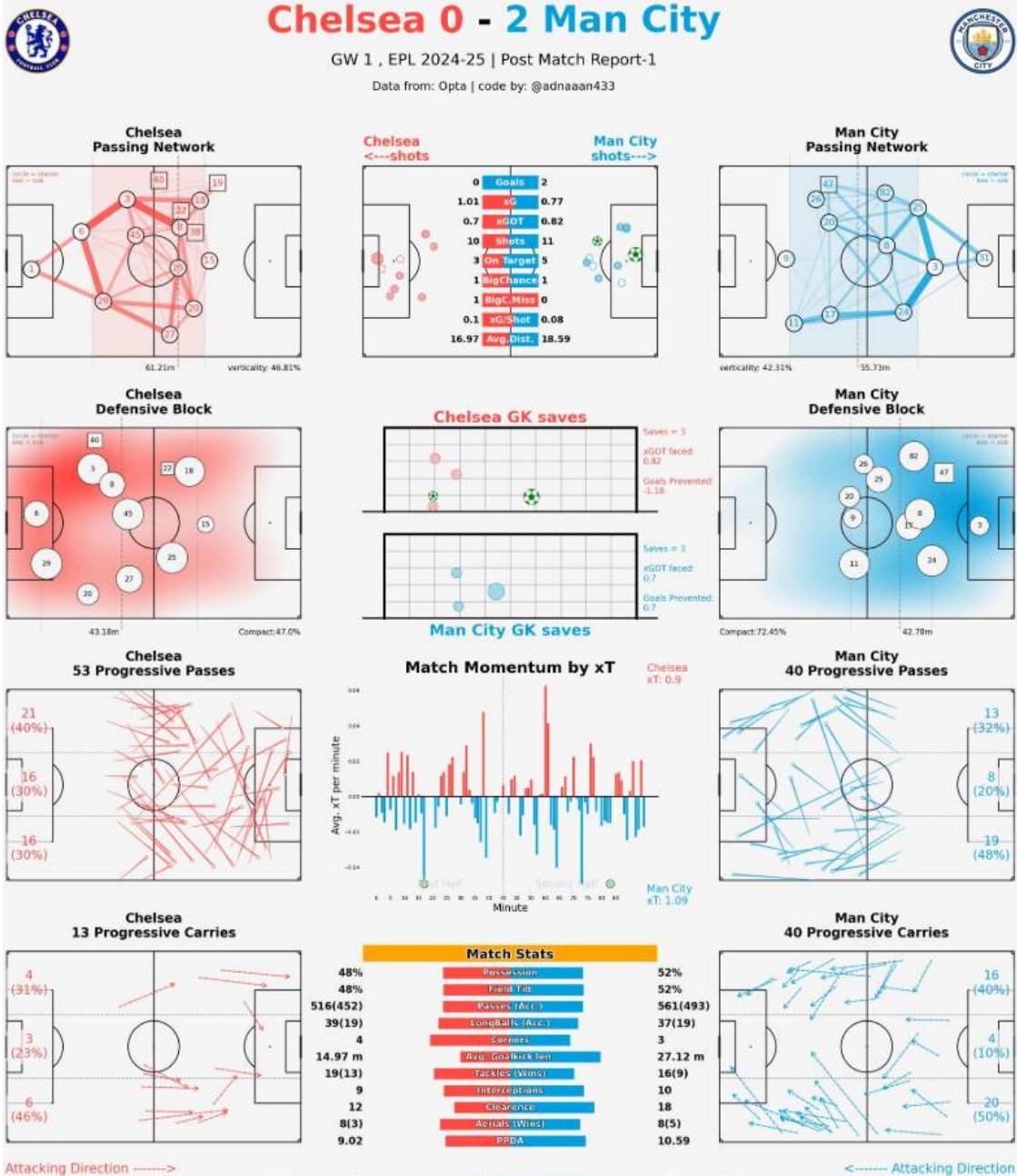
A crucial aspect of successful football data analysis is a deep understanding of the tactical nuances of the game. By combining my technical skills with my extensive knowledge of football, I have been able to uncover valuable insights that might otherwise be overlooked. This synergy between data and tactics is essential for producing impactful analysis.



While I am still learning and growing, I am proud of the progress I have made. I am excited to continue exploring the possibilities of football analytics and contributing to the wider football community!

CJ: Your post match report is a thing of beauty. Could you talk a little as to some of the complexities faced in creating charts like the keeper saves and addressing defensive blocks?

My repo can be found [here](#).



MA: When I learned making my own functions and using that multiple times, I was fascinated by the idea of making an automated process where I will get post match analysis for different matches easily without coding every time.

Then I started to learn gathering multiple plots in a single plot to make a dashboard type. I started to look for other people's dashboards to get some ideas. After a lot of trial and error I had come with the idea of making everything as simple as possible. That's why I didn't use any fancy orientations or arrangements of the individual plots, just put those simply in two sides of the dashboard for the two teams. And also made this usable multiple times for any match analysis, by using some variables in the functions, so that changing the values of those variables will make it possible to use for any match analysis.

Of course I have faced a lot of difficulties doing this. For example, to plot the keeper saves, I needed to resize the data in such a way that it fits the plot. That was a very difficult part for me, because I had to calculate the data coordinates manually for those shots which were hit in the goal posts. From there I got the idea of the range of the values and then organized the plot in that range to plot the keeper saves. The defensive block plot wasn't that much harder, because I used the idea from the

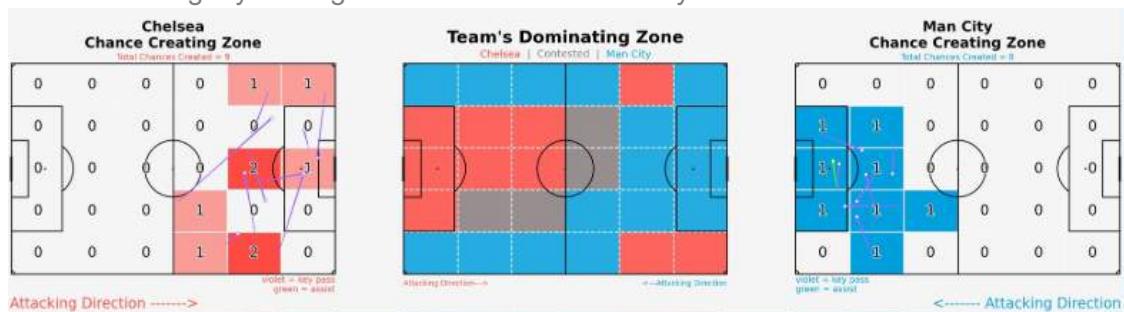
pass network, and it was easier than the pass network plot, because it didn't need any line among the individual players, just the scatters were enough.

You can find the details in my git repo.

CJ: The steps for the post match report are outlined in your github repo, often utilising information from the who scored website and fotmob. It seems like a lot of continuous learning went into creating the report. Which of the charts did you find most challenging? Why is that chart so valuable in the context of football?

MA: The most important thing for a Post Match Report is to get the recent matches event data. Whoscored website provides that, but the scraping process of those data using selenium and chrome web driver was very difficult for me. I would like to Thank [@dizwalski](#) for sharing his beautiful and easy process of data scraping from whoscored website without the complexities of seleniums! That simple process of scraping Whoscored website data kept me in regular football analysis, otherwise I would have already quit the football data analysis. Using his code I could skip learning the complex process of web scraping. Although I think everyone should at least try to learn web scraping skills, this is a very important skill for data analysts. But as I found it very hard for me, I still couldn't learn web scraping.

And the Fotmob website was very easy to scrape, using [@mckayjohns](#) YouTube tutorial, which was a simple 3 lines of code, that's why this was very easy for me. But without learning the basics of web scraping, using other's code is never a good solution. Because recently Fotmob has changed something in their website slightly, the previous process of Fotmob website scraping is not working. I am eagerly waiting for the talented community to see what can be done!



The most difficult chart for me was the 'Team's Dominating Zones'. I had to dig into the `pitch.heatmap()` function of mpl soccer library, to understand how that heatmap works, and how I can fit my data into that manually. After a lot of research, I understood the process and came up with a very unique idea to plot that chart. I didn't find any tutorials to make that chart, so I had to work for every part of the function manually and in my own way. Of course that wasn't an efficient way to code, but all the hard work to understand the heatmap function and using that in my own way was paid off after the successful implementation of that function.

This chart is valuable to understand the team's domination in different pitch areas in a match. I first saw this type of chart on Opta's website, but didn't find any tutorials to make something like that. So I had to make this myself.

CJ: When we chatted, you spoke to how models like chat gpt have helped you in learning to code. Could you tell us a little more about how you tackle that idea of debugging and enhancing your work with prompts?

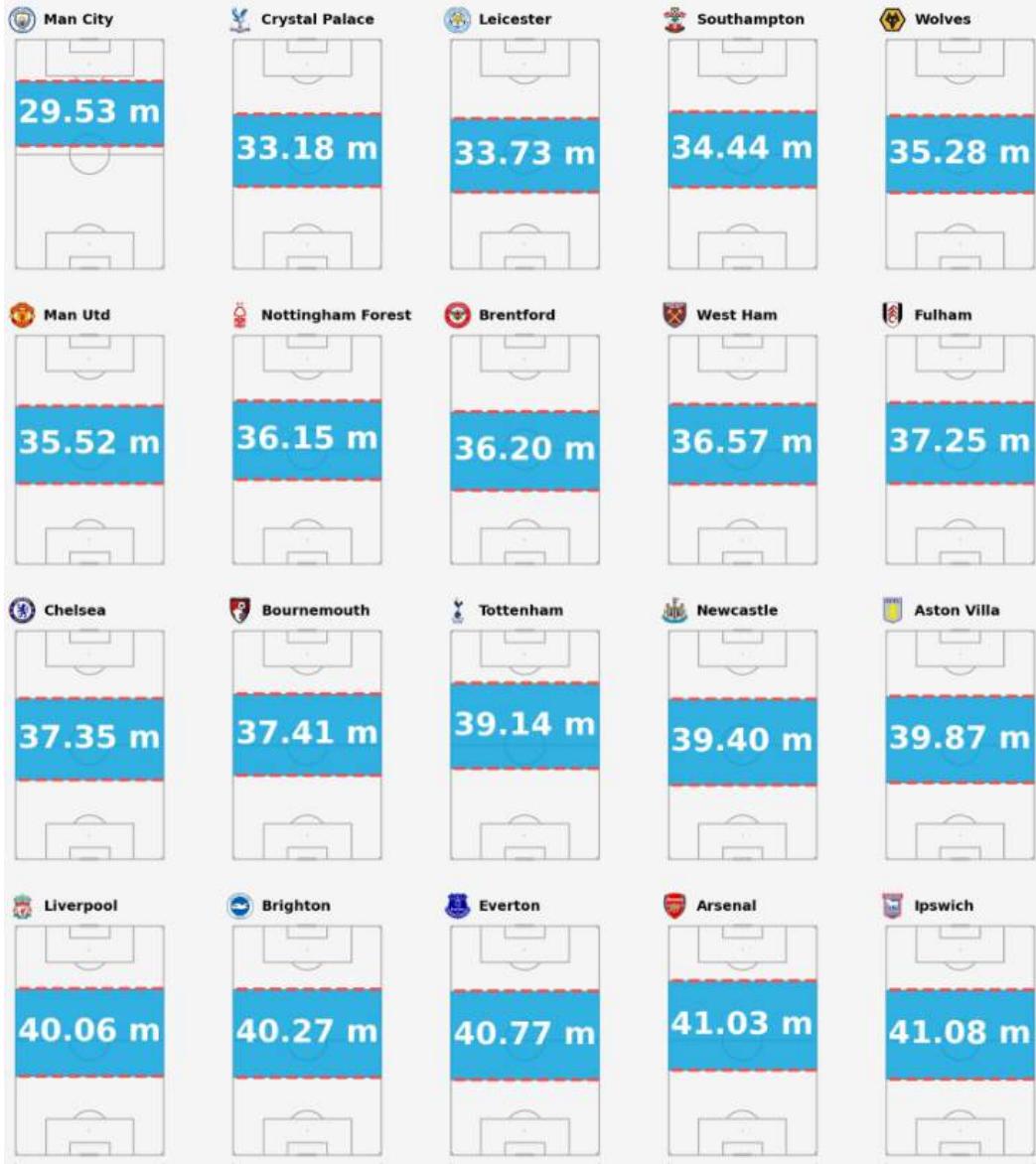
MA: Chatgpt has a huge role in my learning to code. I knew only the basics of the python programming language, but didn't know about different other features like using multiple libraries and functions, making my own functions and using them. I wasn't even familiar with the different filtering process of dataframes to do data analysis! Whenever I needed to filter any specific data from a huge dataframe, I used to copy and paste the whole dataframe into chatgpt and told it to filter that specific data from the dataframe. It gave me the whole python code to filter those specific data and also explained the code line by line! In this way I have learnt many complex processes of dataframe handling from chatgpt. Also whenever I faced any errors, I used to copy and paste the whole error message into chatgpt. Then it explained the error and rewrote the code in the right way. That's why now I can debug 90% of the errors myself.

CJ: Some of your work online is getting huge amounts of attention, such as the on the ball compactness. I hadn't seen a visual like this before! What message were you trying to convey with this piece, if we look specifically at the premier league?



EPL Teams On the Ball Compactness

Teams are sorted by Compactness (Difference between Avg. Height of FWDs & Avg. Height of CBs)
Data from: Opta | till GW11 of EPL 2024-25 | made by: @adnaan433



MA: The term 'On the Ball Compactness' refers to how close (vertically) the teammates stay on the ball, if they are more compact, that means they find themselves with short passes, if they aren't that much compact, they have to go long for ball progression. That's what I was trying to find out in the teams.

So, I took the Difference between avg. height of the Forwards and avg. height of the CenterBacks
(avg. height means avg. On the Ball actions distance from own goal line)

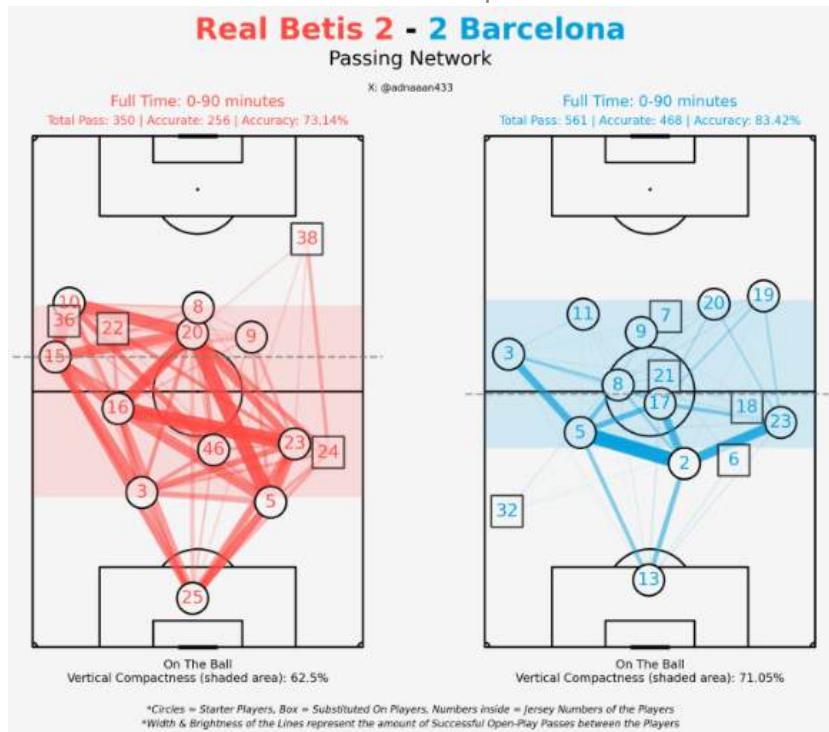
But sometimes in some matches, one forward may stay isolated and too far away from other teammates, Or sometimes we see the Forwards to stay in a much deeper position while receiving the ball. I decided to take the avg. height of the top two locations among the players and using that value as the avg. height of the Forwards.

As I follow Premier League regularly, I have noticed Man City this season, they stay compact on the ball. From there I was interested to know about the other teams' compactness. That's why I made that visualization to compare teams playing style on the ball.

After making that visualization, a lot of people also asked me to make something like this for 'Off The Ball Compactness'. But in event data the only actions with the ball is recorded, so getting the data of

the players location while Off The Ball is not possible in this case. We have to analyse this from Tracking Data which is not available for free.

In this visualization below we can see what I mean by on the vertical compactness, here the shaded area is vertical compactness:



Look at Barcelona, I took the avg height of the 2 CBs (#5 Inigo and #2 Cubarsi) as the lower line. And avg height of the top 2 forwards (#20 Olmo and #19 Lamine). In between area of these 2 lines is what I call On the Ball Compactness!

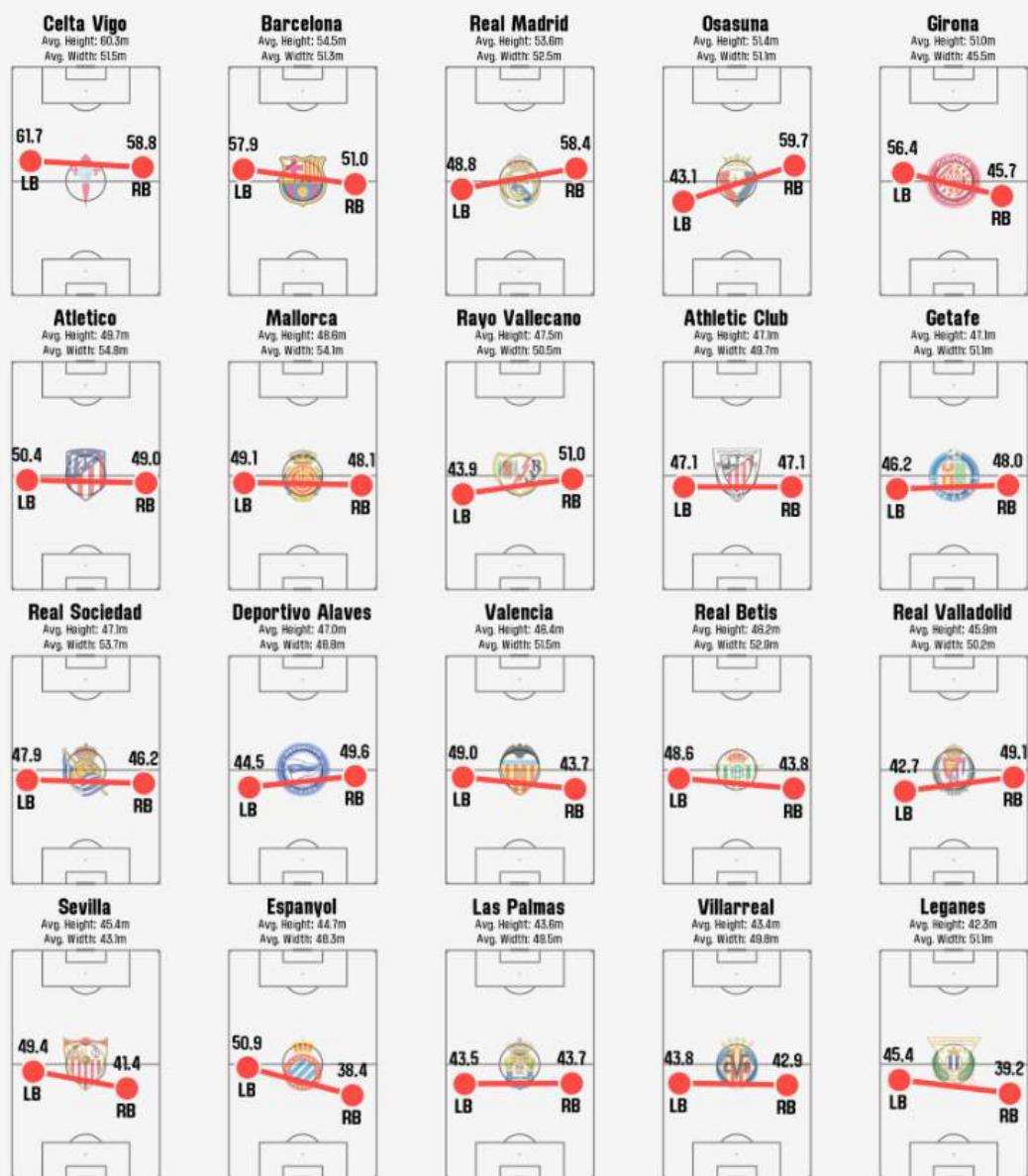
Look, as here Lewa has dropped deeper more often, thus his avg location is much lower. That's why only considering the Forwards location wasn't enough to get the actual on the ball compactness. Thus I prefer to take the avg of top 2 forwards.

CJ: Your work often includes small multiples, looking across either teams, or leagues. Your full back analysis really caught the eye of our SportsVizSunday initiative. How do you think the use of trellis charts favors your analysis in the case of your LaLiga full back viz?



LaLiga Teams FullBack Analysis

Teams are sorted by Avg. Height of the Fullbacks
Data till GW9 of LaLiga 2024-25 | made by: @adnanaan433 | inspired from: @Opta



MA: Thanks a lot for appreciating my work on your platform. Actually that Fullback analysis wasn't my unique idea, I saw a post from the Opta analyst account which was analysing the Premier League Fullbacks. That inspired me to do the same for LaLiga.

We have seen a revolutionary change in the case of the role of a Fullback in modern football. Some coaches like highly attacking Fullbacks, some use them as inverted Fullback to increase the number of players in the midfield during build-up, some prefer to use as defensive Fullback which often becomes a third CB. So if we look at the avg. touch position of the Fullbacks in different teams, we can get some idea about their role in the team. That's why I made that visualization.

In La Liga, Celta Vigo, Barcelona and Real Madrid push up high.

Real Madrid, Osasuna, Girona, Rayo Vallecano, Sevilla, Espanyol asymmetric.

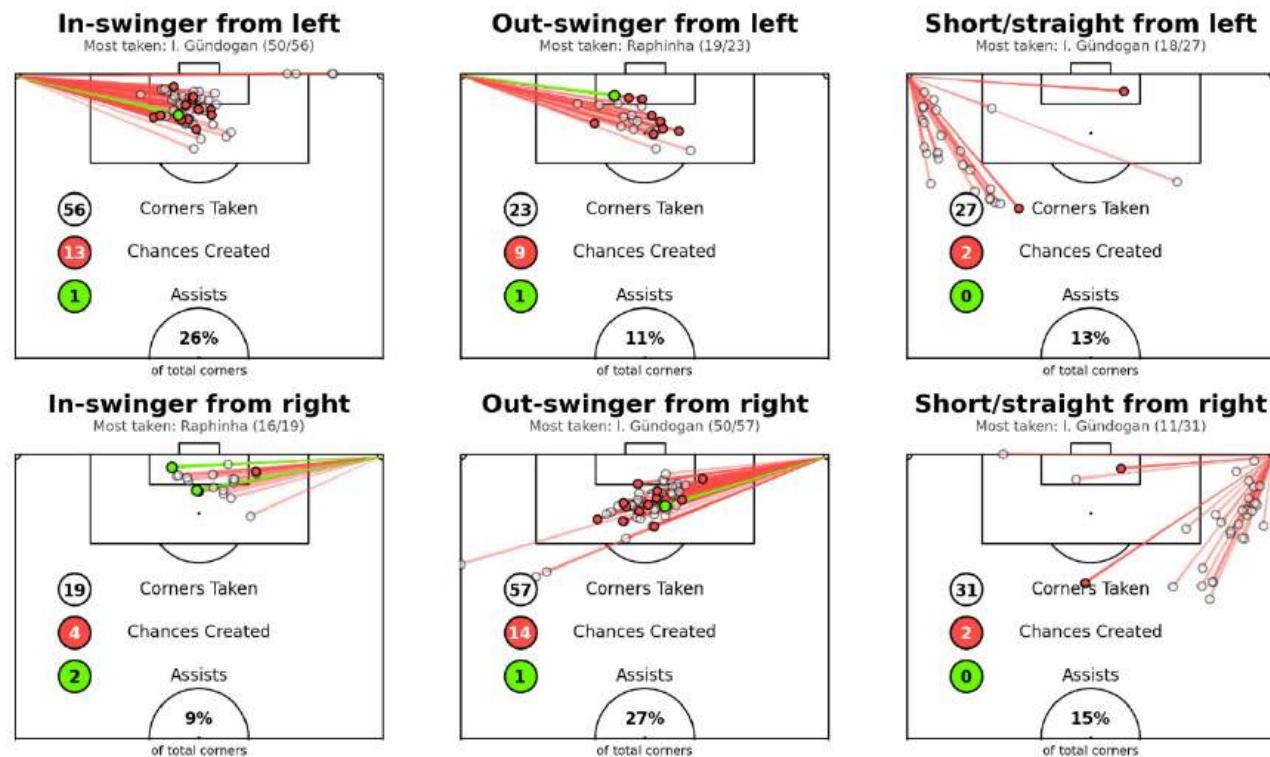
Leganes, Villarreal and Las Palmas very deep.

CJ: I really like your work on looking at corner pieces. If someone was a beginner and wanted help explaining how to recreate something like this, what steps would you outline from the code you wrote?



Barcelona: 213 Corners

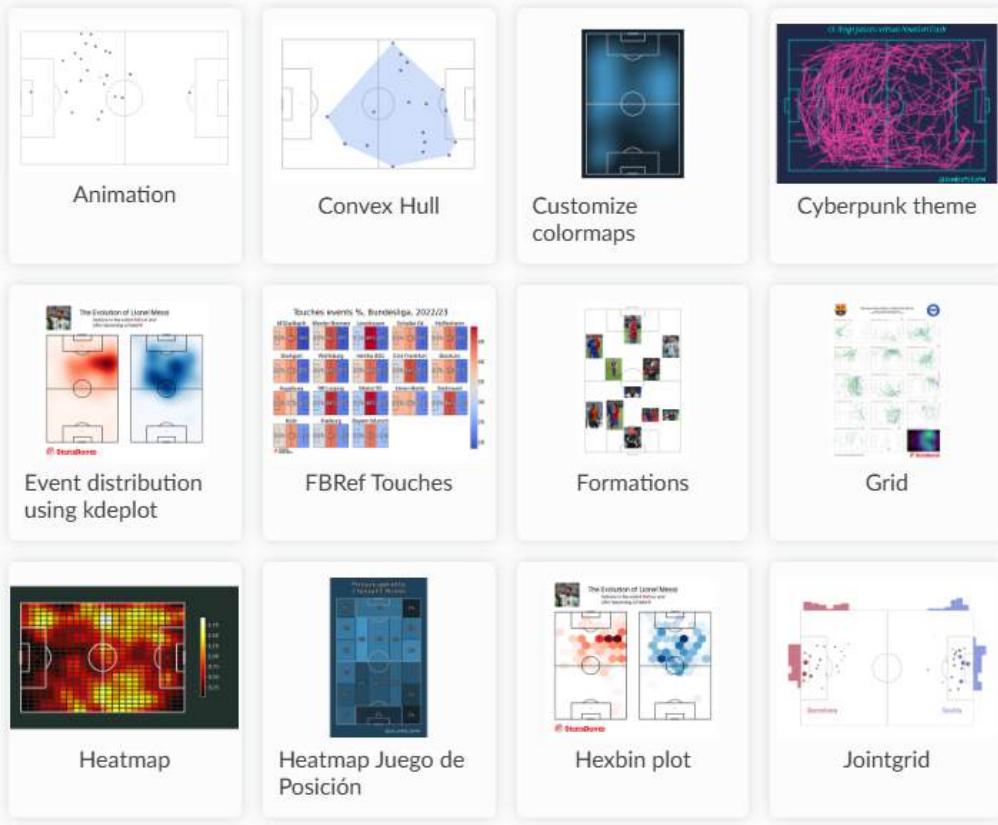
in LaLiga season:2023-24, till Round34 | Data from: Opta | made by: @adnaan433



Thank you very much. Actually for the beginners the most important thing is to learn all the visualisation tools first. For example, as we mostly use the **mpl soccer library**, we have to get an idea about their most common and important functions. They have a beautiful website where every function has been explained with example codes for practice.

Pitches

Examples of the methods for plotting pitches in mplsoccer.



Then you have to understand the event data properly. For example if we get event data from Statsbomb, they have different columns for tagging detailed information about a pass or a shot. But when you scrape event data from whoscored website, they tag all the detailed information about a pass or a shot in a single qualifiers column. You have to learn how to filter the specific data from that qualifiers column.

For making the corners analysis visualization, I didn't get the information about whether a corner kick was in-swinger or out-swinger directly in the qualifiers column. There was information about whether the corner kick was taken with the Left foot or Right foot. According to this information I made the conditions that,

A corner kick will be in-swinger if,

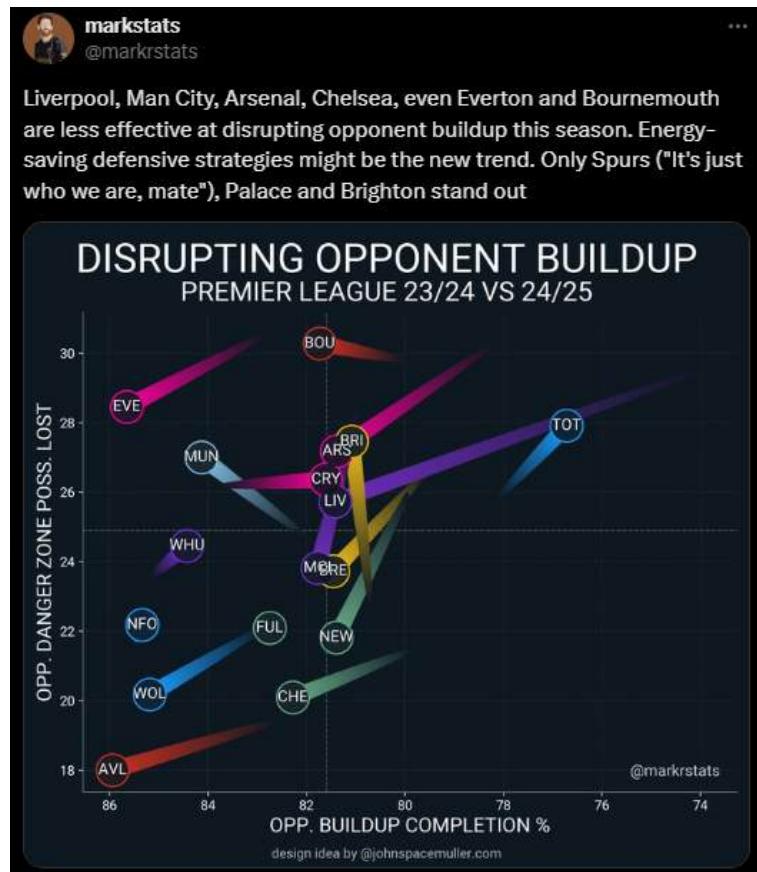
1. It was taken with Left foot from the Right corner of the pitch
2. It was taken with Right foot from the Left corner of the pitch

A corner kick will be out-swinger if,

1. It was taken with Left foot from the Left corner of the pitch
2. It was taken with Right foot from the Right corner of the pitch

This way I filtered the data and the visualised with basic mpl soccer library functions.

CJ: I would say the online x community has a few “big names” when it comes to creating visuals with football data. Was there anyone in particular that influenced your own journey of learning?



MA: [@markrstats](#) is the one who influenced me to start football visualizations. As a football fan I used to follow some of the accounts in X who share different statistics of football ([Sofascore](#), [Fotmob](#), [Squawka](#) etc). One day randomly a post of [@markrstats](#) came to my feed then I realized that football data can be visualised also. From then I started to look at how to visualise those football data. And then I found the [@mckayjohns](#). From then I started to follow the big names who create football data visualizations. I found [@BeGriffis](#) , [@themachineball](#) (Aritra), [@TheDevilsDNA](#), [@pranav_m28](#) as the most consistent analysts. They always bring some new ideas in football data analysis. From them I am also inspired to do something new.

Also there are a lot of big names like [@totalf0otball](#), [@TotalAnalysis](#), [@footballyitics_](#), [@_JKDS_](#), [@Ligandro22](#), [@chunhang7](#) and many more people. I follow them to get new ideas from them. I really admire their ability of making new ideas of visualizations. I am no where near to their level and experience in this field, but they are the main inspiration for me to work in football data analysis.

CJ Round Up: Thank you so much for sharing your journey and work to date. I think many entry level analysts will be inspired from this, and realise the support out there. One of the big takeaways for me is both the number of resources available for free, but also the different metrics the community are currently interested in.

LOGGING OFF,
CJ

SCRAPING BALL-BY-BALL CRICKET DATA WITH YASH SAKHUJA

Hi all,

I am so pleased that Yash joins the site this week to talk through how to scrape cricket data. I met Yash just prior to his masters and to see how he has expanded his toolkit / techstack has been very admirable. I'm extra pleased that Yash gets to share his passion for cricket through the tutorial below, but first a few questions!

Yash Sakuja
Data Analyst | Jamshedpur, Jharkhand, India
Tableau Student Ambassador 2021-2022 | #VOTD X2
MSc Business Analytics, Operational Research & Risk Analytics © The University of Manchester...
[Read more](#)

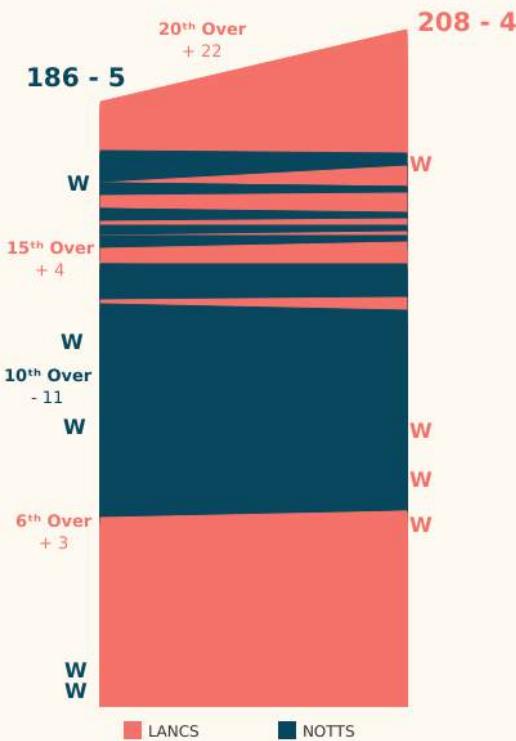
[Follow](#) [Hire Me](#)

Vizzes [11](#) Favorites [14](#) Following [1](#) Followers [1](#)

CJ: Yash, thanks for joining, for those that are unaware, could you tell us a little about your journey to date?

Concept : **Timeline Slope** for Cricket in Tableau

May, 27, 2023 | Emirates Old Trafford | **Lancashire Lightning vs Nottinghamshire Outlaws**



Design: Yash Sakuja | Data: Matchday Collection | Inspiration: Krisztina Szucs

Yash: Thank you so much for having me. Firstly, I'd like to mention that it's a pleasure to be featured on your blog, CJ. I truly admire your work and the impact you've made in the #datafam community.

Talking about my journey, I was set on a path to becoming a "Man in Finance" with my choice of undergraduate degree, but data and Tableau found me just before the COVID lockdowns of 2020.

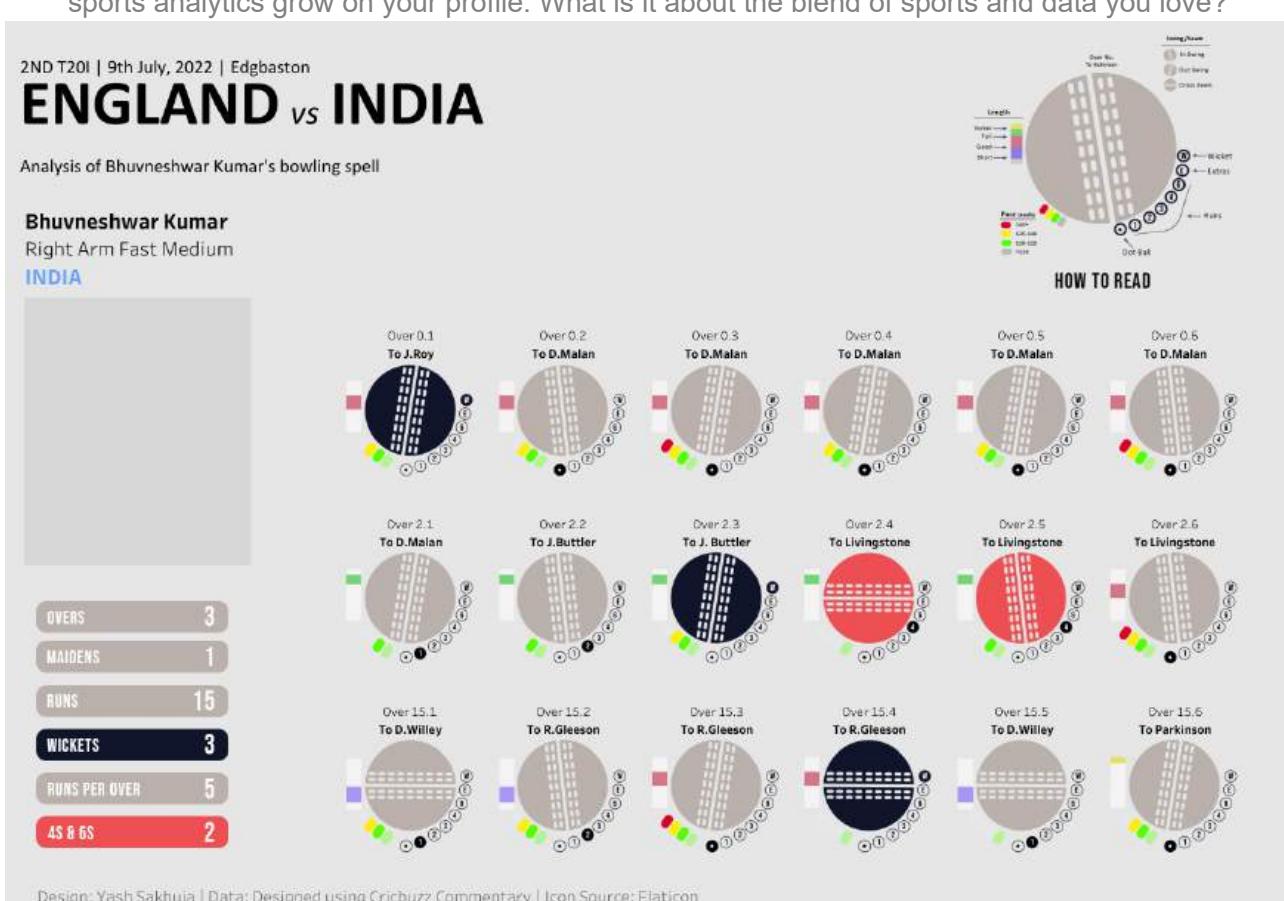
During that time, Tableau and the #datafam community kept me sane as I participated in Makeover Mondays and Sports Viz Sundays, learning from every visualization I created.

Consistency paid off, and my work started gaining attention. A turning point came when **Rajeev Pandey**, a key figure in the community, helped me land my first data role at Brandscapes Worldwide.

This role offered a steep learning curve and exposure to diverse clients, from retail to healthcare.

After completing my MSc in Business Analytics: Operational Research and Risk Analytics at The University of Manchester, I joined Performalytics, an Oxford-based consultancy firm. Now, I work with sustainable retail and sports clients, and in my free time, I love experimenting with Tableau, Python, ML and AI to keep learning and innovating.

CJ: You became active in the community in 2020, and I've had the pleasure of seeing some of your sports analytics grow on your profile. What is it about the blend of sports and data you love?



Yash: The core idea is that when you work on something you love, it feels more like fun than work. Growing up in India, I spent a lot of time playing and watching sports, especially cricket, since football timings weren't always ideal. This is where my passion for both sports and data truly intersected.

I've always been more of a tactician than a player, so sports analytics felt a natural fit. What I find most empowering is that, with data and technology at our fingertips, we can stay one step ahead in the game, analyzing tactical decisions through the lens of data. There's something uniquely satisfying about predicting something just before it happens or before experts/commentators mention it on air.

CJ: Your cricket content has been a massive hit. You really bend tableau to use it in innovative ways, such as in the bowling spell and cricket pitch visuals. For someone that is new to cricket data, what are some data sources that you would recommend?

INDIA 1st Innings

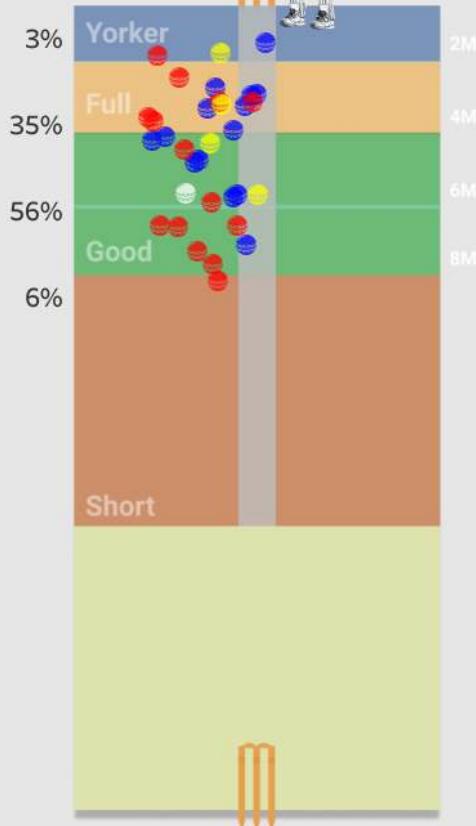
Lunch Break

India- 90/4

Pujara- 20* | Pant-10*



VS



35 of 34 balls



KOHLI- AGAINST ENG BOWLERS THIS SESSION



DOTS



RUNS



WICKETS



BOUNDARIES

Data: All Mock(thanks to randbetween) | Design: Yash Sakhua

Designed using Figma

Yash: That's a great question. As a newcomer, I personally struggled to find structured data for cricket. Historically, structured cricket data hasn't been as readily available as data for sports like football or baseball. However, things are improving with platforms like **Cricsheet**, making ball-by-ball cricket data much more accessible now.

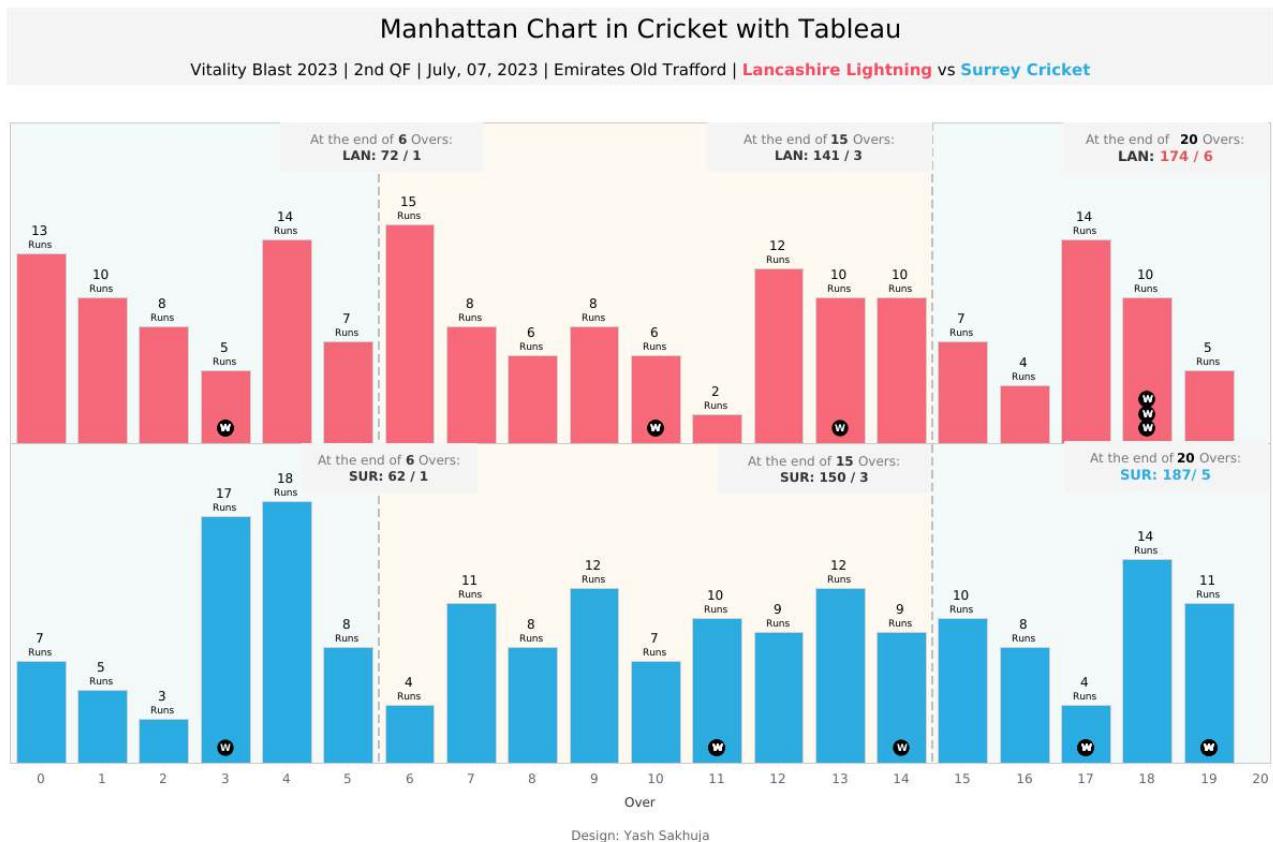
At an aggregated level, for player careers, team, and match records, I find Statsguru by **ESPN Cricinfo** to be useful, though it may require some data preparation. These are the two primary data sources I rely on, alongside scraping ball-by-ball cricket data for matches. I've provided a step-by-step guide on how to do this in this blog. Additionally, for true enthusiasts, there's always a way of collecting data while watching the matches—though it requires a great deal of effort and dedication. CJ: I love that you have continued to blog about how to replicate some of your work in Tableau, such as the worm chart. Is there anyone else in the cricket analytics community, or wider sports community that you often learn from or go to for inspiration?

Yash: When it comes to my inspiration in sports analytics, I truly believe you've played a key role in shaping it. Your work with football and tennis data has been incredibly inspiring, and your page is, without a doubt, the best resource for sports data enthusiasts. Participating in Sports Viz Sunday has also been crucial for connecting with the wider community, and I'm grateful for the learning from the

works of Simon Beaumont, James Smith, Mo Wootten, Chris Westlake, Naresh Suglani and many more.

As for my involvement in the cricket analytics community, I was fortunate to work with Lancashire County Cricket during my masters at Manchester, where I developed FieldAssist, a data collection and reporting tool for fielding data in cricket. This experience exposed me to a variety of coaches and performance analysts from different teams, allowing me to learn from their use of Tableau. Now, my blogs are aimed at helping them create both simple and complex cricket visualizations in Tableau. It's incredibly rewarding to receive messages from them saying that my blog helped them do build something easily in Tableau.

CJ: Recently, you blogged about how to create a Manhattan chart in Tableau. I love the step by step guide. You mention you scraped ball-by-ball data from the **ESPN Cricinfo**. Could you give us a walkthrough of what that looks like?



Yash: Absolutely! I'd be happy to walk you through a step-by-step guide on how I scrape ball-by-ball cricket data from ESPN Cricinfo. I'll also share a framework and python code one can use to apply this process to any match to retrieve the data.

Let's begin by understanding what web scraping is. It's a method used to extract data from websites. While manually copying and pasting data is an option, it's often tedious and time-consuming. Web scraping automates this process, making data extraction more efficient and streamlined. Almost all websites are built using HTML tags and attributes. So, before diving into web scraping, it's important to understand the basics of these tags and attributes, which are crucial for extracting data from websites. This hour-long video on **Web Scraping by Priyanka Dobhal** is an excellent resource for learning how HTML tags and attributes work and is a great starting point for beginners. I recommend that readers watch this video to gain a deeper understanding of how my scraping method works under the hood.

To extract ball-by-ball scores and commentary data from ESPN Cricinfo, I follow a four-step process, which I have outlined in this blog. Before you begin executing these four steps, it would be helpful to

clone this [GitHub Repository](#) and try to replicate the code on your end.

The four step scraping process:

1. Locate and save the webpage.
2. Inspect the web structure and create element list
3. scraping the data (scraper.py)
4. cleaning the data (scraped_data_cleaning.ipynb)

Let's walk through each part.

1. Locate and save the webpage

First, navigate to the specific match or series page on [ESPN Cricinfo](#), where the ball-by-ball commentary is available. Make sure you select the Commentary tab for the match, as shown in the image below. Then, choose the innings you want to extract data for by selecting the batting team from the dropdown list.

P.S. This process needs to be done one innings at a time, as both innings commentary is stored under the same link but displayed under different dropdown selections.

Surrey vs Lancashire, 2nd Quarter Final at Manchester, Vitality Blast, Jul 07 2023 - Ball by Ball Commentary

RESULT
2nd Quarter Final (D/N), Manchester, July 07, 2023, Vitality Blast

Surrey 187/5 (20 ov, T:168) Lancashire 174/6

Commentary

LANCS

END OF OVER 20: 8 runs

19.6 1 Abbott to Mitchell, 1 run

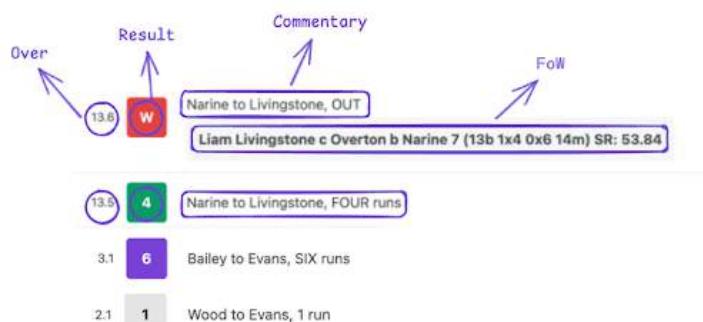
19.6 1 Abbott to Jones, 1 run

Next, scroll down to the section of the page showing the details of the first ball of the innings to ensure the entire webpage has loaded. Then, save the webpage by pressing Ctrl+S (or Cmd+S on a Mac). Rename the file for easy identification of the corresponding innings and save it in the HTML Links folder within the appropriate match folder. Afterward, change the dropdown selection for the batting team and repeat the process to save the HTML file for the other innings. With the HTML files now organized, we can move on to the next step, where we will inspect the files and begin extracting the necessary information.

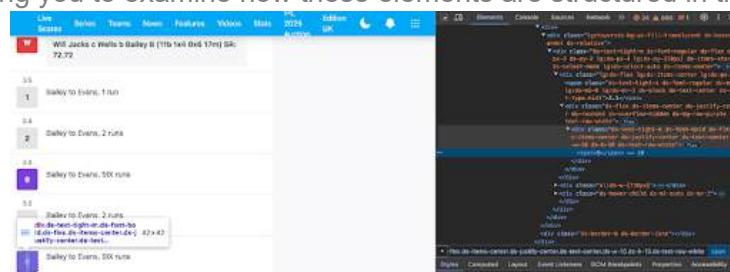
2. Inspect the Web Structure and Create the Element List

When examining how the commentary and scores are displayed, there are four key elements we need to extract: the Over, the Result, the Commentary, and the Fall of Wicket (FoW) (for wicket deliveries), as shown in the image below.

It's also important to note that the boxes have different colors for various events—red for Wickets, green for Boundaries, purple for Sixes, and translucent for other results. These color differences influence how they are represented in the HTML code.



The next steps involve revisiting the webpage, right-clicking on the elements you want to extract and selecting the “Inspect” option. This will open the Developer Tools panel on the right side of the window, allowing you to examine how these elements are structured in the HTML layout.



Refer to the video I mentioned earlier for a more detailed explanation of this process. For our purposes, the elements we need are structured within `<div>` or `` classes, as highlighted in the table and snippets below.

Element	Tags/Attributes
---------	-----------------

Over and Ball	<code></code> Class
---------------	---------------------------------

Result	Outermost <code><div></code> class
--------	--

Commentary	Outermost <code><div></code> class
------------	--

FoW	Outermost <code><div></code> class
-----	--

`` class for Over and Ball

```
<span class="ds-text-tight-s ds-font-regular ds-mb-1 lg:ds-mb-0 lg:ds-mr-3 ds-block ds-text-center ds-text-typo-mid1">18.5</span>
```

Outermost `<div>` class for Results

```
<div class="ds-flex ds-items-center ds-justify-center ds-rounded ds-overflow-hidden ds-bg-red ds-text-white"><div class="ds-text-tight-m ds-font-bold ds-flex ds-items-center ds-justify-center ds-text-center ds-w-10 ds-h-10 ds-text-white"><span>W</span></div></div>
```

Outermost `<div>` class for Commentary

```
<div class="ds-ml-4 lg:ds-ml-3 ds-text-typo-mid1"><div class="ds-leading-[16px] lg:ds-leading-none ds-mb-0.5"><span>Jordan to Vilas<!-- -->, <span>OUT</span></span></div></div>
```

Outermost `<div>` class for FoW

```
<div class="ds-rounded ds-bg-fill-content-alternate ds-ml-4 lg:ds-ml-3 ds-mt-1 ds-inline-block ds-p-3"><strong>Dane Vilas c Curran b Jordan 2 (3b 0x4 0x6 4m) SR: 66.66</strong></div>
```

Finally, we organize the extracted elements into a CSV file, including the batting team to indicate which HTML file the data came from, and linking each field to its corresponding `<div>` or `` class. Wicket balls, 4s, 6s, and other deliveries are listed as separate fields because they are enclosed within outer `<div>` boxes of different colors, which means their HTML coding slightly varies.

After compiling the CSV, save it in the Element List folder with an appropriate file name for easy identification.

Note: Always double-check and inspect the elements before moving to the next steps, as tags and attributes can vary depending on the website's design and structure, which may differ slightly for certain elements. For example, check how in the snippet below the FoW <div> class has subtle differences between the two innings.

Batting Team	Field	Link
SUR	Wicket Balls	ds-flex ds-items-center ds-justify-center ds-rounded ds-overflow-hidden ds-bg-raw-red ds-text-raw-white
SUR	4s Balls	ds-flex ds-items-center ds-justify-center ds-rounded ds-overflow-hidden ds-bg-raw-green-d2 ds-text-raw-white
SUR	6s Balls	ds-flex ds-items-center ds-justify-center ds-rounded ds-overflow-hidden ds-bg-raw-purple ds-text-raw-white
SUR	Other Balls	ds-flex ds-items-center ds-justify-center ds-rounded ds-overflow-hidden ds-bg-ui-fill-default-translucent ds-text-typo
SUR	Commentary	ds-ml-4 lg:ds-ml-3 ds-text-typo-mid1
SUR	Over and Ball Numbers	ds-text-tight-s ds-font-regular ds-mb-1 lg:ds-mb-0 lg:ds-mr-3 ds-block ds-text-center ds-text-typo-mid1
SUR	FoW	ds-rounded ds-bg-fill-content-alternate ds-ml-4 lg:ds-ml-3 ds-mt-1 ds-inline-block ds-p-1
LAN	Wicket Balls	ds-flex ds-items-center ds-justify-center ds-rounded ds-overflow-hidden ds-bg-raw-red ds-text-raw-white
LAN	4s Balls	ds-flex ds-items-center ds-justify-center ds-rounded ds-overflow-hidden ds-bg-raw-green-d2 ds-text-raw-white
LAN	6s Balls	ds-flex ds-items-center ds-justify-center ds-rounded ds-overflow-hidden ds-bg-raw-purple ds-text-raw-white
LAN	Other Balls	ds-flex ds-items-center ds-justify-center ds-rounded ds-overflow-hidden ds-bg-ui-fill-default-translucent ds-text-typo
LAN	Commentary	ds-ml-4 lg:ds-ml-3 ds-text-typo-mid1
LAN	Over and Ball Numbers	ds-text-tight-s ds-font-regular ds-mb-1 lg:ds-mb-0 lg:ds-mr-3 ds-block ds-text-center ds-text-typo-mid1
LAN	FoW	ds-rounded ds-bg-fill-content-alternate ds-ml-4 lg:ds-ml-3 ds-mt-1 ds-inline-block ds-p-3

3. Scraping with scraper.py

Once you've generated the HTML files and the element list, it's time to run the 100-line scraper.py code. This script combines both inputs and uses the element list to extract the necessary data from the HTML files.

Before running the code, make sure to update the necessary details (highlighted in green) in the first few lines of the scraper.py code based on the specific match you are extracting data for and with urls of where your element list and html files are saved.

```

1 match_date="2023-07-07" #yyyy-mm-dd
2 batting_team="SUR"
3 match_name="LAN vs SUR - 2023-07-07"
4
5 # URL where the Element list is saved
6 urls = pd.read_csv("Element List/ESPN Cricinfo Element Scrapping URLs (LAN vs
7 SUR).csv")
8
9 # URL where the respective HTML files are saved
10 if batting_team=="LAN":
11     doc_link="HTML Files/Lancashire vs Surrey - 2023-07-07/Ball by Ball
12     Commentary & Live Score - LANCS vs SUR, 2nd Quarter Final - Lancs
13     Batting.html"
14 elif batting_team=="SUR":
15     doc_link="HTML Files/Lancashire vs Surrey - 2023-07-07/Ball by Ball
16     Commentary & Live Score - LANCS vs SUR, 2nd Quarter Final- Surrey
17     Batting.html"
18 else:
19     print("Wrong Team Name Input")
20     exit()

```

Here's a code snippet for extracting elements using HTML tags and attributes. To view the complete code, click [here](#). The extracted data is stored in lists and then exported as CSV and JSON format within the "scraped Data" folder, under the corresponding match-specific sub-folder.

```

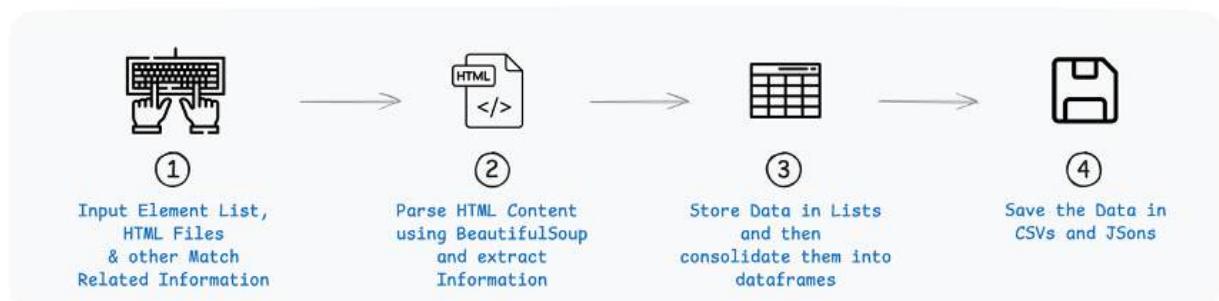
1 # For over and ball
2 for span in innings_soup.findAll('span', {
3     'class': urls[url["Field"]] == "Over and Ball Numbers"]["Link"].iloc[0]}):
4     ball_numbers.append(span.text.strip()) # Extract and clean the text
5
6 class_names = [ #Green for 4s, Purple for 6s, Red for W's, Translucent Else
7
8 urls[url["Field"]] == "Other Balls"]["Link"].iloc[0], # Other Deliveries
9 urls[url["Field"]] == "4s Balls"]["Link"].iloc[0], # 4s
10 urls[url["Field"]] == "6s Balls"]["Link"].iloc[0], # 6s
11 urls[url["Field"]] == "Wicket Balls"]["Link"].iloc[0] # Wickets Results
12
13 # Iterate over all div elements in the soup
14 for div in innings_soup.findAll('div', {'class': class_names}):
15     span = div.find("span")
16     if span:
17         score = span.text.strip() # Extract the score text
18         ball_score.append(score)
19
20
21 # for description of each ball- Commentary
22 for div in innings_soup.findAll('div', {'class': urls[url["Field"]] == "Commentary"]["Link"].iloc[0]}):
23     ball_desc.append(div.text)
24
25
26 # for fall of wickets
27
28 for div in innings_soup.findAll('div', {'class': urls[url["Field"]] == "FoW"]["Link"].iloc[0]}):
29     strong = div.find("strong")
30     if strong:
31         fow = strong.text.strip() # Extract the score text
32         wicket_des.append(fow)

```

After running the `scraper.py`, the output CSV will be generated in the following format, with one file for each innings. Please, note that you would need to run this code twice, once for each innings, after changing relevant information.

	ball	score	commentary
0	0.1	1	Jacks to Salt, 1 run
1	0.2	1	Jacks to Buttler, 1 run
2	0.3	4	Jacks to Salt, FOUR runs
3	0.4	1	Jacks to Salt, 1 run
4	0.5	6	Jacks to Buttler, SIX runs

Here's a brief overview of how the code in `scraper.py` is structured and the sequence of operations it performs to scrape the necessary data:



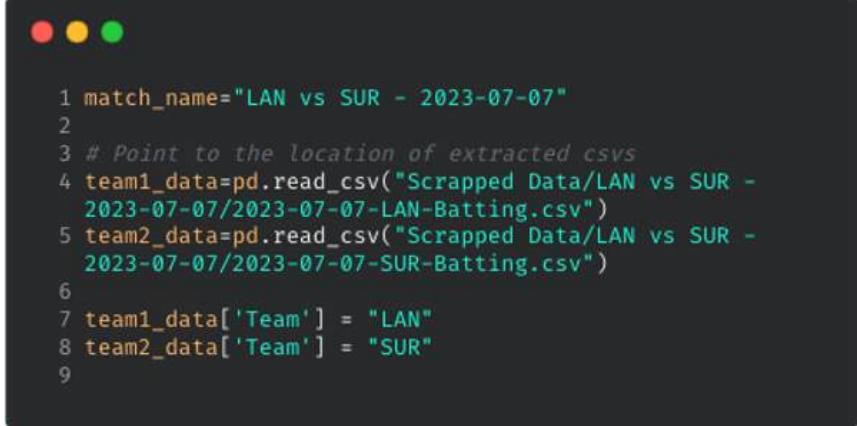
4.

Data Cleaning with `scrapped_data_cleaning.ipynb`

In the two Tableau blogs that I wrote for building a **Worm Chart** and a **Manhattan Chart**, I pulled the data into Tableau in the format as shown below:

Abc	#	#	#	#	#
Sheet1	Sheet1	Sheet1	Sheet1	Sheet1	Sheet1
Team	Over	Ball	Ball ID	Runs	Wickets
SUR	0.0000	1	1	0	0
SUR	0.0000	2	2	0	0
SUR	0.0000	3	3	0	0
SUR	0.0000	4	4	0	0
SUR	0.0000	5	5	2	0
SUR	0.0000	6	6	0	0
SUR	0.0000	6	7	4	0
SUR	1.0000	1	8	0	0

To structure our scraped output, I've also created a **Jupyter Notebook** file in the repository. By pointing to the two extracted CSVs and modifying a few match-related details (as highlighted in green in the code shown below), you can clean the data and process it using basic pandas operations in Python, making it ready for visualization and analysis in Tableau.



```

1 match_name="LAN vs SUR - 2023-07-07"
2
3 # Point to the location of extracted csvs
4 team1_data=pd.read_csv("Scrapped Data/LAN vs SUR -
2023-07-07/2023-07-07-LAN-Batting.csv")
5 team2_data=pd.read_csv("Scrapped Data/LAN vs SUR -
2023-07-07/2023-07-07-SUR-Batting.csv")
6
7 team1_data['Team'] = "LAN"
8 team2_data['Team'] = "SUR"
9

```

In the GitHub repository for this scraping project, I've included the HTML files and an Element List for an additional match (RCB vs SRH – 2024-03-30), along with the corresponding outputs. Additionally, I've provided HTML files for a more recent T10 fixture (Deccan Gladiators vs Team Abu Dhabi T10 – 2024-11-23) for those interested in trying it out.

Thanks for your time, CJ. I hope the readers find this helpful and I'm also excited to see the community expand on this, such as exploring how this approach works for different formats of the game. If anyone has any questions or ideas for improvements, feel free to reach out to me on my

socials & my blog here. Cheers, Yash.

LOGGING OFF,

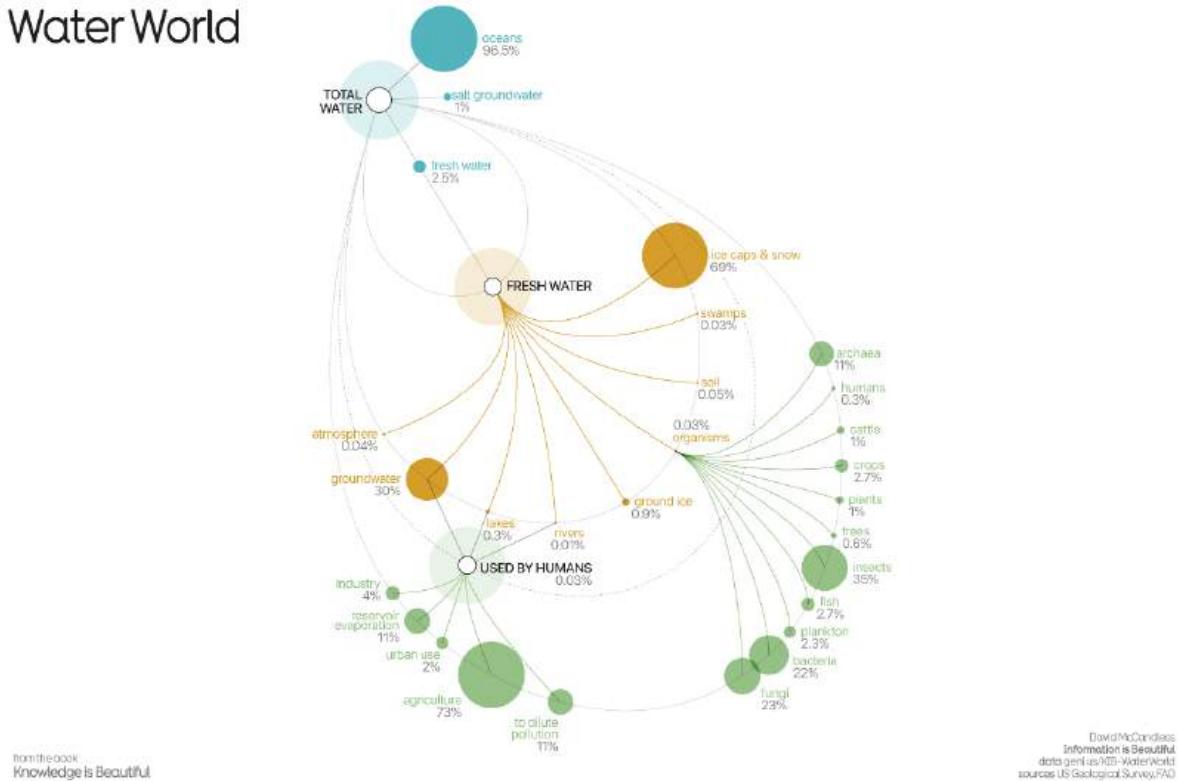
CJ

ROTATING PETAL POLYGONS IN TABLEAU

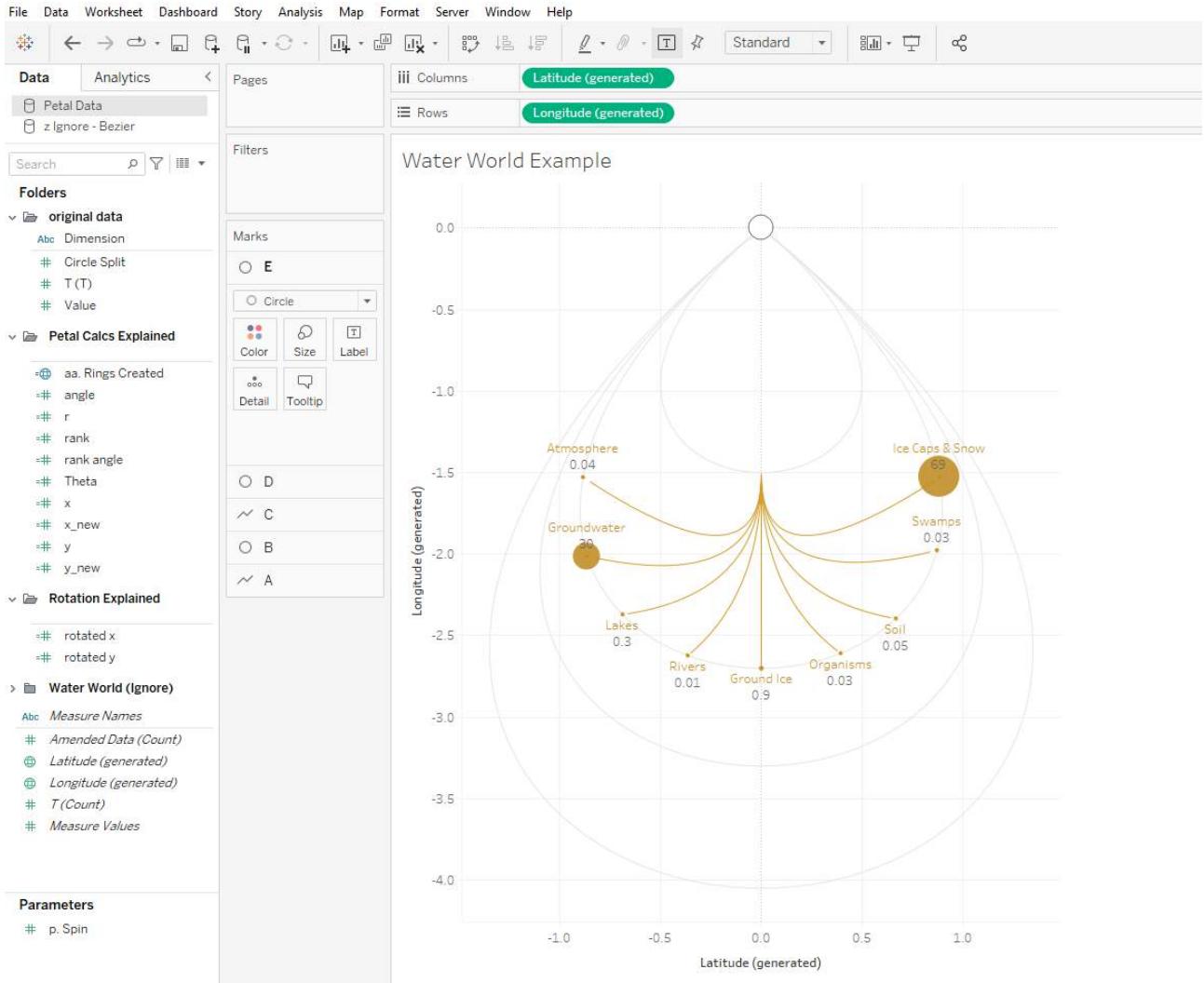
Hi all,

Back this week to walk through some basic petal making maths for Tableau. A few weeks back the Makeover Monday challenge was looking at the Water World distribution chart that was in the information is beautiful.

Water World



Now I was going to recreate it in Tableau in full, got perhaps half way through, realised it was possible and quickly moved on.



It did raise for me however two things that we could probably learn from.

The first is how to create this petal / teardrop shape instead of where we use to create circles. Notice how they all curve towards a singular point at the top instead of have a rounded stretched circle feel.

The second aspect is, the Water world actual design is rotated slightly anti clockwise.

Those two aspects we will learn in this blog!

Lets start with a really basic dataset, that will contain our 4 rings

	A	B	C
1	Dimension	Circle Split	
2	Ring 1		1
3	Ring 2		2
4	Ring 3		3
5	Ring 4		4
6			

and full outer join it to a sheet called T with numbers going to 360, for data densification, using a full outer join of 1=1

	A	B
1	T	
2		1
3		2
4		3
5		4
6		5
7		6
8		7
9		8
10		9
11		10
12		11
13		12
14		13



⌚ Rings+ (Petal)

Rings is made of 2 tables. ⓘ

Rings ▾ 3 fields 1440 rows

Name	Type	Field Name	Physical Table	Remote Field Name
Rings	Abc	Dimension	Rings	Dimension
	#	Circle Split	Rings	Circle Split
	#	T	T	T

Dimension	Circle Split	T
Ring 4	4	1
Ring 3	3	1
Ring 2	2	1
Ring 1	1	1
Ring 4	4	2
Ring 3	3	2
Ring 2	2	2
Ring 1	1	2
Ring 4	4	3
Ring 3	3	3

Next thing we want to do is the calculations to convert the points into a teardrop / petal shape.

First lets create the angle

360/360

This is because we have 360 marks.

next lets create the rank

Similar to a circle shape the we then rank each mark.

Call this calculation rank angle

*

We continue our calculations similar to how we would make a circle,

In this case create x as

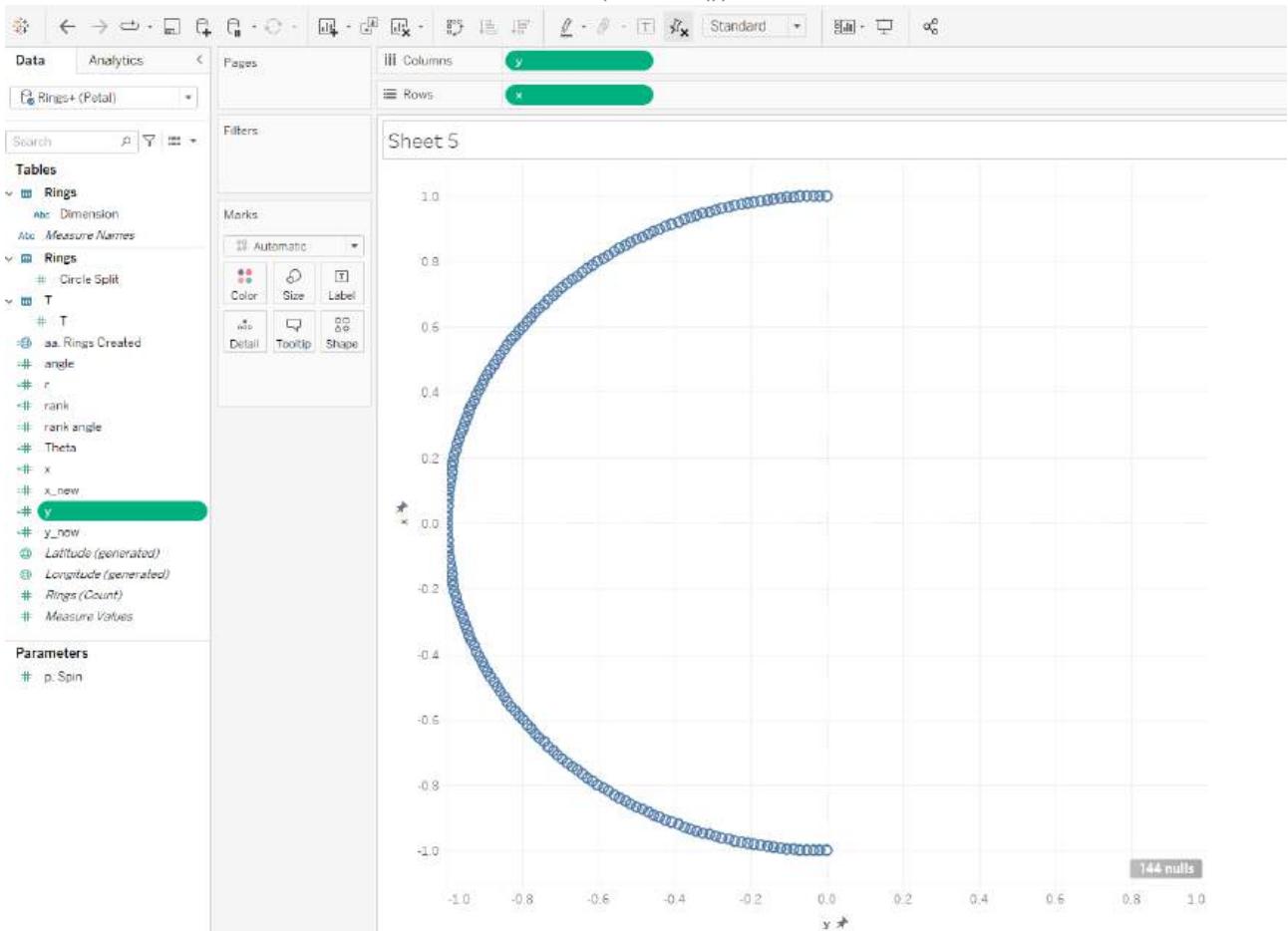
if ≥ 180 then

$\cos(\text{RADIANS}())$

END

and y as

$\sin(\text{radians}())$



You're probably familiar so far with these calcs. We are removing 180 of the points through the x calculation, which gives us a semi circle for now.

Then lets create a calculation called Theta

$\text{ATAN2}(,)$

- It's used to convert Cartesian coordinates (x,y) to polar coordinates (r,θ).
- The angle θ (theta) is crucial for shaping the teardrop, as the transformation applies different effects based on this angle.

Understanding ATAN2 is important because it allows you to work with angles in a way that correctly handles all quadrants of the 2D plane, which is essential for creating symmetrical shapes like your teardrop.

Now we are in a position to make them into our shape.

Y_{new} can be calculated as

$1 * \text{SIN}() - 0.5 * (\text{SIN}() * \text{SIN}())$

This equation determines the y-coordinate of points on the curve.

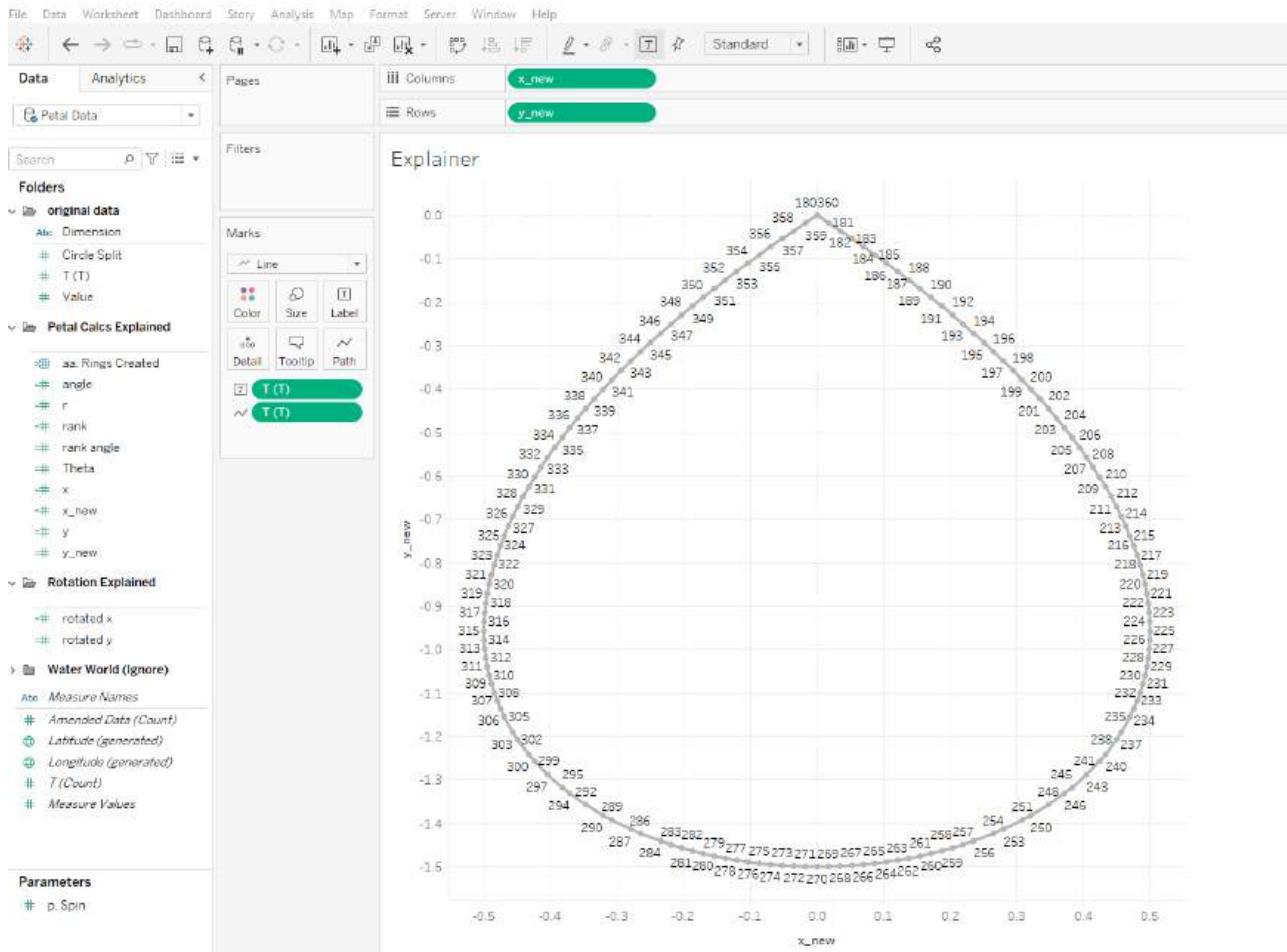
- $1 * \text{SIN}()$: This is a simple sine wave with amplitude 1.
- $0.5 * (\text{SIN}()) * \text{SIN}()$: This is half of the square of the sine function.
- Subtracting the second term from the first creates the asymmetry that gives the cardioid its characteristic shape. X_{new} can be calculated as

$$1 * \text{SIN}() * \text{COS}()$$

This equation determines the x-coordinate of points on the curve.

- $\text{SIN}() * \text{COS}()$: This product of sine and cosine functions is also known as the sine of double the angle ($\sin(2\theta) = 2\sin(\theta)\cos(\theta)$).
- The factor of 1 doesn't change the shape but sets the scale.

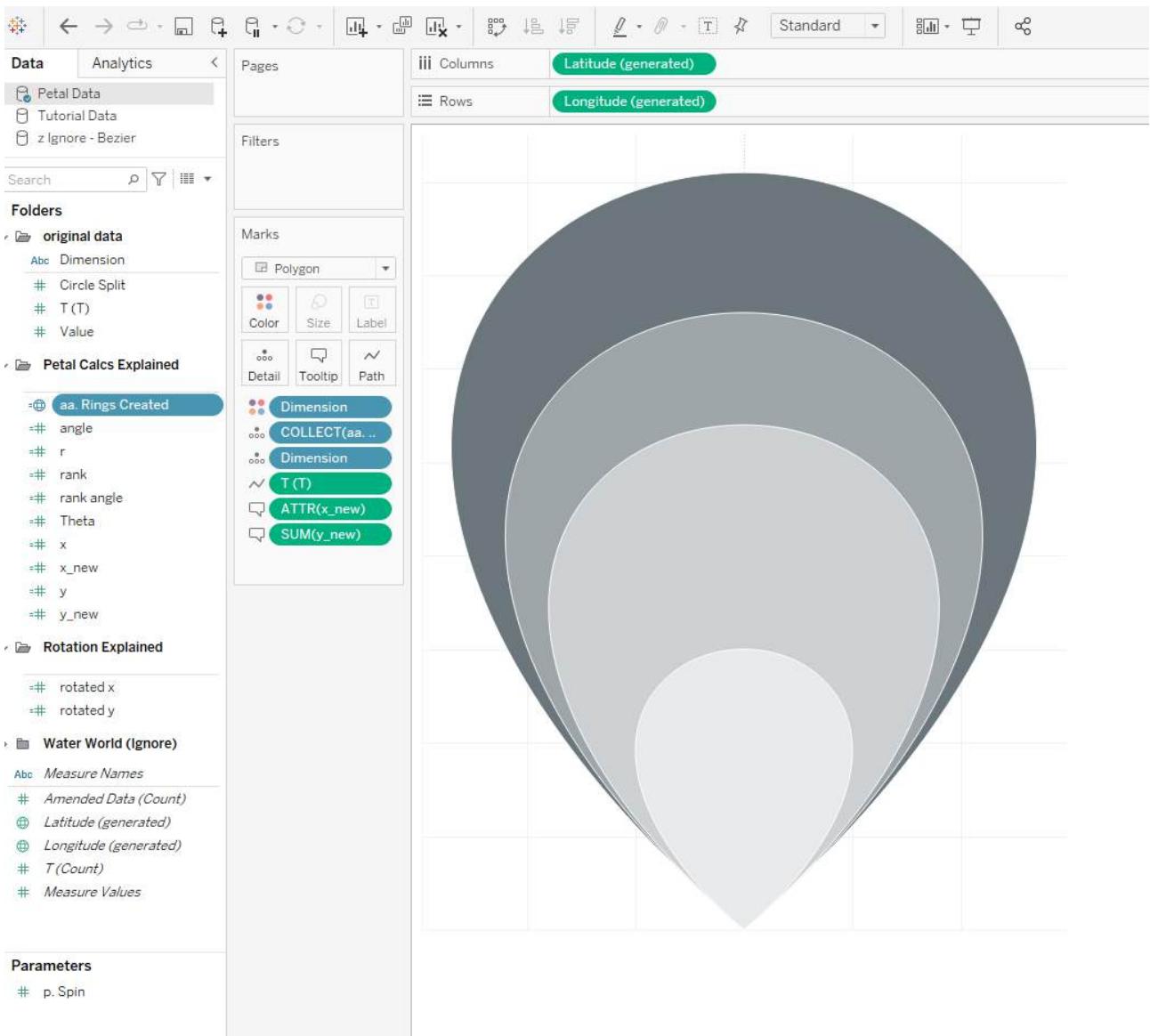
These equations create a shape that starts at the origin (0,0) when $\theta = 0$, loops around to form the rounded part of the teardrop / petal, and then returns to the origin to form the pointed end.



The final thing to recognise is that we can then of course utilise our initial data of the different “rings” to be able to split these out by arbitrary scales.

```

if = "Ring 1" then
    MAKEPOINT()
elseif = "Ring 2" then
    MAKEPOINT(*1.8,*1.8)
        elseif
            = "Ring 3" then
                MAKEPOINT(*2.2,*2.2)
            elseif = "Ring 4" then
                MAKEPOINT(*2.7,*2.7)
        END
    
```



So that is how we have taken logic from previously making circles, to now transform that shape into a petal / teardrop.

Skipping a stage of how to create chords (as I've written blogs on that before) lets go ahead and look at now how we can rotate a chart around a point.

The calculations we need for that are.

$$x' = \cos(\theta) * (x - cx) - \sin(\theta) * (y - cy) + cx$$

&

$$y' = \sin(\theta) * (x - cx) + \cos(\theta) * (y - cy) + cy$$

1.

They rotate a point (x, y) around the center point (cx, cy) by an angle θ .

2.

Positive θ results in counterclockwise rotation, negative θ in clockwise rotation.

3.

The distance from (x, y) to (cx, cy) remains the same after rotation.

•

These equations allow you to rotate shapes or sets of points around any arbitrary center point.

First create a parameter called p. Spin as a integer between 0 and 360. (for a full rotation)

Edit Parameter [p. Spin] X

Name		
<input type="text" value="p. Spin"/>		
Properties		
Data type	Display format	
<input type="text" value="Integer"/>	<input type="text" value="30"/>	
Current value	Value when workbook opens	
<input type="text" value="30"/>	<input type="text" value="Current value"/>	
Allowable values		
<input type="radio"/> All <input type="radio"/> List <input checked="" type="radio"/> Range		
Range of values		
<input checked="" type="checkbox"/> Minimum	<input type="text" value="0"/>	<input checked="" type="radio"/> Fixed
<input checked="" type="checkbox"/> Maximum	<input type="text" value="360"/>	<input type="radio"/> When workbook opens
<input checked="" type="checkbox"/> Step size	<input type="text" value="15"/>	<input type="button" value="Add values from"/>
		<input type="button" value="Cancel"/> <input type="button" value="OK"/>

Next lets create our new x and y calcs.

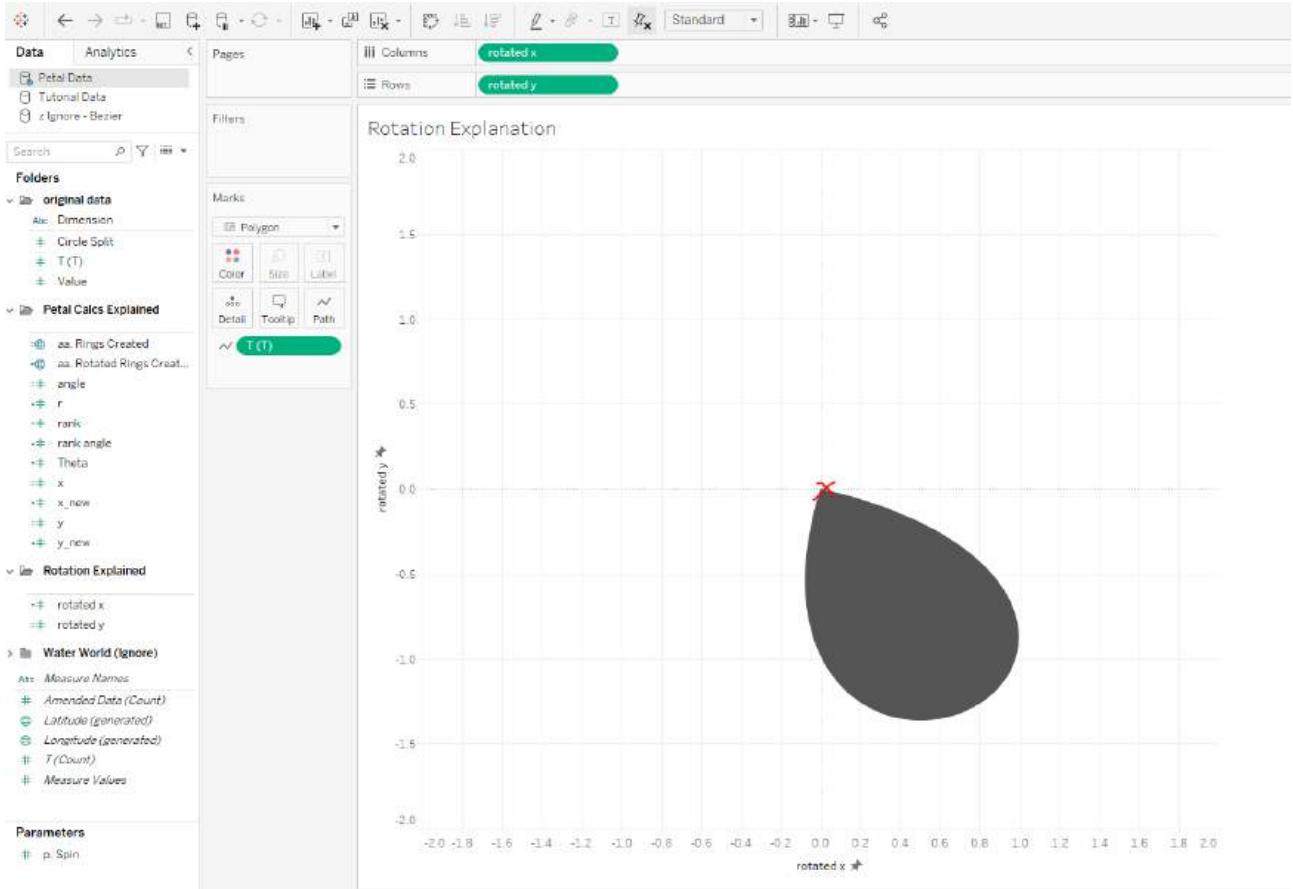
rotated x as

$$\cos(\text{radians}()) * (-0) - \sin(\text{radians}()) * (-0) + 0$$

rotated y as

$$\sin(\text{RADIANS}()) * (-0) + \cos(\text{radians}()) * (-0) + 0$$

These calculations rotate around the point 0,0. But we could change this if we wanted.

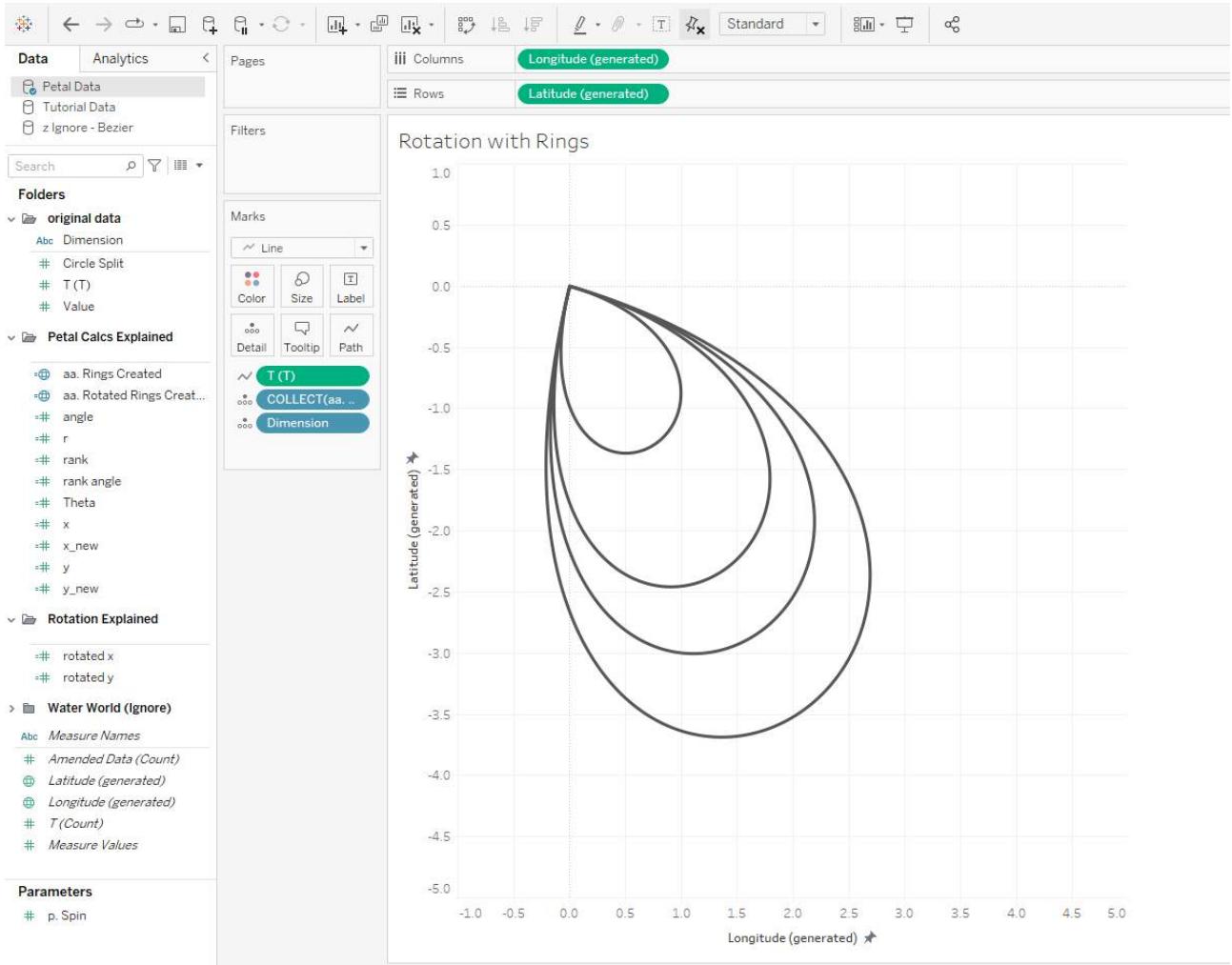


The final step if you want multiple rings would be similar to before but using our newly rotated x and y calcs.

```

if = "Ring 1" then
    MAKEPOINT(,)
elseif = "Ring 2" then
    MAKEPOINT(*1.8,*1.8)
elseif
    = "Ring 3" then
    MAKEPOINT(*2.2,*2.2)
elseif = "Ring 4" then
    MAKEPOINT(*2.7,*2.7)
END

```



Hopefully from the tutorial today you have learnt the maths behind making a petal / teardrop shape at one end of our circle.

You have been also been able to rotate that shape around a chosen point.

The workbook as always, can be found on my profile. Link under the header at the **top of the page**.

LOGGING OFF,
CJ

BUSINESS DASHBOARDS WITH GBOLAHAN ADEBAYO

Hi all,

Welcome back to another spotlight session. I have known Gbolahan for a while now and have really admired both his growth in the community and his desire to pass that passion on to others. Today we sit down to explore how he tailored his content to help build his personal brand.

CJ: For those that are unaware, could you tell us a little about your journey to date? How did joining the community in 2022 help shape your passion for data visualization?



Gbolahan Adebayo He/Him

2x Tableau Public Ambassador | Data Viz. Developer | 6x #VOTD | Tableau Featured Author '23 | 2x Vizzie Awards | Lagos, Nigeria

From Data to Ta-Da!

[Read more](#)

[Follow](#)

[Hire Me](#)



Vizzes: 50 Favorites: 25 Following: 29 Followers: 170

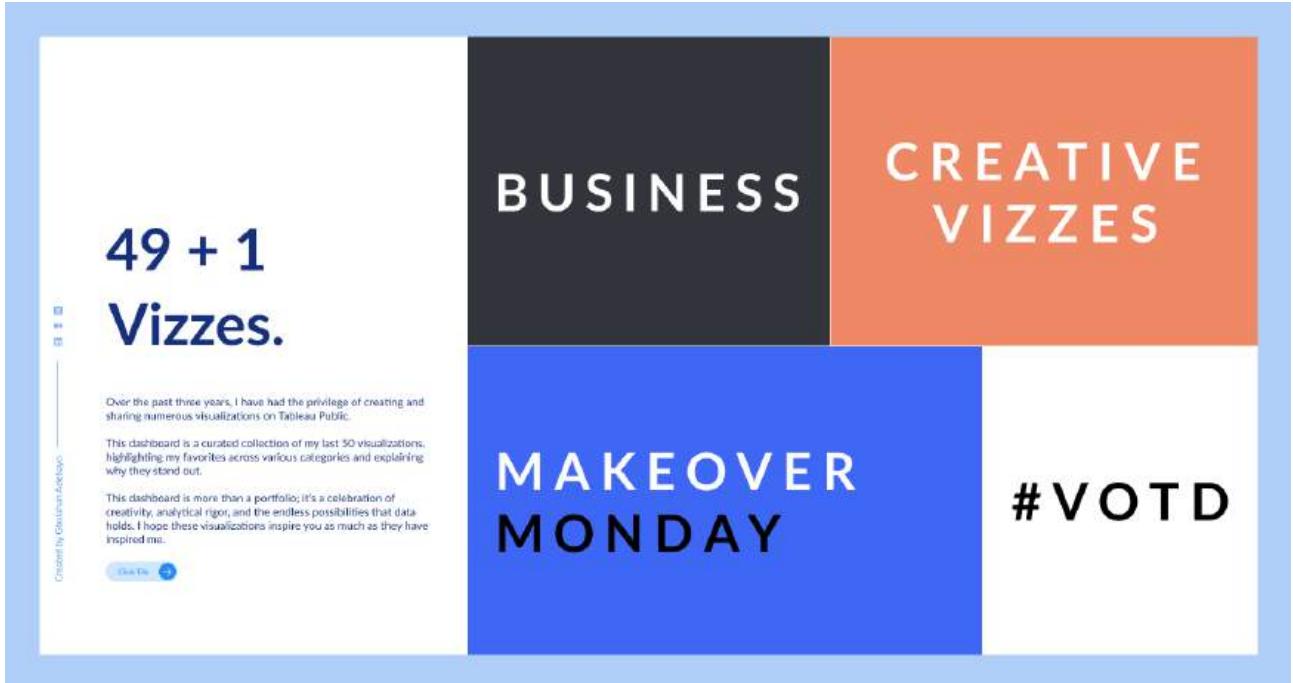
GB: I've been using Tableau for a while, but it wasn't until 2022 that I truly became part of the community, thanks to my friend **Idris Akilapa**. I reached out to him, eager to recreate some of the amazing visualizations I had seen on Tableau Public. We got on a call, where he explained what I needed to do and introduced me to **Makeover Monday**. This became my first significant project in joining the community.

By participating in Makeover Monday, I got access to watching **Andy Kriebel** and **Eva Murray** discuss and create live visualizations. This motivated me to improve my consistency and start sharing my visualizations on Tableau Public, Twitter, and LinkedIn to showcase my progress. Last year, I decided to take things up a notch by pushing out even more projects. I realized the importance of developing a personal style, although initially, I was more into exploring various inspirations.

This year, I focused on creating business dashboards to open up more opportunities, and I had insightful conversations with **Chimdi Nwosu**, who played a pivotal role. We started meeting every Sunday to discuss design and our ongoing projects. This regular exchange significantly contributed to my growth this year, helping me craft a unique style for myself.

One of the most memorable milestones in my journey was winning the Notable Newbie award at the 2022 Tableau Conference. It was an incredibly affirming moment that made me realize the impact of my contributions to the Tableau community. It felt amazing to be recognized as someone who was making a difference. This year, receiving the Vizzes award for Biggest Growth was even more special. It showed me that the community continued to see my efforts and recognized my evolution as I worked towards developing a personal style. It's been an incredible journey, and these awards have only fueled my passion further.

CJ: Over time individuals tend to find their "thing" within the data visualisation space. I often associate with you, really clean design, particularly when it comes to business dashboards. How important is this as part of your personal brand? Is this something you proactively worked towards?



GB: Yes, CJ, this is an important part of my brand, and like I said, I took it upon myself this year to create a style. Last year, if you remember, you and I had a discussion after I won the Notable Newbie Award about how I could move on in my career and in my visualization as well.

I felt that the Notable Newbie was an award to recognize those that are new to the community, and after you win such an award, you have to think about the next steps to elevate your game. You said to me, *"It's just good to find your niche. Choose a particular part of the community where you feel like you excel and deep dive into it."*

I had some deep thoughts about what part of the field I liked the most, and I remembered from the very start that I wanted to highlight the amazing designs of others, particularly the design work of Chimdi Nwosu.

So, I reached out to Chimdi, talked to him about what I was thinking, and told him that I would like to get inspired by some of his processes and what he does. He asked if I would be down to meet once a week to discuss design and the projects he was working on. Through this exchange of ideas, I aimed to improve my work and build my own style.

We started this initiative, and I began to observe how he approached various tasks. I would create visualizations, share them with him, and he would critique my work, providing feedback and suggestions for improvement. One particular dashboard where I really felt his impact was the Press Freedom dashboard, which went on to be shortlisted for the Information is Beautiful Awards. I created an initial idea, sent it to him, and he strongly suggested improvements, even showing me a prototype of what the final result should look like in terms of colors, design, text placements, and formatting.

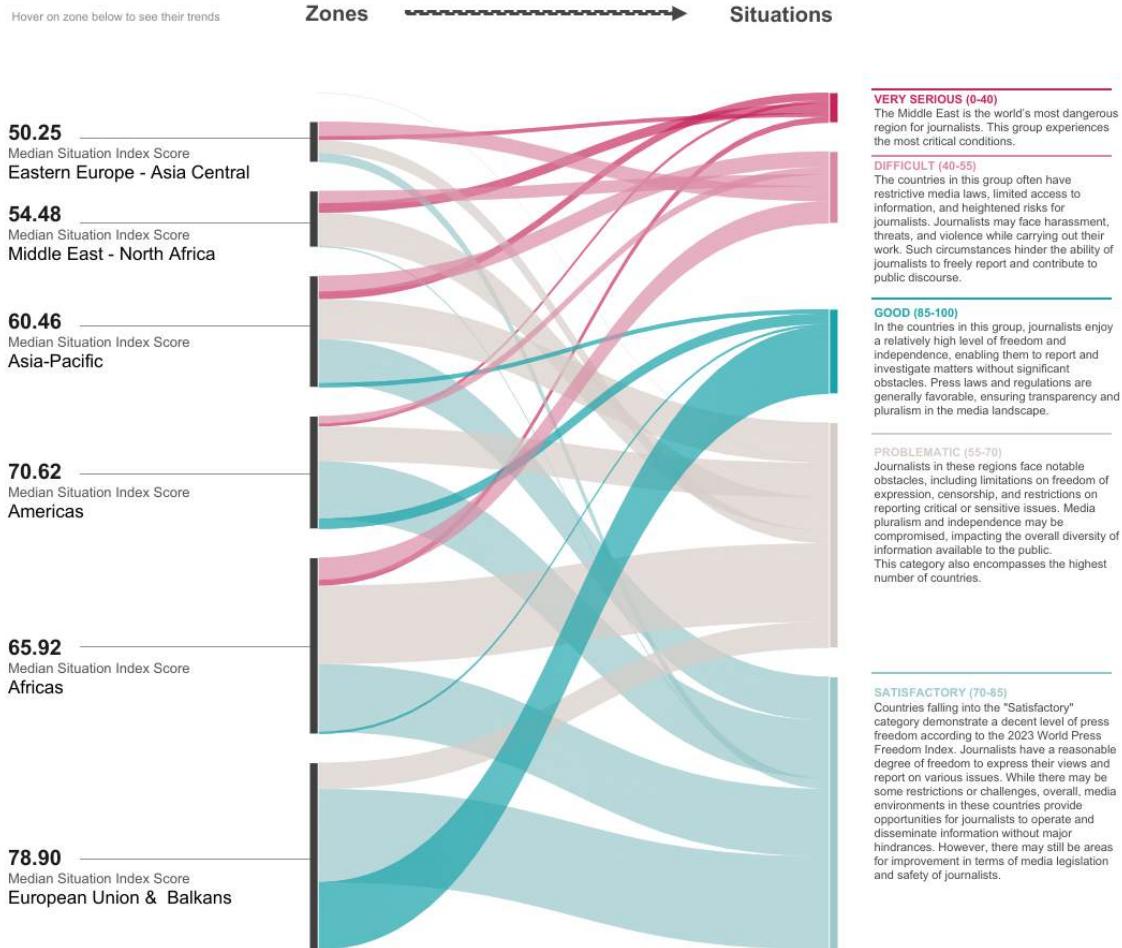
Press Freedom Index

UNDERSTANDING THE WORLD PRESS FREEDOM INDEX

The World Press Freedom Index, compiled by Reporters Without Borders (RSF), ranks 180 countries based on indicators of press freedom. It assesses media independence, pluralism, transparency, and the safety of journalists. It serves as a tool to monitor and compare the state of press freedom worldwide. Lower scores indicate greater press freedom, while higher scores reflect challenges and restrictions faced by journalists. The index is widely used to understand the global landscape of press freedom.

What is happening to Press Freedom?

The free press is under attack from multiple sources. Media outlets are closing their doors, victims to a broken business model. In much of the world, journalism is morphing into propaganda, as governments dictate what can and can't be printed. In the last year alone, hundreds of reporters have been killed or imprisoned for doing their jobs. The UN reports that 85% of the world's population experienced a decline in press freedom in their country in recent years.



Press Freedom Index Scores Across All Countries In 2023

Each circle depicts each country in the view. Hover on each country to see more information.

● Very Serious ● Difficult ● Problematic ● Satisfactory ● Good

DIFFICULT (40-55)
According to the 2023 World Press Freedom Index, Press Freedom situation is difficult in 42 countries

VERY SERIOUS (0-40)
According to the 2023 World Press Freedom Index, Press Freedom situation is very seriou...

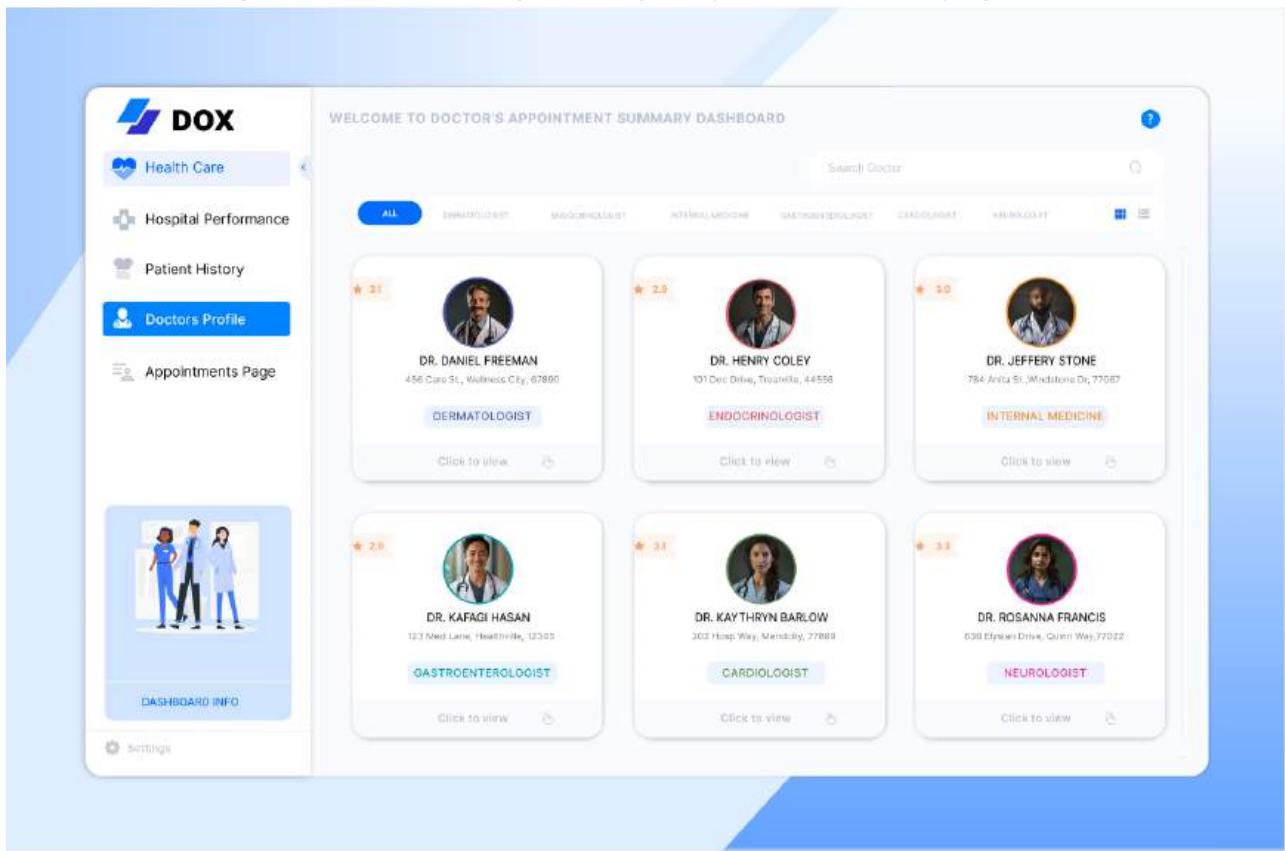
PROBLEMMATIC (55-70)
According to the 2023 World Press Freedom Index, Press Freedom situation is problematic in 55 countries

SATISFACTORY (70-85)
According to the 2023 World Press Freedom Index, Press Freedom situation is satisfactory in 44 countries.

GOOD (85-100)
According to the 2023 World Press Freedom Index, Press Freedom situation is good in 8 countries.

After this experience, I realized how much I had to improve. He also shared how to use Pinterest for inspiration, encouraging me to follow design ideas closely.

My first dashboard towards discovering this new style was my Doctor Appointment dashboard, a collaboration with a friend in the community, Israel Ayola. With this dashboard, I aimed to create something with a nuanced design, looking really clean and conveying a web UI feel.



While working on it, I posted it midway, and Chimdi noticed. He asked why I had shared it unfinished and encouraged me to complete it. We finished it, and it received very positive feedback from the community.

After this project, I felt like I had found my style, so I started gathering inspirations on Pinterest every day. I intentionally visit Pinterest to collect ideas for my dashboards. Continuous practice, trying out new business projects, and the advent of Chargeability allowed me to create datasets and explore various dashboard types, all contributing to my journey of building a personal brand that focuses

heavily on design and UI in business dashboards.

So yes, definitely, this is something I've proactively worked towards.

CJ: How has working as a freelancer, consultant and contractor played a part in understanding how different companies utilize dashboards and the requirements behind creating a business dashboard?

GB: Working as a freelancer, consultant, and contractor has significantly broadened my experience, especially in understanding how companies want to utilize dashboards compared to creating them for my portfolio. I've learned that in a business environment, it's not just about showcasing what you

want, but about delivering what needs to be seen.

Working as a Tableau consultant for large companies has shown me the importance of making sure that even the most basic elements are understood before advancing. It's a rigorous process, starting from gathering and understanding requirements, ensuring the design's usability, to testing, gathering feedback, and iterating to meet stakeholders' needs.

As a freelancer, working on diverse projects has given me confidence in adapting to various requirements and providing tailored solutions. In conclusion, my experiences across freelancing, consulting, contracting, and full-time roles as a Data Visualization Developer have shown me that

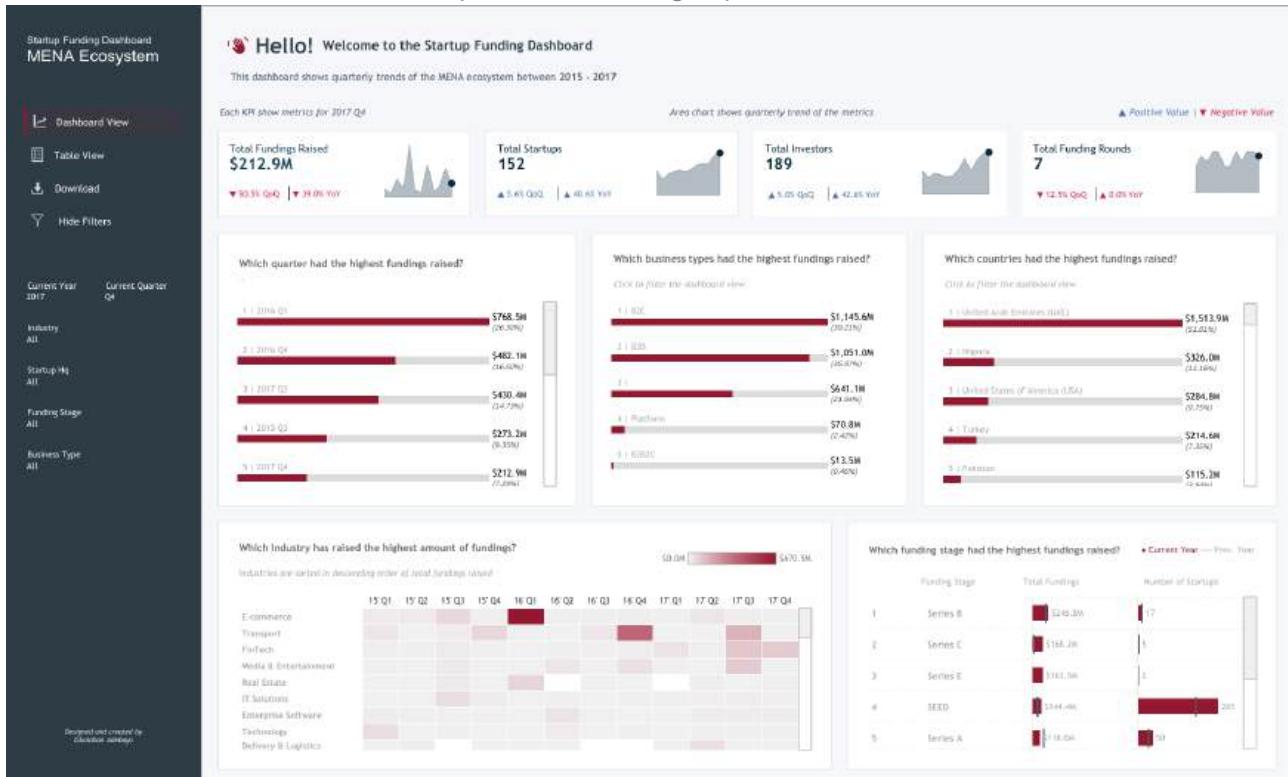
while building business dashboards can be challenging, they are crucial for reporting accurately and provoking insights.

The goal is to create dashboards that prompt stakeholders to ask questions and delve deeper into the data, driving informed decision-making. I've come to appreciate the intricacies of this process and its importance in the business world.

CJ: I'd like to jump into some of the technical behind your business dashboards, from your Tableau Public profile.

CJ: Your startup funding dashboard is wonderful in typography, white space and padding.

You opted to go for titles for each chart that pose as questions. What was the choice behind answering definitive questions? How do you strike the balance for dashboards answering specific questions vs being explorative?



GB: My Startup Funding dashboard was created during my interview process for my current role. The idea was to showcase my skills in the context of startup funding. I found inspiration in the revenue dashboard by **Pradeep Kumar** from the Tableau Public community. This dashboard was meant to demonstrate my approach to unfamiliar datasets and industries, showcasing how I navigate and visualize data.

I focused on showcasing relevant KPIs, conducting research on the types of questions stakeholders in the startup funding industry might ask. This informed the KPIs I included and how I structured them by quarter to provide a comprehensive view. I used a left-sided navigation bar for easy access to different views and included features like highlighting specific months in trend graphs for user interactivity.

To frame the charts, I chose a question-based style to provoke stakeholders to engage with the data, guiding them to answers while encouraging further exploration. This approach helped me focus on specific data points I wanted to highlight, making the dashboard informative even with my limited familiarity with the dataset.

As for whitespace, padding, and typography, I was inspired by **Pradeep Kumar's** work. I learned the importance of ample padding, whitespace, and clear typography for readability and visual hierarchy.

These elements were crucial in making the dashboard legible and allowing users to follow the data story effectively. Striking a balance between answering specific questions and being explorative depends on my familiarity with the dataset. In this case, I focused on specific questions due to my limited knowledge of the industry, aiming to demonstrate thorough research.

With more familiarity, I would incorporate more explorative elements to connect different data points and provide deeper insights.

CJ: Dashboards such as the Doctor Appointment dashboard bring with it a modern, website-like feel to the reporting. The booking / landing page style compliments the appointment theme. What thoughts go into designing the dashboards with “themes”?

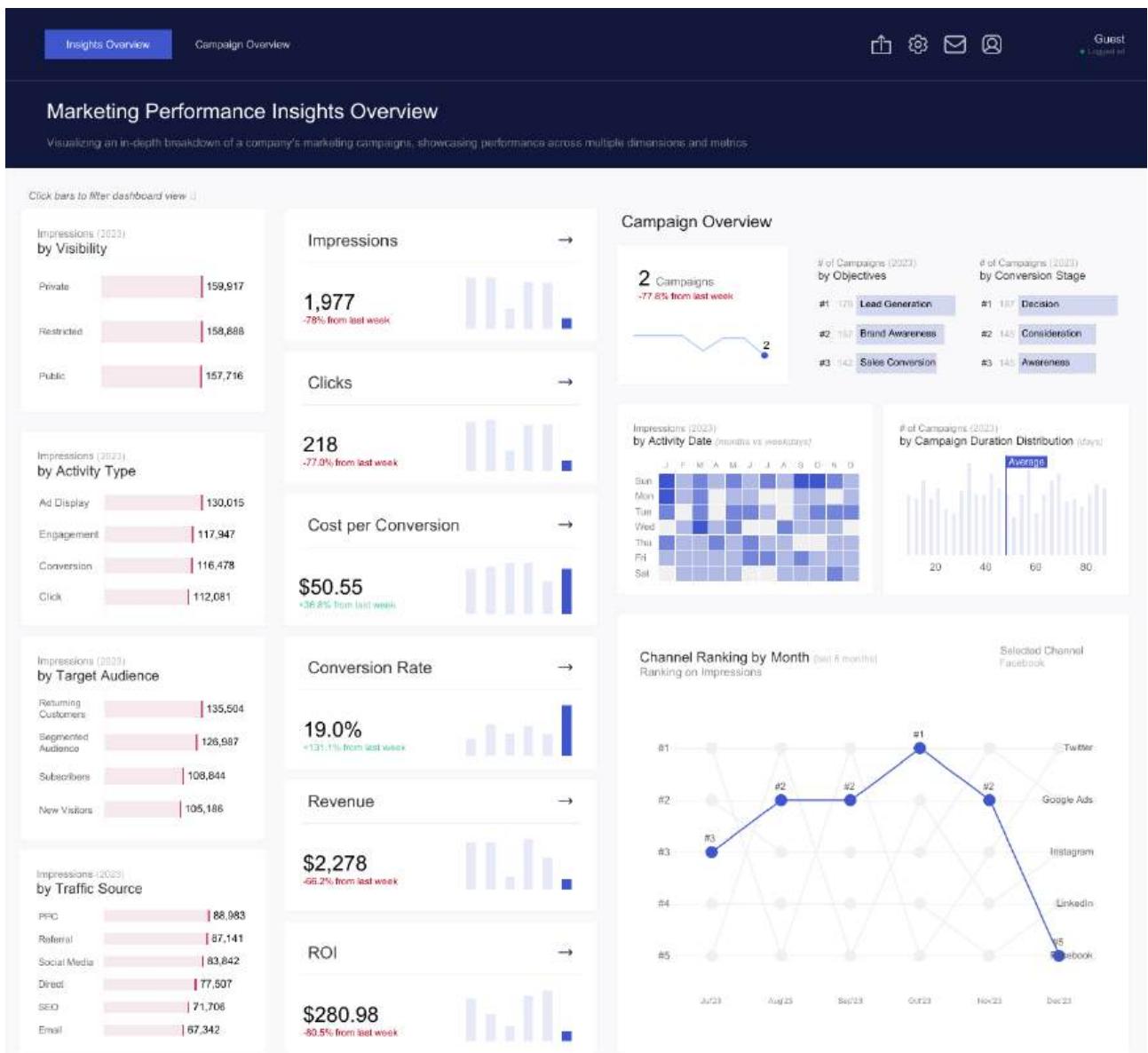
Booking ID	Patient ID	Patient Name	Gender	Avg Satisfaction Rating	Total Duration	Appt. Resolution
BK39411052381	10523	Noah Johnson	Female	4.8	59	No
BK109210829209	10829	Noah Yamamoto	Female	2.5	59	Yes
BK117810659678	10659	John Khan	Female	0.5	59	No
BK194110159133	10159	Jack Nguyen	Male	2.5	59	Yes
BK194110694240	10694	Amarachi Khan	Male	0.5	59	No
BK200110089799	10089	Haruto Yamamoto	Female	1.5	59	Yes
BK200110686539	10686	Amelia Hassan	Male	0.5	59	No
BK200110702686	10702	Amarachi Johnson	Female	5.0	59	Yes
BK200110963130	10963	Amelia Garcia	Male	5.0	59	Yes

GB: When I first start building dashboards, I try to imagine what it is I'm going to want to build. Then, the most important thing I think about, before even deciding what's next, is to look for inspiration. I'm someone who constantly thrives on inspiration and wants to see what others have done, either in the data space or the UI/UX space. I explore these dashboards on platforms like Pinterest, Dribbble, and Behance to find my next source of inspiration.

Once I find a dashboard that inspires me, I choose the industry I want to apply this style to. In terms of design, I consider the theme of the dashboard in a way that fits the concept of what it's about, especially for business dashboards.

For example, with the doctor appointment dashboard, I felt that focusing on user interaction and minimizing native features to give it a modern website-like feel would suit a doctor use case. Doctors typically don't have a lot of time for extensive dropdowns, so buttons that help them quickly achieve their goals are ideal. I consider who the dashboard is targeting and how the end-product user will perceive it. The inspiration I draw from, the intended user, and my vision of the final result all play crucial roles in designing dashboards with themes.

CJ: Some of your dashboards vary in terms of layout with filters and headers, left, right or at the top in the case of your marketing campaign visual. Do you think there are cases where one takes precedence over another in terms of visual hierarchy?



GB: Yes, definitely. I think there are cases where one takes precedence over another in terms of visual hierarchy, and here's my approach to it. I approach it in terms of three things: Firstly, based on choice. Sometimes, I choose to have filters or navigation panes at the top, left, or right, based on the inspirations I work with. If I see something inspiring with a unique layout, I might incorporate that into my design purely by choice.

Secondly, the number of pages in the dashboard. If my dashboard has more than three pages, I tend to use a left-sided navigation pane to provide enough space to stack all the pages and allow easy navigation. If there are three or fewer pages, a top navigation pane can work well to show the switch between pages clearly.

Thirdly, the number of filters. For dashboards with many filters, I might use a left or right navigation that floats in the filters when selected or a top navigation that floats them from the top. It depends on what suits the layout and user experience best.

So, when building a dashboard and determining what should take precedence in terms of layouts and filters, I consider these three factors: the number of pages, the number of filters, and choice or inspiration. This approach helps ensure that the final product is not only visually appealing but also functional and user-friendly.

CJ: Your logistics and shipments dashboard is one of your best to date. From a design perspective what are you most pleased about? What subtle design techniques do you think help particularly

elevate the visual?



Dashboard

Drivers Overview

Trucks Breakdown

How-to-read

Gbolahan Adebayo
gbolahan@movepoint.com

[Logout](#)

Drivers Overview

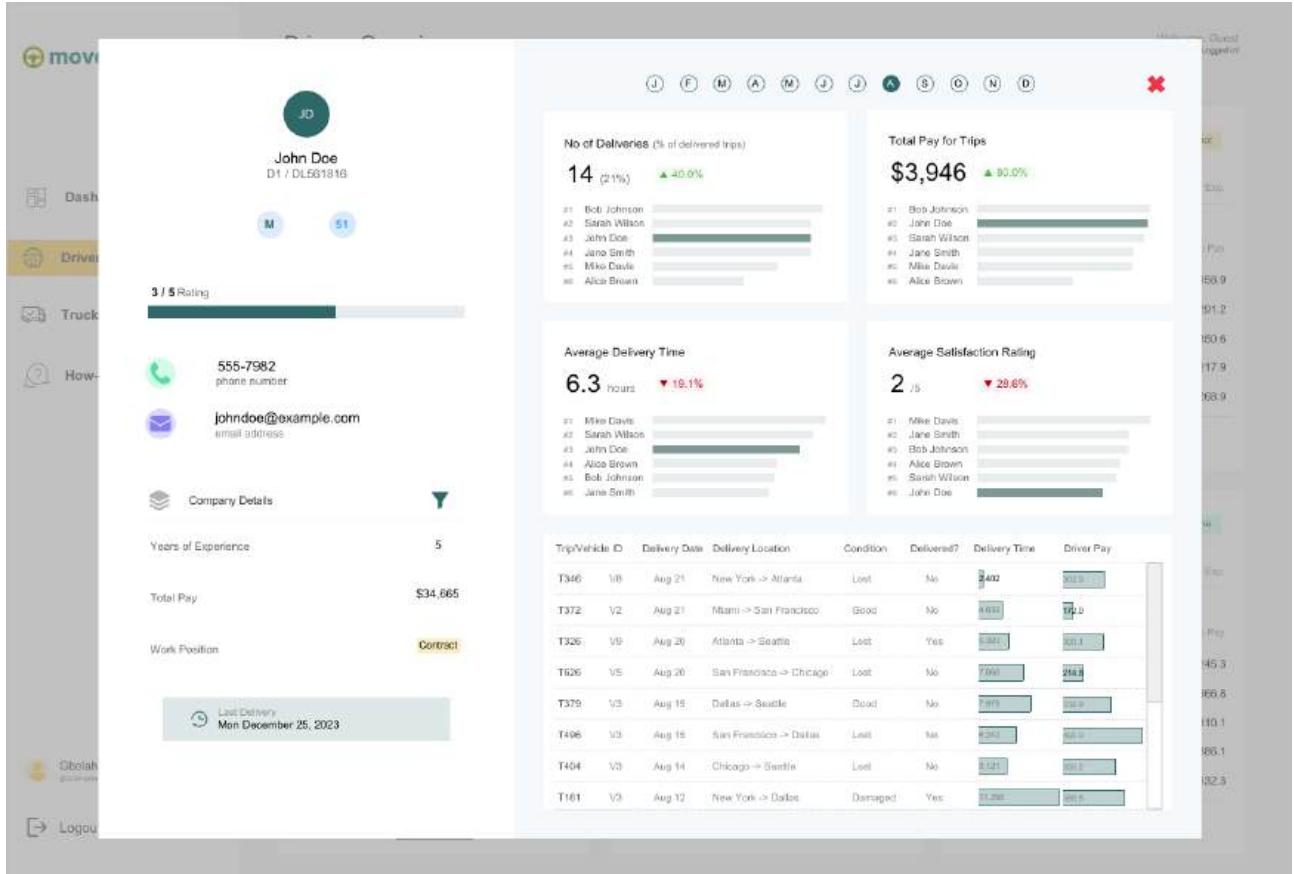
Welcome, Guest [Logout](#)

View each driver information from their trips delivered to satisfaction rating. Expand driver card for in-depth view.

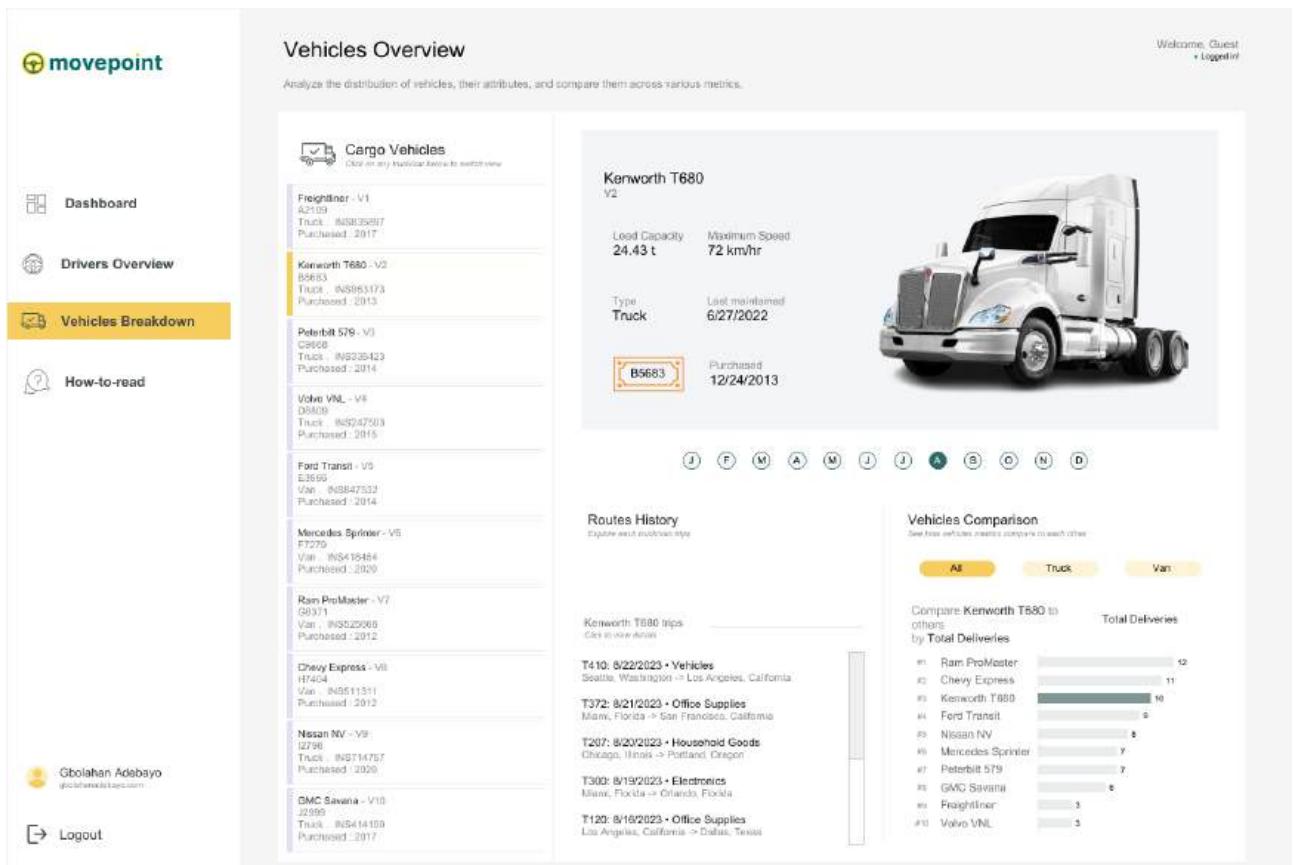
Driver	Information	Status	
AB Alice Brown 04 / DL79160	Tue, March 12, 2013 Date of Hire	17 Years of Exp.	Fulltime
BJ Bob Johnson 03 / DL75634	Sun, January 28, 2018 Date of Hire	25 Years of Exp.	Fulltime
JS Jane Smith 02 / DL40847	Wed, January 25, 2006 Date of Hire	11 Years of Exp.	Contract
<i>Last 5 Trips</i>			
Trip ID	Goods Information	Trip Pay	
T122	Sports Equipment 12345678... New Product	\$236.0	Expand
T400	Office Supplies 12345678... New Item	\$309.7	Expand
T818	Building Materials 12345678... Heavy RV	\$255.6	Expand
T142	Machinery 12345678... Chevy Express	\$341.5	Expand
T140	Office Supplies 12345678... Mercedes Sports	\$241.8	Expand
<i>Last 5 Trips</i>			
Trip ID	Goods Information	Trip Pay	
T137	Clothing 12345678... Fragrance	\$144.0	Expand
T210	Furniture 12345678... Farm Tract	\$292.0	Expand
T48	Toys 12345678... Ram ProMaster	\$158.8	Expand
T7	Machinery 12345678... Kenworth T800	\$322.9	Expand
T85	Machinery 12345678... Freightliner	\$334.3	Expand
<i>Last 5 Trips</i>			
Trip ID	Goods Information	Trip Pay	
T642	Toys 12345678... New Product	\$168.9	Expand
T203	Furniture 12345678... Volvo VNL	\$291.2	Expand
T293	Electronics 12345678... Vehicles 57%	\$360.6	Expand
T344	Toys 12345678... Kenworth T800	\$217.9	Expand
T478	Electronics 12345678... Palisade 57%	\$268.9	Expand
<i>Last 5 Trips</i>			
Trip ID	Goods Information	Trip Pay	
JD John Doe 01 / DL881818	Fri, April 16, 2004 Date of Hire	5 Years of Exp.	Contract
SW Sarah Wilson 08 / DL974392	Sat, July 11, 2020 Date of Hire	7 Years of Exp.	Contract
MD Mike Davis 05 / DL438503	Wed, March 23, 2016 Date of Hire	15 Years of Exp.	Fulltime
<i>Last 5 Trips</i>			
Trip ID	Goods Information	Trip Pay	
T779	Toys 12345678... Kenworth 57%	\$216.5	Expand
T828	Sports Equipment 12345678... GMHD Service	\$239.9	Expand
T269	Toys 12345678... Volvo VNL	\$360.2	Expand
T564	Sports Equipment 12345678... Palisade 57%	\$312.2	Expand
T113	Office Supplies 12345678... Kenworth T800	\$355.3	Expand
<i>Last 5 Trips</i>			
Trip ID	Goods Information	Trip Pay	
T239	Clothing 12345678... Freightliner	\$274.9	Expand
T800	Toys 12345678... GMHD Service	\$322.6	Expand
T260	Furniture 12345678... Freightliner	\$312.3	Expand
T199	Furniture 12345678... Farm Tract	\$159.0	Expand
T261	Sports Equipment 12345678... Volvo VNL	\$237.5	Expand
<i>Last 5 Trips</i>			
Trip ID	Goods Information	Trip Pay	
T362	Furniture 12345678... Mercedes Sprint	\$245.3	Expand
T421	Furniture 12345678... Kenworth T800	\$366.8	Expand
T334	Vehicles 12345678... Volvo VNL	\$310.1	Expand
T183	Sports Equipment 12345678... Volvo VNL	\$386.1	Expand
T817	Vehicles 12345678... Mercedes Sprint	\$132.3	Expand

GB: I absolutely love my logistics and shipment dashboard because it's one of the dashboards I've enjoyed building the most. I had so much fun with it and drew inspiration from various dashboards, bringing everything together into one cohesive design.

The first source of inspiration came from a restaurant calculator that compared different customers. I loved the breakout view and thought it would work perfectly for any industry needing to compare multiple clients or customers, allowing for an expanded view of individual performance. The expanded view was inspired by another dashboard that broke down a customer agent's activities over a year, which I felt suited the expanded view when a driver is selected.



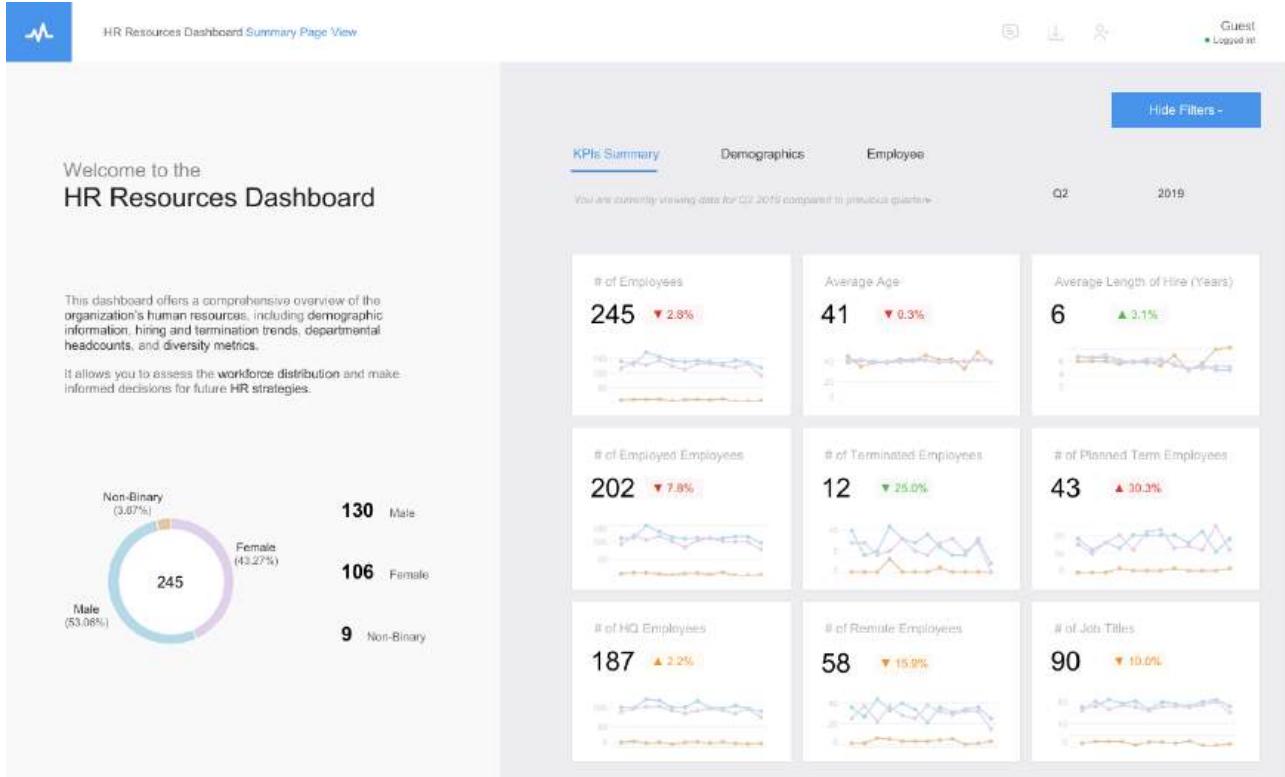
The second page drew inspiration from a transportation business dashboard comparing trucks and vans, where selecting a vehicle switches the displayed information and image. After gathering inspiration from about four to five different dashboards, I set out to build my logistics and shipment dashboard.



From a design perspective, this is my best work because I had a clear goal and was able to achieve it, even though it took time to create dynamic zone visibility for each driver. Subtle design techniques, like maintaining good white space to display a lot of information on a single page and ensuring visible call-to-action buttons, like ‘Click here to see the full driver’s information,’ elevate the visual experience. I also applied the golden ratio rule for color selection, making sure my call-to-action buttons had the deepest color to immediately draw attention.

Finally, the vehicle breakdown page stands out with a scrollable interface inspired by Chimdi’s email views. It allows users to click on each vehicle to see its picture, license plate, and other information, along with comparisons to other trucks or vans. All these elements together make this dashboard my favorite from a design perspective.

CJ: Your dashboards often come with a range of interactivity, be it parameter changes, filter views or navigation. What techniques do you use to help the user understand the difference? What are some of your favorite call to action design styles?



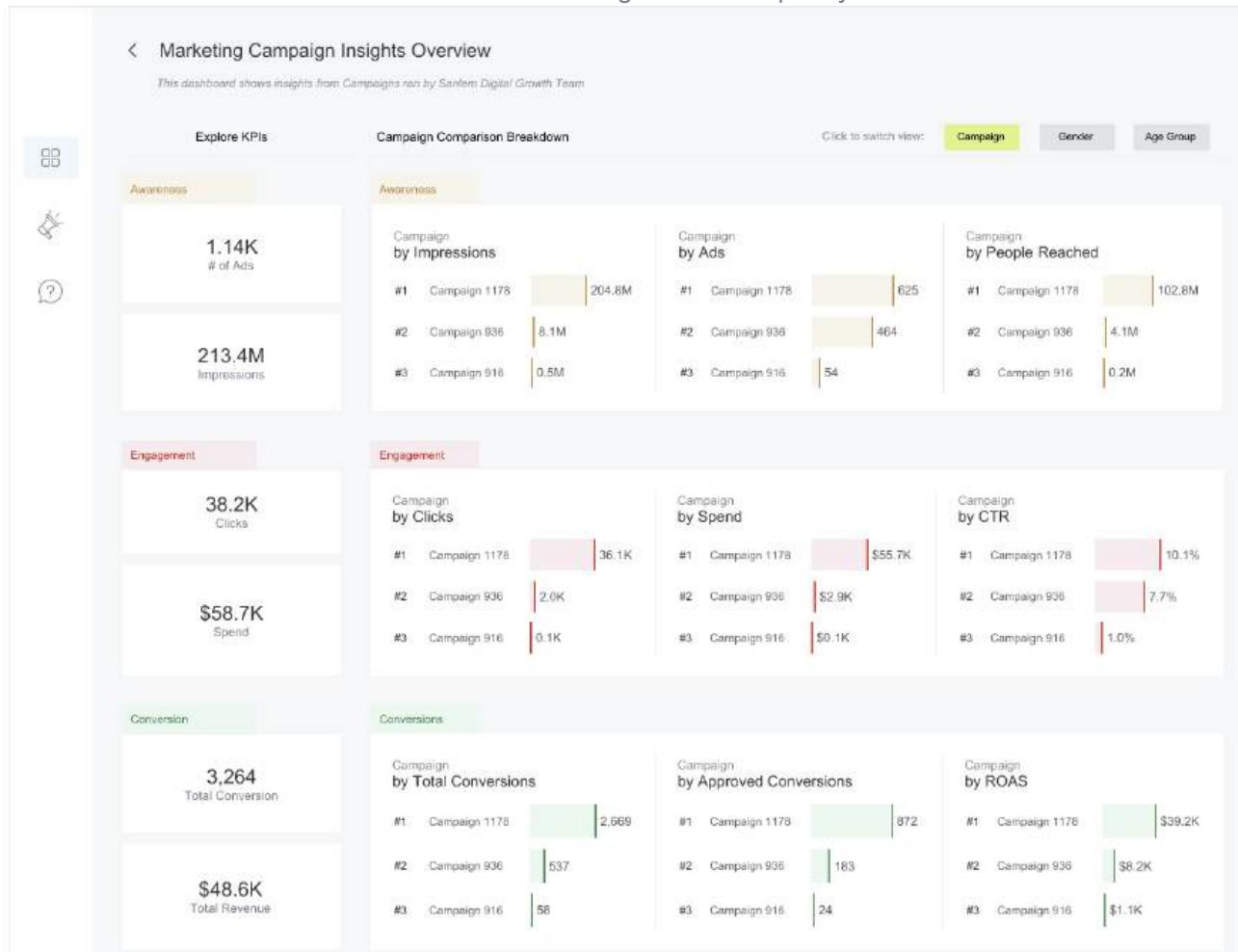
Absolutely! I love incorporating interactivity into my dashboards, as it enhances their power and depth of insights beyond initial impressions. For interactivity, I leverage features like parameters (through dropdowns or parameter actions), filter views, navigation, and especially dynamic zone visibility, which is a favorite of mine in Tableau.

To help users differentiate between these interactive features, I maintain consistent iconography and text hierarchy. Familiar icons, like a funnel for filters or arrows for navigation, make it easier for users to understand their functions at a glance.

I also employ color and contrast to distinguish interactive elements, using softer tones for filters and bolder colors for navigation buttons. Tooltips and instructive callouts further guide users, explaining actions when hovered over, enhancing their understanding of dashboard interactivity.

As for my favorite call-to-action design styles, I enjoy using sleek, rounded buttons that emulate a web app feel, guided walkthroughs with clickable hints or step-by-step guides, and straightforward instructional text like “click here to filter” or “select to change the view,” ensuring users grasp each interactive element’s purpose without disrupting the flow.

CJ: Your design style has slowly changed in color palette to use more toned down colors, only tending to use 3-4 colors at a time on a page – with maybe one striking bold color. How has your knowledge of color utilization changed over the past year?



GB: Over the past year, I've tried to really grow in terms of design, and there's definitely no design without appropriate color usage.

My understanding of color has evolved from working with a lot of inspirations. I've observed that many of these sources already have a strong grasp of color usage, often following the principle of having one bold color paired with two to three undertones. I feel this approach is especially important for business dashboards, as a consistent color theme can often relate to company branding or personal portfolio branding.

Generally, my understanding of color has grown to focus more on functionality. I think about how each color supports the story, and I find that limiting the palette to three to four colors helps reduce clutter and guides the user's focus on the most important insights without feeling overwhelmed. Bold colors are particularly useful because they can highlight key metrics or call-to-action elements, so I make sure to reserve bold colors for the most critical parts of the data.

In terms of accessibility, I've also put effort into understanding color contrast and standards. I regularly read articles or watch videos on these topics, and, most importantly, I learn from the work of others. I go through many dashboards to see how colors are matched, and I think it's been a valuable blend of insights.

CJ: I really admire the amount of effort you've put into paying it forward with leading the Lagos TUG. What prompted you to get involved? How have the sessions been going, and what do people have to look forward to in the coming months?

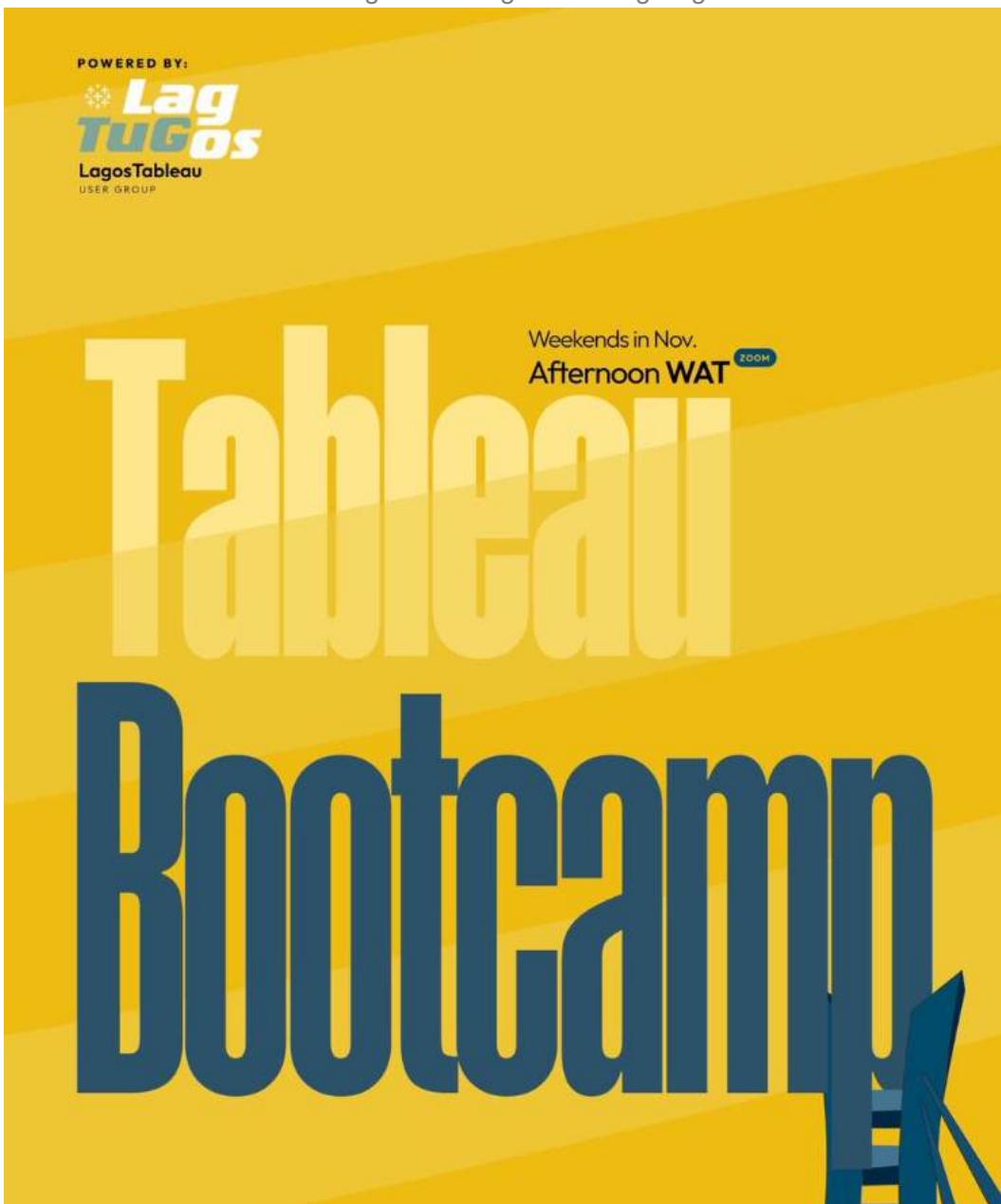
Thank you, CJ, for appreciating the work the Lagos TUG has been putting back into the community.

Getting involved felt natural because, for myself—and I'm sure for others in the Nigerian community—there's always been a strong desire to have a vibrant Tableau user group right in the heart of Nigeria's tech space, which is Lagos. We wanted one that's active and impactful.

Abisola Oni, a Tableau Visionary, did an amazing job in the lead-up to the official launch of the Lagos TUG. Let me emphasize "official," because it has existed before us, and Abisola put in significant work to bring people together to learn Tableau, visualize data, build portfolios, and refine their work.

Big kudos to her for that.

However, we felt the need for a group that could be even more active and easily recognizable, with a dedicated page and a central hub where community activities are accessible. This vision birthed the Lagos TUG. The original idea didn't come from me but from Omokehinde Ayodeji, who reached out and shared the concept of building this community together as co-leads. We also brought on two incredible co-leads, Maureen and Idris. We held initial meetings, discussed our vision, and since then, many community members have been incredibly supportive by joining the team. I think anyone who has benefited from the community naturally wants to give back, completing the cycle of learning and sharing—receiving and then giving.



What prompted me to get involved was precisely that: the opportunity to give back. I received so much support from amazing people as I was growing in Tableau, learning the platform, and advancing my skills. I believe it's invaluable for new members to find a strong community of their own, especially here in Nigeria, where you might not see many people in your immediate environment using Tableau. Having a familiar space, where you can connect with others right here, is empowering. Seeing others use Tableau to achieve great things, get jobs, and create opportunities is exactly why I wanted to be involved. When you join an exceptional community, you can build your skills, grow, and realize that

the sky is truly the limit.

In terms of the sessions, we've held three virtual ones so far, and we're planning our first in-person session at the end of November. The sessions have been incredible. We've divided our tasks, and

Maureen has done an amazing job reaching out to speakers for our events.

We've aimed to cover a broad spectrum of topics related to Tableau—Tableau Prep, Tableau Desktop, Tableau Cloud—and even larger themes like design, color usage, and general visualization and data modeling concepts. The sessions have been fantastic, with excellent speakers and good turnout. People are genuinely excited to join our community, contribute insights, and speak, which has been

incredibly helpful.

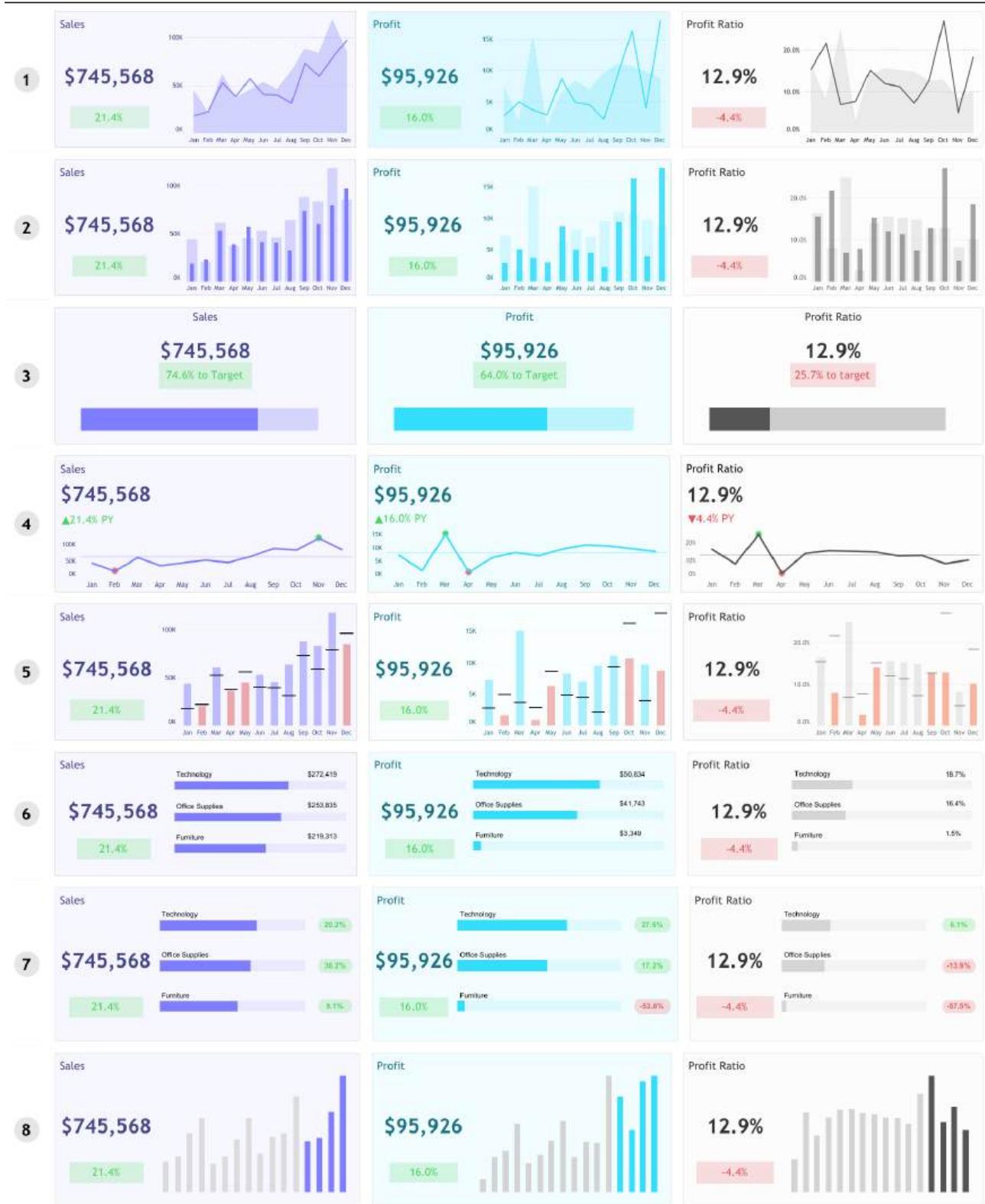
As for what people have to look forward to in the coming months, we're working on a boot camp this November focused on live dashboard building. Community members will be able to follow along, create these dashboards for their portfolios, and gain hands-on experience—not just with theoretical concepts but also practical application. They'll learn how to use different aspects of Tableau to build business dashboards and creative visualizations.

Next year, we have even more exciting sessions planned, and our upcoming physical session promises to be a memorable event. So, there's a lot to look forward to!

CJ: Is there anyone in the community that has particularly impacted your journey from a technical and design perspective? What was it about their work that you admired?

20 ways to design your KPIs

"20 ways to design your KPIs" is a dashboard that showcases various creative and effective ways to display Key Performance Indicators (KPIs). KPIs are measurable metrics that reflect the success or progress of a business or organization. However, KPIs can be more than just numbers, they need to be contextualized to be meaningful. This dashboard aims to provide users, especially those who build business-related dashboards, with ideas for visually compelling ways to display KPIs. With 20 different examples, users can explore a variety of techniques and designs to help improve the outlook and summarization of their KPIs, ultimately making their dashboards more informative and engaging for stakeholders.



GB: Yes, definitely. There have been people in the community who have significantly impacted my journey from various perspectives. Honestly, it's possible I might miss a few names because there's so much value in the data viz community. You literally learn from everyone, whether actively or passively. Just being on Tableau Public every day allows you to learn from so many people—from how they design their dashboards, the choices they make in visualizing data, to the articles they write and the videos they share. Everyone inspires me in some way and has had an impact on me.

If I were to mention a few individuals specifically, one would definitely be Idris Akilapa. He was the first person I reached out to when I was starting, and he was incredibly supportive. He helped me understand how to connect with the community, feel comfortable sharing my work publicly, ask questions, and sometimes even land gigs. We also worked together on the "20 Ways to Build KPIs" project, which was an incredible experience and has become one of the most viewed dashboards on Tableau Public, with over one hundred thousand views—an amazing accomplishment.

Another person is Chimdi Nwosu. Chimdi has been pivotal, especially from a design perspective. His visualizations were the first ones I saw on Tableau Public that truly inspired me to step up my design skills and think more deeply about how I present information.

In terms of design, Chimdi has been particularly supportive. Earlier this year, we decided to set up a weekly Sunday meeting where we'd spend an hour discussing design, any questions I had, and the projects I was working on at the time. We'd share ideas, talk about the best ways to approach certain challenges, and work through technical or calculation issues. Chimdi's been a huge help, both creatively and technically.

My good friend Michael Fajemilehin has also been an incredible resource, especially on the technical side. He's very skilled with Tableau calculations, and we often bounce ideas off each other whenever I'm working on something complex. He's been instrumental in my journey from a technical perspective.

Then there are others like Zainab Ayodimeji and Abisola Oni, along with other talented Tableau people in Nigeria. Seeing what they can accomplish from here has been a huge motivation, reminding me that I can push my limits too.

On a broader scale, I'm a big fan of Judit Becker's work. I love seeing how her ideas come to life in Tableau; she's truly inspiring. And of course, CJ, you've been a huge influence as well. From the start, we've had meaningful conversations that go beyond just technical or design advice. We've discussed the data visualization field and Tableau community as a whole, shaping my perspective on what to pursue next. I remember the incredible advice you gave me after the Vizzes award for Notable Newbie—it really shaped my direction after that.

There are so many others, like Andy Kriebel, Flerlage Twins, Adedamola Ladipo and Sarah Barlett. I could probably keep listing names all day. Each of these amazing people has been incredibly helpful in different ways—Chimdi for his design insights, Idris for showing me what's possible here, and you, CJ, for your guidance and constant willingness to offer advice. This community has been a huge part of my journey, and I'm grateful to everyone who's played a role.

LOGGING OFF,
CJ

IRONVIZ SEASON

Hi all,

Hope everyone is doing well. I just got back from a wonderful holiday break which included turning my emails and teams notifications off. Wow, it was refreshing. We often forget that working in data isn't life or death, and whilst we love data – that time away with friends, family, and not 12 hours in front of

a computer each day, is much needed time off. That time also gave me some moments of reflection on some current on-going community activities.

As most of you are aware it is Iron Viz “season” – with the competition closing in around a weeks time. Having been fortunate enough to present on stage and wear one of the awesome black chef jackets, the competition holds a soft spot in my heart.

The competition has become more transparent in recent years around what the definition of a “good” entry looks like. You can read what that looks like in terms of design, analysis and storytelling below.

Judging Criteria

Design

- Visual elements add to the overall understanding of the visualization rather than distract
- Interactivity and layout are user-friendly, instructed/specify, and purposeful
- Charts are clearly presenting the data
- Charts contain a title, summary, and/or caption
- All charts contribute to the story
- Accessibility is applied - colorblind / low vision (contrast) friendly palette, limited use of images to convey text, font size larger than 12pt

Analysis

- Viz topic aligns with contest theme
- Dataset and calculations appear to be accurate and clean
- Analysis illustrates profound insights grasped from visualizations
- Analysis supports the story being told
- Analysis has been mostly produced within Tableau
- Analysis highlights a broad range of Tableau capabilities

Storytelling

- A clear story is being told
- Story flows through visualizations and guides consumers from question to insight
- Visualizations and animations support the story being told
- The story includes a unique idea or perspective
- The story being told has complexity/nuance that elevates the visualizations.
- Storytelling captures and maintains interest throughout the entire viz

For my blog today, I thought we could look at some of the previous years finalists entrants, and comment on a few stand out elements that really helped elevate their visual. (*Not all of them will have impacted their final score, I only include the screenshot above as a reference point*)

The things I call out below within the visuals are just my opinion on what makes that viz stand out to me, and will of course include some technical elements that aren't necessarily included in marking!

The gallery is the best place to start.

It goes back to 2011. For today, we will take just the last 5 years. Lets start with 2020.

Alex Jones – Coruna

Coruna

Lockdown started here in England on the 24th March 2020.

Being stuck inside day in day out, running became the escapism I needed to stay sane during Lockdown.

Caveat:

I do not consider myself a good runner by any stretch, this is just all good fun.

The purpose of this visualisation is to chronicle and analyze my running experiences throughout lockdown.

01 when

I typically ran 3 lunchtimes on alternate days during the week and once in the morning on the weekend.

Running regularity

"For most beginner runners, running three or four days a week on alternating days."

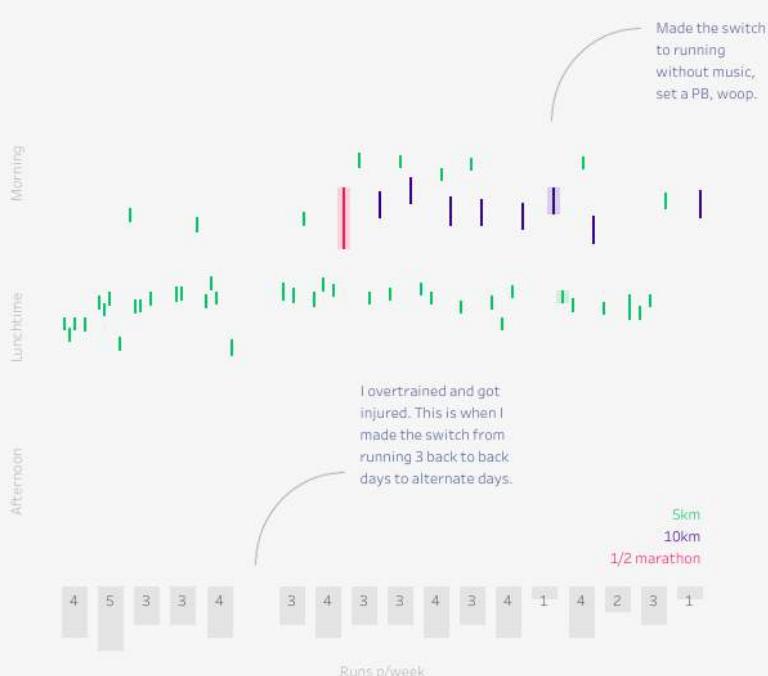
published on RunnersWorld

Optimal time of day

"Run when your core body temperature is high, typically the late afternoon."

"Everybody's different. If you've been running at a fixed hour every day... that's probably best for you."

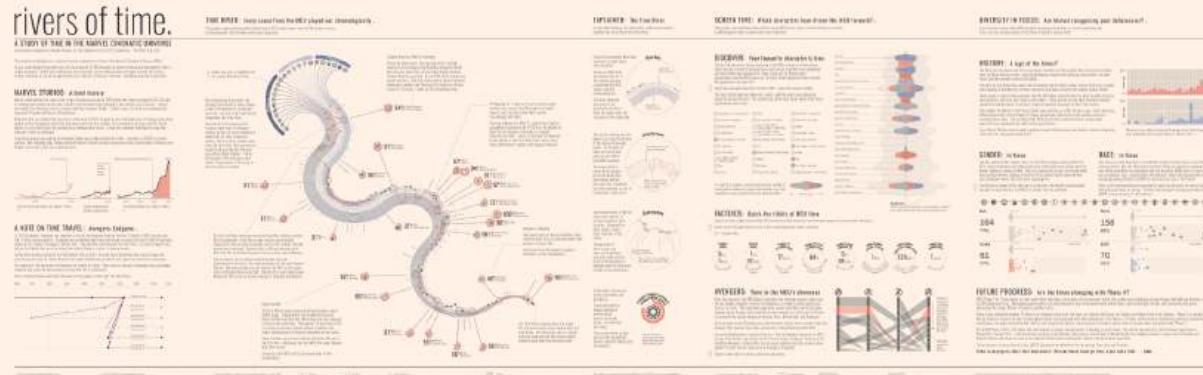
both published on Medium.com



02 pace

Alex's coruna viz was one I saw when I was first joining the online community. The bold title making for a wonderful thumbnail, then with the offset of more toned down subtle colouring. The high contrast between text and background, along with clear labeling, enhances readability and accessibility. Alex's visual has clear hierarchy. (Title, when, pace, conditions, splits, summary, future, to finish) to organize information logically, guiding the viewer through the data.

I really enjoy Alex's balance of story and context to each visual, such as the half marathon attempt with the new shoes, but how that stand out effort gets carried through to future charts to know it was 10 degrees at the time, and the adjustment in pace needed for a longer distance.



03 Time

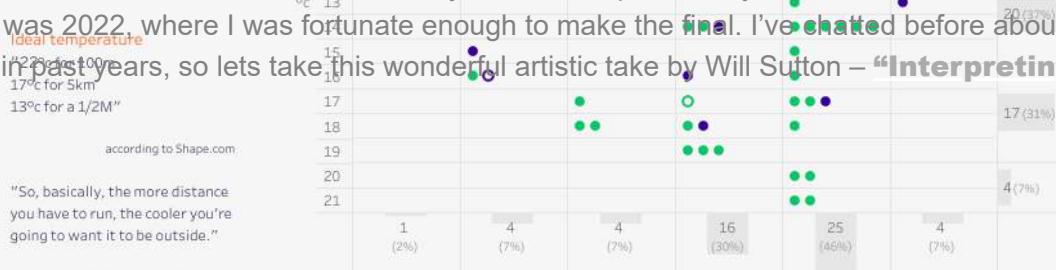
Sam had done a wonderful job of creating a visual metaphor. The "central river" graphic cleverly represents the flow of time in the Marvel Cinematic Universe, reinforcing the title concept and providing an intuitive way to understand the timeline.

Now, of course, technical skill isn't part of the contest, but I have much admiration for the detail that will have gone into some of the chart types like the curved chart with line and shape overlay. The streamgraph chart and the parallel bars with polygon join.

Sam's visual was one that broke the trend of vertical longform, opting for a newspaper roll-out feel. The design follows a consistent grid, creating a sense of order and balance across the viz.

A clear favourite for many on Tableau public for years to come.

Next up was 2022, where I was fortunate enough to make the final. I've chatted before about my own visual in past years, so let's take this wonderful artistic take by Will Sutton – "Interpreting Art"



So, they say, 17°C is the ideal 5k running temp.
I hit my PB at it, therefore correlation = causation for once.

04 splits



Interpreting Art



WHAT IS ART?

Experts believe that art, in its broadest sense, is a **form of communication**. It means whatever the artist intended it to mean, be it telling a story, an expression of feelings, or a simple observation.

HOW SHOULD WE INTERPRET ART?

To interpret a work of art is to make sense of it. To interpret is to see something as "representing something, or expressing something, or being about something, etc.". So when we interpret art we are deriving meaning from it.

TAKE THE QUIZ

Below are six drawings from an aspiring artist, captured during a game of Drawful 2. The artist was given a prompt to draw, other players in the game create their own interpretation of the drawing. Can you guess what the artist's intended meaning was?

HOW DO YOU INTERPRET ART?

Pick an interpretation from the options below

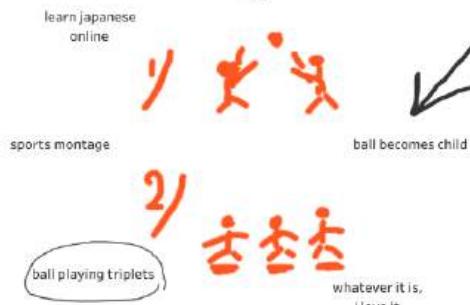
Q1



Q2



Q3



For each image, click the text to select an answer

Q4



Q5

© 2013 Drawful LLC. All rights reserved.

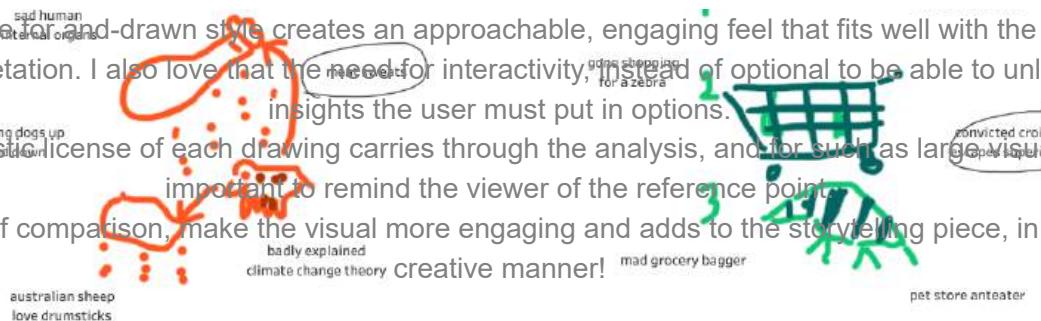
Q6

shopping weevil

Will's choice for hand-drawn style creates an approachable, engaging feel that fits well with the topic of art interpretation. I also love that the need for interactivity, instead of optional to be able to unlock the insights the user must put in options.

That artistic license of each drawing carries through the analysis, and for such a large visual it is important to remind the viewer of the reference point.

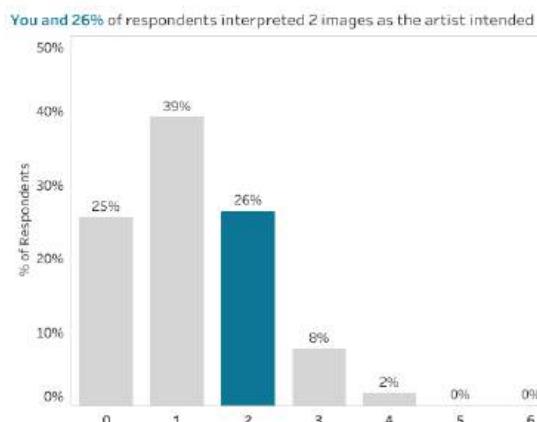
The act of comparison, make the visual more engaging and adds to the storytelling piece, in a fun, creative manner!



SUBMIT YOUR ANSWERS



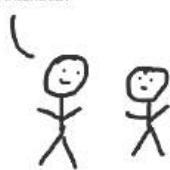
HERE'S HOW YOUR ANSWERS ALIGNED WITH ARTIST AND OTHER RESPONDENTS



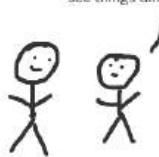
Question Breakdown

Q.	Image	Your response	Artist's interpretation	% of responses that agreed with you
Q1		all up hill	walk of shame	28%
Q2		hostile takeover	hostile takeover	31%
Q3		ball playing triplets	sports montage	26%
Q4		poisoned guitar	ted bundy's banjo	22%
Q5		meat sweats	meat sweats	15%
Q6		convicted croissant escapes supermarket	mad grocery bagger	14%

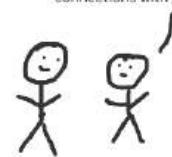
So why do we struggle to interpret art as it was intended?



Because our backgrounds are unique, meaning we think and see things differently.

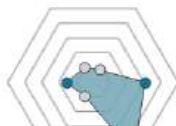


This is actually not a bad thing as this can help us build connections with people.



YOUR ANSWERS SHOW YOU ARE MOST ALIGNED TO JANET & CASSIE

JANET
THE ASPIRING ARTIST



CASSIE
HERE FOR THE WIN



BOB
RETIRIED AND TIRED



GOOGLE LENS
KNOWS A LITTLE TOO MUCH



WILL
TRIED TO BE FUNNY ONCE



Spiel des Jahres

How one German competition helped me and can help you choose your next favorite game

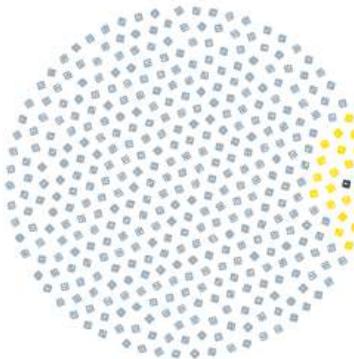
What is the Spiel des Jahres competition?

Spiel des Jahres is an annual German competition recognizing outstanding games created the past year.

Evaluation criteria include:

- Does the game have a new concept?
- Are the rules clear and easy to understand?
- Is the design attractive and components well built?

The purpose of the award is to promote playing quality games with family and friends. **There is no cash prize!**



Last year nearly 400 entries were submitted. Of those 22 games made the finals, aka 'recommended'.

Only one game receives the award in 3 categories:

Spiel des Jahres - Overall game

Kinderspiel des Jahres - Children's game

Kennerspiel des Jahres - Connoisseur game

Note: Kennerspiel winners are for folks that are already familiar with learning board games. These games may not be more difficult to play, but are more complex to learn.

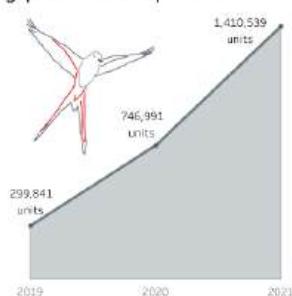
If there's no cash prize, what's the point?

While there's no cash prize, you do get bragging rights, and you can put the logo of the award on your box or website.

Games that have won the award have gone on to sell millions of copies.

This is remarkable in an increasingly crowded game industry.

Wingspan sales soar post win



Sales data for board games, whether in units or in dollars is difficult to find.

Various reports indicate that success for a board game that does not hit the mainstream is in the thousands of copies, which would be right around here.

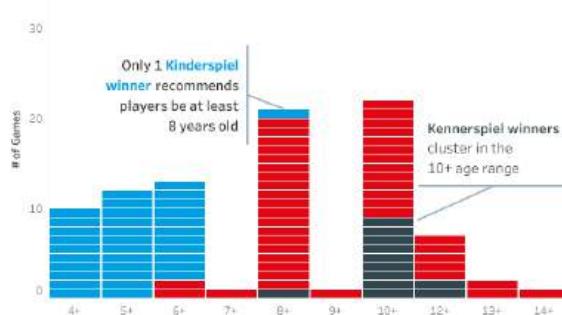
Besides great design, what can I expect from a Spiel des Jahres winner?

Looking for a game you've played? Search in the box to the right to highlight a game across the viz

No items highlighted

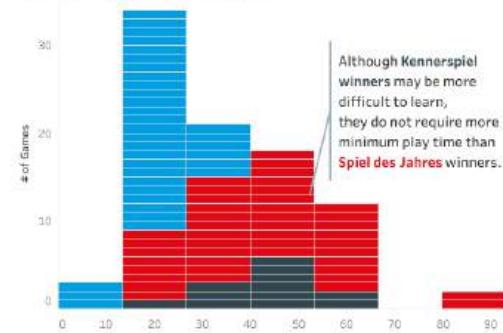
Generally, the 3 categories can be split by age range.

Spiel des Jahres winners have the most variation for minimum age.



Kinderspiel winners have the lowest minimum play time.

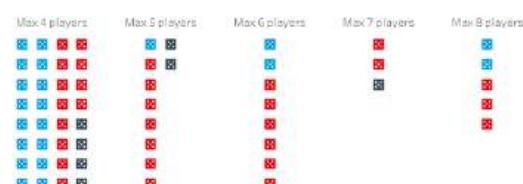
Hover over any game for more detail.



Most games require at least 2 players- but 10 include a solo mode



40% of winning games accommodate more than 4 players



Brittany's was my favourite visual from the 2023 feeder finalists. A consistent color scheme is used throughout, with different colors representing various categories or types of games, aiding quick visual comprehension. Something about using the primary colours I felt associated well to the games theme, simple, almost child like?

Brittany concludes with a section encouraging the viewer to explore and choose games based on their preferences, making the information actionable. You'll see this call to action has become fairly popular in recent years. For me, this is a prime example of where it works well, due to its interactivity of filtering out other choices, but doesn't necessarily force you to go load an external site.

I also really enjoyed the personal touch. It wasn't over baring and is light. The inclusion of Brittany's personal experience with two winning games adds a relatable, human element to the design. Again, this personalisation is something that has become an added feature of recent visuals.

Finally its worth mentioning Brittany perfectly balances hand drawn images in a way that doesn't become over powering. Notice how they are hand drawn adding to the youthful games like feel, as oppose to opting for something like the image of a games box lid etc.

Ticket to Ride



Winner of **Spiel des Jahres** in 2004, Ticket to Ride is a family game where the objective is to build and complete train routes between cities.

Ticket to ride is easy to learn, but because of the route variations no two games are alike. It's the perfect combination of luck and strategy - even if my little sisters beat me every time.

Mastered the original? Ticket to Ride now has over 25 versions:

- 9 Full
- 8 Map Expansions
- 6 Mini Expansions
- 4 Full Specialty

Now it's your turn: what winning games should you try?

Click on the icons on the left to filter games from the circle of dice. Hover over dice for game details.

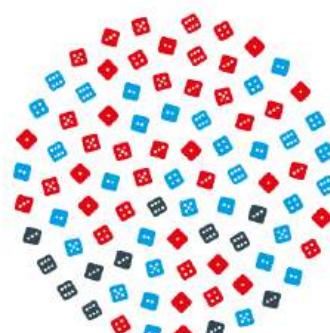
Which category game would you like to play?



How much time do you have?



How many people will be playing?



90 games
meet your criteria

For more on Spiel des Jahres and the featured games,
Click the icons below:



LOVE FOR FOOD

HUNGER FOR CHANGE

a data visualisation by @westlake_cjw

Throughout this visualisation, click the  icon to uncover further analysis.

I love cooking food

I am at my most relaxed when I spend time in the kitchen with

I love eating food

My heart is happy when I am enjoying a delicious meal, and a

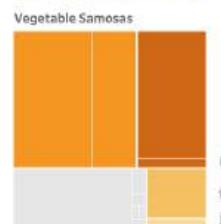
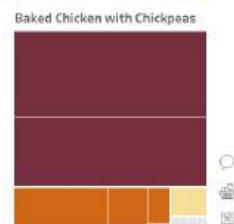
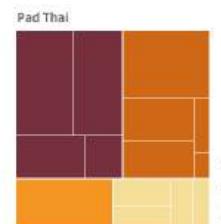
I love sharing food

Sharing food and making 'food memories' with the people

Below are 6 of my favourite recipes to cook, eat, and share.

Each area represents an ingredient, sized by weight and coloured by food group

● Protein ● Fruit and Vegetables ● Starchy Carbohydrates ● Dairy ● Oils, spreads and sauces ● Other
Hover on the icons to find out  why I love them,  what the ingredients are, and  how to make them yourself.



 Look out for this icon below to learn more about the countries that inspired these dishes

Whilst I enjoy eating these recipes,

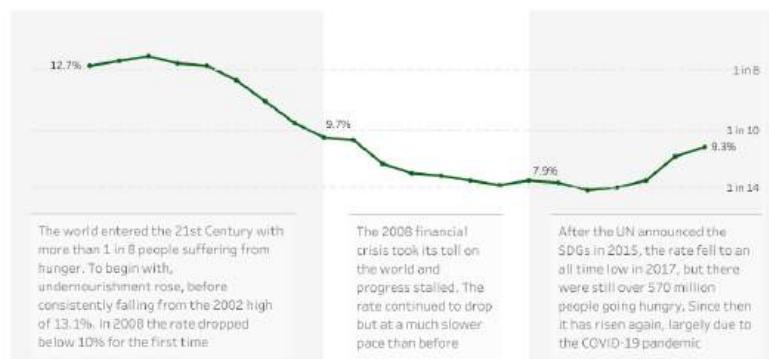
738,900,000

9.3%

people do not have this luxury.

The number above is an estimate of the number of people globally thought to be suffering from undernourishment in 2021. The prevalence of undernourishment, or hunger, measures the share of the population that has a daily food intake that is insufficient to provide the amount of dietary energy required to maintain a normal, active and healthy life.

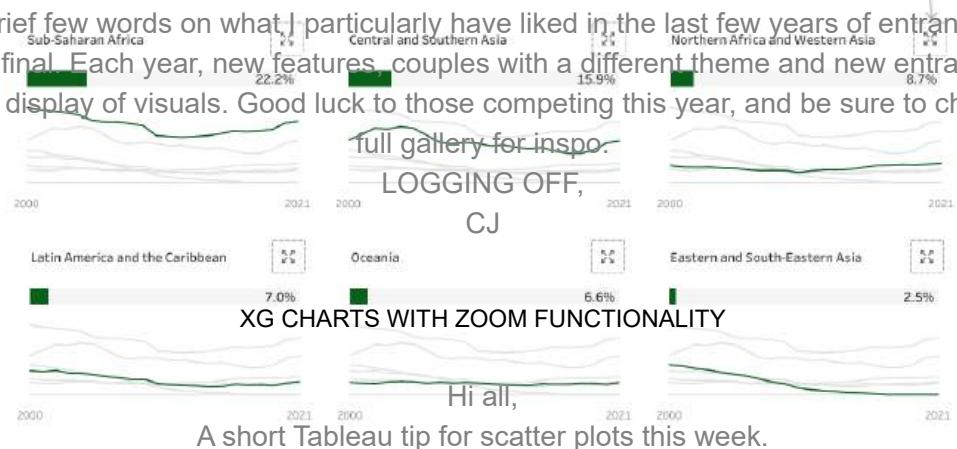
In 2015, the United Nations introduced 17 Sustainable Development Goals. The aim is to meet all of these targets by 2030, and they act as a global call to end poverty and inequality, protect the planet, and ensure that all people enjoy health, justice and prosperity. The second of them is to "create a world free of hunger by 2030", and that makes my heart full.



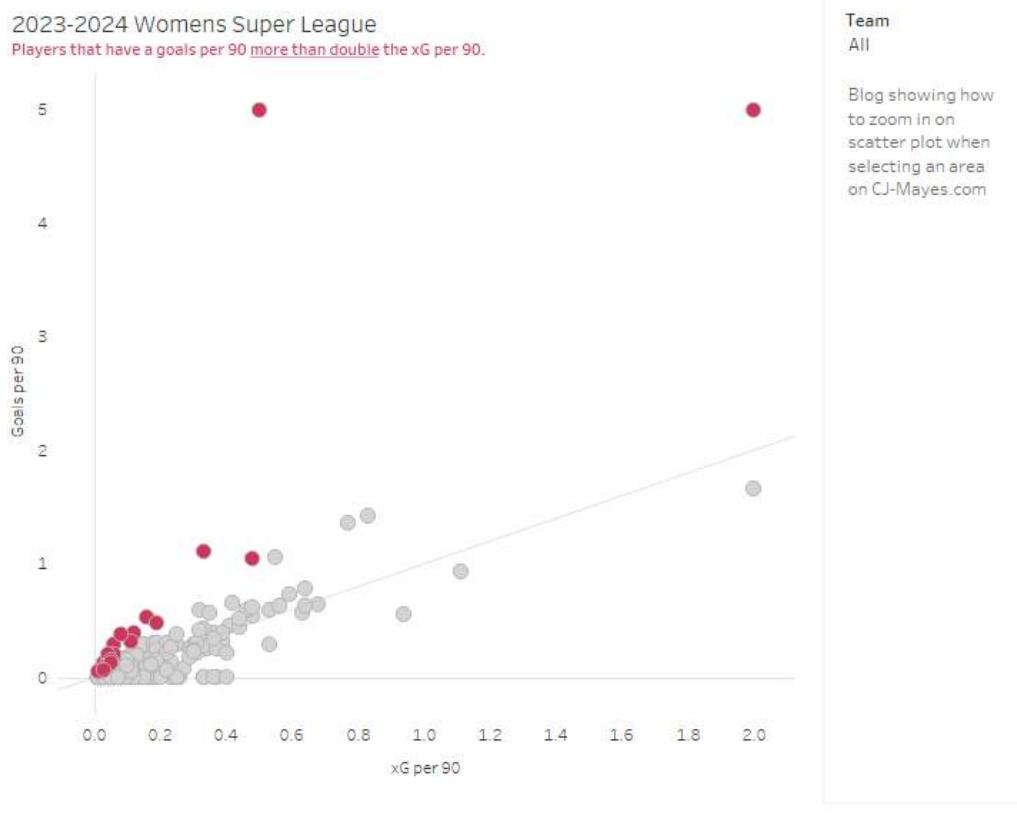
In 2024, Chris Westlake made Iron Viz royalty with his “Love For Food” visual – like many, going for the more standard long form. Each section has a numbered, question-based heading that guides the reader through the content. It helps dice up the visual, or draw attention to a newly posed question. Chris utilises a new feature that year of dynamic zone visibility popping out additional context where needed. He cleverly balances what NEEDs to be on the page, vs added information if interested.

Chris uses various chart types (Treemaps, line graphs, scatter plots) to present data in an easily digestible format. Note everything has to be flashy, crazy radial charts, especially when it comes to the Iron Viz competition. But when you do make these line and bar charts, the importance is heavily put on its readability. For example, where Chris uses light grey to fade other lines into the background. Clean, elegant and quickly interpretable.

That's it a brief few words on what I particularly have liked in the last few years of entrants that made the Iron Viz final. Each year, new features, couples with a different theme and new entrants make for an exciting display of visuals. Good luck to those competing this year, and be sure to check out the



A short Tableau tip for scatter plots this week.

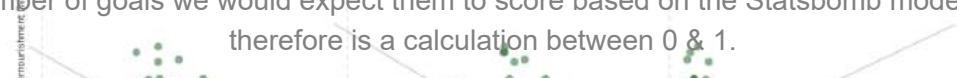


Team
All

Blog showing how to zoom in on scatter plot when selecting an area on CJ-Mayes.com

Are you ever faced with a scatter plot where a lot of marks are clustered together?

The above chart looks at the womens super league football performance from last year. We plot every players goals per 90. (number of goals they score per 90 minutes on the pitch) against what the xG per 90. (number of goals we would expect them to score based on the Statsbomb model) Each shot therefore is a calculation between 0 & 1.

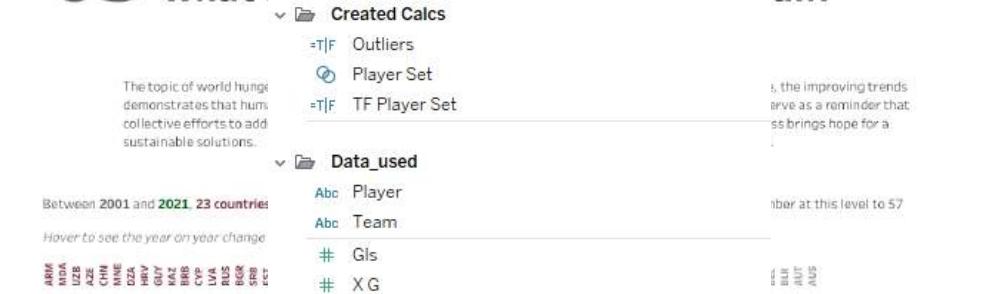


But when it comes to trying to find stories in the data, it's okay at high level. Of course anyone above the line is scoring more per 90 than they were expected. That's great.

But when we try and find the deeper stories that aren't the likes of Isabella Hobson and Denny Draper with a goals per 90 of 5, then we need to start zooming in to some of the more clustered together marks.

These are the steps below to create a zoom functionality on your scatter plot. The data is available in

03 what about this makes my heart full?



Firstly, create the set of the players.

Data_used

- Abc Player
- Abc Team
- # Gls
- # XG

Add to Sheet

- Show Filter
- Duplicate
- Rename
- Hide
- Aliases...
- Create
 - Calculated Field...
 - Group...
 - Set...**
 - Parameter...
- Transform
- Convert to Measure
- Change Data Type
- Default Properties
- Geographic Role
- Image Role
- Group by
- Folders
- Hierarchy
- Replace References...
- Describe...

The World Food Programme (WFP) is the leading humanitarian organization fighting hunger worldwide, changing lives, delivering food assistance in emergencies and working with communities to improve nutrition and build resilience.

WFP focus on emergency assistance, relief and rehabilitation, development aid and special operations. Two-thirds of their work is in conflict-affected countries where people are three times more likely to be undernourished.

Oxfam is a global community of people who believe in dignity, equality and justice for everyone. They work in over 20 countries around the world where everyone has the power to thrive, not just survive.

Oxfam focuses on addressing the root causes of poverty and inequality, which includes tackling hunger and food insecurity. They work on various initiatives to improve food access, support sustainable agriculture, and advocate for policy changes.

CARE works around the globe to save lives, feed the hungry, combat disease, provide education and empower women. They are a global leader within a worldwide movement dedicated to ending poverty.

CARE emphasises the importance of long-term solutions such as empowering women and promoting sustainable agriculture in the fight against hunger and giving all people equal rights and opportunities.

Click to donate

Click to donate

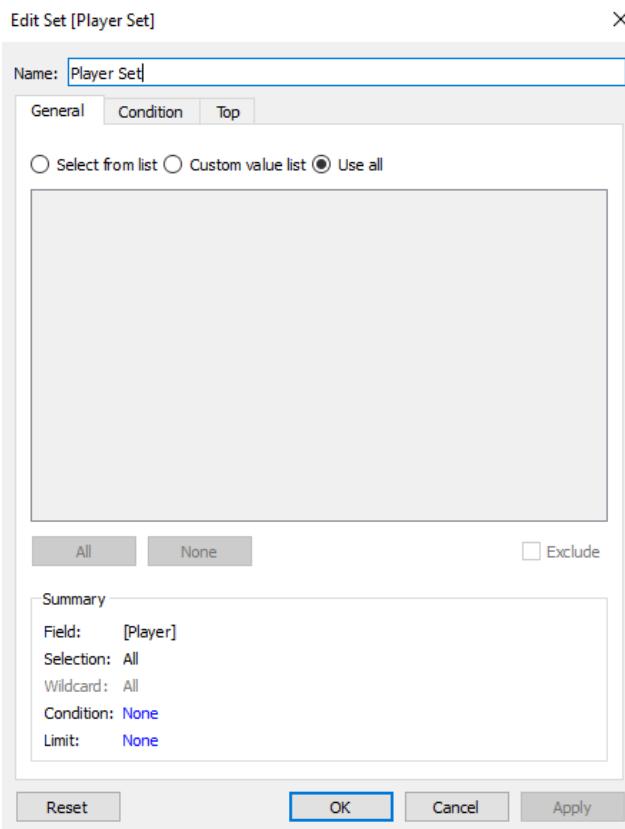
Click to donate

Together we can work towards a world where undernourishment is a thing of the past

Sources

Undernourishment Data: UN SDG Indicators Database
 Global Education Index: Our World in Data | Global Education
 Women, Peace, and Security Index: Georgetown Institute for Women, Peace and Security
 World Risk Index: World Risk Report 2021

Kongl. Icons8

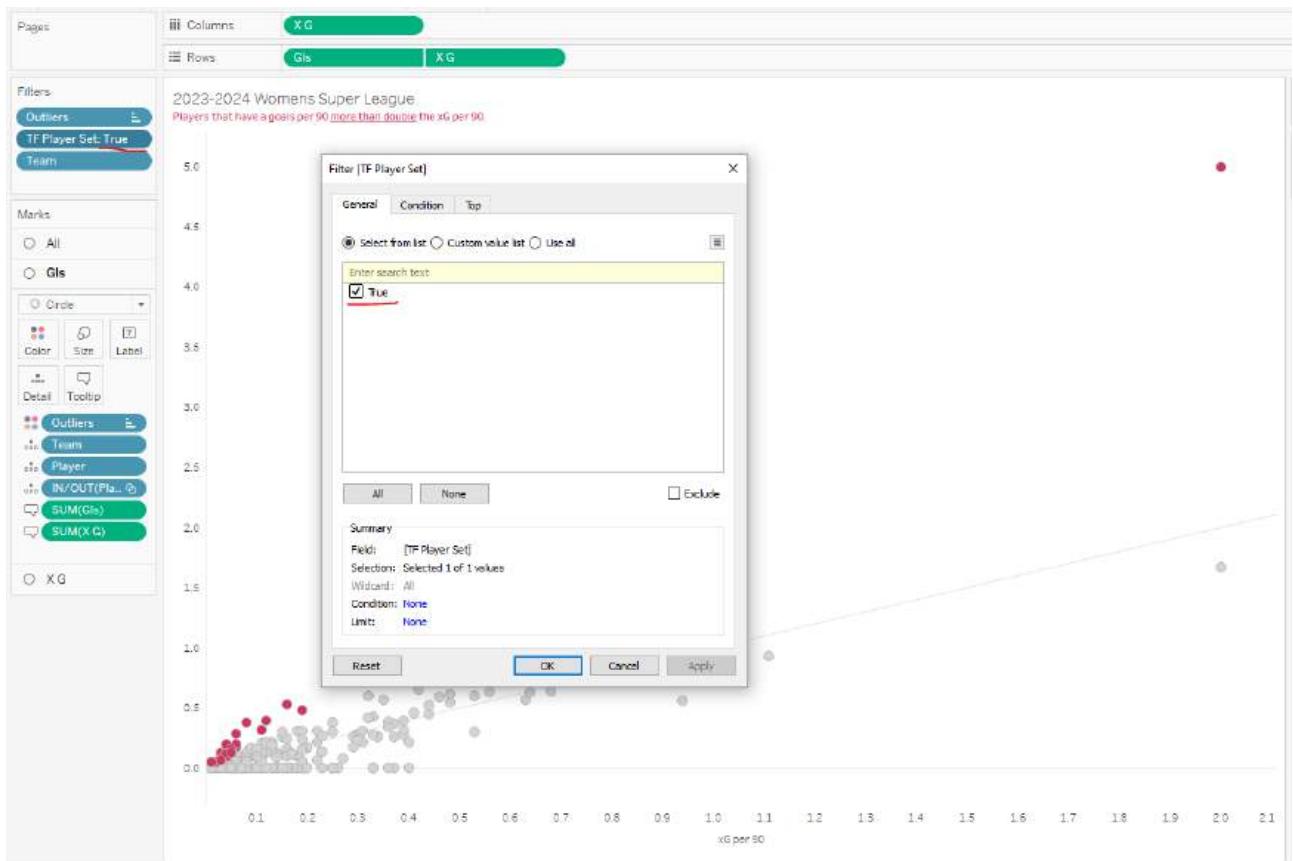


Create a boolean field, to say whether the player is in the set.

TF Player Set

[Player] in [Player Set]

Then add that calculation to the filters page and set it = True.



On the menu select Worksheet > Actions to add a Set Action

Edit Set Action

Name: Insert ▾

Source Sheets: ▼

Scatter

Run action on:

- Hover
- Select (highlighted)
- Menu
- Single-select only

Target Set: ▼

Running the action will:

- Assign values to set (highlighted)
- Add values to set
- Remove values from set

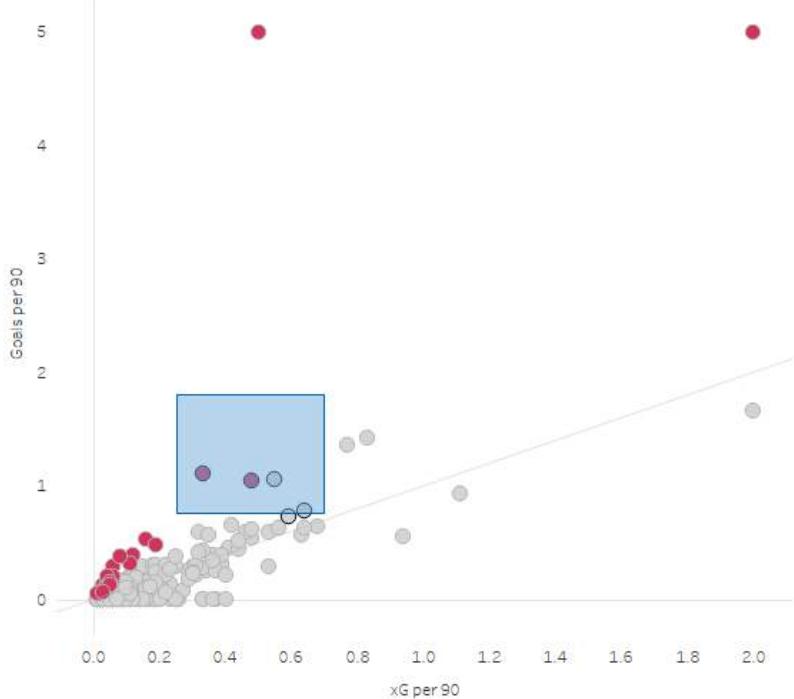
Clearing the selection will:

- Keep set values
- Add all values to set (highlighted)
- Remove all values from set

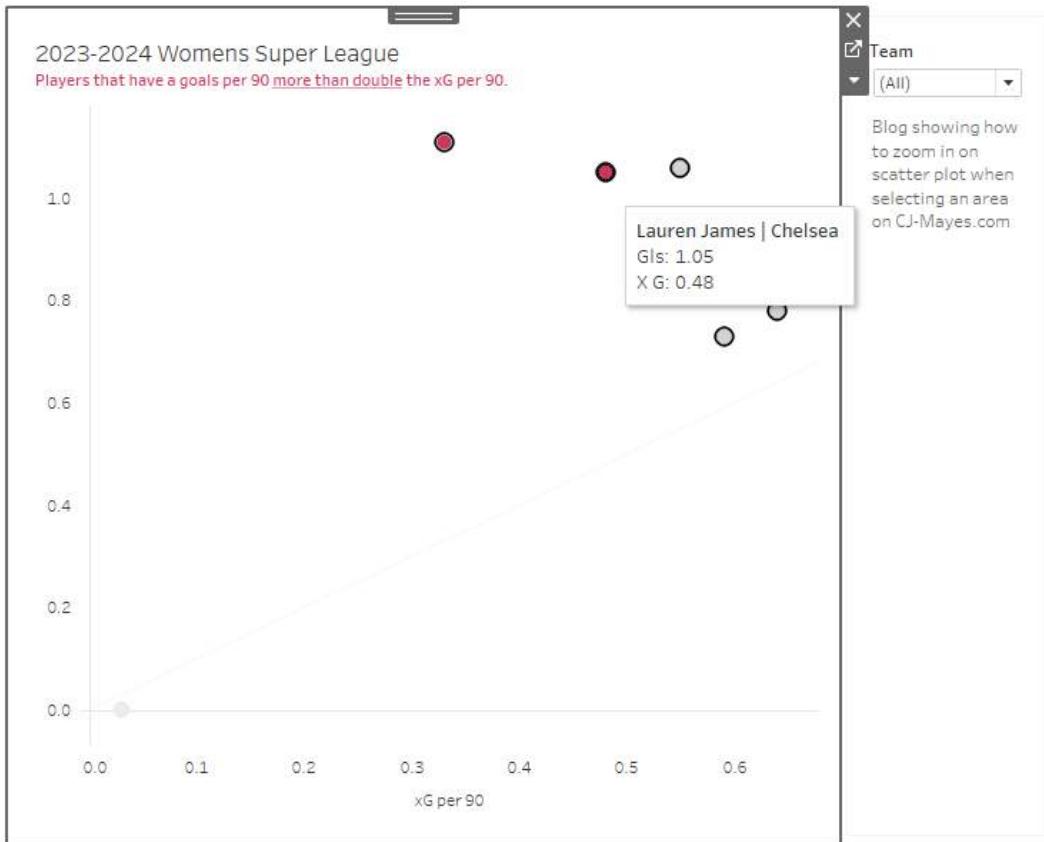
Cancel OK

Now when you select part of the scatter plot dragging a rectangle around the marks

2023-2024 Womens Super League
Players that have a goals per 90 more than double the xG per 90.



You'll see it zoom in, and we can look at reviewing other players.



A few other things to note before we close,

- It won't work if you fix your axis, for obvious reasons.
- You can remove clutter from the page by probably removing players that have 0 goals, goal keepers too.
- xG per 90 and goals were used for demonstrative purposes, of course use metrics that you feel are suitable for what you're trying to show!

LOGGING OFF,
CJ

SPORTS DESIGN WITH RYAN SOARES

Hi all,

I am stoked that I get to have Ryan Soares join the blog today. If you missed it last week Ryan was named as part of the Tableau Public Ambassadors cohort for this year. Congratulations Ryan!

You can follow Ryan on Tableau Public, [here](#).

Ryan Soares
Visual Analytics at Deloitte Canada | Toronto, Ontario, Canada

Follow Hire Me

Vizzes 124 Favorites 163 Following 138 Followers 1165

Lighting the Flame for Change | Deloitte Viz Games 2024
Ryan Soares
★ 25 (1,355)

PGA Courses 2024
Ryan Soares
★ 19 (387)

Olympic Medallists Age Distributions
Ryan Soares
★ 3 (70)

Summer Olympic Medals
Ryan Soares
★ 8 (650)

Margins of Victory
Ryan Soares
★ 3 (234)

The Perfect Tournament
Ryan Soares
★ 0 (96)

Balancing Age and Experience at Euro 2024
Ryan Soares
★ 0 (35)

UEFA Euro 2024 Squads
Ryan Soares
★ 3 (343)

CJ: Ryan, thanks for joining. I am so pleased to have another avid sports fan join the site. For someone that is new to your profile. How would you personally describe your style? Has your style developed from a combination of your public work and time in developer / consultant roles?

R: Thanks for having me CJ! The style of my work has changed over the 6 years I've been an author on Tableau Public. At first, my public work was mostly experimentation with different visualizations and designs as I was learning the tool. Over time I've really honed in on simplifying my visualizations so that they can be consumed and understood easily, which shows in the designs of my recent work. I want to make life easy for the audience so I try to reduce complexity where I can to speed up their time to insight. This does follow from my professional work as a visual analytics consultant, as quick insights are always desired. To me, data visualization is about effective communication of information and insights, so I try to balance both creativity and simplicity in my work. Sometimes striving for an eye-catching viz might take away from how understandable it is, so I always start simple and be creative when appropriate.

CJ: You have a fantastic array of sports visuals that I'm hoping to dive into, to really understand some of your thoughts behind each one.

Let's start with the Euro's. A personal favorite of mine is your Euros 2024 squad visual. Were you surprised by any country's domestic league representation? In your opinion how much do you think this impacts how well teams gel when it comes to the euros?

UEFA Euro 2024 Squads

Player representation by domestic league system

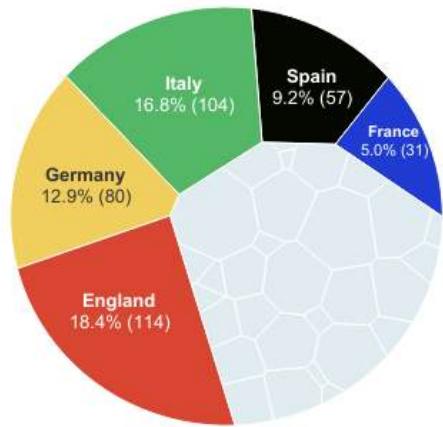
Over 600 players will play in the 2024 UEFA European Football Championships. Over 60 percent of these players currently play club football in countries with a top 5 professional football league: the Premier League (England), Serie A (Italy), Bundesliga (Germany), LaLiga (Spain), and Ligue 1 (France).

Domestic professional and semi-professional clubs are an important factor in the development of players and international success. Looking at the nations with a top 5 league, four out of five have the majority of their national team squads playing club football within the country. France is the lone exception with players in both its domestic leagues and other top leagues.

Other top ranked nations such as Portugal, Belgium, and the Netherlands, do not have a top league and have a majority of their players playing outside the country. Without a top league, these countries have a different strategy which is to develop talent domestically and eventually transfer them to a higher quality league to develop further.

The stacked bar charts below visualize the share of each nation's players playing club football worldwide. Nations are sorted by their FIFA world ranking as of April 4, 2024. Hover over the charts to view the player list.

Percentage of players at UEFA Euro 2024 by domestic league system



Group A

Germany (#16)
Germany | Spain |

Switzerland (#19)
Germany | Italy | France | Spain |

Hungary (#26)
Hungary | Germany | Spain | Italy |

Scotland (#39)
England | Scotland |

Group B

Spain (#8)
Spain |

Italy (#9)
Italy |

Croatia (#10)
Croatia | Italy |

Albania (#66)
Italy |

Group C

England (#4)
England |

Denmark (#21)
England | Belgium | Italy |

Serbia (#33)
Italy | Spain |

Slovenia (#57)
Italy |

Group D

France (#2)
France | Italy | Spain |

Netherlands (#7)
England | Netherlands | Germany | Italy | Spain |

Austria (#25)
Germany | Austria | Italy |

Poland (#28)
Italy |

Group E

Belgium (#3)
England | Spain | Italy |

Ukraine (#22)
Ukraine | Spain | England |

Romania (#46)
Romania | Italy | Turkey | Spain |

Slovakia (#48)
Italy |

Group F

Portugal (#6)
England | Portugal | France | Italy |

Czech Republic (#36)
Czech Republic | Germany |

Turkey (#40)
Turkey | Italy | France |

Georgia (#75)
England | Italy |

By Ryan Soares | Data from UEFA as of June 9, 2024

R: It was interesting to see the different strategies that nations have for developing players internationally. Team's with a top domestic league have the advantage since they have a higher base for player development (England is a great example with the Premier League). Countries without a domestic league have to get players into the top leagues so they can play higher quality football and then contribute to the national team. This can be seen in the viz with France, Netherlands, and

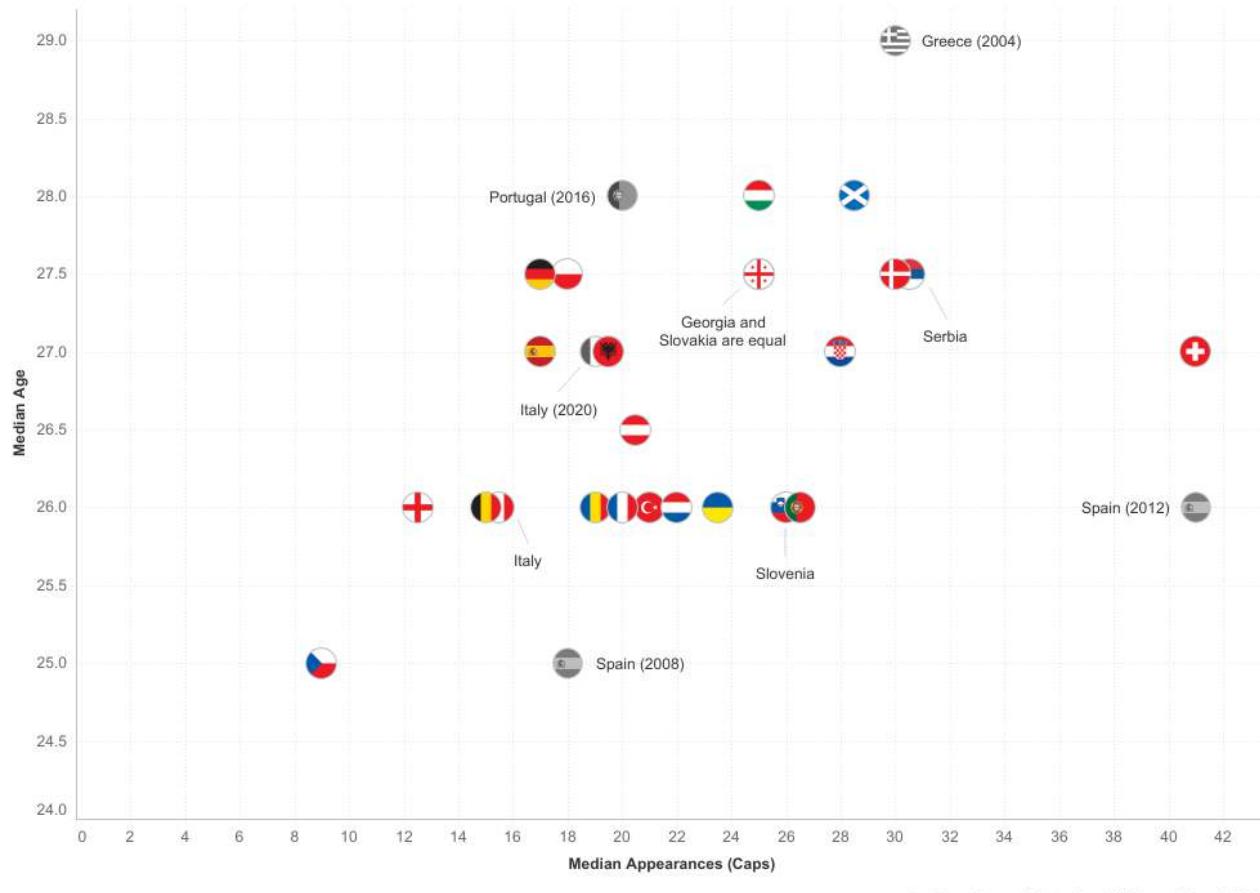
Portugal. As a Canadian, our men's national team is continuing in this direction with more players securing transfers to European leagues after the Copa America this summer.

In my opinion, I don't think where players play has a major impact on team cohesion since players often progress together through the national team ranks. The golden generations of Croatia and Belgium are good examples since the players came into the national team together but many did not play in the country's domestic league. Even Argentina only had one player in their domestic league in their World Cup 2022 squad. Having a top domestic league doesn't directly translate to success, but it does provide a solid base for developing players.

C: Your visual on age and experience also caught my eye. It is a wonderful way of letting someone understand a story without having to add written verbiage. The grayed previous competition countries give additional context. Why did you choose to use a scatter plot in this manner?

Balancing Age and Experience at Euro 2024

How do Europe's best compare across the age and experience of their squads?



By Ryan Soares | Data from UEFA as of Jun 9 2024

R: I thought the data really fit well in a scatter plot because it was spread out enough where you can see the differences but also group together similar countries. It was interesting to me that many of the tournament favorites had the same median age, but differed in the experience of their squads with England being the least experienced (but still made the final!). I wanted to add in the previous European champions to see if there were any similarities with current squads, but rather this highlighted some outliers in Greece '04 and Spain '12 (kudos to my colleague Steven Scott for suggesting the greyscale color). There were many stories to pick out and comparisons to make with the scatter plot, which is why I went with it!

CJ: Your PGA Tour Toughest Courses is one of the best functioning table views on Tableau Public. You spoke on LinkedIn on how it uses Tableau parameter actions and dynamic zone visibility to create

a drill-down within the table. Could you give us some more in depth technical detail as to how someone could create the same?

PGA Tour Toughest Courses

Courses played during the 2024 season ranked by average score **over** or **under** par. Click rows in the table below to view course details and hole-by-hole breakdown.

Courses that hosted major championships are indicated by ..



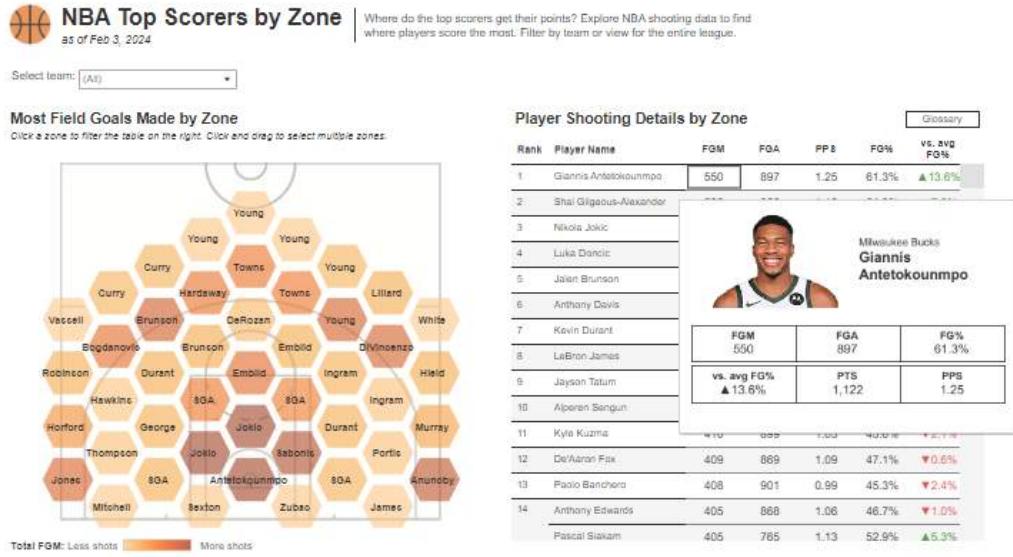
RANK	COURSE	PAR	YARDAGE	PAR	YARDAGE	Avg Score	+/-													
1	Pinehurst Resort & Country Club (Course No. 2) ★	70	7,548	70	7,548	72.89	+2.89													
#1 Pinehurst Resort & Country Club (Course No. 2)																				
PAR	YARDAGE	HOLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
70	7,548	PAR	4	4	4	4	5	3	4	4	3	5	4	4	4	4	3	4	3	4
SCORING AVG	+/-	Avg	4.1	4.4	4.0	4.4	4.7	3.3	4.1	4.4	3.2	5.0	4.2	4.2	3.9	4.2	3.3	4.3	3.2	4.1
72.89	+2.89	+/-	+0.1	+0.4	0.0	+0.4	-0.3	+0.3	+0.1	+0.4	+0.2	0.0	+0.2	+0.2	-0.1	+0.2	+0.3	+0.3	+0.2	+0.1
2	Royal Troon ★	►	71	7,385	73.85	+2.85														
3	Augusta National Golf Club ★	►	72	7,555	73.91	+1.91														
4	Muirfield Village Golf Club	►	72	7,569	73.50	+1.50														
5	Colonial Country Club	►	70	7,289	70.82	+0.82														
6	Quail Hollow Club	►	71	7,538	71.73	+0.73														
7	TPC San Antonio (Oaks Course)	►	72	7,438	72.51	+0.51														
8	Innisbrook Resort (Copperhead Course)	►	71	7,340	71.41	+0.41														
9	Torrey Pines Golf Course (South Course)	►	72	7,765	72.40	+0.40														
10	Arnold Palmer's Bay Hill Club & Lodge	►	72	7,466	72.33	+0.33														
11	Hamilton Golf & Country Club	►	70	7,084	70.29	+0.29														
12	Memorial Park Golf Course	►	70	7,435	70.25	+0.25														
13	Valhalla Golf Club ★	►	71	7,609	70.57	-0.43														
14	TPC Sawgrass (THE PLAYERS Stadium Course)	►	72	7,275	71.47	-0.53														
15	TPC Twin Cities	►	71	7,431	70.40	-0.60														
16	Castle Pines Golf Club	►	72	8,130	71.38	-0.63														
17	The Riviera Country Club	►	71	7,322	70.10	-0.90														
18	PGA National Resort (The Champion)	►	71	7,147	70.10	-0.90														
19	Vidanta Vallarta	►	71	7,456	70.07	-0.93														

Source: PGA | By Ryan Soares

R: I have to credit **Lorna Brown** with the approach for the drill-down table as she used it for her **Workout Wednesday challenge visualization**. I built upon this by adding a dynamic zone container in the drill-down.

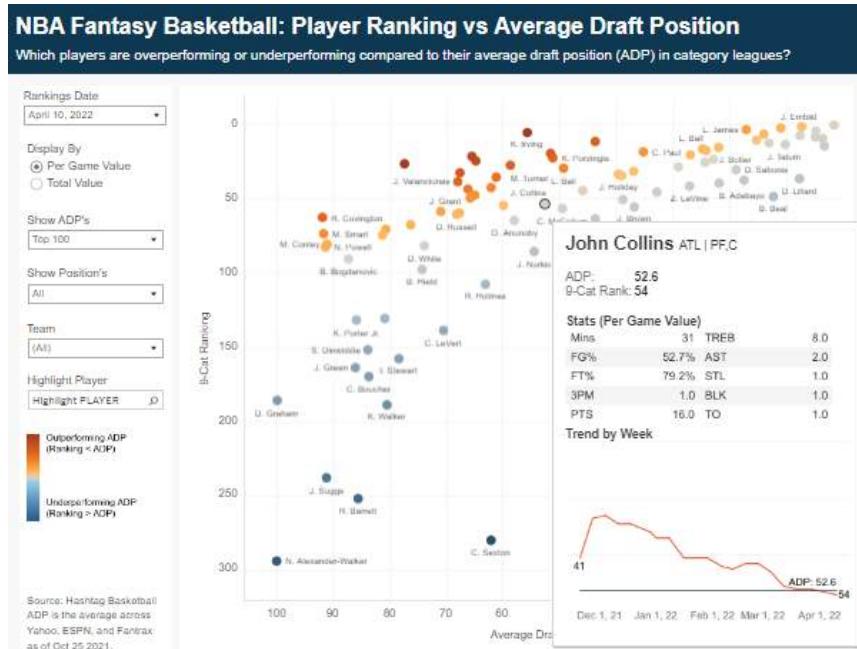
While it may seem like one table (I hope), there are actually three components that are stacked on top of each other: a table, a dynamic zone container, and a second table. There is one parameter for the course name and a parameter action sets it to the course selected by the user. On each of the table sheets, a calculation determines the rank of the course selected in the parameter and accordingly the sheets are filtered to display courses below and above the selected rank. The dynamic zone container is set to show when a value is selected in the parameter and the series of sheets are filtered for the selected course in a dashboard action filter, creating the drill-down effect. I highly recommend downloading Lorna's workbook to see all this in more detail!

CJ: I am guessing that basketball may be one of your favourite sports given the number of visualisations on it! You don't shy away from a well designed viz in tooltip. How do you decide the amount of detail that needs to be on the page, vs shown in the tooltip for your top scorers by zone and fantasy basketball visual?



R: I approach this decision by thinking about what information is critical for the user to draw insight from the viz for the particular use case. This information stays on the page so it can be consumed immediately by the user. Any secondary information that is not critical goes into the tooltips and the user can view it if they need extra detail. I treat tooltips as an interactive component that the user can choose to use to enhance analysis.

I designed both of these visualizations to function as tools for the user to explore. I included a key question in the sub-title that I am trying to answer through the viz. Each viz has a main visual up front (scatter plot and heat map) which is the basis for the viz and provides the main insight answering the key question. The tooltips contain additional details on the stats and trends, but this is not crucial information for the user.



CJ: I had such a choice when looking for great trellis charts on your public profile. You had visuals on the olympics, NBA and the tour de france. What was it about these stories that made a comparative view between teams or athletes displayed in this way?

Greatest Comebacks in NBA Playoff History

Minute-by-minute point differentials of the ten largest playoff comebacks



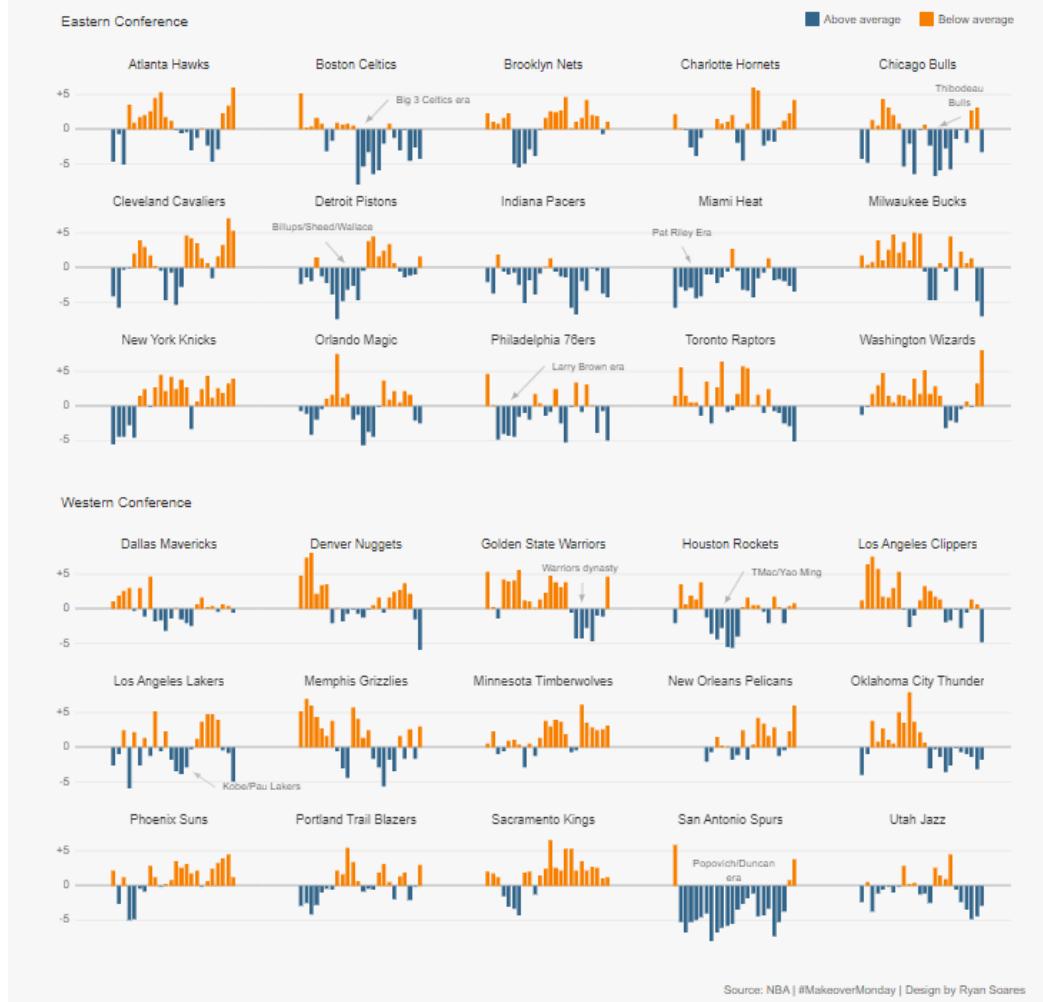
Note: Data not available for Lakers v Sonics 1989 West Finals Game 4 (29 point comeback)

Source: Basketball Reference | Design by Ryan Squires

R: As I've mentioned in a previous question, my style is all about simplicity. I really love trellis charts because they are so simple to understand and compare across different dimensions. There is only one chart type, so it is easy from a user comprehension perspective as they only have to understand a single chart. Sports data fits well because there are many dimensions to analyze across such as by year, sport, team, time, and so on. In these three visualizations, comparisons can be made very easily and the user can pick out specific points of interest. Particularly in the Margins of Victory viz, I enjoyed being able to pick out specific stages and watching back highlights of how the race unfolded. An important point to mention here on trellis charts is that fixed axes are important to have accurate comparisons and avoid misleading the audience.

NBA Defensive Rating Since 1996-97

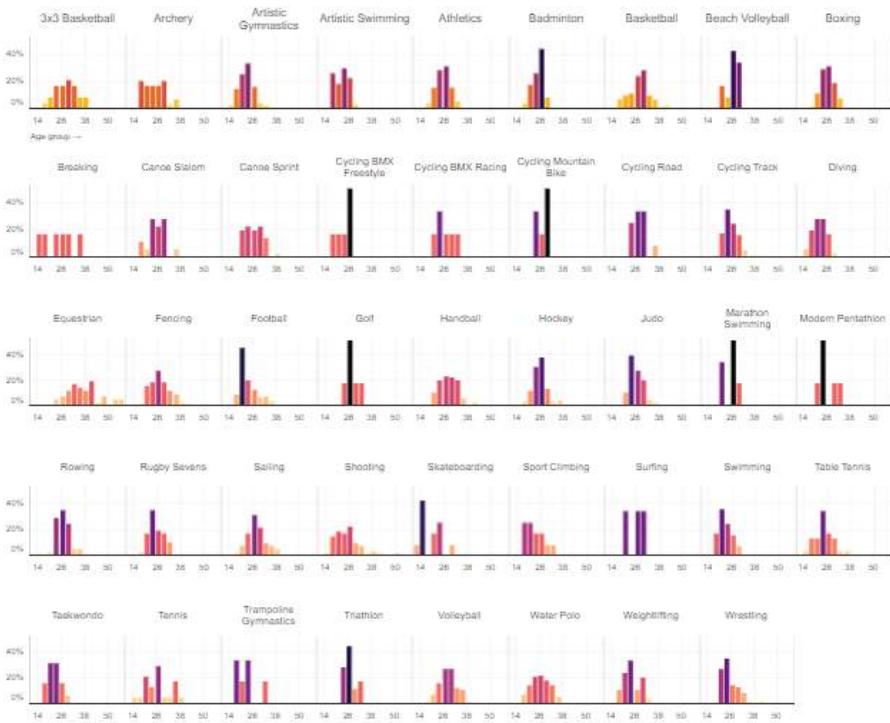
Defensive rating measures the number of points allowed per 100 possessions. The charts below show the difference between each team's defensive rating and the league average since the 96-97 season.



C: Looking specifically at the Olympics visual of medallists age distribution – I loved some of your call outs that most medalists are in their early to mid 20's for most disciplines, with a few outliers when we look at perhaps equestrian or skateboarding. What I'd love to know though, is maybe your top 3 moments from the games as a collective?

Olympic Medallists Age Distributions

How do the ages of Olympic medallists differ by sport? The charts below show the percentage of total medallists by age group for all sport disciplines at the Paris 2024 Olympics. Ages are grouped in buckets of 3 years.



By Ryan Soares | Source: IOC

R: It's hard to pick just 3 moments since the Games were great, but here they are:

1. Djokovic winning the gold medal in men's tennis over Alcaraz. It was a great match and a huge achievement for Djokovic, so it was awesome and exciting to watch.
2. USA vs Serbia semi-final in men's basketball. That was one of the best basketball games I've seen with how much passion was shown by the stars of Team USA. The Americans were seriously threatened by Serbia and at risk of going out, so seeing them come back and winning the way they did was very entertaining.
3. Remco Evenepoel double gold medals in the time trial and road race. He had a podium finish in this year's Tour de France and followed it up with two Olympic golds. The picture of him celebrating after crossing the finish line with the Eiffel tower in the background was iconic.

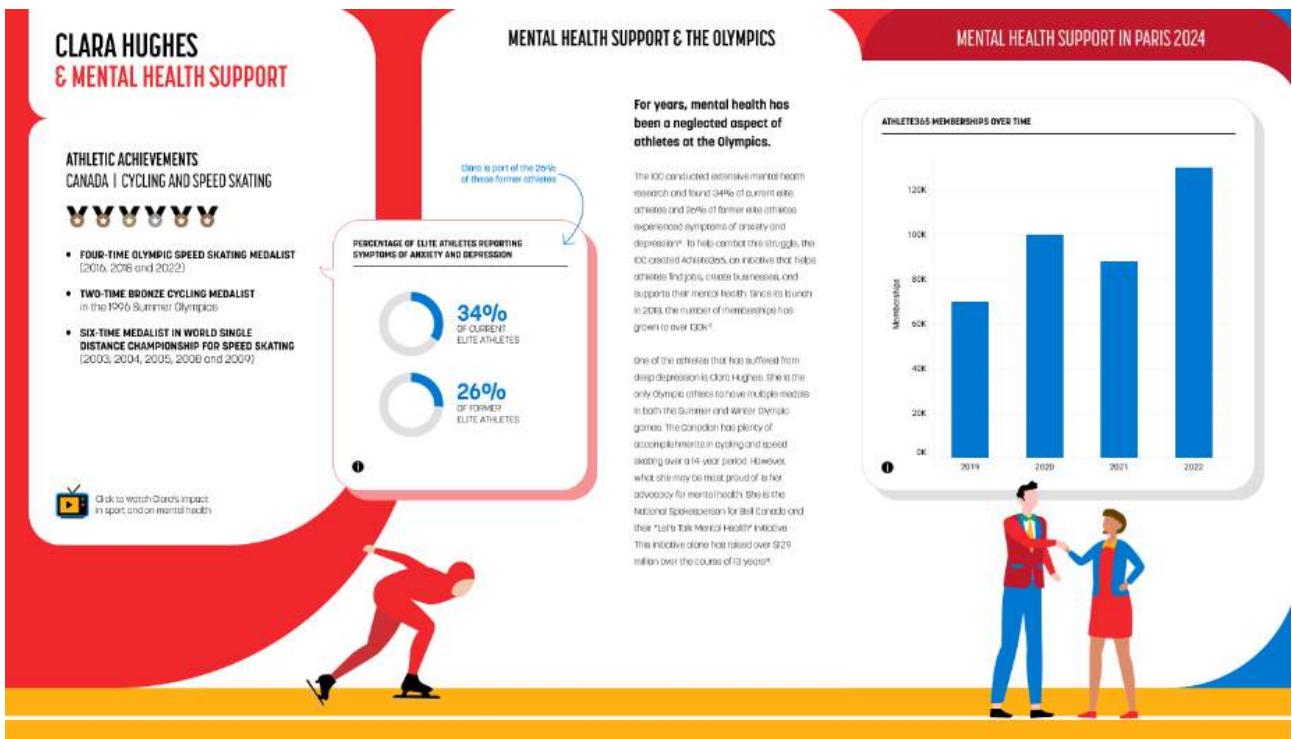
Apart from this, I really liked the venues for some of the events. I'm not a follower of fencing but I watched it since the Grand Palais was a great backdrop. Beach volleyball and road cycling in front of the Eiffel Tower were also nice locations.

CJ: As part of a Viz Games run internally at your company, you collaborated on a wonderful infographic design for the olympics and progressing global issues. Can you speak a little to your choices behind the wide print style and the order you chose to showcase the different issues?

R: Firstly, I have to mention my amazing colleagues and teammates Maggie Eng and Steven Scott as I couldn't have done this alone. It was a true team effort with Steven focusing on storytelling, Maggie on design, and myself on Tableau development.

We decided on the horizontal scroll layout for the story to be read in a linear fashion. It also allowed the design elements to flow nicely into one another and for the tabbing functionality to work well. The choice of athlete order was based on how strong we felt the connection would be with the Paris 2024 Olympics and the audience. Gender equality was a big achievement at this year's Games, so we started with this issue and closed with environmental sustainability which was a focus for Paris 2024 and will be for future Olympic Games. Despite the order, each issue holds the same importance as they are priorities for the IOC to address now and progress towards in the future.

We did a walkthrough of the design approach at the August Sports TUG event which can be found [here](#).



CJ: You've recently started the Sports Tableau User Group alongside Mo, Fred and Steven. What exciting things can we expect from it in months to come?



Industry

Sports Tableau User Group

[Join](#)

About

Here at the Sports Tableau User Group we want to create a global Tableau-user community of sports enthusiasts and professionals. Our goal is to share our passion and insights within the sports industry. We want to increase accessibility to the sports industry by sharing professionals' journeys.

We'll be discussing:

- How sports enthusiasts and professionals think about the game - on and off the field - through data / best practices
- How design is an important part of communicating ideas within dashboards
- How functionality can add and improve the user experience
- Much more!

Due to this Tableau User Group being a global community, we will be rotating the time of each event. But don't worry! We will record all the sessions so you can catch up whenever is convenient for you.

R: We had our first Sports TUG event in early August and it was great to get started and interact with the Tableau sports community! Our events are going to be a mix of hearing experiences from professionals in the sports industry and sharing Tableau know-how related to sports visualizations. Our goal is to increase accessibility to the sports industry by sharing the community's passion and insights on data analytics in sports. We'll soon be sharing details about our next event taking place in November!

CJ: Thank you for talking through some of your visuals! Are there individuals in the community that you look to for inspiration, both sports and wider?

R: The community has always been a great source of inspiration so there are too many people I could mention here. There is lots of good content being produced on a daily basis, so kudos to everyone who shares their work publicly as I'm sure it has inspired many other authors like me. I want to thank the **Makeover Monday** organizers, both current and former, for the platform they provide for learning and getting involved in the community. I started with Tableau through Makeover Monday so it has been an important part of my journey. Tableau aside, I enjoy data-driven journalism type work from publications such as **The Pudding**, **FlowingData**, **Reuters**, and the **New York Times**. I've also recently discovered **SportsBall** on Instagram which does some very fun hand drawn sports data visualizations.

LOGGING OFF,

CJ

CUSTOM METRIC WHEEL IN TABLEAU

Hi all,

A small bonus blog to the blog last week on radials in Tableau.

Even More Radials In Tableau

This blog is only intended for those that are wanting to build on that radial chart, I won't be re-explaining any of the calculations as a heads up.

This blog:

- Adjusts our visual to be able to plot a "before" and "after" comparison. Instead of one value.
- We customise the layering to toggle labels and colours.
- We take a look at some cosmetic amendments to change the feel of the radial / wheel.

This is what we will create.

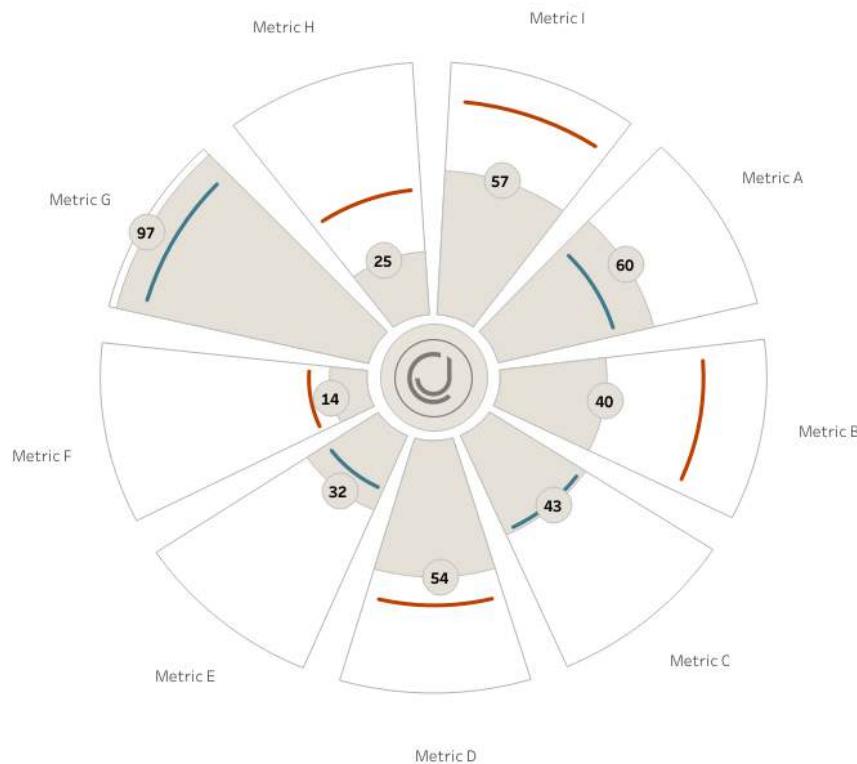
● Current Value vs Previous Year

Increased PY to CY | Decreased PY to CY

Previous Year Marks
On

Value Labels
On

Cosmetic Parameter
1.2



DATA PREP

First we use a slightly amended dataset, that you can find in the git repo.

	A	B	C	D	E	F
1	Player Name	Year	Sub Grouping	Sub Group Rank	Value	
2	Player A	Current Year	Metric A	1	60	
3	Player A	Current Year	Metric B	2	40	
4	Player A	Current Year	Metric C	3	43	
5	Player A	Current Year	Metric D	4	54	
6	Player A	Current Year	Metric E	5	32	
7	Player A	Current Year	Metric F	6	14	
8	Player A	Current Year	Metric G	7	97	
9	Player A	Current Year	Metric H	8	25	
0	Player A	Current Year	Metric I	9	57	
1	Player A	Previous Year	Metric A	1	45	
2	Player A	Previous Year	Metric B	2	76	
3	Player A	Previous Year	Metric C	3	41	
4	Player A	Previous Year	Metric D	4	65	
5	Player A	Previous Year	Metric E	5	23	
6	Player A	Previous Year	Metric F	6	22	
7	Player A	Previous Year	Metric G	7	87	
8	Player A	Previous Year	Metric H	8	50	
9	Player A	Previous Year	Metric I	9	85	
0						

ADDITIONAL LINE FOR PREVIOUS YEAR

if
= "On" and
= "Previous Year"
and > 13 and < 24

```
    then  
    MAKEPOINT(.,)  
    end
```

A small amendment to our previous chart, we now only look for the T points between 13 and 24.

That's because these are the outer edge of our radial.

COLOUR

Secondly we want to amend our colour on our chart to be able to account for both years.

We will want to end up comparing the current and previous year for each metric.

```
    if  
    { Fixed : max()} > then "Smaller"  
    else "Maximum"  
    END
```

With this calculation we will be able to see of the current and previous year which is larger.

That's the colouring sorted on the new chart.

SEGMENT DESIGN

The third main thing I did was add in some flexibility to the wheel design. You can see the segments slightly unwind from one another using the parameter. The calculation that drives this is calculation 4:

*

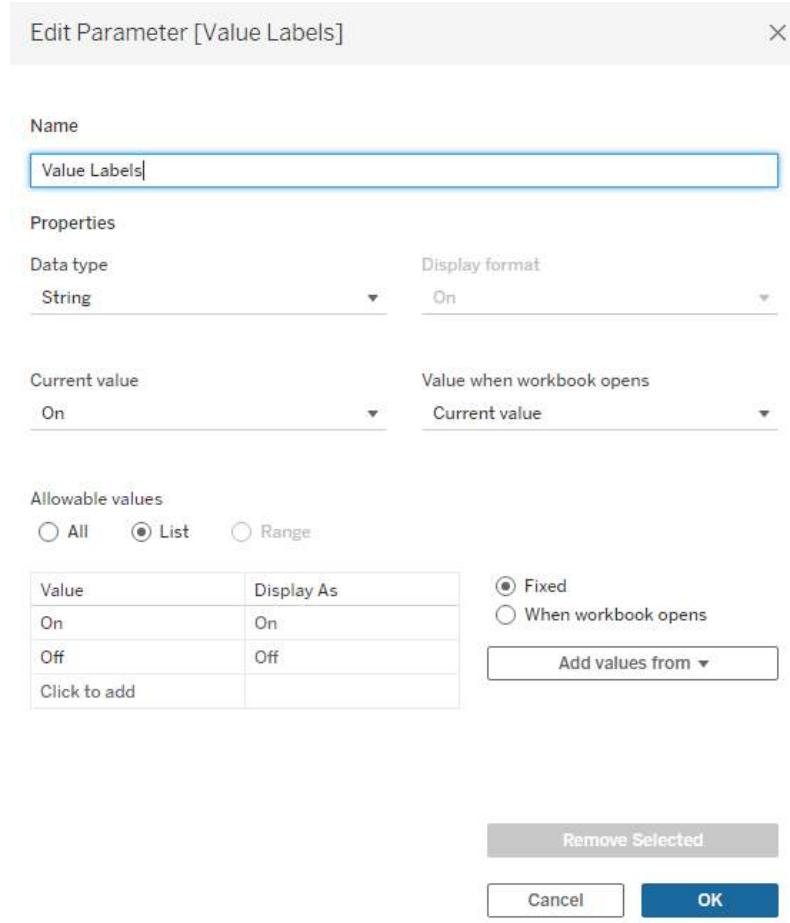
/ ()/

Previously we space our points out perfectly and then divide it by the number of segments, here we just add in a slight increase to what that divisible amount is, making the segments pull away from one another.

PARAMETERS WITH MAKEPOINT

The final aspect I wanted to influence was the idea of being able to toggle on and off different parts of the chart.

Lets take a simple example of turning off the value labels.



I can then write a calculation that says, if the parameter is on, then use the makepoint calc.

if = "On" and

= 18

and = "Current Year"

then

MAKEPOINT(,)

END

These small amendments take me from

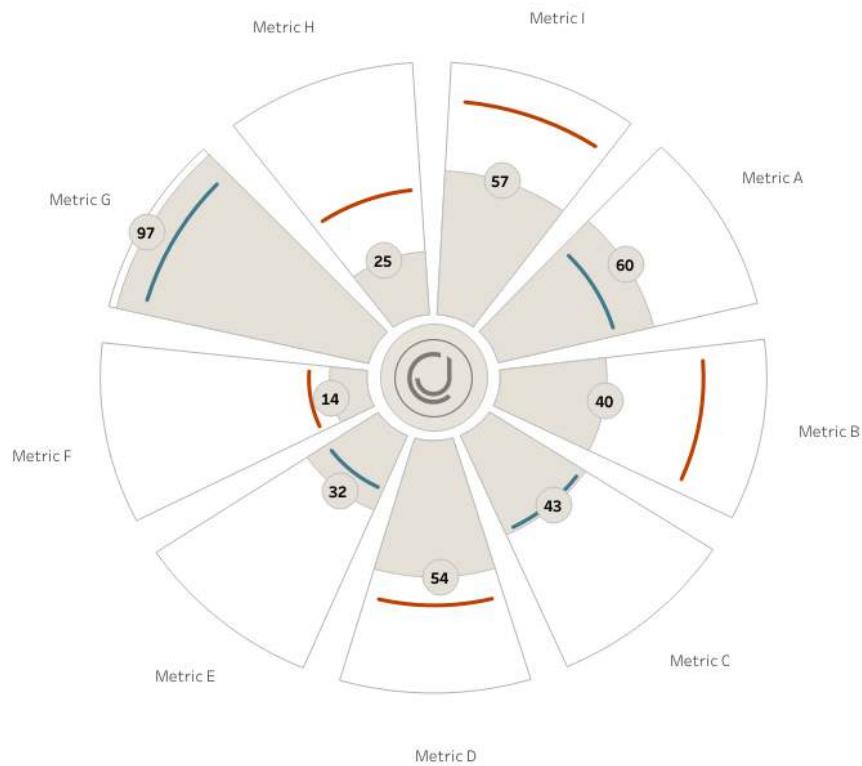
● Current Value vs Previous Year

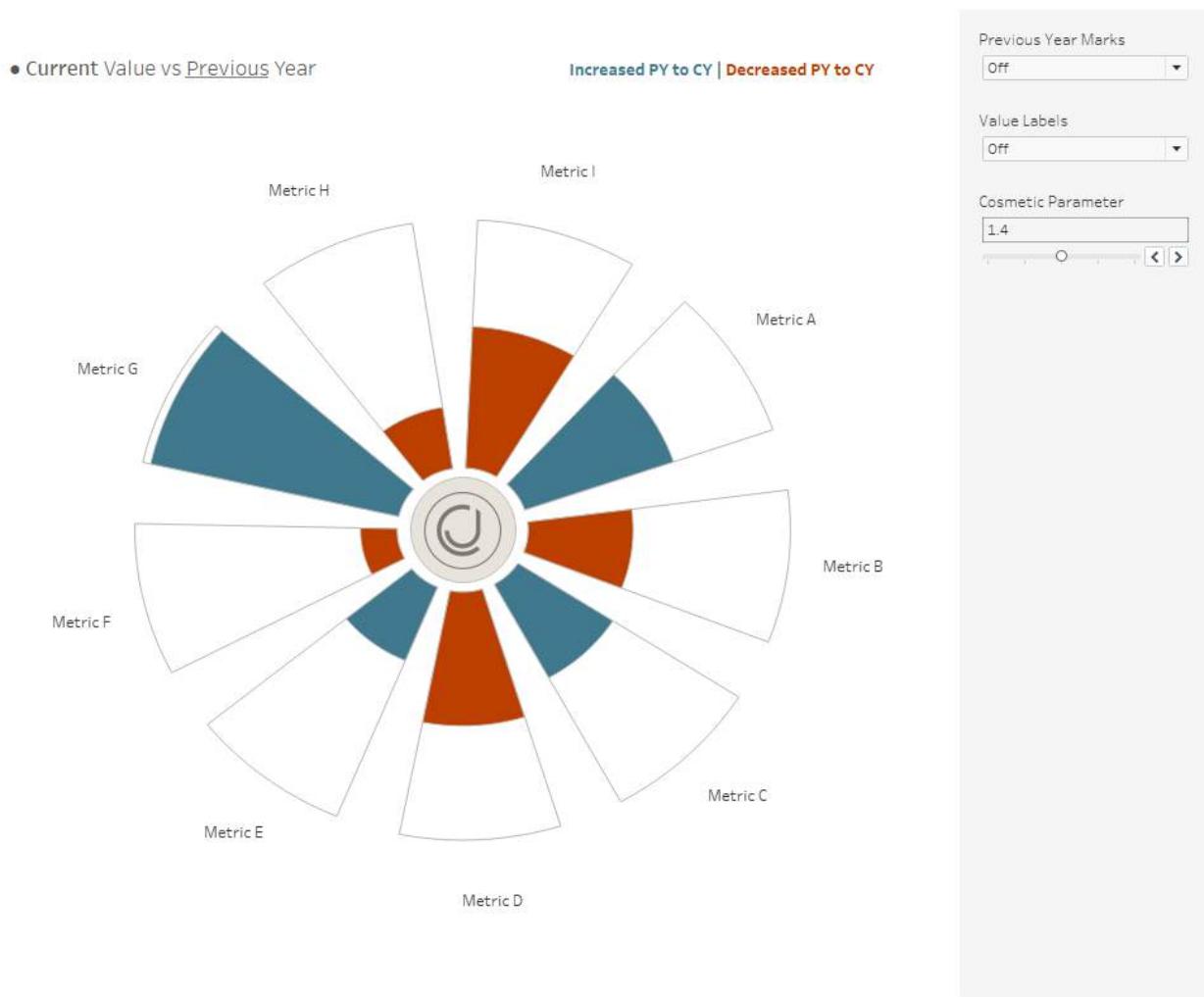
Increased PY to CY | Decreased PY to CY

Previous Year Marks
On

Value Labels
On

Cosmetic Parameter
1.2





in just a few clicks.

Now.... if none of the above made sense, simply replace the data source from the template with your own data or reach out on socials.

LOGGING OFF,
CJ

EVEN MORE RADIALS IN TABLEAU

Hi all,

Hoping everyone is keeping well and enjoyed the summer months? I personally have loved it. My puppy turns one in a months time, he has been a wonderful addition to the family and a welcomed distraction from sitting at the computer long into the evening.

I haven't done tutorials for a while, mainly in anticipation to see what happens with viz extensions but I've also had the fortune to showcase 10 fantastic guest blogs to date this year. But this weekend I was itching to get pen to paper to share another radial design. So what's different with this radial that others don't have?

I've written extensively about radial charts in Tableau in the past.

Player Metrics

Pizza Plot Player Comparisons

Each time learning a little more about maths, design and Tableau.

Where this radial will stand out is because:

We don't look to create the origin at (0,0) – we scale it to allow for an image in the center.
 We build it completely in Tableau, not needing to design anything outside as images.
 We utilise relationship blends instead of a forced join, hoping to improve performance.
 You have the flexibility to add or remove as many additional groups to the radial.

Even More Radials



With all my work, it can be downloaded for free from Tableau Public.

The template dataset can be found in the Git repository. Both these links are at the top of the page.
 Of course, the easy route would be, fill in the template, replace my dataset and hit refresh. But we will still walk through all the calculations if you ever want to go beyond what I've built.

A	B	C	D	E	F	
1	Player Name	Grouping	Main Group Rank	Sub Grouping	Sub Group Rank	Value
2	Player A	Group 1	1	One A	1	60
3	Player A	Group 1	1	One B	2	40
4	Player A	Group 1	1	One C	3	43
5	Player A	Group 2	2	Two A	4	54
6	Player A	Group 3	3	Three A	5	23
7	Player A	Group 3	3	Three B	6	92
8	Player A	Group 3	3	Three C	7	100
9	Player A	Group 3	3	Three D	8	32
10	Player A	Group 4	4	Four A	9	48
11	Player A	Group 4	4	Four B	10	71

⊕- Data (Multi-Level Groups)



Data — T ▾

How do relationships differ from joins? Learn more

	# T	=# Calculation 002. Rank	*# Calculation 007. polypath	
T (T)	1	1	1	
	2	2	2	
	3	3	3	
	4	4	4	
	5	5	5	
	6	6	6	
	7	7	7	

All the calculations are in order.

001. Max Dimensions

{ Fixed : max()}

This is the number of records, in this case 10. You could do a count, but I just find the maximum subgroup rank.

002. Rank

```

if <= 12 then elseif <=24 then -12
      end
  
```

The value T is brought in to 'densify' the data. Don't ever change this value, PLEASE. This is joined as a relationship, to save on records of data. T will be used to create the curvature seen in the chart.

Important difference from other radial charts: We split T into two sections as we want to make the inner part of the polygon and the outer part.

003. Angle

$360/(12-1)$

360 degrees in a circle... then split by the number of T values

i.e 12 because we have 12 on the inner ring, 12 on the outer ring.

This won't change so is hard coded in.

The angle is consistent throughout and will always be based on number of dimensions

iii Columns				
iii Rows	T (T)	001. Max Dimensions	002. Rank	003. Angle
1	10	1	32.727272727	
2	10	2	32.727272727	
3	10	3	32.727272727	
4	10	4	32.727272727	
5	10	5	32.727272727	
6	10	6	32.727272727	
7	10	7	32.727272727	
8	10	8	32.727272727	
9	10	9	32.727272727	
10	10	10	32.727272727	
11	10	11	32.727272727	
12	10	12	32.727272727	
13	10	1	32.727272727	
14	10	2	32.727272727	
15	10	3	32.727272727	
16	10	4	32.727272727	
17	10	5	32.727272727	
18	10	6	32.727272727	
19	10	7	32.727272727	
20	10	8	32.727272727	
21	10	9	32.727272727	
22	10	10	32.727272727	
23	10	11	32.727272727	
24	10	12	32.727272727	

004. Rank Angle

*

/()

Once we have the T values Ranked and the angle (based on the segments) we want to then space them out.

By dividing it by the number of categories we split the circle into the perfect segment size. i.e a 360 degree circle, divided into 2 dimensions leaves each segment as a half.

So far, we have just created where dots would be spaced out, now we need to make them circular.

005. x

```
if <= 12 then cos(radians())*1
elseif < 25 then cos(radians())*5
END
```

This calculation is trigonometry.

cos our value will help make the circle, it is used alongside the sin calc in Y.

You will see that the standard circle is radius 1.

We make the 100% mark of radius 5.

These multiplier values are important in later calcs.

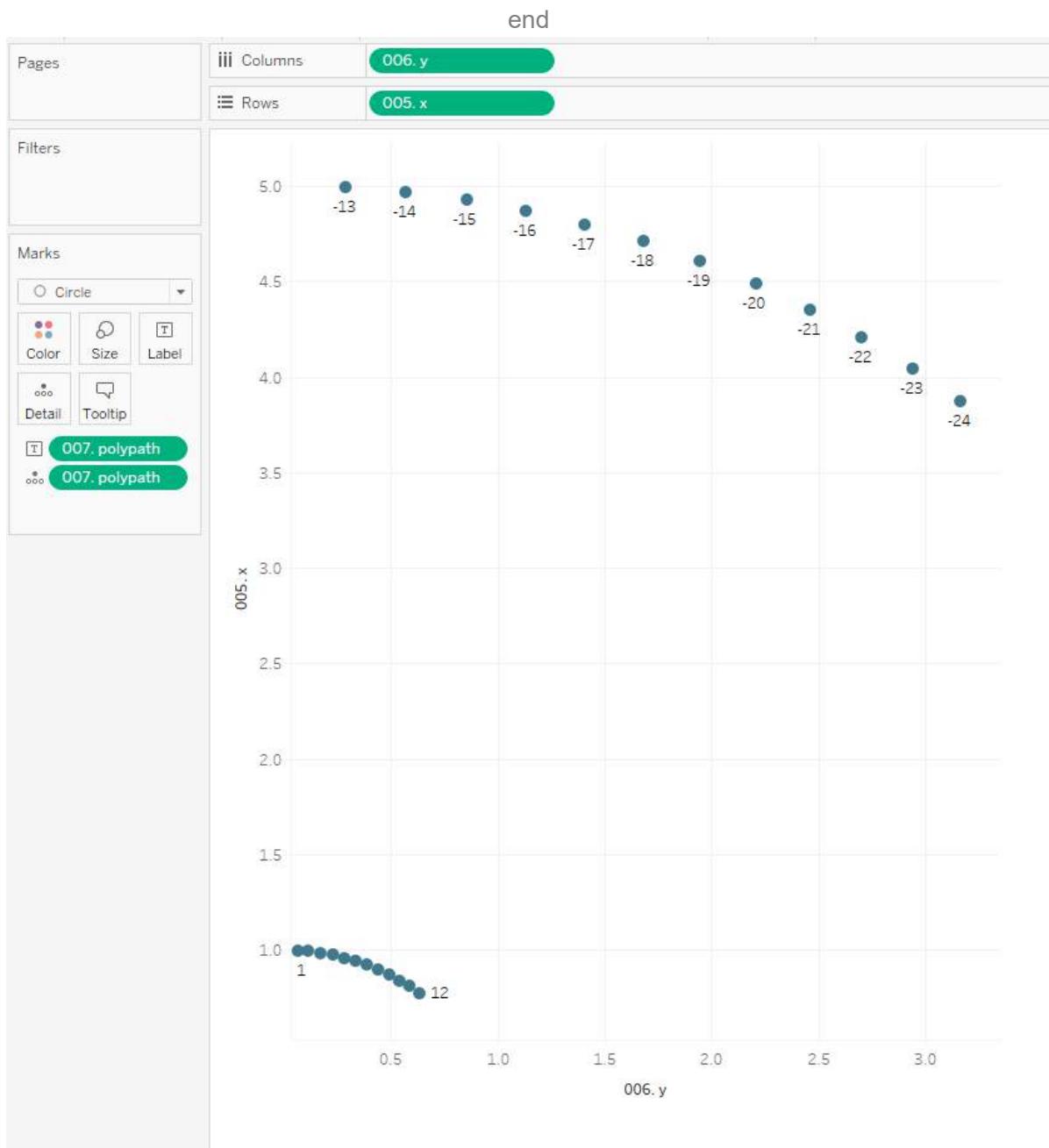
006. y

```
if <= 12 then sin(radians())*1
elseif < 25 then sin(radians())*5
END
```

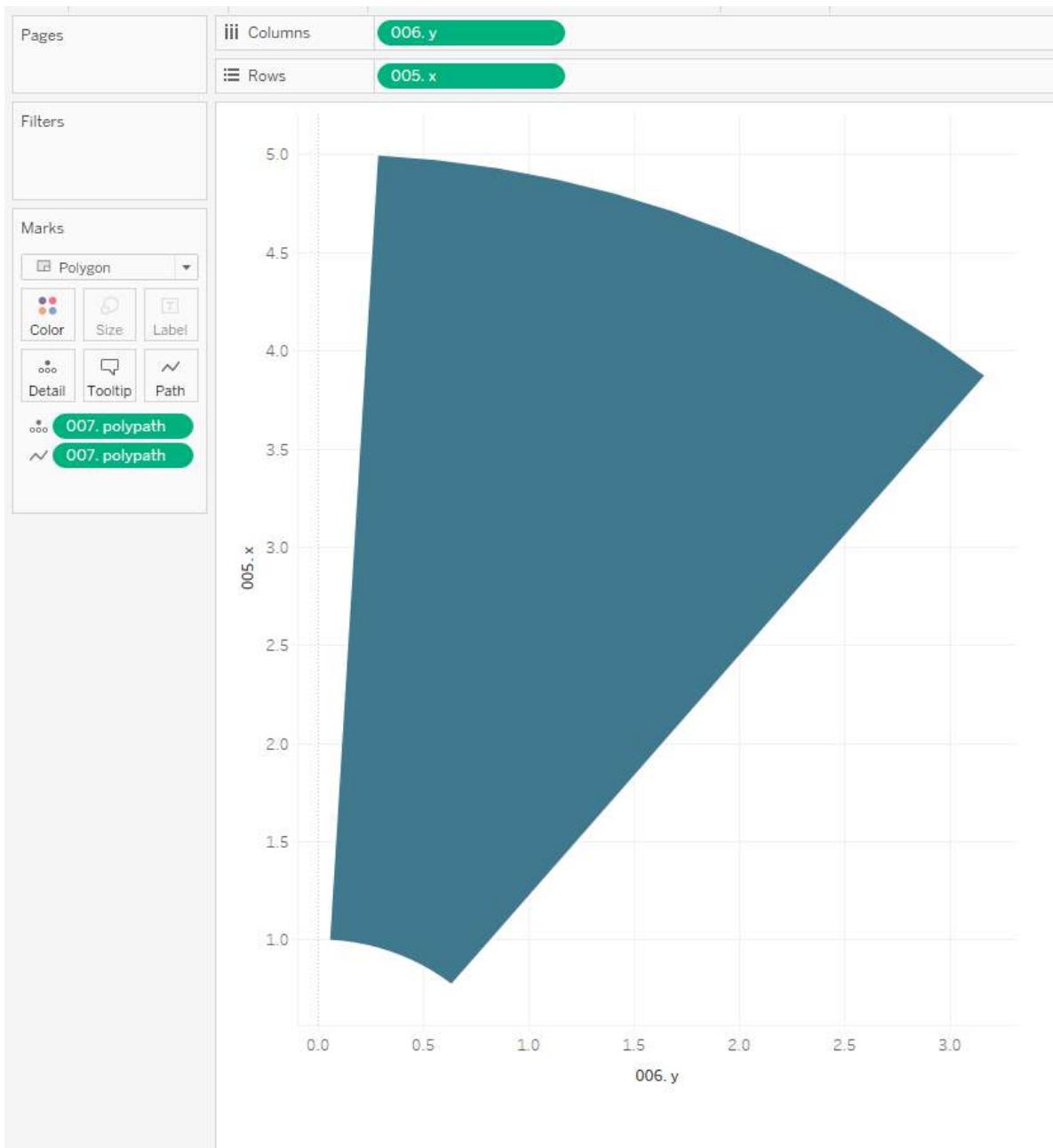
same as x calc above, but sin and cos work in unison to make it a circle.

007. polypath

```
if <= 12 then
else -
```



This calculation is new. The reason we need a path is because we need a route to join our lines up. We will want to join along the inner edge of the polygon, before going to the outer edge. By adding this in it'll join them in the correct order. In this case -13 would join to 1.



So far we have every polygon built, but they are overlaid on top of one another.

008. x circle

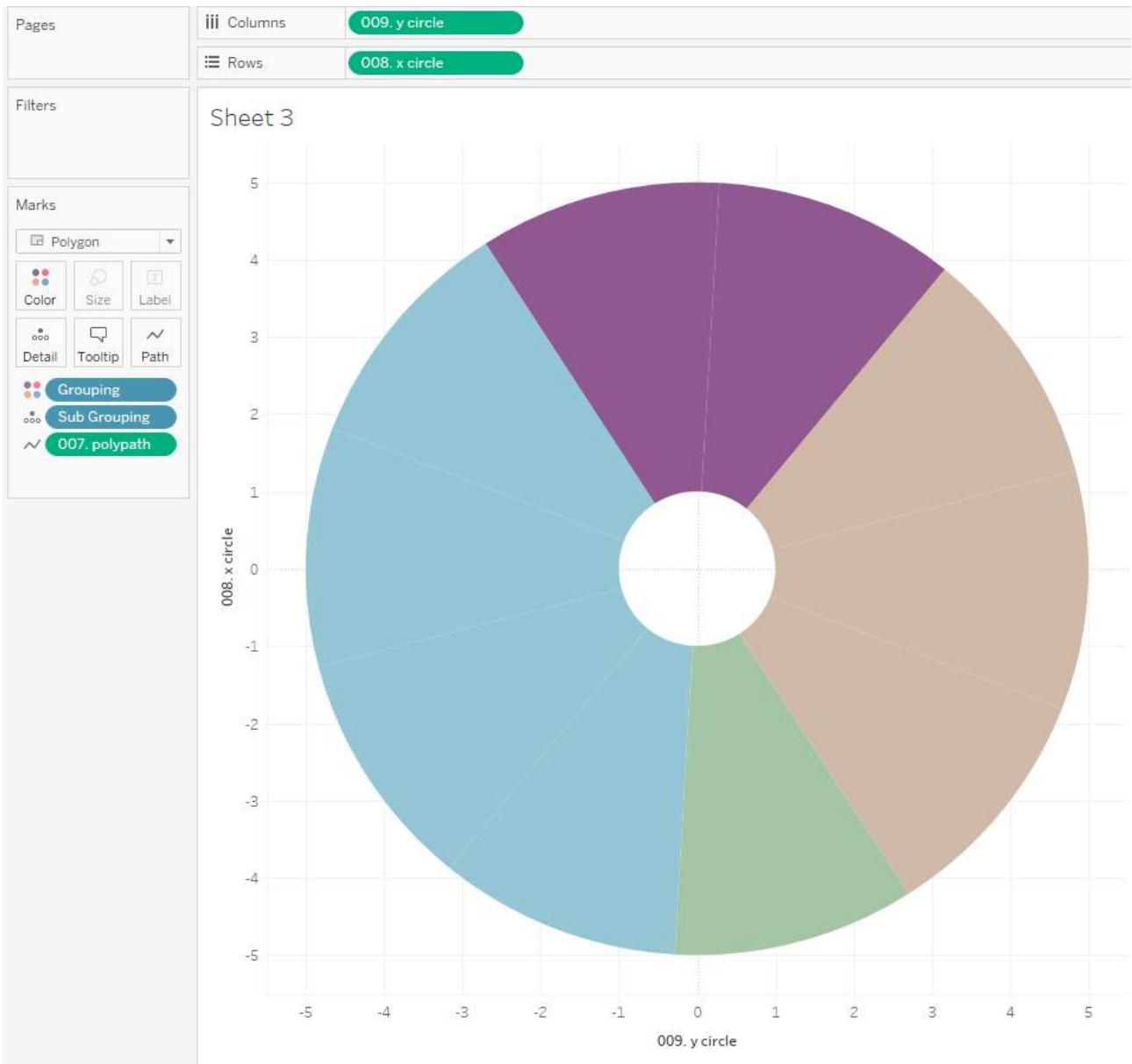
$$\begin{aligned} & * (\cos(\text{radians}(360/*))) \\ & - * \sin(\text{radians}(360/*)) \end{aligned}$$

In this calculation we rotate the segments around a central point 0,0.

009. y circle

$$\begin{aligned} & * (\sin(\text{radians}(360/*))) \\ & + * \cos(\text{radians}(360/*)) \end{aligned}$$

We do the same to y. The syntax for this calculation is $y1 = x0\sin(\theta) + y0\cos(\theta)$ -> arbitrary point rotation. Either take my word for it, or simply google the formula for rotating objects around a coordinate.



We've now rotated the segments, leaving our gap in the middle. What you'll now realise though is that the segments aren't actually scaled to our value in our template.

```

010. resized x
if > 12 then /5
* (1 + (/100) * (5-1))
else
*1
END

```

Here we take the strength value and make it proportional to the outer edge. $(R1 + (P/100) * (R2 - R1))$

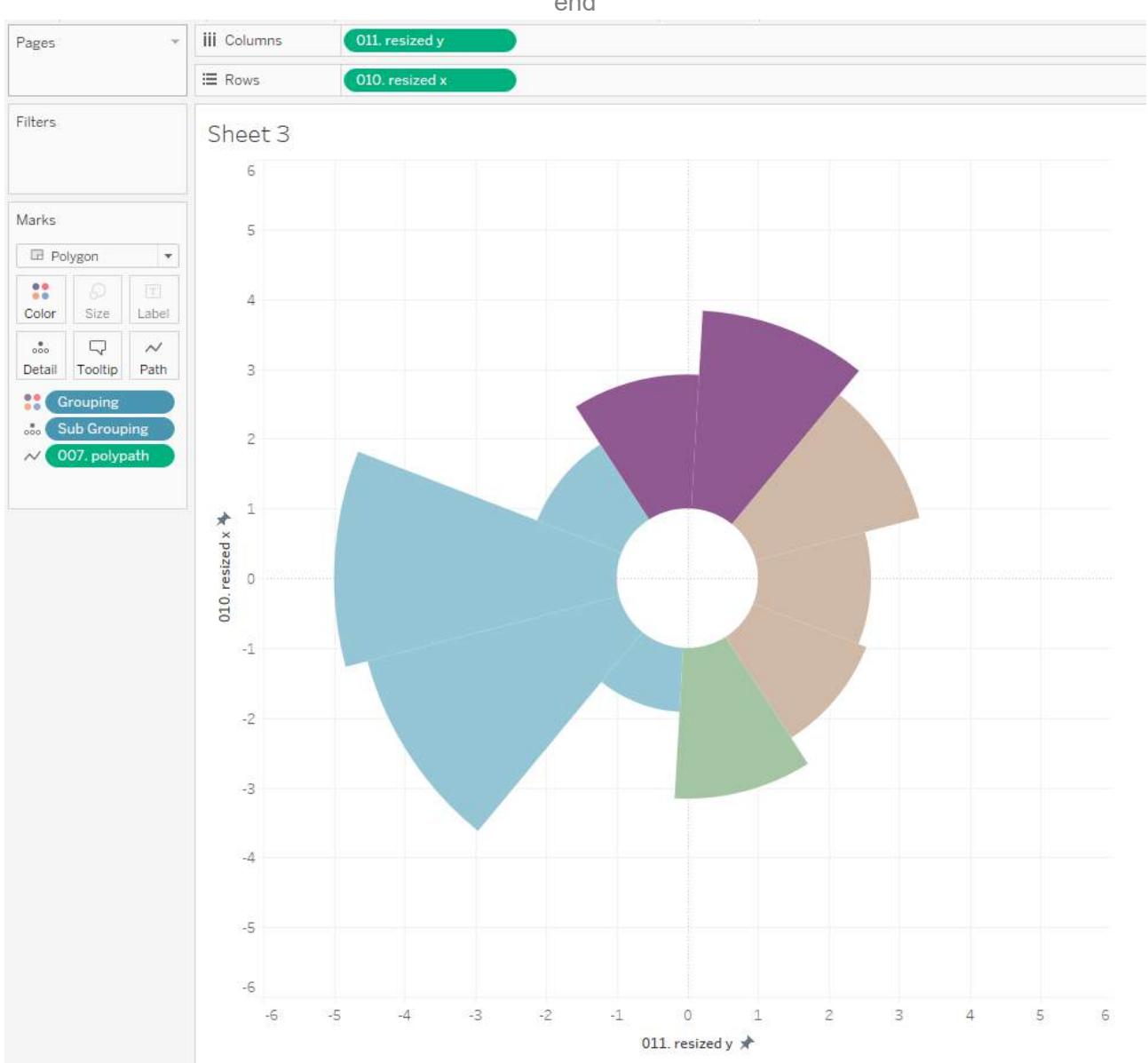
These are the values we end up plotting.

You will see the value 5 in here because we made our 100% radius 5 in earlier calculations.

```

011. resized y
if > 12 then
/5
* (1 + (/100) * (5-1))
else *1

```



Awesome. Our radial chart is now complete.

Last thing to do is plot it.

I create various layers.

makepoint(0,0) To be able to plot my shape in the middle of the radial.

MAKEPOINT(,) To be able to plot a frame for the radial.

MAKEPOINT(,) To be able to plot the actual chart

if = 6

then

MAKEPOINT(*6,*6)

END

This is for the labels, we want the titles just outside the outer radius. Remember the radius of our chart is 5.

Then finally, bring in the values for the chart itself. 18 is used because its our middle number for the outer edge.

if = 18

then

MAKEPOINT(,)

END

Even More Radials



If you have any questions please reach out on socials.

LOGGING OFF,
CJ

DORLING MAPS IN TABLEAU WITH RAISA HANNUS

Hi all,

I always love when community members find new ways of using tools, or adapting charts in innovative ways, so when I came across Raisa Hannus' World Population visualisation I reached out to let her know how amazing it really is. Raisa has been kind enough to share a little more around the design process of that visualisation.



X [f](#) [in](#)

Raisa Hannus

Key Largo, Florida, United States

DataViz Enthusiast #B2VB #WOW2022 #MakeoverMonday | Tableau Desktop Specialist | Business Analyst | MSc in Applied Math | Diving Instructor | DIY & Interior Design Geek | Yogi

Follow

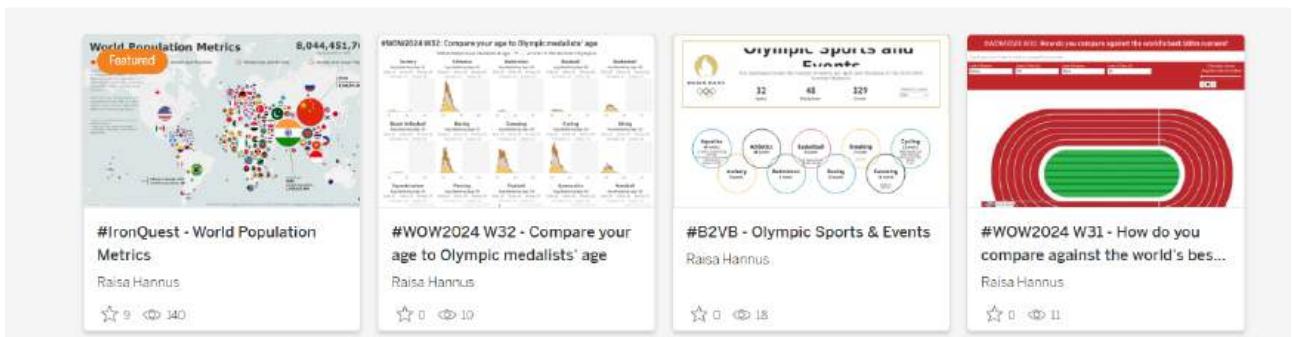
Hire Me

Vizzes

Favorites

Following

Followers

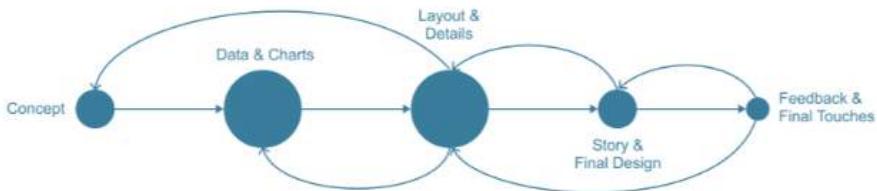


Thank you, CJ, for featuring my guest blog, it's a great honor! My design process for the "World Population Metrics" viz for #IronQuest didn't quite follow a typical route for a defined dataviz assignment. Although I made it in 5 distinct phases, the setup was more like "I want to make a cool chart, let's see if it works and what it could be used for".

You can find my World Population Metrics visual, [here](#).



The Design Process for World Population Metrics (size of the circle approximates time/effort for each step).



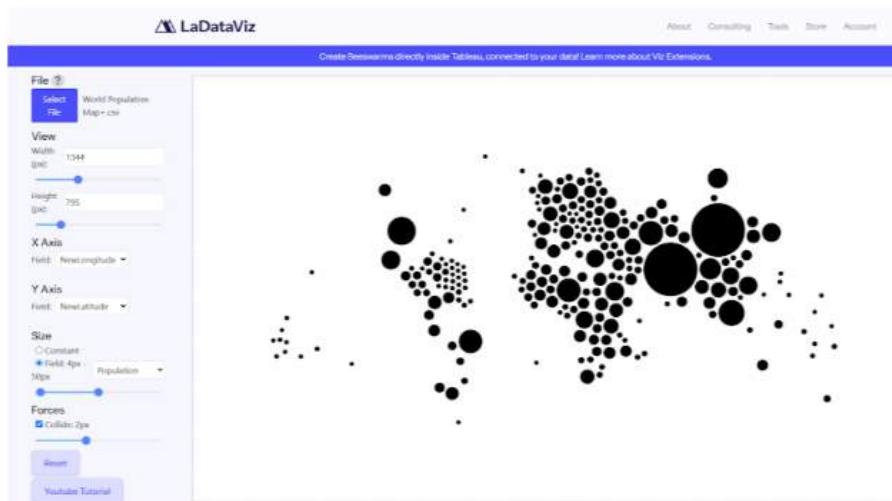
Phase 1 – Concept

The last time I used Tristan Guillevin's [AdvViz Beeswarm generator](#), I realized that you could place dimensions on TWO axes, like in a scatter plot rather than just a beeswarm but without overlapping marks. And if you put map coordinates on X and Y, and a measure on size, it would be a Dorling map!

Dorling map/cartogram = map that uses circles to represent geographic regions and their corresponding data values.

I had seen some impressive Dorling maps from [Visual Capitalist](#), [Ansgar Volsing](#), and [Neil Richards](#), wishing there was a way to make one without coding skills. To test the idea, I downloaded the most recent population data from the United Nations (UN) and geocoded map coordinates for the countries in Google Sheets.

Once I uploaded the data into the Beeswarm generator, I noticed that negative latitudes and longitudes did not appear, meaning I only had the Northeastern quadrant of the world. I modified my dataset by adding 90 degrees to all latitudes and 180 to longitudes and uploaded it again, this time with success. Adjusting the circle size up to 50px and colliding forces down to 2px seemed closest to an actual map.



With a downloaded Tableau Public extract, all I had to do was come up with a design. I tested it with some round country flags in my Shapes folder and made the background dark to bring up the colors. Pretty cool, I thought, but so what?

Phase 2 – Data & Charts

The #MakeoverMonday 2024 Week 6 challenge 'World Population' was a perfect topic for my Dorling map!

The dataset had 2023 population by country, and six other population-related metrics from Worldometer, which again used UN data. I brought the data and a csv version of my beeswarm onto Tableau and created a relationship calculation for countries with mismatched names.

In addition to my flag map, I combined the six population metrics into pairs totaling three additional maps. Each would have the population on size, one metric on color, and another as a symbol.

Based on some research, common logic, and a few scatter plots, I found that the fertility rate lowers median age and immigration contributes to population growth. However, population density and urbanization didn't seem directly related to each other or the other metrics, at least in the short term. Once set on my pairings, I created a parameter, a basic Geometry, and a geometry for each set of metrics. With these fields, I could make almost the entire map.

Metrics (parameter)

Float 1,2,3,4 with display names Population, Growth & Migration, etc.

Geometry (using the Beeswarm XY coordinates)

MAKEPOINT(- / {MAX()}, / {MAX()})

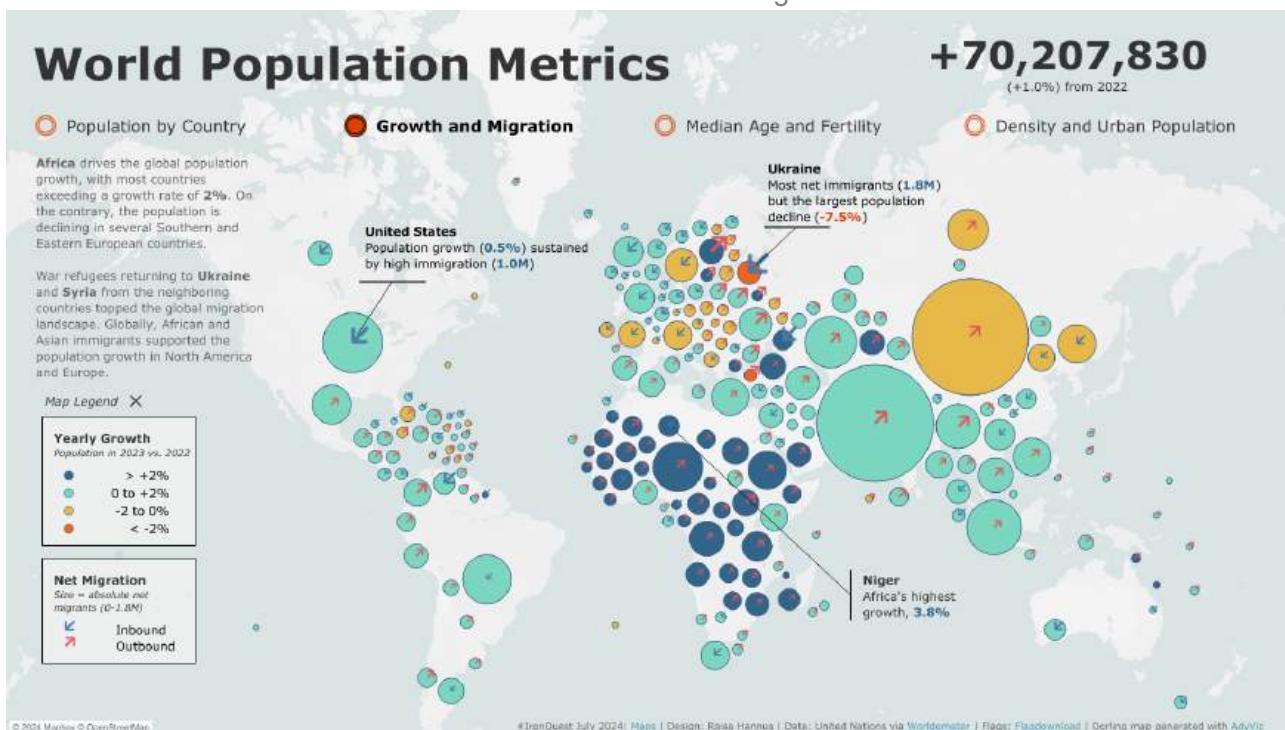
Geo 1 (Repeat for Geo 2-4)

IF Metrics=1 THEN END

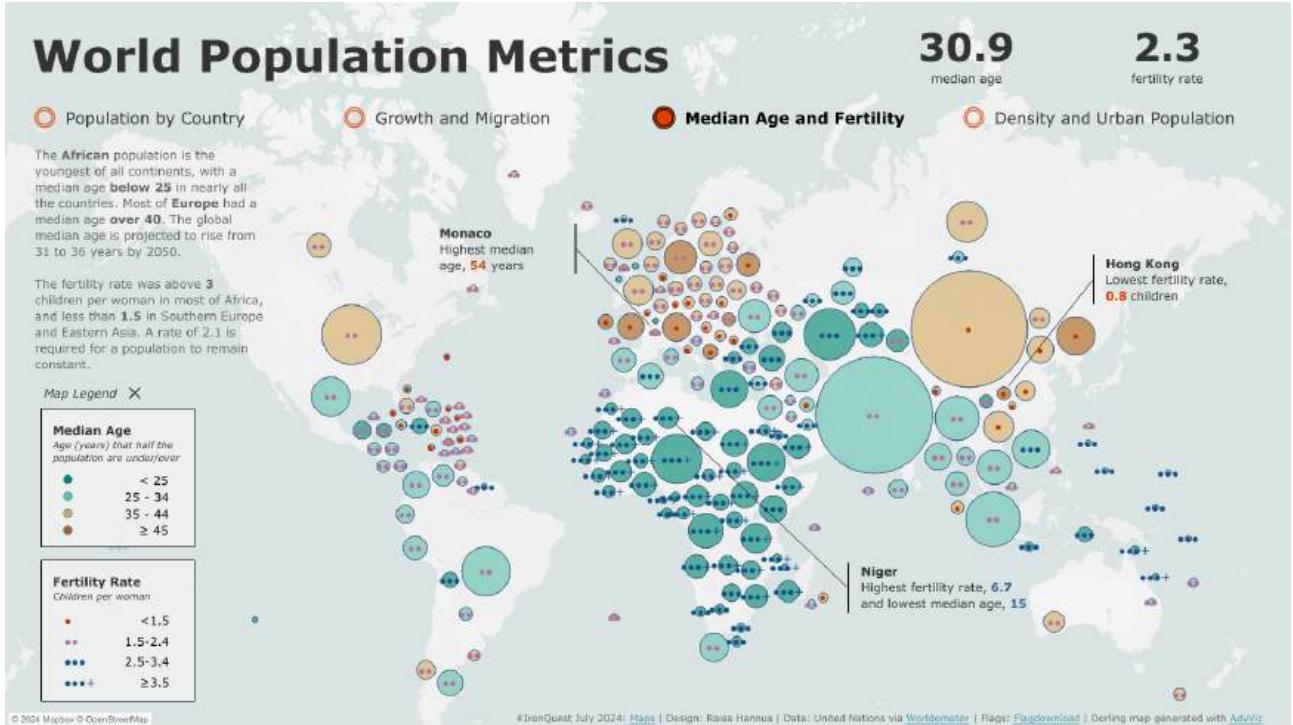
Colors: I chose slightly different but uniform colors for each metric, using color palettes I already had. Blue/green tones represent population growth and values that reflect growth, such as immigration and high fertility. On the flip side, red/orange/brown tones represent population decline, emigration, high median age, etc.

The colors showed clear patterns, the African population is growing while the European is slowly declining, etc. With so many countries, the differences in shade were hard to distinguish, so I changed all the color ranges to four steps.

Shapes: The small size of many countries steered me to simple Unicode shapes. An arrow shows the direction of net migration (IF >1 THEN 'Inbound', etc.) and is sized by the absolute number of migrants. To make an off-centered arrow, I created two transparent shapes in Figma with '←' & '→' closer to the circle's edge.



Again, I started seeing interesting patterns. Besides the common migration trend towards Europe and North America, the movement around Ukraine and Syria stood out and made me do more research. For the fertility rate, I simplified the wide range to 1, 2, 3, or 3+ children, as a rate of 2.1 equals zero population growth. The symbols are text marks with Unicode dots (IF ROUND(=1 THEN '●' etc.), again colored on a blue/red palette. Urban Population is shown as a percentage of the total population, leaving the remaining share as Rural (). I placed the two measures on a Pie Chart mark behind the Population Density layer, and made the Density mark smaller by replacing the Population size with another copy of the same.

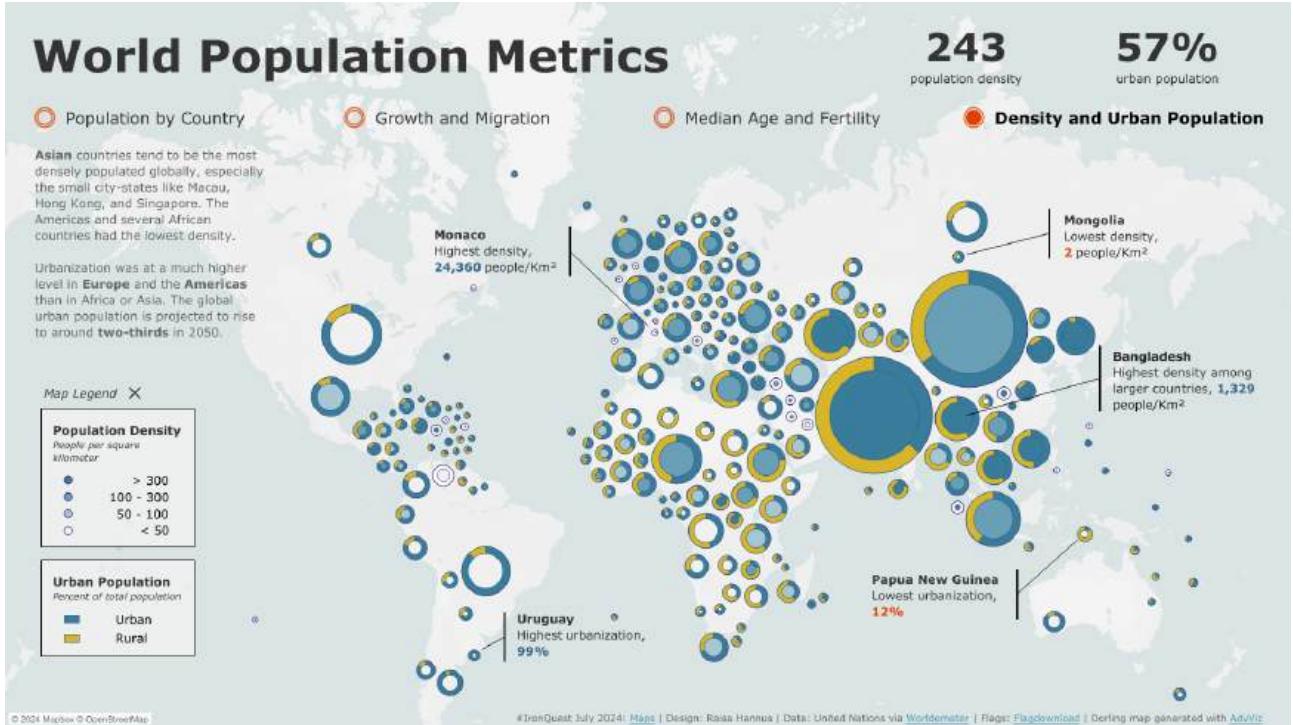


In addition to the seven parameter-controlled layers for the metrics, I created a background layer with a white circle in case of null values. Finally, I labeled the minimum and maximum values of each metric and formatted the tooltips to display both metrics of the selected pair.

My map started looking good, but many more details were needed for a publishing-worthy visualization, and most importantly to make myself happy with it. I wanted custom legends, Big Aggregate Numbers (BANs) for world totals, and text/conclusions for each pair of metrics. And it was going to take some time.

Phase 3 – Layout & Details

The dashboard has a relatively simple layout with the map filling most of it, while text & legends are in floating containers on the less busy left side. A horizontal menu at the top of the map lets users select the metrics via parameter action and BANs above are shown only for the selected metrics. The parameter also controls the text and legends via dynamic zone visibility. Finally, I added a hide/show button for the legend container since it covers a group of Pacific Islands



Menu: I imported a new data source with an Index (1-4) to create the menu. The headers are defined by a case statement using the same names as the parameter (CASE WHEN 1 THEN 'Population by Country' etc.). The selected set of metrics is highlighted with a bold font and a filled circle, using a true/false statement (=).

Legends: Even though the default legends with four-stepped colors and black symbols didn't look bad, I wanted them to match the visual appearance of the map. For each color-coded metric, I created a calculation to show the range of values as text labels (IF >0.02 THEN '> +2%' ELSEIF, etc.). Then I placed each metric on a new sheet with dual axes; MIN(0) for colored circles and MIN(6) for transparent shapes with the labels. This way I could align all the text to the right. Similarly, I created a calculation for the fertility rate labels, while migration and urban population just needed a colored shape.

BANs: Like the legends, I created a new sheet for the world total of each metric. The total population and net growth are the sum of each country, and percent growth equals net growth vs. total population in 2022. For all the other metrics, I used a weighted average (wavg).

$$\text{Growth \%}$$

$$\text{SUM() / (SUM() - SUM())}$$

Median Age wavg (Repeat for fertility, density, urban population)

$$\{ \text{FIXED : AVG()} \} * \{ \text{FIXED : SUM()} \} /$$

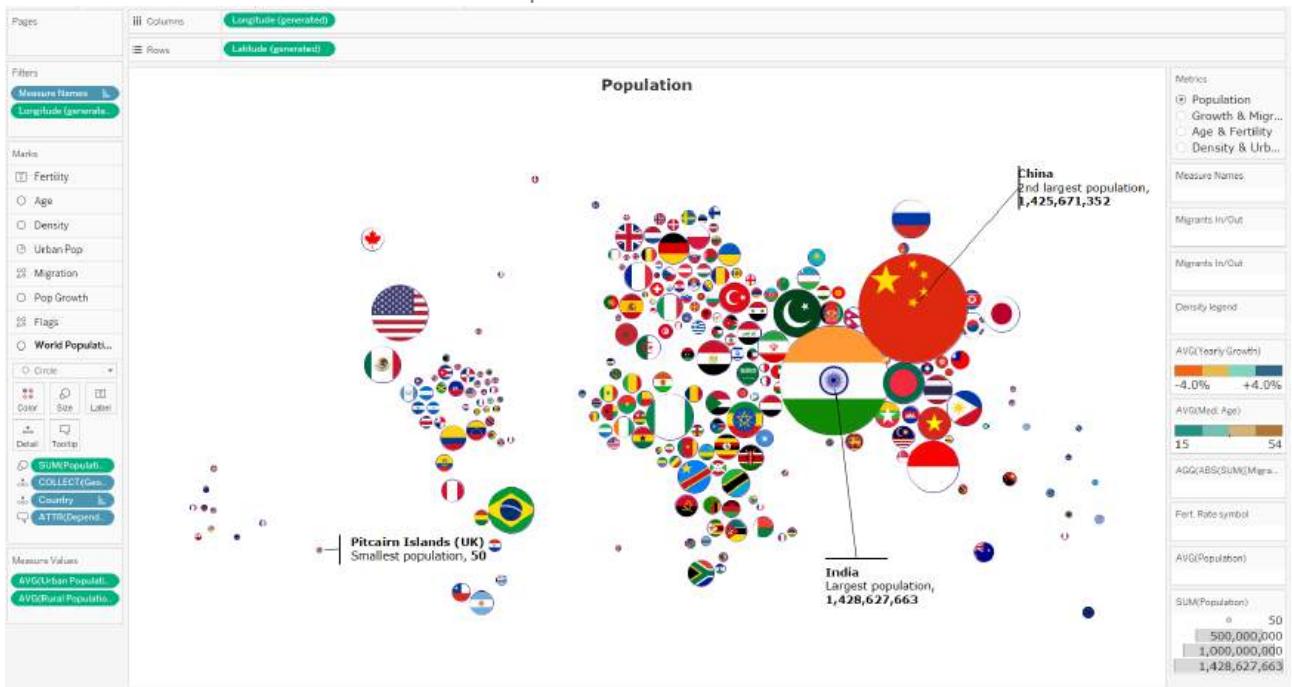
Labels: As the last step for the map, I switched from the min/max labels to more flexible and descriptive annotations. Annotations often disappear when any dimensions are changed or removed, so adding them later is safer.

Phase 4 – Story & Final Design

Story: After finalizing the details, it was time to zoom out and look at the big picture again. As the map pointed out the smallest and largest values for each metric and the BANs showed the world total, I focused on the regional trends in the storyline. I also wanted to highlight some key points that are not visible without historical data, such as India surpassing China as the most populated country, and the migration patterns around countries at war. Finally, I added some notes about future prospects that I found in the UN reports.

Design: With everything else lined up, the map was bothering me – the countries were floating in the air and it was visually tasking to place them. An actual map in the background would fix the issue, but

I didn't think it would work because the countries were shifted. Other options seemed time-consuming, like a freehand-drawn image or a cylindrical map projection, so I decided to test a regular Tableau map first to see how far off it was.



I added the on a new sheet, then removed it leaving just a background map. On the dashboard, I floated my main container, resized it to 1600×900, and added another floating container of the same size for the map. I changed the background map from 'Light' to 'Normal' to make the outline more visible, and its neutral blue/grey background fit well with all my colors.

After a few rounds of switching the containers' floating order and adjusting the map, surprisingly it was not too far off! I needed to shift the Americas closer to Africa and Europe, so I duplicated the map and added it to the same container, then aligned the Western and Eastern hemispheres separately with my Dorling map. Flipping the map's Base layer on/off while keeping the Land Cover layer in place helped fine-tune the exact position.

In hindsight, building the Dorling map on top of a background would have been smarter, but who knew I could do that too!

Phase 5 – Feedback & Final Touches

July's #IronQuest Maps challenge motivated me to finish the viz that had been sitting in my draft folder for too long. However, I wasn't sure if it was #IronQuest-worthy, as it was "just a dashboard" rather than a grand story. I also had the common feeling of missing something important completely while tweaking all the little details.

Signing up for #VizOfficeHours with Michelle Frayman and Nicole Klassen was a great way to rid those fears and doubts, and to get help with details. They spotted some inconsistencies in the annotations and pointed out marks that were hard to see. Talking through the viz also reminded me of the whole purpose of this viz, and after a few tweaks, I could call my map experiment successful!

LOGGING OFF,

CJ

WEB SCRAPING FBREF IN ALTERYX WITH TOBY HORNE

Hi all,

An exciting guest blog this week from Toby Horne, utilising Alteryx to retrieve FBREF data!



in

Toby Horne

Tableau & Alteryx Consultant at The Information Lab

Follow

Vizzes 12

Favorites 48

Following 31

Followers 68

The dashboard grid displays eight distinct Tableau visualizations:

- Player Scouting Dashboard** by Toby Horne: A complex dashboard with multiple cards showing player statistics and performance metrics.
- Stock Performance Dashboard** by Toby Horne: Shows stock market trends and performance across various sectors.
- Sales Pipeline Dashboard | #WOW2023wk5** by Toby Horne: A pipeline management dashboard with bar charts and progress indicators.
- England Women's Euro Passing Analysis Dashboard** by Toby Horne: Analyzes passing data for the England Women's national football team.
- FIFA World Cup Gender Pay Gap #MOM** by Toby Horne: Compares gender pay gaps in the FIFA World Cup context.
- Crime in Wellington** by Toby Horne: Visualizes crime data in Wellington, showing geographical distribution and trends.
- EU SUPERSTORE KPIs** by Toby Horne: Monitors key performance indicators for a European supermarket chain.
- Do You Consume Too Much Caffeine?** by Toby Horne: An interactive dashboard asking users if they consume too much caffeine, with a bar chart showing results.

Resources can be found at the top of the page through the github link.

CJ: Toby, great to have you join to share your expertise on soccer analytics. Let's start with a little about you – how did you get into data? Tell us more about your background.

Thanks for having me CJ! During my previous roles in Finance and Recruitment, I could see businesses leveraging data to automate processes and drive performance which is what inspired me to get into Data Analytics. Having no real previous experience in data, a friend of mine told me about the Information Lab – a data consultancy offering a unique training program known as the 'Data School'. At the DS, you undergo 4 months of full-time training followed by 4 6-month client engagements. During my training, I did some work on Tableau and Alteryx for a Premier League team. This was my first exposure to Data Analytics in sports, having never realised it to be a possible career path. After training, I was fortunate enough to have my first client engagement as Tableau consultant for the FA!

CJ: You clearly have a love for sports analytics – and I love your recent player scouting dashboard. Did your time at the FA help you think differently about what metrics are important in soccer?



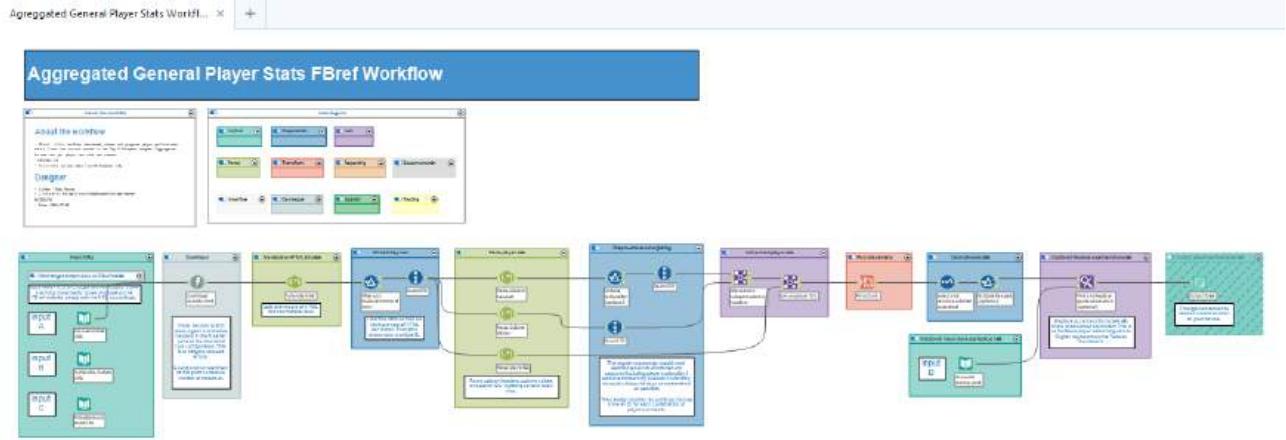
Definitely, metrics in football vary widely depending on who you ask. For example, different clubs and departments have unique questions that require answering.

One thing I would say is that I was amazed by the sheer amount of data the FA both collect and purchase from third parties. From the hundreds of metrics in provider match data, to gym pressure plates collecting 140 touchpoints per second. The biggest challenge is presenting the correct information to the right stakeholders in a meaningful way.

My time at the FA taught me the importance of tailoring data for different departments. For example, performance analysts (broadly speaking) tend to focus on event data like goals and assists, whilst physios and sports scientists would prioritise physical metrics like measuring physical fatigue in order to increase sharpness or reduce load to mitigate injury where required. One factor I had never previously considered was nation eligibility risk, which was critical to being translated to members of the recruitment department.



CJ: It looks like your Player scouting dashboard is built off FBREF data through an Alteryx workflow. Would this be something you could walk us through with details around some of the complexities? Sure! For my dashboard, I built two different Alteryx workflows, one that scrapes metric data at an aggregated level for the current season (one row per player), and another for match-level metric data (one row per player per match). In this blog, I'll walk through the simpler of the two, the aggregated flow, shown in the screenshot below.



This workflow captures the data shown in the table below, which can also be found on the Fbref website [here](#).

Rk	Player	Nation	Pos	Squad	Comp	Age	Born	NP Starts		Min	90+	Gls		Ast		G+A		G+PK		PKatt		CrdY		CrdR		xG		mpxG		xAG		npxG+nAG		PrG		PrGp		PrGr		G+Ast		G+A		G+PK		G+A+PK		Per
																																						Per										
1	Max Aarons	ENG	DF	Bournemouth	Premier League	23	2000	20	13	1,237	13.7	0	3	1	0	0	0	1	0	0.0	0.0	0.8	0.9	22	43	26	0.00	0.07	0.07	0.00	0.07																	
2	Kendren Aaronson	USA	MF	FPW	Union Berlin	Germany	22	2000	30	14	1,287	14.1	2	2	4	2	0	0	3	1	2.0	2.0	1.9	3.8	37	55	91	0.14	0.14	0.28	0.14	0.25																
3	Paxton Aaronson	USA	MF	Eintr Frankfurt	Bundesliga	19	2003	7	1	101	1.1	0	1	1	0	0	0	0	0	0.1	0.1	0.1	0.2	2	5	7	0.00	0.89	0.89	0.00	0.86																	
4	Keylano Abdallah	FRA	FW	Marseille	Ligue 1	17	2006	1	0	4	0.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0.00	0.00	0.00	0.00	0.00																	
5	Yusuf Ahluwalia	MAR	DF	Elche	Ligue 1	35	1987	31	31	3,781	30.8	4	0	4	3	1	1	5	0	3.4	2.6	6.3	3.9	36	137	9	0.13	0.00	0.13	0.10	0.11																	
6	Talal Ahluwalia	GHA	MF	Lens	Ligue 1	23	2000	27	17	1,539	16.9	0	0	0	0	0	0	2	0	0.8	0.8	0.5	1.3	9	79	20	0.00	0.00	0.00	0.00	0.01																	
7	Habib Ahmedi	FRA	DF	Getafe	La Liga	20	2002	2	2	189	2.0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0.00	0.00	0.00	0.00	0.01																	
8	Laurent Achong	FRA	MF	Lorient	Ligue 1	30	1993	33	32	3,850	31.0	2	3	3	2	0	0	4	0	1.1	1.1	2.2	3.3	38	194	51	0.00	0.03	0.09	0.05	0.05																	
9	Mathias Ahling	FRA	FW	Nantes	Ligue 1	20	2003	22	12	1,044	11.6	5	0	5	5	0	0	1	0	3.8	3.8	2.0	5.8	28	20	67	0.43	0.00	0.43	0.43	0.42																	
10	Ahmed	BRA	DF	Betis	La Liga	23	2000	23	15	1,400	15.6	0	1	1	0	0	0	5	0	0.1	0.1	1.0	1.1	14	35	58	0.00	0.05	0.06	0.00	0.00																	
11	Zakaria Ahouéchiali	MAR	MF	FW	Toulouse	France	23	2000	13	9	754	8.4	3	0	3	2	1	1	2	0	2.8	2.0	0.2	2.2	23	20	75	0.36	0.00	0.36	0.24	0.24																
12	Abdel Ahsan	MAR	DF	Aljazira	La Liga	24	1995	27	27	2,312	25.7	0	0	0	0	0	0	10	0	0.5	0.5	0.2	0.6	7	49	1	0.00	0.00	0.00	0.00	0.01																	
13	Tammy Abraham	ENG	FW	Roma	Serie A	25	1997	8	2	242	2.7	1	0	1	1	0	0	1	0	0.6	0.6	0.1	0.7	6	4	15	0.27	0.00	0.37	0.37	0.37																	
14	Francesco Acerbo	ITA	DF	Inter	Serie A	35	1988	29	26	2,388	26.5	3	1	4	3	0	0	1	0	1.6	1.6	1.8	3.4	21	87	24	0.11	0.04	0.15	0.11	0.15																	
15	Joshua Achempong	ENG	DF	Chelsea	Premier League	17	2006	1	0	6	0.1	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0.00	0.00	0.00	0.00	0.01																		
16	Marcos Acuña	ARG	DF	Sevilla	La Liga	31	1991	21	18	1,292	14.4	3	2	3	1	0	0	7	0	0.3	0.3	1.5	1.8	23	50	42	0.07	0.14	0.21	0.07	0.21																	
17	Kévin Adams Traoré	CIV	FW	FWN	Sheffield Utd	Premier League	20	2002	8	3	367	4.3	0	0	0	0	0	0	0	0.3	0.3	0.5	0.8	7	9	14	0.00	0.00	0.00	0.00	0.01																	
18	Kévin Adams Traoré	CIV	FW	FWN	Stade Rennais	Ligue 1	20	2002	14	4	504	5.6	0	1	1	0	0	0	1	1	0.7	0.7	0.9	1.6	23	15	45	0.00	0.18	0.18	0.00	0.11																
19	Ekor Adams	IND	FW	Montpellier	Ligue 1	23	2000	32	27	2,252	25.0	8	3	9	8	0	0	2	0	12.4	12.4	1.9	14.3	40	17	127	0.32	0.04	0.36	0.32	0.35																	
20	Tyler Adams	USA	MF	Bournemouth	Premier League	24	1999	3	1	121	1.3	0	0	0	0	0	0	0	0.0	0.0	0.1	0.1	4	5	1	0.00	0.00	0.00	0.00	0.00																		

To download the html, connect a download tool to a text input tool, containing the desired URL. In this example, I'm only looking at the standard stats URL, but if you want to scrape any other metrics provided on the Fbref website, simply change the URL in the text input tool accordingly.

The next step is optional but may be useful if you find that you are returning an error stating “*too many requests*”. In the ‘Header’ section of the download tool configuration, you’ll need to add a ‘User-Agent’ header and a ‘Cookie’ header. This is to help mimic a human user loading the website in a browser. Many websites implement security measures to detect and block bots. Using realistic User-Agent strings and valid cookies can help circumvent these measures, reducing the likelihood of getting blocked or banned.

To find your User-Agent and cookie details on a Google Chrome browser, right-click the FBref website that you are scraping and click *Inspect*. Here we can see the HTML that our download tool web scrapes. Then click the *Network* tab highlighted in the red box at the top of the screenshot below and refresh the web page.

Big 5 European Leagues History

More 2023-2024 Big 5 European Leagues Pages ▾

Competition Name	MP	Performance			Expected			Progression								
		Gls	Ast	PKatt	PKts	PW	CrdY	CrdR	G-PW	G+PK	nG	npxG	xG	xG+xDG	npxG+xDG	PrG
Bundesliga	1.57	1.37	0.17	9.17	0.13	2.07	0.08	1.44	2.01	1.51	1.19	1.35	2.60	2.44	17.0	36.9
La Liga	1.79	0.91	0.14	9.14	0.11	2.46	0.12	1.16	2.09	1.30	1.19	0.93	2.23	2.13	16.8	36.9
Ligue 1	1.30	0.87	0.17	9.17	0.14	1.95	0.12	1.17	2.05	1.37	1.23	0.96	2.32	2.19	17.7	42.3
Premier League	1.58	1.13	0.14	9.14	0.13	2.16	0.08	1.45	2.01	1.55	1.44	1.13	2.68	2.57	18.8	36.6
Serie A	1.27	0.90	0.17	9.17	0.13	2.15	0.09	1.14	2.04	1.24	1.11	0.97	2.11	1.98	17.0	35.9

Network

Fathy2018 Dev CSS JS Font Img Media Manifest WS Wasm Other Network Requests Headers Response Inspector Timeline Cookies

Request URL: https://fathy2018.com/mw5/api/big5/statistics/Big-5-European-Leagues-Stats

Required Method: GET

Status Code: 200 OK

Remote Address:

Referrer Policy: strict-origin-when-cross-origin

Response Headers (11)

Request Headers

Accept: */*

Accept-Encoding: gzip, deflate, br, zstd

Accept-Language: en-US,en;q=0.9

Cache-Control: max-age=0

Cookie: fathy2018=175267880041214291682; isAdm=true; related=true; refid=24000000000000000000000000000000; id=16-AU02-0100

Timeline

48 requests | 2.4 MB transferred | 17.3 M

Console: What's new ▾

Historical from the Chrome 125 update

Then click on the website request, highlighted in the second red box in the screenshot above. Once you've done that, you'll be able to find your User-Agent and Cookie information under the *Request Headers* section as highlighted below.

The screenshot shows the Network tab of a browser developer tools. A specific request to `https://fbref.com/en/comps/Big5/stats/players/Big-5-European-Leagues-Stats` is selected. The Headers section is expanded, showing the following request headers:

- :authority: fbref.com
- :method: GET
- :path: /en/comps/Big5/stats/players/Big-5-European-Leagues-Stats
- :scheme: https
- Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.7
- Accept-Encoding: gzip, deflate, br, zstd
- Accept-Language: en-US,en;q=0.9
- Cache-Control: max-age=0
- Cookie: hubspotutk=2aac133a8a4fd784e7e902083103ec1b; _fbp=fb.1.1702474800491.2142958982; fs.admiral.whitelisted=true; cr_id=a260068-33f610a29d428f7ea00a4e717; aut.1d=AU1D-0100-

Then we want to add this information to the *Headers* section of the download tool as shown below (note: for the User-Agent, we only want to input the information before the parentheses). **If you are downloading the provided workflow, be sure to change these values to your own.**

Name	Value
Cookie	hubspotutk=2aac133a8a4fd784e7e902083103ec1b
User-Agent	Mozilla/5.0

If you chose to skip the previous step, welcome back. Now, it is important to *Cache and Run* the workflow so that you don't keep sending requests to the FBref website each time you press *Run* in Alteyryx Designer.

Then we want to tokenise our Download Data so that we start to formulate our desired data structure. We can do this by using a Regex Tool, configured as shown in the screenshot below.

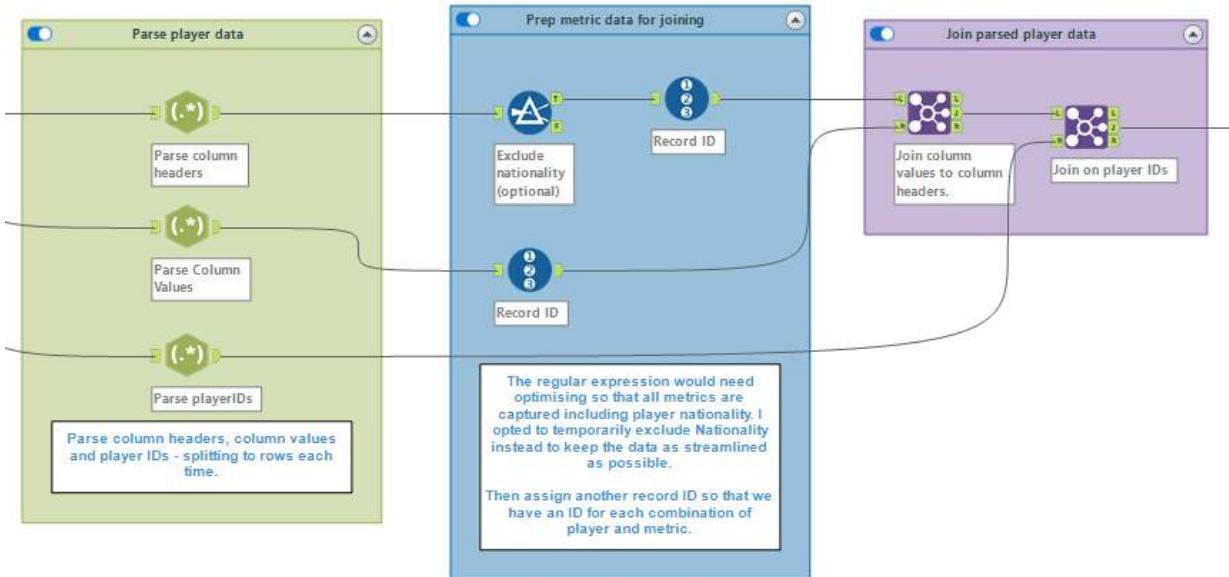
Column to Parse
DownloadData

Format to Convert
Regular Expression [?](#)
<tr .?</tr>
 Case Insensitive

Output
Output Method
Tokenize
 Split to Columns
 Split to Rows

The next step is to filter our results so that we only return rows of data that contain `<tr><th scope="row" class="right">`. By doing so, we will only keep the rows of HTML that contain player metrics. At this point, connect a Record ID tool as we have one row of HTML for each player (we'll use this when it comes to joining later).

Now we have our desired structure, we just need to parse out the relevant information for each row of HTML. We achieve this using a series of Regex tools with tokenised outputs (shown below).



Notice in the HTML how each metric header name comes directly after the text, “`data-stat=`”. Knowing this, we can use the regular expression `data-stat="(.*)"` to capture all text that appears after `data-stat=`. By choosing to tokenise our output whilst splitting to rows, we don't need to specify how many times to expect this pattern in the HTML. Instead, a new row will simply be added to our results after each occurrence of the pattern. See the screenshot below for the configuration of the regex tool.

Build an expression to parse, match, or replace data.

Column to Parse
DownloadData

Format to Convert

Regular Expression [?](#)
data-stat="(.*?)" [+](#)
 Case Insensitive

Output

Output Method
 Tokenize
 Split to Columns
 Split to Rows

The configuration for the regex tool parsing the metric values is `(?:>(*)</|(?::td|a)>)`. Without going outside the scope of this blog, all it's doing is capturing our metric values, which occur between `>` and `</td>` or `` tags.

The final regex tool (configuration shown below) simply captures the Player ID, which occurs every time `data-append-csv=` appears.

Now we have these three outputs with all our relevant information, all we need to do now is join on record ID.

After the join, your output will look something like this

Record	Columns	Values	PlayerID
1	ranker	Max Aarons	774cf58b
2	player		774cf58b
3	position	DF	774cf58b
4	team	Bournemouth	774cf58b
5	comp_level	Premier League	774cf58b
6	age	23	774cf58b
7	birth_year	2000	774cf58b
8	games	20	774cf58b
9	games_starts	13	774cf58b
10	minutes	1,237	774cf58b
11	minutes_90s	13.7	774cf58b
12	goals	0	774cf58b
13	assists	1	774cf58b
14	goals_assists	1	774cf58b

The final step is to pivot the data and rename our `ranker` column to `player`.

I also added an optional step which is to replace any accented characters found in a player's name with its unaccented equivalent – for example, 'Martin Ødegaard' becomes 'Martin Odegaard'. This facilitates searching for players on my Tableau Dashboard using an English keyboard. I achieved this

by connecting a find and replace tool to our output along with a character lookup table (courtesy of Chat GPT) containing the target accented characters, and their unaccented replacements.

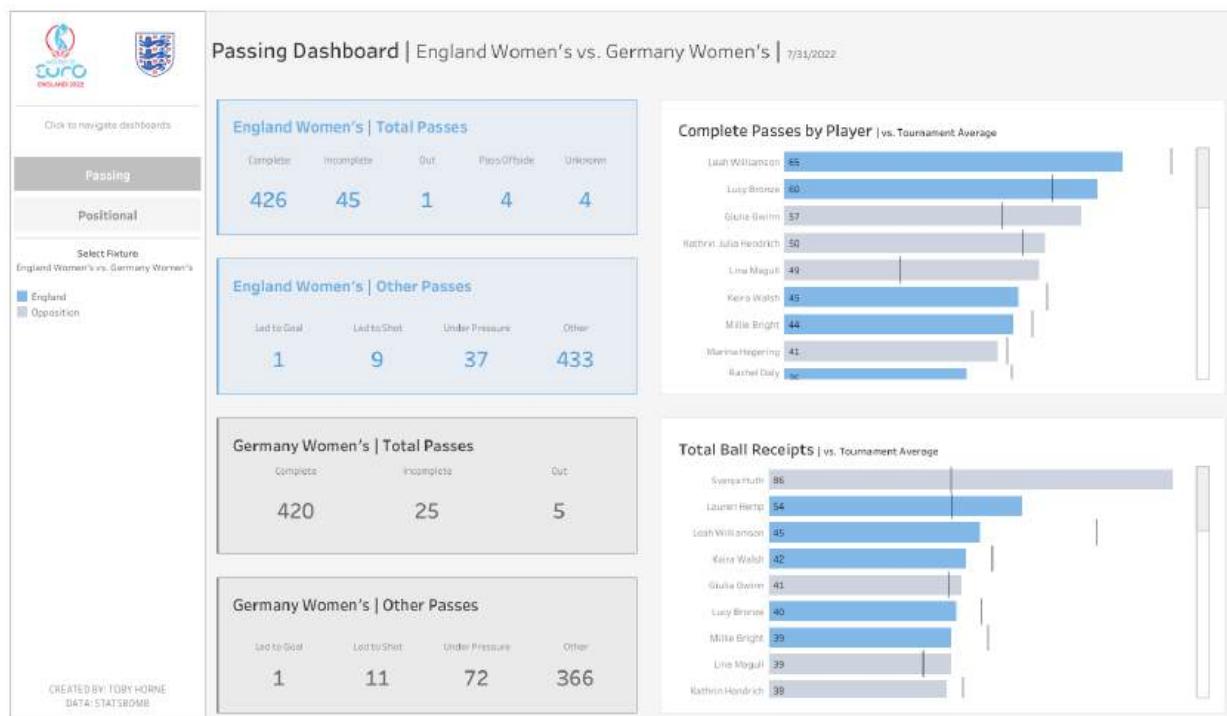
And that's it! You now have clean, aggregated performance metric data for every player that has made an appearance across the 'Big 5' European leagues this season!

Record	PlayerID	player	age	team	comp_level	minutes	minutes_90s	assists	assists_per90	birth_year	cards_red	cards_yellow	games	games_
1	00242715	Moussa Niakhaté	27	Nottingham Forest	Premier League	1	16.2	0	0.00	1996	1	5	21	15
2	00242675	Luandro Barreiro Martínez	23	Mainz 05	Bundesliga	2	29.0	1	0.03	2000	0	8	31	29
3	00344141	Jayden Darnis	17	Liverpool	Premier League	9	0.1	0	0.00	2006	0	1	2	0
4	00459419	Marco Arnautović	34	Inter	Serie A	801	8.9	3	0.34	1989	0	0	27	5
5	00756717	Nemanja Gudelj	31	Sevilla	La Liga	1	19.4	1	0.05	1991	0	7	22	20
6	00953490	Mikkey Moore	15	Tottenham	Premier League	4	0.0	0	0.00	2007	0	0	2	0
7	00953611	Konstantinos Mavropanos	25	West Ham	Premier League	1	16.7	0	0.00	1997	0	2	19	16
8	00954460	Julian Ryerson	25	Dortmund	Bundesliga	1	19.6	1	0.05	1997	0	3	21	20
9	00a062210	Przemysław Frankowski	28	Lens	Ligue 1	2	24.4	2	0.08	1995	0	4	30	24
10	00b28772	Nenadža Maksimović	28	Celta	La Liga	3	36.3	0	0.00	1995	0	4	37	37
11	00b774cd	Nathan Fraser	18	Wolves	Premier League	186	2.1	0	0.00	2005	0	1	7	1
12	00552483	Hugo Larsson	19	Eintracht Frankfurt	Bundesliga	1	21.0	1	0.05	2004	0	1	29	21
13	00e36323	Florian Pusk	27	Heidenheim	Bundesliga	327	3.6	1	0.28	1995	0	2	22	1
14	00e5a5c9	Selim Amallah	26	Valencia	La Liga	604	6.7	1	0.15	1996	1	5	20	7
15	00f518d8	Franck Alloua	22	Köln	Bundesliga	1	14.2	0	0.00	2001	0	4	26	14
16	0113c54c	Anthony Caci	26	Mainz 05	Bundesliga	2	27.6	3	0.11	1997	0	0	31	29
17	01162a32	Trevor Dugard	18	Lille	Ligue 1	23	0.2	1	3.91	2005	0	1	4	0
18	0127419a	Ivanis Hagi	24	Alavés	La Liga	804	8.9	2	0.22	1998	0	3	22	8
19	01947f50	Valentino Lazaro	27	Torino	Serie A	1	22.2	4	0.18	1996	0	3	35	21
20	017018d1	Fabian Holland	33	Darmstadt 98	Bundesliga	1	20.4	2	0.10	1990	1	4	23	23
21	0174ba21	Robin Hack	24	Gladbach	Bundesliga	1	15.0	2	0.13	1998	0	2	29	14

A copy of the results as well as the workflow can be found in the git repo at the top of the page!

CJ: Really appreciate the careful effort and consideration put into the workflow. Can I just say how well formatted is that workflow??? I just love that with Toby's efforts it really is a case of amending a few file paths and clicking run!

CJ: To close Toby, Is there anyone in the wider sports data community you think particularly creates compelling visuals and stories?



My first exposure to data visualisation in sport would have been a combination of Tableau public authors and Statsbomb's social media pages. I still consider Statsbomb's visualisations to be the gold standard of sports data communication and I often read their blog's for inspiration on design, metric analysis and to improve my analytical understanding of the game.

As for Tableau Public authors, I've enjoyed following Nick Van Lieshout's work ever since I came across his **Jong PSV Performance dashboard**. I would say that his clear focus on stakeholder needs throughout his dashboards and their clean design has inspired both my personal work and some of projects at the FA. Your blog starring Nick is great read as well!

Rob Carroll would be another inspiration of mine. He's got some fantastic tutorial-based content [Tableau Public](#) and [Youtube channel](#) which I couldn't recommend enough to anyone who is looking to get started in sports analytics.

LOGGING OFF,

CJ

F1 CIRCUIT MAPS WITH DAN WADE

Hi all,

You may or may not know that we run a weekly round up blog on Sports called #SportsVizSunday each week. Through that I've had the delight of following **Dan's growing portfolio** and passion for F1. I am so pleased that he has offered to share some of that passion with us today.

Dan Wade
The Information Lab / The Data School | United Kingdom

Tableau Certified Data Analyst...
[Read more](#)

[Follow](#)

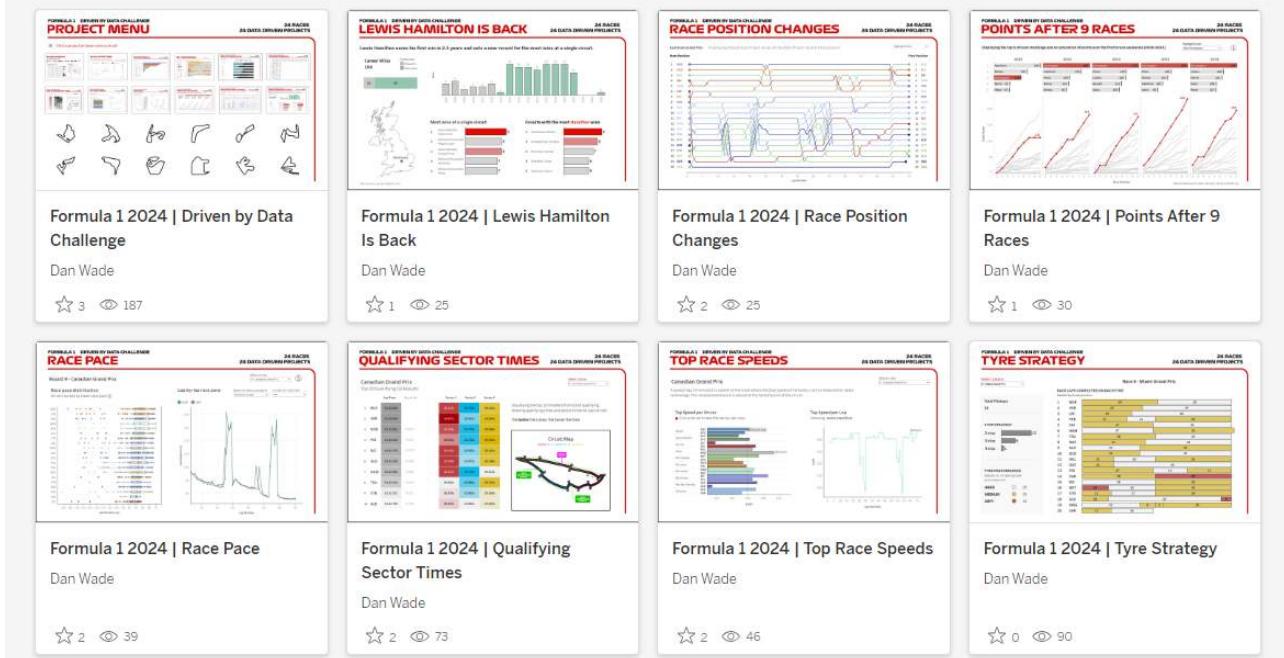
Vizzes 8 Favorites 64 Following 11 Followers 51

Latest 8 Vizzes

- Novak Djokovic: Putting the 'Ten' in Tennis**
Novak Djokovic | Grand Slam Finals #SportsvizSunday
Dan Wade
0 17
- UEFA Euro 2024 Tournament Bracket**
UEFA Euro 2024 Tournament Bracket
Dan Wade
5 81
- Formula 1 2024 | Driven by Data Challenge**
Formula 1 2024 | Driven by Data Challenge
Dan Wade
3 188
- Formula 1 2024 | Lewis Hamilton Is Back**
Formula 1 2024 | Lewis Hamilton Is Back
Dan Wade
1 26

CJ: Dan, thanks for joining. I am so pleased to have a #SportsVizSunday regular, join the blog today. Can you talk a little about your journey to date in data?

Formula 1 14 Vizzes



D: Thanks CJ. Your website has been a lot of help on my data journey so it's great to be featured on your blog.

I graduated university with a degree in sports science in 2020 and a masters degree in exercise physiology in 2021. The plan was to go into sport but covid put a stop to all opportunities so I had to pivot. I had no idea what I wanted to do but found a graduate scheme at a genomics research institute where I rotated around different teams. It was related to health so it felt like a good option after my degree.

In my first month there, I was supporting a team that was consolidating data from hundreds of spreadsheets and visualising that data in a tool called Tableau. I was really impressed and my manager booked me onto an internal Tableau training course, which happened to be led by a **Matt Francis**, a respective member of the Tableau community and Tableau Zen Master (That was the previous title before Tableau Visionary).

After that session I was hooked and found myself spending the next 18 months of my graduate scheme trying to find ways to introduce Tableau to other teams and practice my data skills.

During that time, I developed dashboards for several teams and my passion for data analysis began. I came to the conclusion that I wanted to get into data full-time which was when I came across The **Information Lab** and their Data School Program. After an intense 4-month training programme learning Tableau, Alteryx, SQL and Power BI, I now find myself 6-months into working with my first client as a Data Analytics consultant.

CJ: What sparked your initial interest in visualising sports data?

D: Having studied sport science for many years and not entering the industry after university, I think I had a natural pull to working with sport data. When starting out with Tableau I came across #sportsvizsunday which has and still is a great source of inspiration. Plus, now I'm working with business data 9-5, it's refreshing to switch to sport data analysis in the evening and weekends to push my skill development in a fun way.

CJ: You visualise predominantly F1 data. I love that you are aiming to do 24 rounds of visuals in alignment to the races. Is this also your favourite sport to follow in terms of watching? How has analytics shaped your view when watching F1?

D: Yes, I would say Formula 1 is my favorite sport. As a Spurs fan I needed to attach my emotions to another sport to keep me sane! Formula 1 is heavily influenced by data. Entire data teams are behind

the decision of each upgrade, pitstop and tyre choice. As a data guy this leaves me with a million questions when watching live.

You often see data visualisations on screen which adds to the viewing experience. However, beyond the TV broadcast there's a whole community of fans who build and share visualisations and stats. This has not only helped to answer some of my questions but has also taught me so much more about the sport which has improved my viewing experience.

So, it felt right to start a project where I go off to get the data and answer some of these questions myself. I began the challenge because I wanted a long-term project. I wanted a bigger goal to work towards that would keep me learning. I can now spout useless F1 information to friends and family at the dinner table.

CJ: Where do you recommend going for F1 datasets? I see a variety of different references in your work!

D: So I would start anyone off using this **Kaggle dataset** at Kaggle. Here the data is clean and structured in tables that are ready to use. This contains all the historical information up to the previous season.

Next you can webscrape the data from websites.

- <https://www.formula1.com/en/results.html/2024/races.html>
- <https://pitwall.app/>
- <https://www.statsf1.com/en/statistiques/pilote.aspx>

It's easy to web scrape in Excel as a beginner. **Check out this tutorial**

However to get data for the current season or even near-live data you'll want to head to <https://ergast.com/mrd/> or <https://openf1.org/> which are free and open-source API's. These are the sources I've been using for my projects. I've been using Alteryx to do this because I have a license, but you can do so for free in Python. I plan to write a blog on how to get hold of this data during the F1 summer break, so stay tuned for that!

CJ: I particularly love your race schedule visual looking at the season route. Interestingly enough starting in Middle East / Asia, moving towards Europe and then out towards North and South America.

Although there are a few outliers. What did you want to highlight most with this visual?

F1 2024 SCHEDULE

RACE SCHEDULE ACROSS THE SEASON

Click a mark to highlight the table and map



DISTANCE BETWEEN RACES

Click a race to highlight the map

Round	Start Date	Race	Next Race	Distance (km)
1	24 Feb	Bahrain Grand Prix	Saudi Arabian Grand Prix	2,519
2	07 Mar	Saudi Arabian Grand Prix	Australian Grand Prix	12,815
3	22 Mar	Australian Grand Prix	Japanese Grand Prix	8,095
4	05 Apr	Japanese Grand Prix	Chinese Grand Prix	1,480
5	19 Apr	Chinese Grand Prix	Miami Grand Prix	13,268
6	03 May	Miami Grand Prix	Emilia Romagna Grand Prix	8,189
7	17 May	Emilia Romagna Grand Prix	Monaco Grand Prix	351
8	24 May	Monaco Grand Prix	Canadian Grand Prix	6,139
9	07 Jun	Canadian Grand Prix	Spanish Grand Prix	5,907
10	21 Jun	Spanish Grand Prix	Austrian Grand Prix	1,175
11	28 Jun	Austrian Grand Prix	British Grand Prix	1,258
12	05 Jul	British Grand Prix	Hungarian Grand Prix	1,536
13	19 Jul	Hungarian Grand Prix	Belgian Grand Prix	1,020
14	26 Jul	Belgian Grand Prix	Dutch Grand Prix	239
15	23 Aug	Dutch Grand Prix	Italian Grand Prix	820
16	30 Aug	Italian Grand Prix	Azerbaijan Grand Prix	3,323
17	13 Sep	Azerbaijan Grand Prix	Singapore Grand Prix	6,942
18	20 Sep	Singapore Grand Prix	United States Grand Prix	15,853
19	18 Oct	United States Grand Prix	Mexican Grand Prix	1,197
20	25 Oct	Mexican Grand Prix	Brazilian Grand Prix	7,421
21	01 Nov	Brazilian Grand Prix	Las Vegas Grand Prix	9,775
22	21 Nov	Las Vegas Grand Prix	Qatar Grand Prix	13,043
23	29 Nov	Qatar Grand Prix	Abu Dhabi Grand Prix	339
24	06 Dec	Abu Dhabi Grand Prix	"End of Season"	0

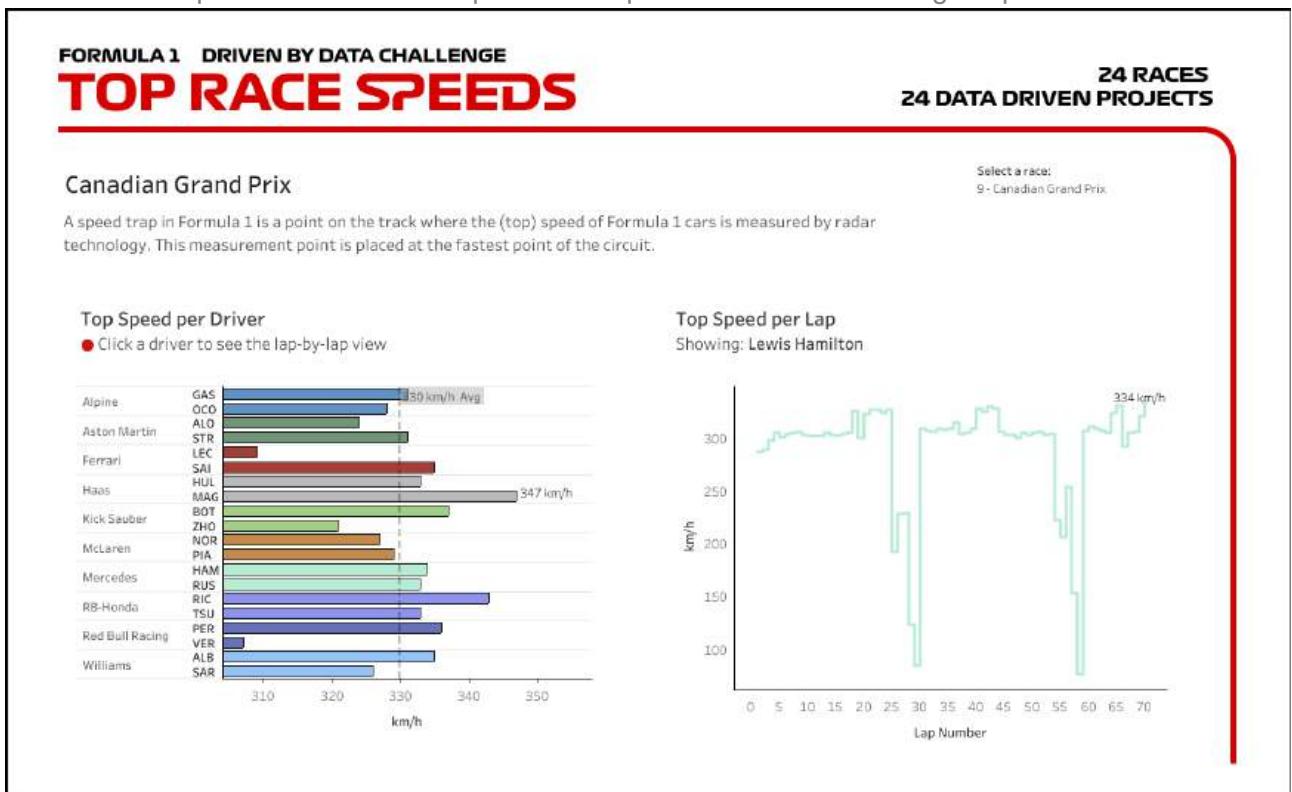
SEASON ROUTE

Use cursor to pan and zoom or hover for more info



D: Thanks. It's definitely one of my favourites. I find myself coming back to this one regularly to plan when I need to keep a Sunday free to watch a race. I wanted to highlight how intense the season is. We're up to 24 races a season now which is demanding on teams and drivers, particularly the triple header weekends where teams travel from one circuit to the next on back-to-back weekends. It's a logistical challenge to move people and equipment around the world crossing oceans, continents and time zones. It's difficult to put it into perspective when looking at a calendar so I thought this would make good use of Tableau's spatial features to plot the locations and routes.

CJ: What impact does tyre type tend to have on performance? Do you consider it the one of the most important factors that impacts race speed – what else is of high importance?



D: Tyre type and tyre strategy is a huge factor. There are five tyres compounds available for each race, two are used in wet conditions but the three compounds used the most are the soft, medium and hard. Every team is given the same number of tyres each race. It's almost the only thing that's standardised between each car. The hard tyre is harder-wearing so it can be run for longer but is generally slower, whereas the soft tyre is the faster tyre but degrades more easily and the medium sits between the two. In a race setting, each team has to change their tyre compound at least once so teams have to use all kinds of data to run prediction models to figure out the best tyre combination to complete the race.

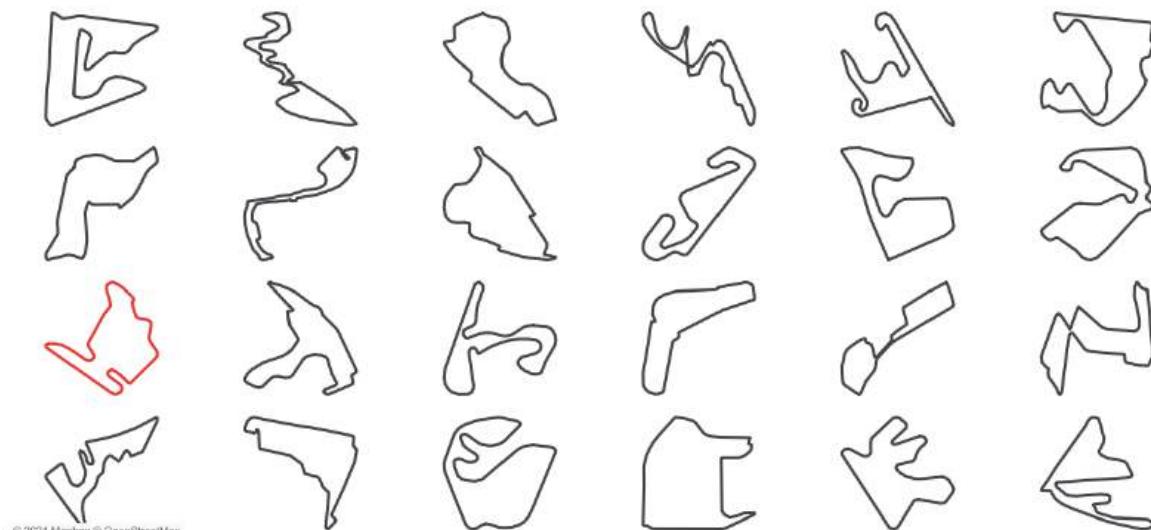
It's certainly one of the most important factors that impacts the race because a lot of the strategy sits on the performance of each tyre like when to pit or the grip of the tyre when overtaking into a corner.

When it comes to race speed however, the aerodynamics of each car is the most important factor. How the flow of air moves up, down and around the car will determine how much downforce or drag is created for example. If you have a close look at a formula 1 car you'll see all sorts of tiny pieces of carbon fibre, which all play a role in improving the aerodynamics. The aerodynamic parts are changed from circuit to circuit and get upgraded throughout the season as the teams collect more data. In 2021, a lot of teams had an aerodynamic issue where there car would bounce up and down at high speed. [See here.](#)

CJ: Could you give us a walkthrough of how to visualise a circuit map?

F1 CIRCUITS 2024

Next race: Jul 19-21
Hungarian Grand Prix



© 2024 Mapbox © OpenStreetMap

By Dan Wade
Source: github.com/baderger/F1-circuits/ & [F1.com](https://f1.com)

D: Of course. I did my prep in Alteryx and I'll talk you through the challenges I came up against. To begin with, you need the circuit data. You can find most of all circuits from [here](#).

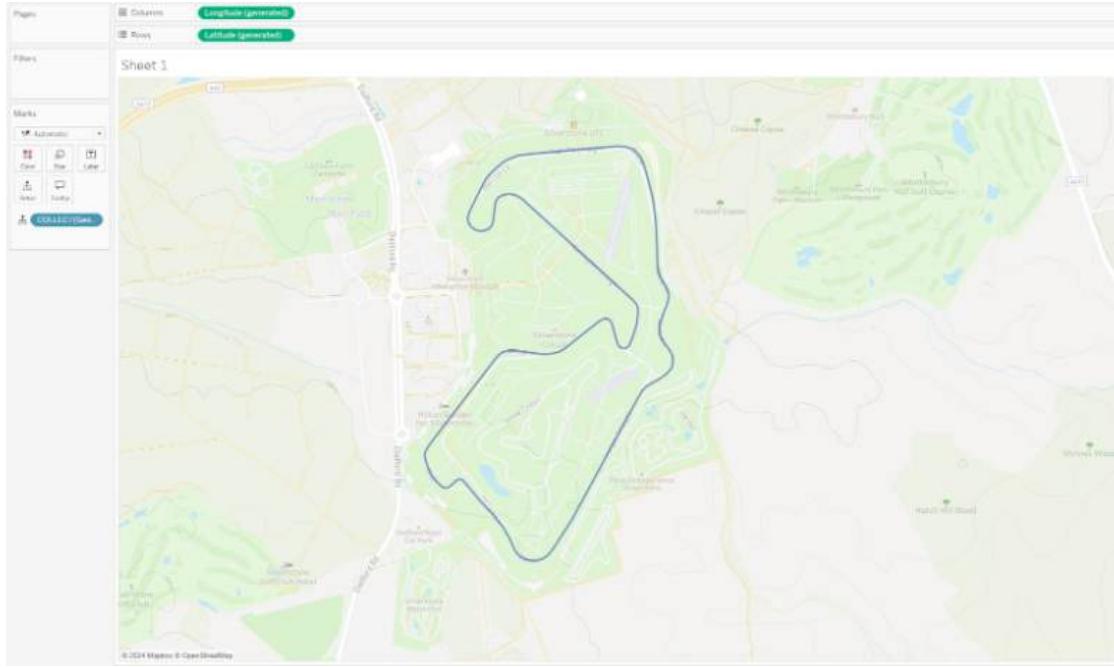
All my resources can be found in the [Git repository](#).

The screenshot shows a GitHub repository named "f1-circuits". The repository page lists several files:

- F1 Circuits.tmb
- Normalised Circuits.json
- Readme.md
- f1-circuits.geojson
- normalising circuits.yaml

The "f1-circuits.geojson" file is selected, and its preview is displayed below. The preview shows a world map with green lines representing the locations of Formula 1 circuits.

Download a copy of that and you'll have a geo.JSON file which contains all the coordinates which are stored within a single string. Tableau can handle this format natively by selecting “More” from the connection window. From here you can load up your circuits as normal by adding longitude ion columns and latitude on rows and from here you can see that Tableau has connected all these points to form a line of the circuit.



This is great if you're only interesting in one circuit, but not useful if you want to visualise all circuits side by side. One option is to filter to one circuit per sheet and add all these sheets to a dashboard but that's not ideal either and can run into performance issues.

Instead we can create a trellis chart, which plots all the circuits maps on columns and rows, essentially creating a grid of maps. We can create a trellis chart using the following calculated fields.

Trellis X:

`(index()-1)%(INT(SQRT(size()))))`

Trellis Y:

`int((Index()-1)/ (int(sqrt(size())))))`

Next, drag ID onto detail, Trellis X to columns and Trellis Y to rows, make both of these discrete by right clicking on the pill and under edit table calculation, you need to select Specific Dimensions and click ID (do this for both the Trellis X and Trellis pill).

Table Calculation

Trellis X

Compute Using

- Table (across)
- Table (down)
- Table (across then down)
- Table (down then across)
- Pane (across)
- Pane (down)
- Pane (across then down)
- Pane (down then across)
- Cell

Specific Dimensions

Id

At the level

Restarting every

Sort order Specific Dimensions

Show calculation assistance

You should then have something like this. Where we have a grid of individual maps. The issue you might spot is that whilst we can see Silverstone, we cannot see the circuits across the rest of the trellis. This is because we are working with one map layer across the same x and y coordinates. Meaning, we can only pan to one circuit at time. If I zoom out you can see each grid still contains the circuit map, but it's not possible to show each one individually in this way.



To rectify this, we need each circuits coordinates to be normalised which basically means to have the coordinates within the same range, or in this instance, fit within the same grid.

To do this I jumped across into Alteryx to transform the coordinates. You can get a copy of the workflow from the [github repo](#).

A geojson file is not a natively supported file type in Alteryx, so we have to connect via the file path. Alteryx will give an option to resolve the file type and we can read it in as a JSON. We then split the string into 6 columns.

The screenshot shows the Alteryx interface with the following components:

- Input Data (1) - Configuration:** A dialog box where "Connect a File or Database" is selected, pointing to "C:\Users\DanWade\Downloads\f1-circuits (3).geojson". Other options like "Set Up a Connection" and "Use Data Connection Manager (DCM)" are shown.
- Resolve File Type dialog:** A modal window titled "Resolve File Type" stating "The selected file is not a recognized type C:\Users\DanWade\Downloads\f1-circuits (3).geojson". It has a radio button selected for "Read it as a built in type" with "JSON (*.json)" chosen. Other options include "Read it as a fixed width text file", "Read it as a delimited text file" (with Delimiter sub-options like Comma, Tab, Space, None, or Other), and a checkbox for "First Row Contains Field Names". Buttons at the bottom are OK, Cancel, and Help.
- Workflow Canvas:** Shows a green "Input json" tool connected to a yellow "Split Text" tool. The input file is labeled "f1-circuits.geojson". The output of the Split Text tool feeds into a "Results - Input Data (1) - Output" viewer.
- Results - Input Data (1) - Output:** A viewer showing a preview of the data with 0 fields.

In the fourth column that's generated, we have our field headers, we can filter this column to only show the coordinate fields so we can then work on our normalisation.

Record	JSON_Name	JSON_ValueString	1	2	3	4	5	6
1	features.0.geometry.coordinates.0.0	144.968644	features	0	geometry	coordinates	0	0
2	features.0.geometry.coordinates.0.1	-37.849757	features	0	geometry	coordinates	0	1
3	features.0.geometry.coordinates.1.0	144.967627	features	0	geometry	coordinates	1	0
4	features.0.geometry.coordinates.1.1	-37.848964	features	0	geometry	coordinates	1	1
5	features.0.geometry.coordinates.2.0	144.966121	features	0	geometry	coordinates	2	0
6	features.0.geometry.coordinates.2.1	-37.847751	features	0	geometry	coordinates	2	1

Next we use a formula tool to assign our lat and long fields. You can use the following formula or an if statement on the 0 and 1 values:

Formula (8) - Configuration		⋮	⋮
	Output Column	Data Preview	
	header	lon	
	SWITCH([6], 'lat', '0', 'lon', '1', 'lat') //create latitude and longitude headers		
	Data type:	V_WString	Size: 1073741823

We then use the Select tool to change our data types and rename our fields, before pivoting our data using the Crosstab Tool.

Cross Tab (10) - Configuration

Select data to transform.

Group data by these values:

- JSON_ValueString
- row
- seq
- header

Select All

Change Column Headers

header

Values for New Columns

JSON_ValueString

Method for Aggregating Values

- Sum
- Average
- Count (Without Nulls)
- Count (With Nulls)
- Percent Row
- Percent Column

Select All

Now onto normalisation:

First, for each circuit we need to return the minimum and maximum values for both latitude and longitude.

Summarize (12) - Configuration

Field	Type
row	Int64
seq	Int64
lat	Double
lon	Double

Actions:

Field	Action	Output Field Name
row	Group By	row
lat	Min	Min_lat
lon	Min	Min_lon
lat	Max	Max_lat
lon	Max	Max_lon

We join this back to the original coordinates so we can complete the following calculations using the Formula Tool.

Formula (14) - Configuration

Output Column	Data Preview
Normalised Lat	0.646351931330535
$([lat]-[Min_lat])/([Max_lat]-[Min_lat])$	
Normalised Lon	2.72698859196619e-02
$([lon]-[Min_lon])/([Max_lon]-[Min_lon])$	

Here are all the parts of the calculation with a definition:

: This represents a specific latitude value.

: This represents the minimum latitude value in a given range or dataset.

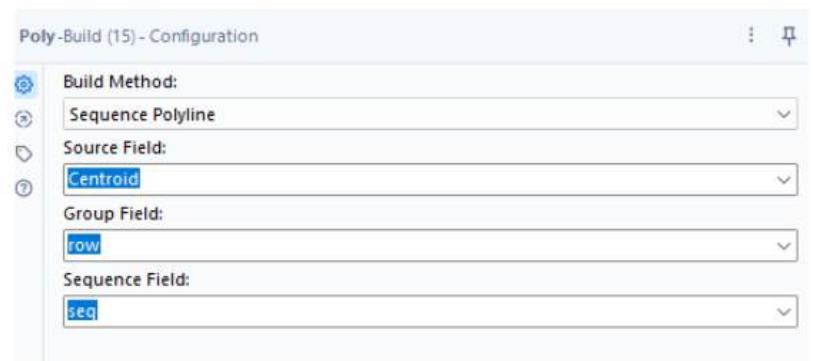
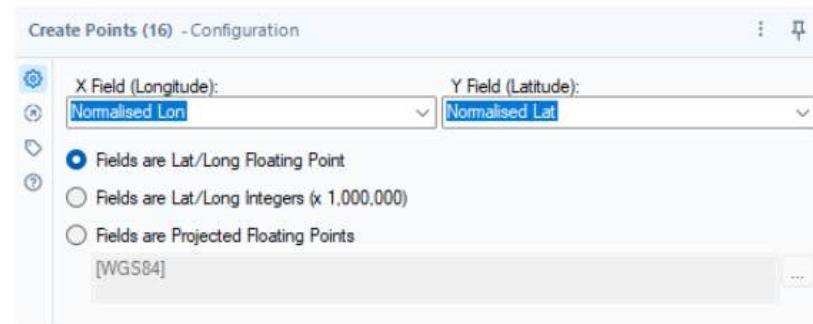
: This represents the maximum latitude value in the same range or dataset.

The calculation - finds the difference between the specific latitude value and the minimum latitude value. This difference represents how far the specific latitude value is from the minimum latitude value.

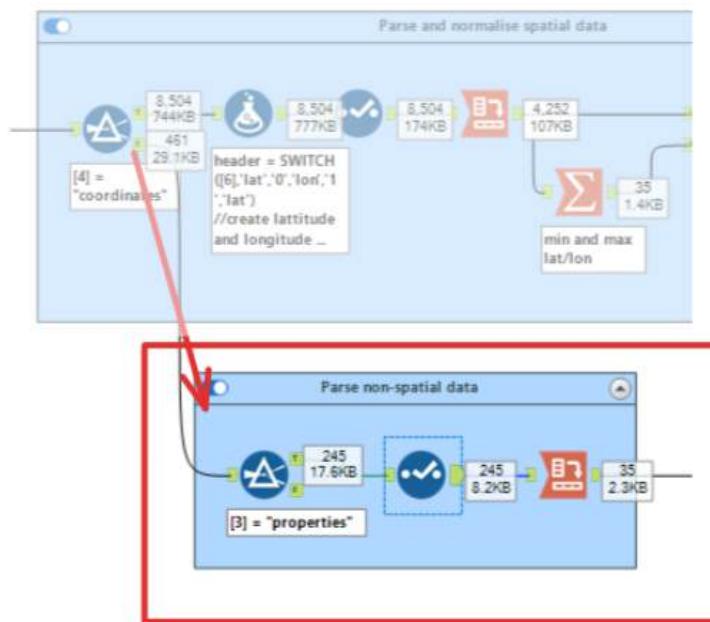
Similarly, - finds the range of latitude values in the dataset.

Dividing the difference between the specific latitude value and the minimum latitude value by the range of latitude values normalizes the latitude value to a range between 0 and 1.

We then rebuild the circuit route between the coordinates using Alteryx's spatial tools; Create Point and Poly-build to create our connected circuit lines.



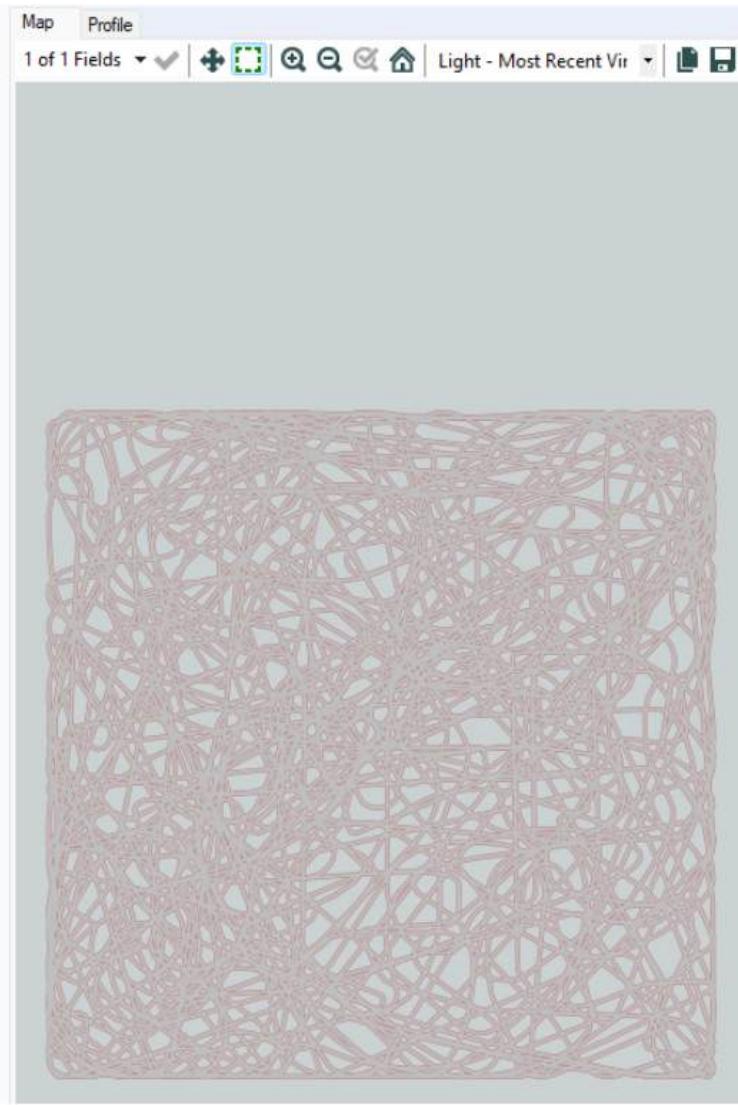
Now, going back a few steps, we need to parse our non-spatial data like circuit name etc.



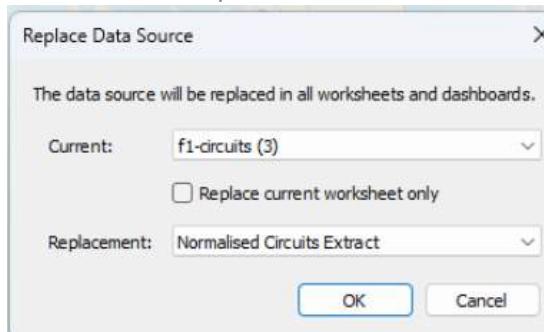
To do this, from the false anchor, we filter the “properties” string from column 3. Followed by updating our datatypes and header names ready to pivot again.

That's our transformation done and we just need to join these two datasets together on our row field before outputting our data as a Tableau Hyper file.

A quick look at the browse tool shows we've successfully normalised our coordinates and all circuits are now sitting on top of each other rather than in their individual countries.



Once you have output your file, we can easily replace our previous data source without rebuilding our calculations. To do this, head back to Tableau and select Data > New Data Source Right click on the previous data source > Select replace data source. Make sure you see both files as current and replacement and click ok.

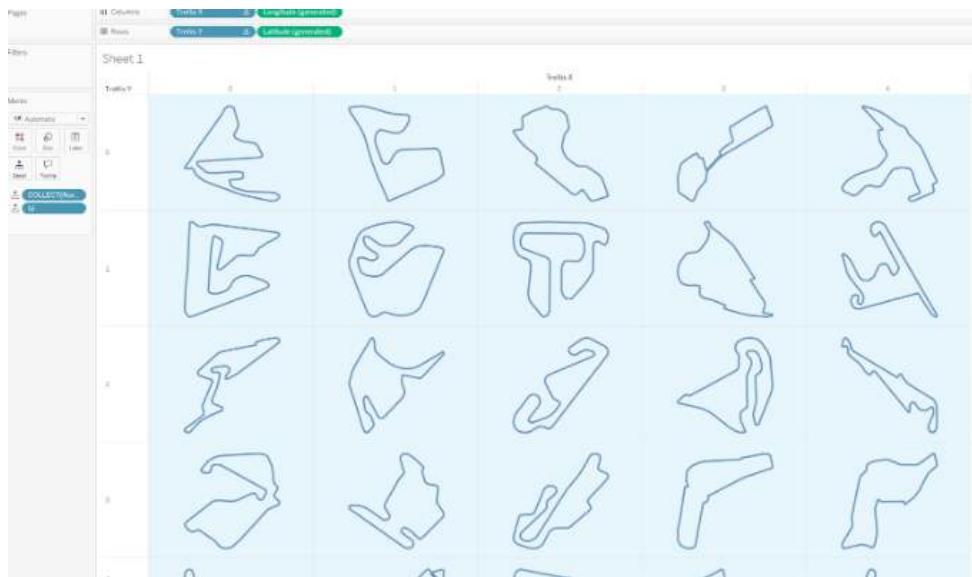


You'll get an error as Geometry is no longer found as we names our new column "Normalised Tracks", we can replace this field following a similar method.

Right click the field displaying the error which is "Geometry" > Click Replace Reference and select our "Normalised Tracks field".

The chart will refresh and if you zoom out you'll find all the tracks located on Null Island which is the location at zero degrees latitude and zero degrees longitude.

Now you can zoom in and format as you please.



Credit has to go to Andy Kriebel and his work on Strava normalisation which inspired this work, as well as Ollie Clarke from The Information Lab who helped with this one.

To download all the resources, Go to the **[Github Page located at the top of the page](#)**.

CJ: Has there been any other visuals in the wider analytics community on F1 that you've particularly liked? What did you appreciate about them?

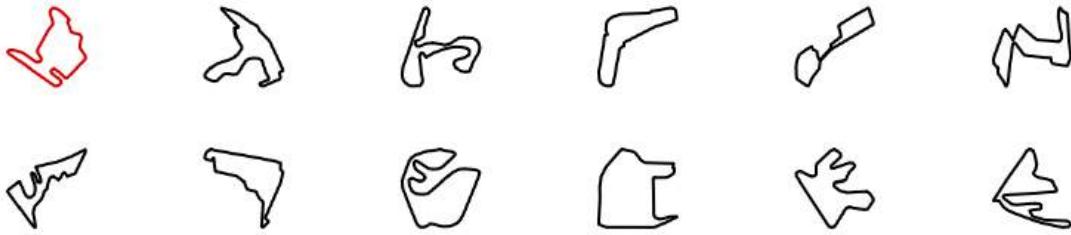
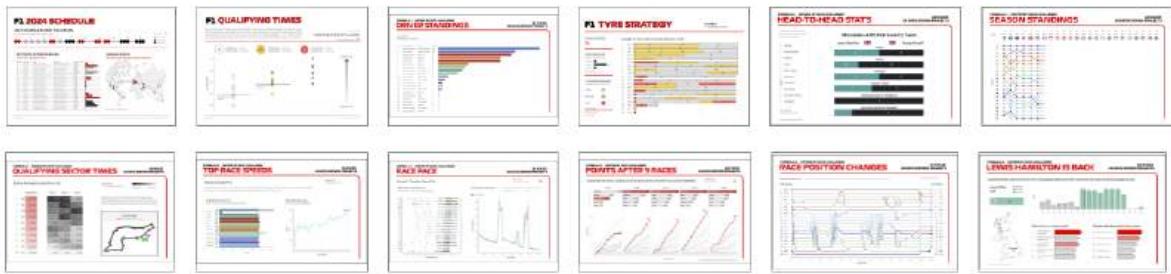
D: My inspiration comes from a number of sources. **Mateusz** has some incredible F1 projects that are updated weekly, I've tried to recreate some of these myself. **Naresh Suglani** has some awesome projects and **Anna** have some amazing work too. Recently, Rob Taylor is inspiring me with some of his F1 work, check out what Rob has done here to visualise identical lap times during qualifying; [link](#)) Outside of the Tableau community you have to check out **Formula Data Analysis**, **Formula Addict** and **F1 Holiness** who all post great content that's informative and entertaining. I also take a lot of inspiration from the wider #Sportsvizsunday weekly round-ups. Although they are different sports presented, I always look to see how others are visualising their data and telling stories, as ways to present my work.

CJ: What's next?

FORMULA 1 DRIVEN BY DATA CHALLENGE
PROJECT MENU

24 RACES
24 DATA DRIVEN PROJECTS

Click a project or hover over a circuit



D: I'm now halfway through the season, so that's 12 projects completed and another 12 to go. It's been a real challenge so far in terms of managing my time but I'm finding it rewarding. I still have plenty of ideas that I'm yet to tackle and that includes the telemetry data. I would say this is more technical from a physics perspective, so I'm going to use the summer break to get my head around it all and learn some python. I also plan to share some of the methods used to prep the data in Alteryx so can get involved in Formula 1 data too.

CJ: Any advice?

D: For those starting out are looking to begin in data, definitely build your skills with personal projects that interest you. When wanting to transition into data you have to put the hours in to upskill. You'll feel less friction and be more motivated to work on projects that interest you like a hobby, a sport, a tv series and that's where you can put the hours in. My second piece of advice is reach out. The Tableau community is amazing and everyone is so supportive. If you're stuck on a problem, need help gathering some data or simply want to expand your network the #datafam will make you feel welcome.

LOGGING OFF,

CJ

TENNIS PLAY BY PLAY #SPORTSVIZSUNDAY

Hi all,

A short one this week. Recently I wanted to look at some tennis data, and my usual go to is the tennis slam point by point data repo by **Jeff Sackmann**. Sometimes however the information you are after isn't always in the repository so it means venturing further to the tennisabstract.com website he hosts that contains thousands of different match reports that have been compiled together.

I am always on the hunt for “cool tennis stories” so when I come across them I want to be able to find the associated data.

Recently, I saw two stories that I thought I might want to later visualise.

The first was Iga Swiatek lost ten points in R16 of the French open. Total. That matches the most dominant tour-level performance—any tournament, any round—since 2010.

A	B	C	D	E	F
Points lost	Tournament	Rd	Winner	Loser	Score
10	2024 Roland Garros	R16	Iga Swiatek	Anastasia Potapova	6-0 6-0
10	2021 Australian Open	R128	Ashleigh Barty	Danka Kovinic	6-0 6-0
12	2019 Madrid	R16	Simona Halep	Viktoria Hruncakova	6-0 6-0
12	2010 Estoril	R32	Anabel Medina Garrigues	Magali De Lattre	6-0 6-0
13	2021 Rome	F	Iga Swiatek	Karolina Pliskova	6-0 6-0
13	2017 Tokyo	R32	Aleksandra Krunic	Kimiko Date Krumm	6-0 6-0
13	2015 Wimbledon	R128	Petra Kvitova	Kiki Bertens	6-1 6-0
14	2021 Cleveland	R16	Katerina Siniakova	Nagi Hanatani	6-0 6-0
14	2020 Auckland	R32	Caroline Wozniacki	Paige Mary Hourigan	6-1 6-0
14	2017 Bol	R32	Tereza Mrdeza	Prarthana G Thombare	6-0 6-0
14	2015 Miami	R32	Serena Williams	Catherine Cartan Bellis	6-1 6-1
14	2011 Tokyo	R64	Maria Kirilenko	Erika Sema	6-1 6-0
14	2011 Luxembourg	R32	Anabel Medina Garrigues	Anne Kremer	6-0 6-0
15	2023 US Open	R32	Iga Swiatek	Kaja Juvan	6-0 6-1
15	2019 US Open	QF	Serena Williams	Qiang Wang	6-1 6-0
15	2017 Bastad	R32	Barbora Krejcikova	Annika Beck	6-0 6-0
15	2015 Toronto	R64	Roberta Vinci	Karin Knapp	6-0 6-0
15	2014 US Open	R128	Agnieszka Radwanska	Sharon Fichman	6-1 6-0
15	2014 Baku	R32	Jana Cepelova	Nazrin Jafarova	6-1 6-0
15	2013 Australian Open	R64	Maria Sharapova	Misaki Doi	6-0 6-0
15	2012 Cincinnati	R32	Sofia Kenin	Mona Barthel	6-0 6-0

The second was that Roger Federer has the title of most Aces in a singles grand slam final. Which was done in the 2009 Wimbledon final, hitting 50 aces during the match.

So both these matches were hard to come by the data in terms of being in the repository but they do exist on tennis abstract.

Let's walk through the Iga example.

2024 Roland Garros R16: [Iga Swiatek vs Anastasia Potapova](#)

Iga Swiatek d. Anastasia Potapova 6-0 6-0

Use the links below to see dozens of tables displaying detailed data on every aspect of this match. For further context, tour and player averages are visible for most cells when you move your cursor over them. These figures are based on other charted matches, including [6283 WTA matches](#), 1623 WTA matches on clay, [266 Iga Swiatek matches](#) (86 on clay), and [37 Anastasia Potapova matches](#) (13 on clay). The more charted matches in the database, the more valuable this project becomes. Please try [charting a match yourself](#).

[Stats Overview](#) | [Serve Statistics Overview](#) | [Serve Influence](#)

[Key point outcomes](#) | [Point outcomes by rally length](#) | [Point-by-point description](#)

[Iga Swiatek](#): [Serve Breakdown](#) | [Return Breakdown](#) | [Net Points](#) | [Shot Types](#) | [Shot Direction](#)

[Anastasia Potapova](#): [Serve Breakdown](#) | [Return Breakdown](#) | [Net Points](#) | [Shot Types](#) | [Shot Direction](#)

STATS OVERVIEW	A%	DF%	1stIn	1st%	2nd%	BPSaved	RPW%	Winners (FH/BH)	LFE (FH/BH)
Iga Swiatek	11.1%	0.0%	59.3%	93.8%	81.8%	0/0	77.4%	13 (5/5)	3 (0/3)
Anastasia Potapova	0.0%	3.2%	38.7%	25.0%	21.1%	2/8	11.1%	5 (3/2)	15 (7/7)
SET 1									
Iga Swiatek	0.7%	0.0%	0.0%	0.0%	0.0%	0/0	0.0%	7 (3/3)	2 (0/2)
Anastasia Potapova	0.0%	6.7%	46.7%	28.6%	12.5%	0/3	20.0%	2 (0/2)	7 (2/4)
SET 2									
Iga Swiatek	16.7%	0.0%	58.3%	100.0%	100.0%	0/0	75.0%	6 (2/2)	1 (0/1)
Anastasia Potapova	0.0%	0.0%	31.3%	20.0%	27.3%	2/5	0.0%	3 (3/0)	8 (5/3)

This match was charted by [Zindaras](#).

[Find out how to chart matches yourself](#) | [Main Menu](#)

For some of the more standard point by point metrics we will need to refer to the “point-by-point description”

[Stats Overview](#) | [Serve Statistics Overview](#) | [Serve Influence](#)

[Key point outcomes](#) | [Point outcomes by rally length](#) | Point-by-point description

Iga Swiatek: [Serve Breakdown](#) | [Return Breakdown](#) | [Net Points](#) | [Shot Types](#) | [Shot Direction](#)

Anastasia Potapova: [Serve Breakdown](#) | [Return Breakdown](#) | [Net Points](#) | [Shot Types](#) | [Shot Direction](#)

Server	Sets	Games	Points	
Iga Swiatek	0-0	0-0	0-0	1st serve wide; forehand return down the middle (deep); backhand crosscourt; backhand slice down the middle, forced error .
Iga Swiatek	0-0	0-0	15-0	1st serve to body; backhand return down the middle, forced error .
Iga Swiatek	0-0	0-0	30-0	1st serve to body, fault (long). 2nd serve to body; backhand return inside-out (very deep), winner .
Iga Swiatek	0-0	0-0	30-15	1st serve to body; forehand return inside-in (very deep); forehand crosscourt; forehand down the middle; forehand inside-out; backhand crosscourt (net), unforced error . (5-shot rally)
Iga Swiatek	0-0	0-0	40-15	1st serve wide, ace .
Anastasia Potapova	0-0	0-1	0-0	1st serve to body; forehand return crosscourt (very deep), forced error .
Anastasia Potapova	0-0	0-1	15-0	1st serve to body, fault (long). 2nd serve to body; backhand return crosscourt (very deep); backhand crosscourt; backhand crosscourt; backhand down the line (wide and long), unforced error . (4-shot rally)
Anastasia Potapova	0-0	0-1	15-15	1st serve wide; forehand return down the middle (deep) (net cord); forehand inside-out; backhand down the line, forced error .
Anastasia Potapova	0-0	0-1	30-15	1st serve down the T; forehand return down the middle (very deep); backhand down the middle; forehand crosscourt; forehand crosscourt; forehand crosscourt; forehand down the middle; forehand inside-out; backhand lob down the middle; smash inside-out, winner . (10-shot rally)
Anastasia Potapova	0-0	0-1	30-30	1st serve to body, fault (long). 2nd serve to body; backhand return down the middle (deep); backhand crosscourt; backhand slice down the middle; forehand inside-out; backhand down the line; forehand lob crosscourt; forehand down the line; backhand slice down the middle; forehand at net crosscourt, winner . (10-shot rally)
Anastasia Potapova	0-0	0-1	30-40	1st serve down the T, fault (long). 2nd serve to body; backhand return crosscourt (deep); backhand down the middle; backhand crosscourt; backhand crosscourt; backhand down the line, winner . (6-shot rally)
Iga Swiatek	0-0	2-0	0-0	1st serve to body; forehand return down the middle, forced error .
Iga Swiatek	0-0	2-0	15-0	1st serve to body, fault (net). 2nd serve to body; backhand return down the line, forced error .
Iga Swiatek	0-0	2-0	30-0	1st serve down the T, fault (wide). 2nd serve down the T; backhand return down the middle (shallow); backhand crosscourt; backhand crosscourt; backhand crosscourt; backhand slice crosscourt; backhand slice at net down the line; forehand lob down the middle; smash inside-out; backhand lob down the line, forced error . (9-shot rally)
Iga Swiatek	0-0	2-0	40-0	1st serve to body, fault (net). 2nd serve to body; backhand return crosscourt (shallow), winner .
Iga Swiatek	0-0	2-0	40-15	1st serve to body; backhand return down the middle (very deep); backhand down the middle; backhand crosscourt; backhand down the middle (net), unforced error . (4-shot rally)
Iga Swiatek	0-0	2-0	40-30	1st serve down the T, fault (long). 2nd serve wide; backhand return down the middle (deep); forehand crosscourt; forehand crosscourt; forehand down the middle; forehand crosscourt; forehand crosscourt; forehand down the line; backhand crosscourt; backhand slice crosscourt, forced error . (9-shot rally)

Amazing. It's not hugely detailed like the dataset in the standard repo's but at least it contains the point winners.

So let's take a look at the code...

```
# Make sure you have downloaded the appropriate version of Chrome WebDriver from
# https://sites.google.com/a/chromium.org/chromedriver/downloads

DRIVER_PATH = r"C:\Users\Cj.Mayes\PycharmProjects\pythonProject1\chromedriver.exe"
import html

from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from bs4 import BeautifulSoup
import pandas as pd

options = webdriver.ChromeOptions()
# Set any additional options as needed
# For example, to run Chrome headless:
# options.add_argument("--headless")

driver = webdriver.Chrome(executable_path=DRIVER_PATH, options=options)

driver.get("https://www.tennisabstract.com/charting/20240602-W-Roland_Garros-R16-Iga_Swiatek-Anastasia_Potapova.html")

# Wait for 10 seconds until the element is visible and clickable
wait = WebDriverWait(driver, 10)
button = wait.until(EC.element_to_be_clickable((By.ID, "pointlog")))

# Click the button element
button.click()

# Now that the page content has loaded, capture the page source
html_content = driver.page_source
driver.quit() # Ensure the driver is closed after fetching the data

soup = BeautifulSoup(html_content, 'html.parser')

# Find the table. Assuming 'forecast' is the correct id based on the context you've provided; if not,
# adjust as needed
forecast_table = soup.find('span', id='forecast').find('table')

data = [] # Will hold our extracted data

for row in forecast_table.find_all('tr')[1]: # Process the rows in the table, skipping the header row
    cols = row.find_all('td')

    if len(cols) == 5: # Ensure the row has the expected number of columns
        server = html.unescape(cols[0].text.strip())
        sets = html.unescape(cols[1].text.strip())
        games = html.unescape(cols[2].text.strip())
        points = html.unescape(cols[3].text.strip())
        commentary = html.unescape(cols[4].text.strip())

        data.append({
            'Server': server,
            'Sets': sets,
            'Games': games,
            'Points': points,
            'Commentary': commentary
        })

# Convert the data into a pandas DataFrame
df = pd.DataFrame(data)

# Optionally, before exporting, you can further clean up if you suspect additional HTML entities or
# specific formats
# For spaces, the html.unescape should generally cover it, but here's an example of manual replacement
# df = df.applymap(lambda text: text.replace(u'\xa0', ' ') if isinstance(text, str) else text)

df.replace({'\xa0': ' ', ' ': ' '}, regex=True, inplace=True)

# Now, exporting the cleaned DataFrame to a text file
df.to_csv('points_data.txt', index=False, sep='\t', encoding='UTF-8')

# Notes:
# - Ensure Selenium, BeautifulSoup, and pandas are installed in your environment.
# - Confirm the correct WebDriver path and web page URL.
# - This script clicks a button to load the table, so ensure the ID used for clicking is correct.
```

The full code can be found in the repository

1. The code starts by setting the path to the Chrome WebDriver. This WebDriver is necessary to automate the Chrome browser. (Make sure you have the webdriver that suits the version of chrome you use)
2. Next, it imports the required libraries: html, selenium, BeautifulSoup, and pandas.
3. The code creates an instance of the Chrome WebDriver using the specified driver path and options.
4. It navigates to a specific URL using driver.get().
5. The code waits for a button element with the id "pointlog" to become clickable using WebDriverWait and the expected conditions. I put this step in because my chrome wasn't loading quick enough.
6. Once the button is clickable, it is clicked using the click() method.
7. The page source is captured using driver.page_source.
8. The WebDriver is then closed using driver.quit().
9. The captured page source is parsed using BeautifulSoup, creating a BeautifulSoup object named soup.
10. The code finds a table element with the id "forecast" using soup.find(). On the site tennis abstract website all the tables had the same structure.
11. It initializes an empty data list to store the extracted data.
12. It iterates over the rows of the forecast table (excluding the header row) using a for loop.
13. For each row, it extracts the data from the columns using cols = row.find_all('td').
14. If the row has the expected number of columns (5 in this case), it extracts the server, sets, games, points, and commentary data.
15. The extracted data is appended to the data list as a dictionary.
16. After all rows have been processed, the data list is used to create a pandas DataFrame.
17. The DataFrame is cleaned up by replacing specific HTML entities or formats using df.replace().
18. Finally, the cleaned DataFrame is exported to a text file named "points_data.txt".

The screenshot shows a code editor with two tabs: 'main.py' and 'points_data_RG_R16_24.txt'. The 'main.py' tab contains Python code for extracting tennis data. The 'points_data_RG_R16_24.txt' tab displays the resulting CSV-like data. The data includes columns for Server, Sets, Games, Points, and Commentary. The commentary column provides detailed descriptions of each serve and return, such as 'serve wide', 'forehand return down the middle', and 'forced error'.

	Server	Sets	Games	Points	Commentary
1	Iga Swiatek	0-0	0-0	0-0	1st serve wide; forehand return down the middle (deep); backhand crosscourt; backhand slice down the middle, forced error.
2	Iga Swiatek	0-0	0-0	15-0	1st serve to body; backhand return down the middle, forced error.
3	Iga Swiatek	0-0	0-0	30-0	1st serve to body, fault (long). 2nd serve to body; backhand return inside-out (very deep), winner.
4	Iga Swiatek	0-0	0-0	30-15	1st serve to body; forehand return inside-in (very deep); forehand crosscourt; forehand down the middle; forehand
5	Iga Swiatek	0-0	0-0	40-15	1st serve wide, ace.
6	Anastasia Potapova	0-0	0-1	0-0	1st serve to body; forehand return crosscourt (very deep), forced error.
7	Anastasia Potapova	0-0	0-1	15-0	1st serve to body, fault (long). 2nd serve to body; backhand return crosscourt (very deep); backhand
8	Anastasia Potapova	0-0	0-1	15-15	1st serve wide; forehand return down the middle (deep) (net cord); forehand inside-out; backhand down
9	Anastasia Potapova	0-0	0-1	30-15	1st serve down the T; forehand return down the middle (very deep); backhand down the middle; forehand
10	Anastasia Potapova	0-0	0-1	30-30	1st serve to body, fault (long). 2nd serve to body; backhand return down the middle (deep); backhand
11	Anastasia Potapova	0-0	0-1	30-40	1st serve down the T, fault (long). 2nd serve to body; backhand return crosscourt (deep); backhand do
12	Iga Swiatek	0-0	2-0	0-0	1st serve to body; forehand return down the middle, forced error.
13	Iga Swiatek	0-0	2-0	15-0	1st serve to body, fault (net). 2nd serve to body; backhand return down the line, forced error.
14	Iga Swiatek	0-0	2-0	30-0	1st serve down the T, fault (wide). 2nd serve down the T; backhand return down the middle (shallow); backhand
15	Iga Swiatek	0-0	2-0	40-0	1st serve to body, fault (net). 2nd serve to body; backhand return crosscourt (shallow), winner.
16	Iga Swiatek	0-0	2-0	40-15	1st serve to body; backhand return down the middle (very deep); backhand down the middle; backhand crosscourt;
17	Iga Swiatek	0-0	2-0	40-30	1st serve down the T, fault (long). 2nd serve wide; backhand return down the middle (deep); forehand crosscou
18	Anastasia Potapova	0-0	0-3	0-0	1st serve wide; forehand return crosscourt (deep); forehand down the line (wide), unforced error.
19	Anastasia Potapova	0-0	0-3	0-15	1st serve down the T, fault (net). 2nd serve to body; backhand return crosscourt (deep) (wide), unfo
20	Anastasia Potapova	0-0	0-3	15-15	1st serve to body; forehand return down the middle (shallow); backhand crosscourt (net cord); forehand
21	Anastasia Potapova	0-0	0-3	15-30	1st serve down the T, fault (net). 2nd serve to body; backhand return crosscourt (deep); backhand down
22	Anastasia Potapova	0-0	0-3	15-40	1st serve down the T; backhand return down the middle (shallow); forehand crosscourt; forehand crossco

There we have it, a quick way to extract the information from the tennis abstract site.

GOING FURTHER

Remove the row line spaces,

Indicate whether it was a first or second serve,

Indicate point winners,

Flag for each row what the score then turned to.

That's it for this week. Enjoy accessing more tennis data.

LOGGING OFF,

CJ

ROMANIA EUROS JOURNEY #ROMANIANDATATRIBE

Hi all,

Last week I got tagged in a linkedin post from Claudiu Ionuț Anghelescu about soccer data available for their #RomanianDATA Tribe challenge.

It was a collaboration with Federatia Romana de Fotbal Intelligence Centre and Andrei Angelescu. Their aim to analyze the road of our national team that culminated in the qualification for the Euros

2024.

The challenge:

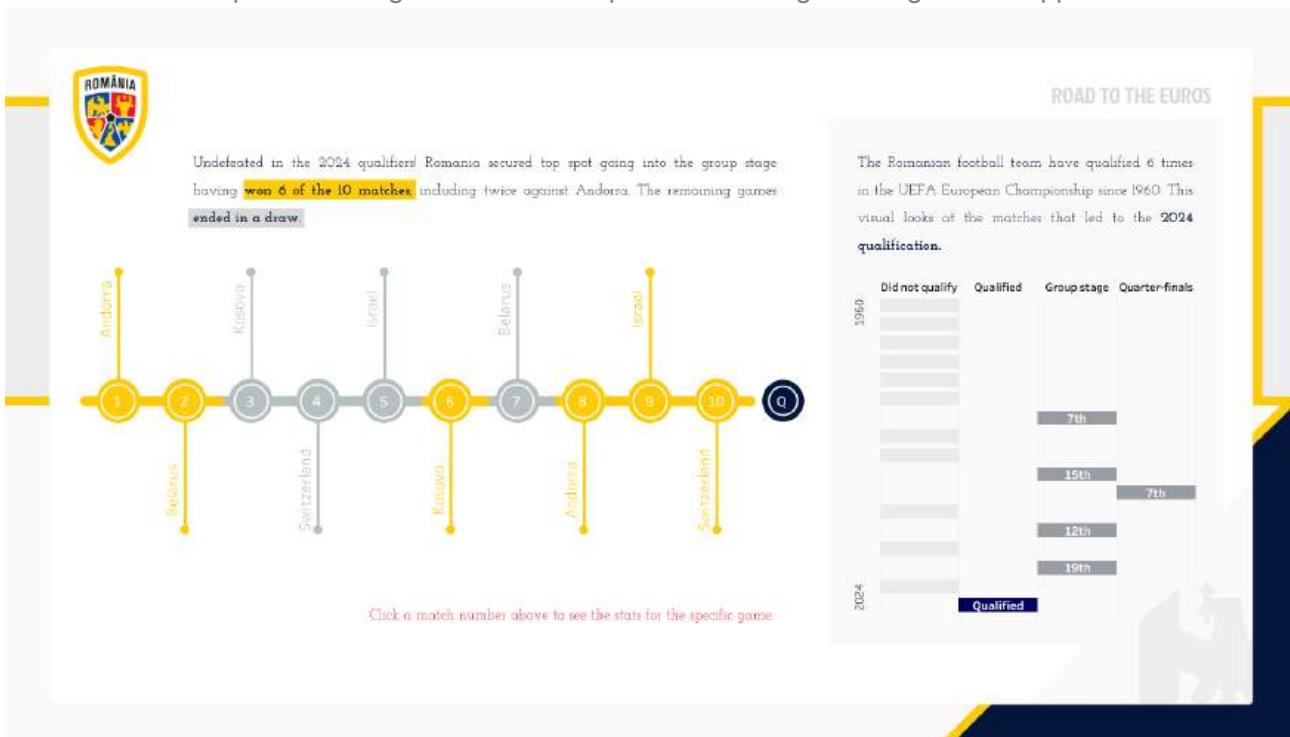
"Create your version of data visualization using the attached data source and/or additional data you want to explore from Flashscore, Transfermarkt GmbH & Co. KG, UEFA, Wyscout, etc.

You can analyze the matches, the audience, the performance of the individual players or the entire team, the increases in market value of the players, the events during a match or the opponents of the Romanian national team."

I love things like this because unless you are coding savvy and can web scrape data, then you have to rely on paid memberships to soccer data.

I've re uploaded the datasets in the repo. They are a fantastic resource for those that want to experiment with different chart types.

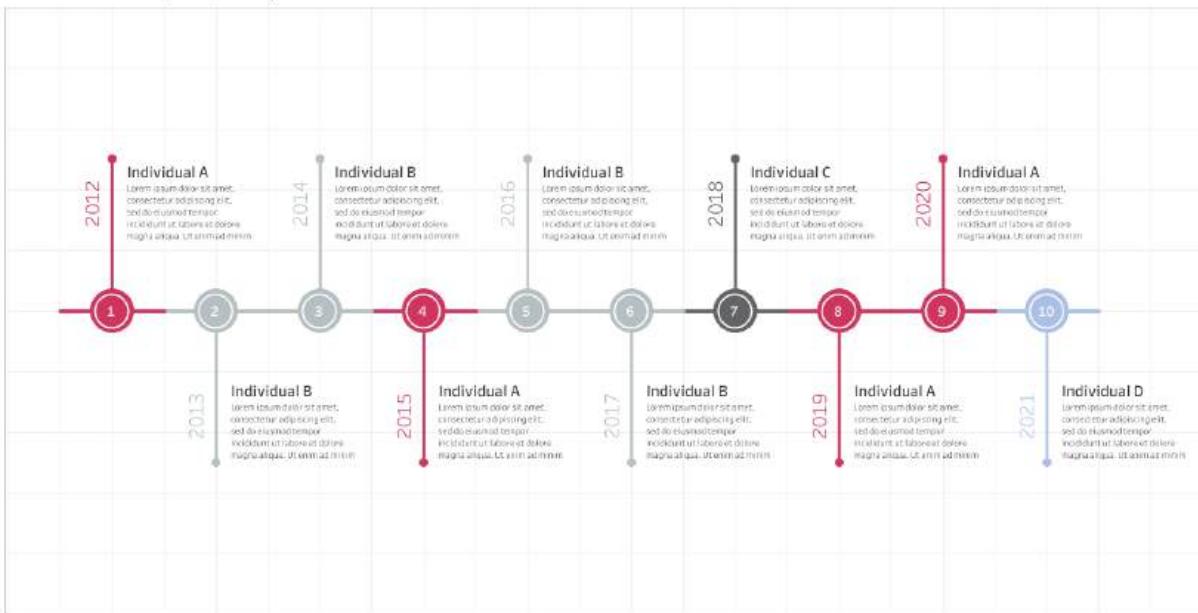
Despite not being Romanian....I spent an evening entering out of support.



The first view is fairly simple. it is a timeline of the matches by game week using color to show the matches won, and grey to indicate the draws. So impressive that Romania won 6 of the 10 games to reach the group stages. This is the 6th time making the group stages, so added in the chart to the right hand side for context of highlighting what an exciting time it is.

I built a timeline template a while back but never ended up writing about it, you can find it [here](#).

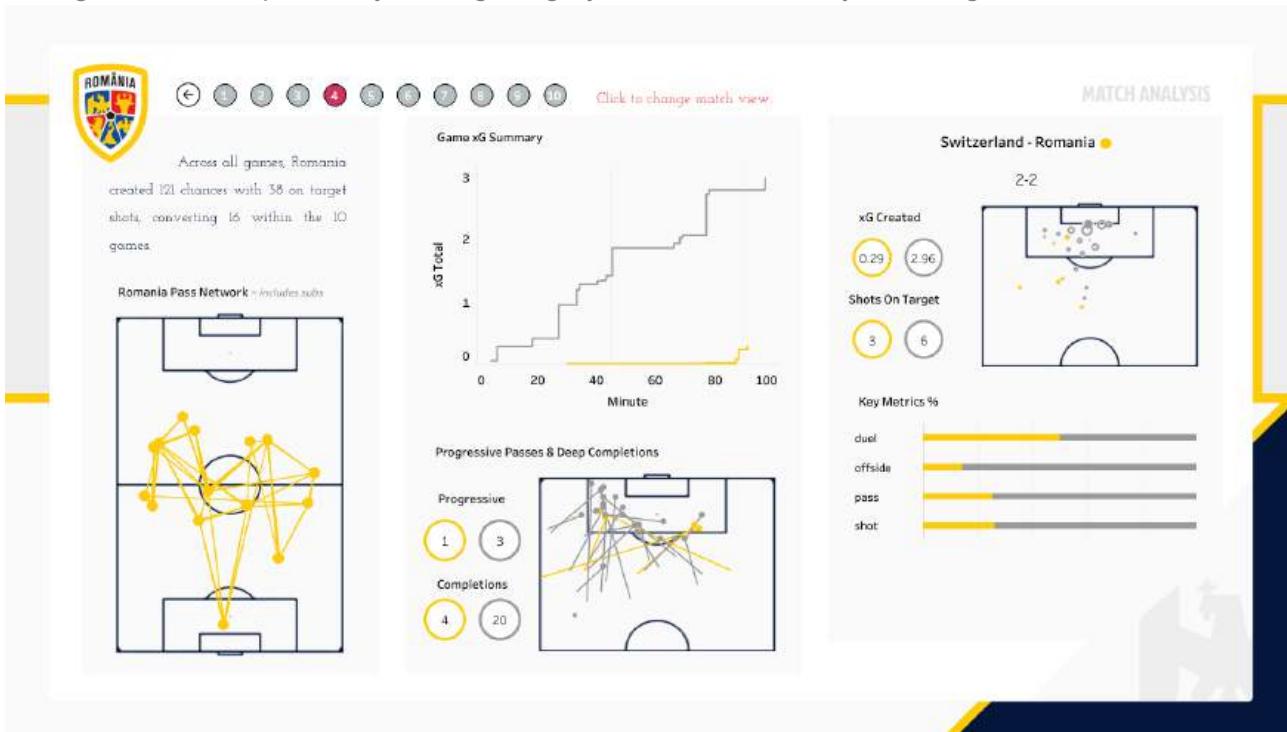
A Timeline Mock-Up Inspired by Khalid Amribet's Visual "The Best of the Best"



The second page to the view clicks through from the timeline and show the details for that specific match.

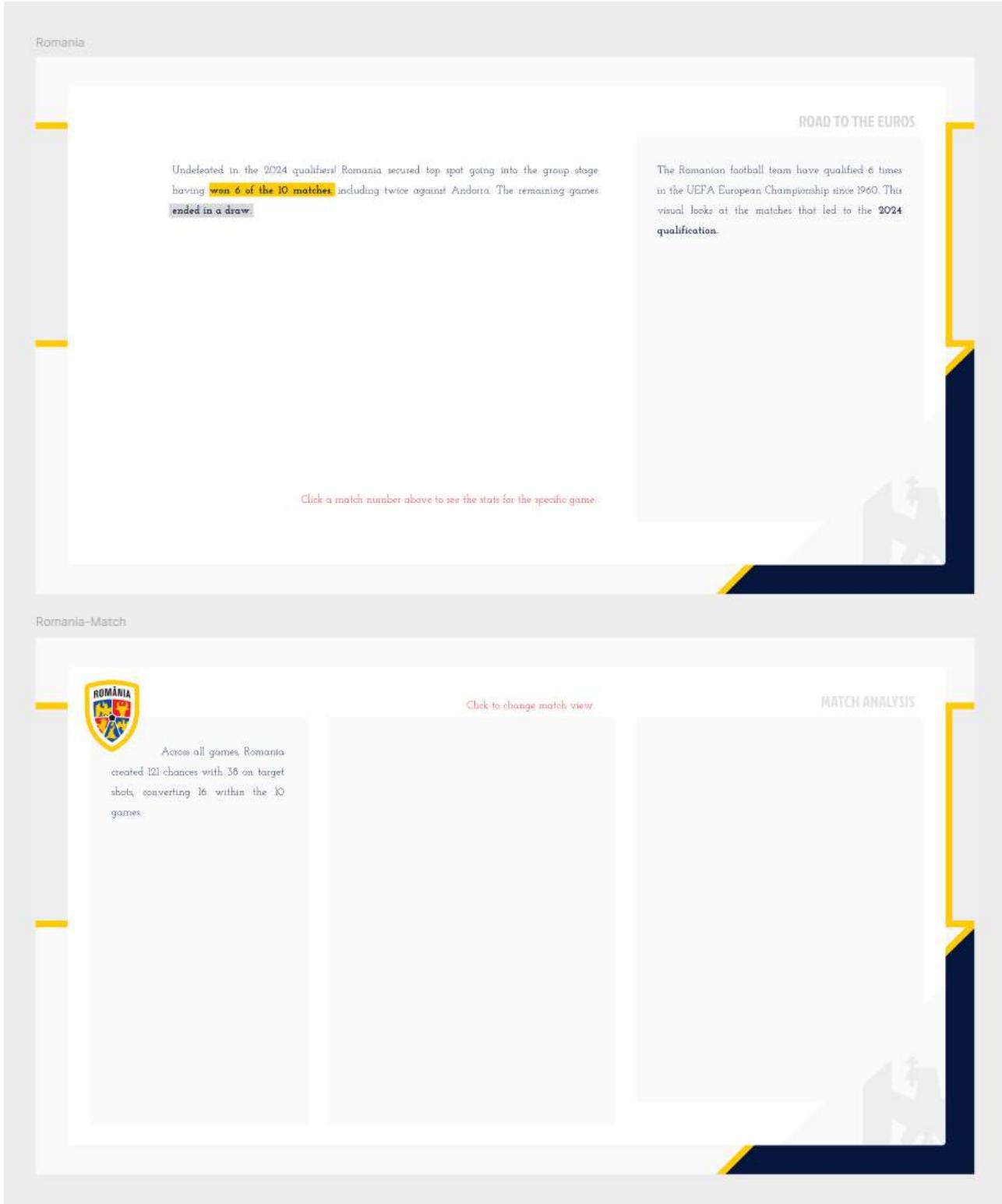
Mainly looking at attacking metrics such as xG, number of shots, number on target as well as some progressive or deep completions and then some possession type stats.

The game can be updated by clicking the grey match buttons to cycle through the different matches.



All the charts on this view have been created using logic from previous tutorials.

**Plotting xg
Pass networks
Soccer shot maps**
The background is fairly static,



They were made in Figma. No doubt you could create this in PowerPoint it's nothing too extravagant. Just a few overlays. Originally I wanted to put some more images into it but was becoming quite the uphill battle of images vs statistics and only had the one evening to spend on it.

Below are a few other individuals whos work stood out to me from the challenge.

Firstly I came across this piece by Guilhem – The hex map really shows dominance in certain parts the pitch. The box area seems relatively mirrored for both teams but when you can see much greater differences in attempts when outside the box and attacking angles.



Guilhem Figueres-Sarda • 2nd
MSc Student, National School for Statistics and Data Science - ENSAI ...
4d •

Follow ...

Last week, [Claudiu Ionuț Anghelușcă](#) and the Romanian Football Federation launched a data viz challenge on [RomanianDATA Tribe](#) guided by the beginning of [#Euro2024](#).

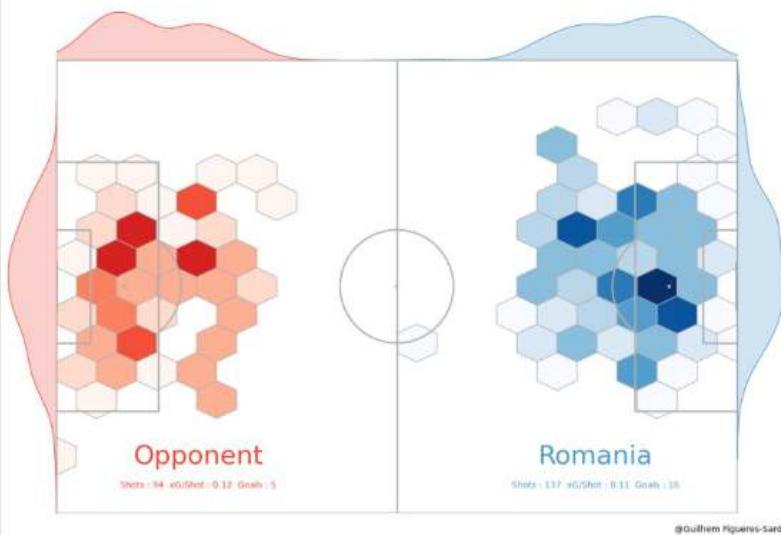
It was a fun challenge where I isolated shot quantity as a key point in order for the Romanian National Team to win the Euro Qualifiers.

Good luck other participants!

#RomanianDATA #RomanianDATATribe

Shot selection in Romania's matches during the Euro Qualifiers

With 6 victories and 4 draws, Romania has won its ticket to Euro through shot quantity not quality



I also enjoyed this piece from **Sergiu Rotaru** – He uses just two colors to be able to distinguish easily the shots that have led to a goal. The part to whole ratio is very telling for looking at conversion %. Sergiu's title and added context of number of games won / drawn adds cohesion to the piece whilst allowing for really only 1-2 charts being used!

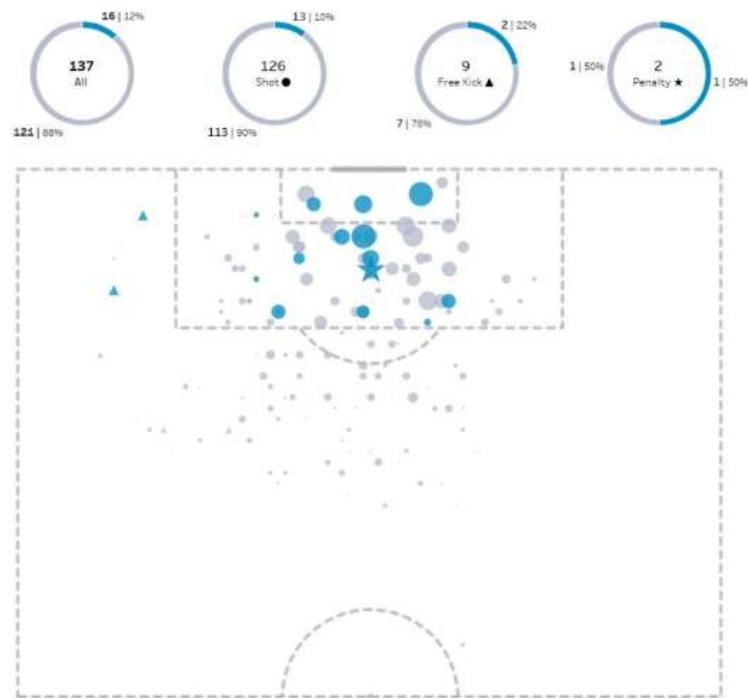
 ROMANIA | UEFA EURO 2024 QUALIFYING
Design: Sergiu Rotaru | Project: #RomanianDATATribe

10 games UNBEATEN Won Drawn

ATTEMPTS ANALYSIS click the donuts to filter the attempts type

GOAL NO GOAL

SHAPE SIZE:XG (EXPECTED GOALS) LOW>>HIGH



Hope to see many others get involved in future and looking forward getting back into creating more sports analytics pieces again!

Wishing the #RomanianDATA Tribe team all the best with future challenges.

LOGGING OFF,

CJ

SOCER ANALYTICS WITH NICK VAN LIESHOUT #SPORTSVIZSUNDAY

Hi all,

We have another guest blog this week. **Nick Van Lieshout** joins to talk about his journey within soccer analytics and some of his public facing portfolio.



[Twitter](#) [LinkedIn](#)

Nick van Lieshout

Data Visualization at Parma Calcio 1913 | Parma, Emilia-Romagna, Italy

[Follow](#)

Vizzes 6 Favorites 0 Following 1 Followers 80

The grid displays six examples of data visualization:

- Goalkeeper Restarts | CL Quarter-Finals**: A dashboard showing goalkeeper restarts during the CL Quarter-Finals.
- Serie A Historical Point Benchmarks**: A dashboard showing historical point benchmarks for Serie A.
- Dynamic Heat Maps #VOTD**: A dashboard showing dynamic heat maps for a VOTD (Player of the Day).
- Player Monthly Performance Report #VOTD**: A dashboard showing monthly performance reports for a VOTD.
- Coppa Italia 23/24 Predictions**: A circular visualization showing predictions for the Coppa Italia 23/24 tournament.
- Football (Soccer) Performance Dashboard | Jong PSV (live)**: A live performance dashboard for Jong PSV.

CJ: Nick, thanks for joining. I am so pleased to have a #SportsVizSunday regular contributor to join the blog today. You are in a fortunate position of being able to do sports analytics for both work and ‘play’ – Have you always had an analytical mind when looking at sports?

N: Thanks for having me, CJ! I guess I have always had quite an analytical approach to it. As a kid I used to cut out pieces from magazines and newspapers of football players and teams, together with every stat and piece of information I could find about them, to create booklets supplied with some of my own notes and thoughts. For me as a kid, these booklets served as some kind of ground truth to base my football opinions on, as a kind of reference. What I do as a job today is honestly not too dissimilar to creating such booklets, minus the scissors and the glue, so I guess part of the work still feels like ‘play’.

Looking back at it, kid-me was using cherry picked versions of these booklets to point out to my brother why his favorite team was bad and my favorite team was, in fact, good, thereby only using data fitting my opinion agenda. It’s a kind reminder that data does not strictly equate to objectivity, because people have a decision on what they show and how they show it.

CJ: Many individuals reading this blog will be inspired by your journey of entering the soccer community. Could you talk a little as to your role at Parma Calcio and how the industry is changing?

N: At Parma Calcio I am responsible for creating a fairly big chunk of the reports and dashboards on the football side: for match analysis, scouting, performance and leadership across first team and youth academy. In my work I use Tableau and Figma a lot, and occasionally R for ad hoc analysis. I am lucky to have colleagues dedicated to data architecture and engineering, so I can fully focus on optimizing the packaging, delivery and communication of data and analysis to staff members. All to make it easy and intuitive for them to understand and integrate in their workflow. Conscious and careful data visualization and communication is a very effective means to this end, and I think it is

being more and more recognized also in the industry as a specialty instead of just one of many tasks of a data analyst or scientist.

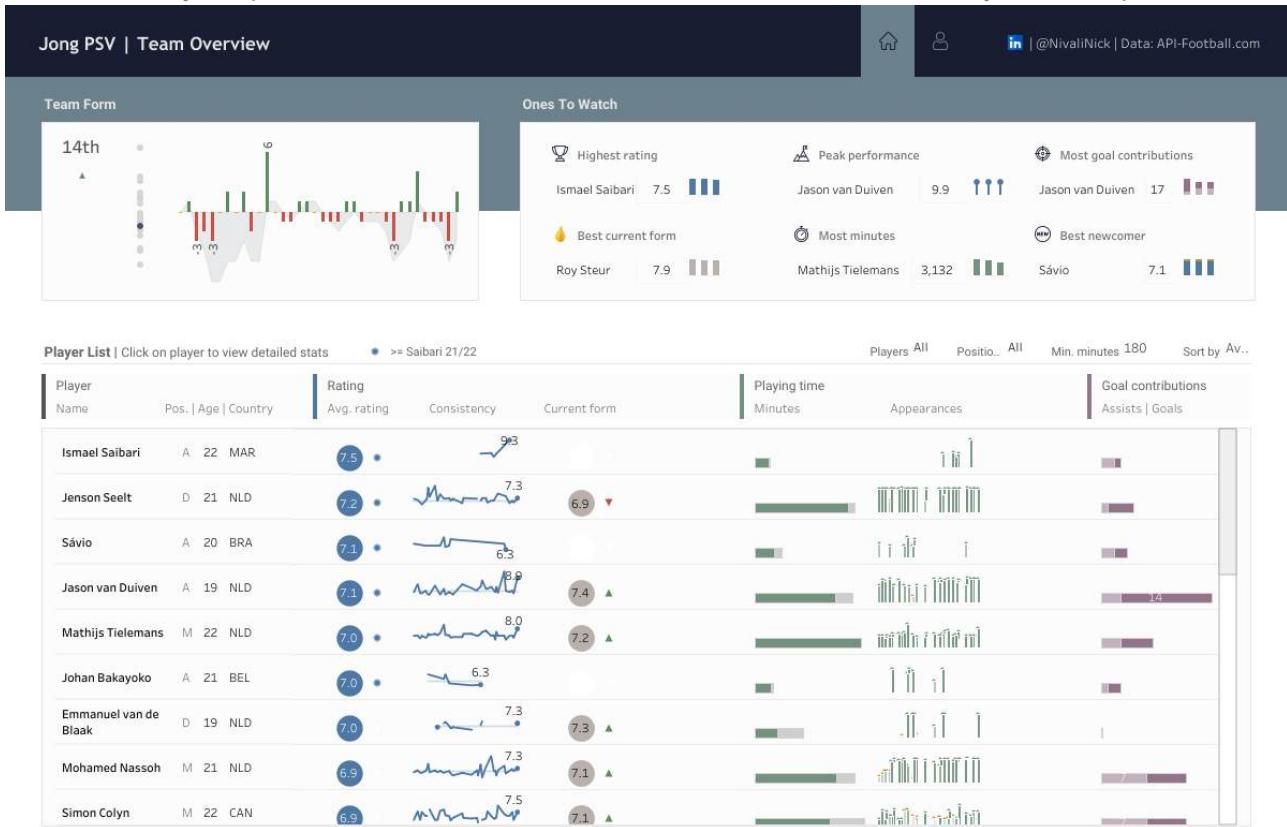
CJ: Do you have any tips for those wanting to expand their portfolio for match analysis and scouting?

What resources can people turn to that are free and publicly available which help individuals start their portfolio for match analysis or scouting analytics? How important is the data collection as part of your visuals?

N: There are quite a few open data sets available that can get you started. **StatsBomb** for instance has very detailed event data available for free that you could use to experiment with football data analysis. These are extensive enough to create quite a detailed analysis of the playing style of a team, or to create a deepdive on a player, and there are plenty of tutorials available on how to retrieve and play with their data in Python and R (also on this blog)

Of course clean, nicely prepped data just makes life easier as an analyst in general. It's usually the most time consuming part if you are starting a viz project from scratch. Also, good prep allows you to pull off specific advanced types of charts in Tableau much easier, and helps you optimize performance.

CJ: It's not often we see a business style framing design to a sports visual. Your performance dashboard has some incredible features in it technically. I particularly like the layers in the Team Form section. Can you speak a little about the decisions of the area and bar overlay and tooltip decisions?

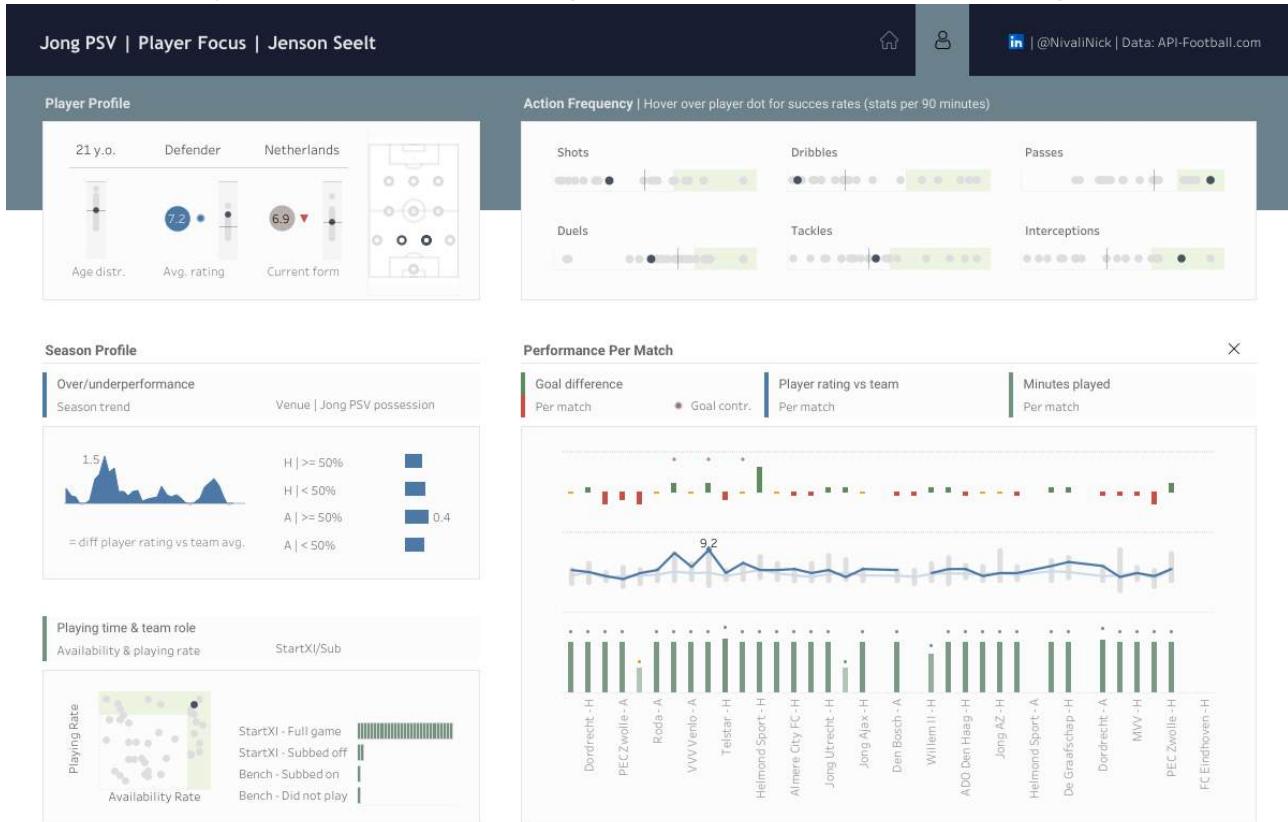


In this dashboard the envisioned end users are fans that want to get a quick pulse on the youth team of their favorite club, and want to know who the emerging talents are. The team form and ones to watch sections are prominently placed, because they directly address these questions.

Specifically about team form; given the constraints of the data at hand, the running goal difference was a fairly reasonable proxy of the ‘form’ of a team. Although, when available, I think xG difference would be a better alternative for this. I got the idea of overlaying an area on a bar chart in this context from one of Sam Parsons (legend) dashboards on Rugby. It gives you a quick view of the waves of

form the team has had, while being able to compare individual match performances, and detect which performances might have turned the tie in a season.

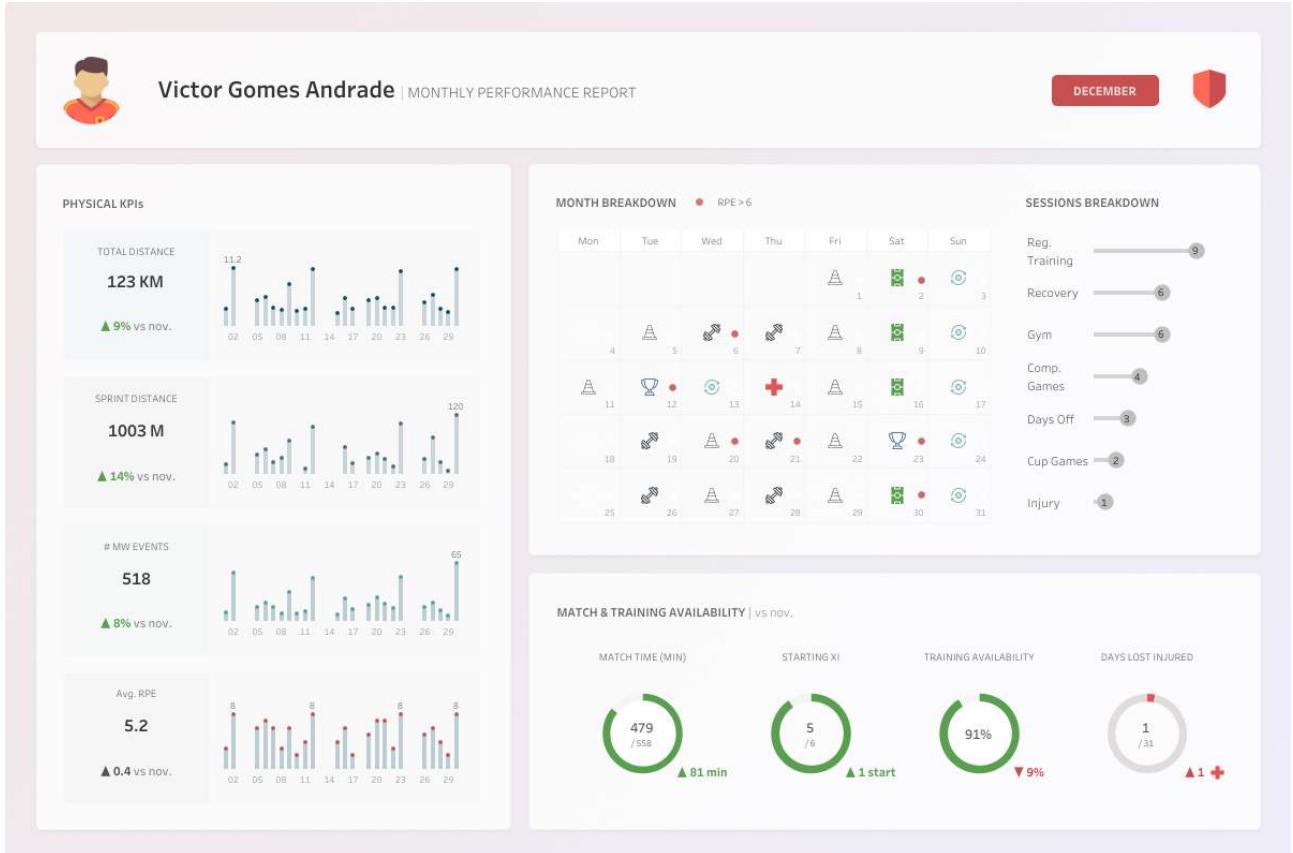
CJ: Again, on the player profile focus page I love the design choices behind the player profile. Can you speak a little to the design choices of color, shapes and framing?



The idea was to create separate sections summarizing performance (i.e a rating given by API-Football), playing time, technical/tactical kpis all on the season level, as well as one single chart breaking down the form of the player throughout the season, and his performance relative to his teammates from match to match.

If you dissect it in terms of Gestalt Principles (principles of human perception), I used enclosure and proximity to create distinct sections (via borders and plenty of whitespace), and similarity in the form of color to create some coherence across different sections.

CJ: Another business style framed dashboard you created that got VOTD is your player monthly report. Can you talk through some of the complexities creating the calendar?



N: Here I used 2 sheets. One sheet that only shows the custom shapes of the sessions, hiding everything else (e.g. transparent background, hide field labels, remove lines separating the days in the calendar). And another sheet containing the calendar with the RPE (rate of perceived exertion) indicator.

It was a quick and easy solution to integrate RPE nicely in the calendar, in which I wanted color and shape to encode the session type, and sizing the shape according to RPE was just not ideal in my view.

There is also an **excellent step by step example by Andy Kriebel** that shows how to do create calendars.

CJ: Your Serie A visual and Coppa Italia visual both showcase your advanced Tableau experience – using both polygons and chart layering. How do you approach new projects that don't necessarily conform to “out of the box” chart types?

SERIE A

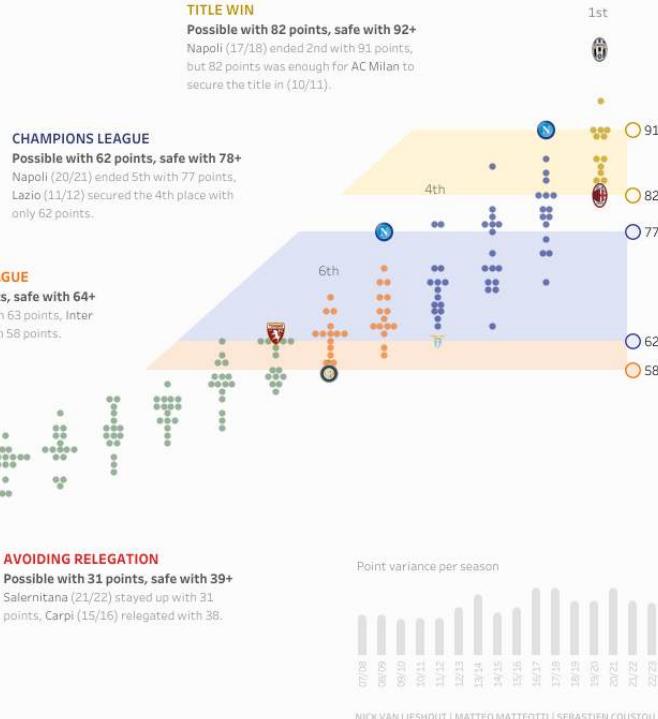
POINT BENCHMARKS

How many points do you need at minimum to have a chance at winning the title, qualify for European football or to avoid relegation, translated to the current year of Serie A? When are you safe?

The shaded areas indicate point intervals of contest, in which reaching one of the above objectives is a possibility but not a 'certainty' based on seasons 07/08 until 22/23.

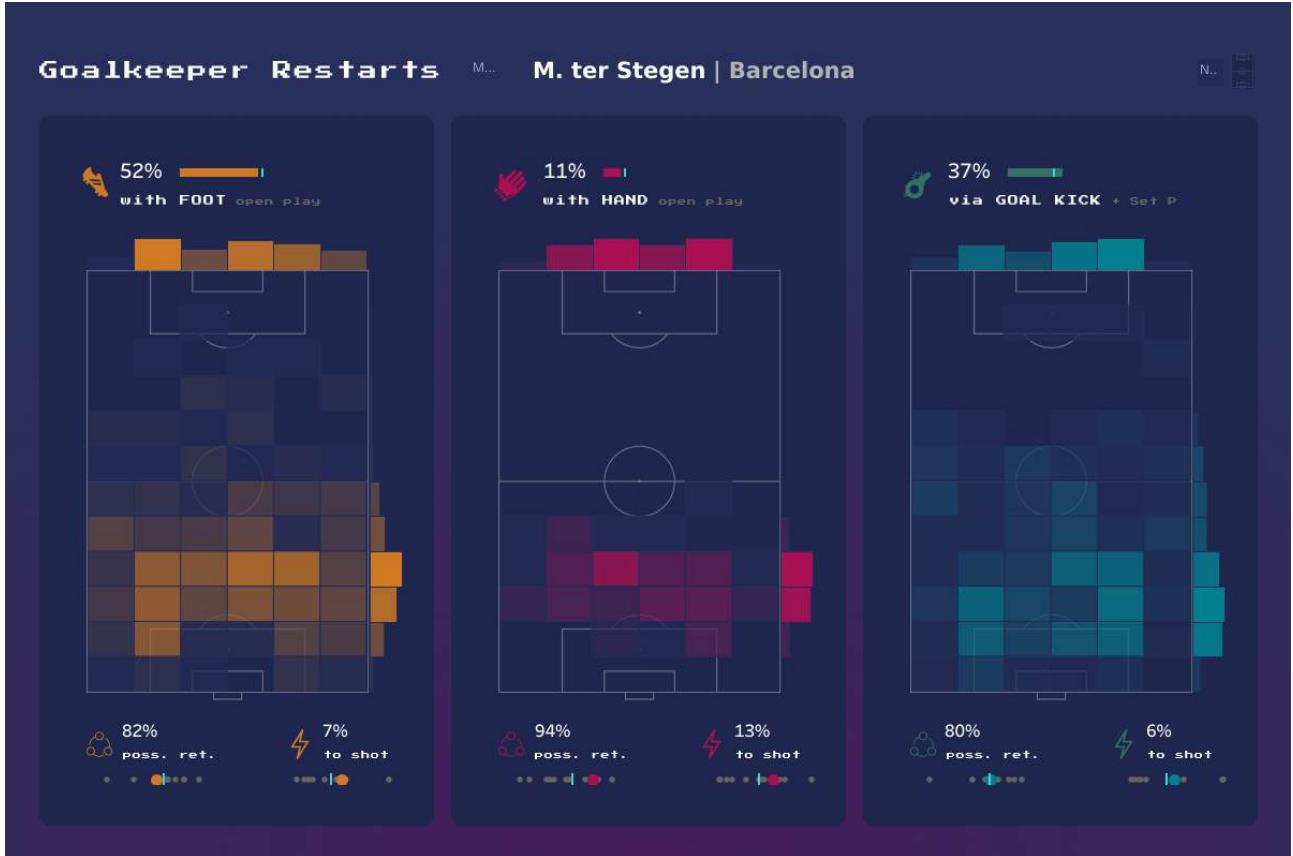
Hover over a dot to see the team, season and the number of points.

Data from Transfermarkt.com

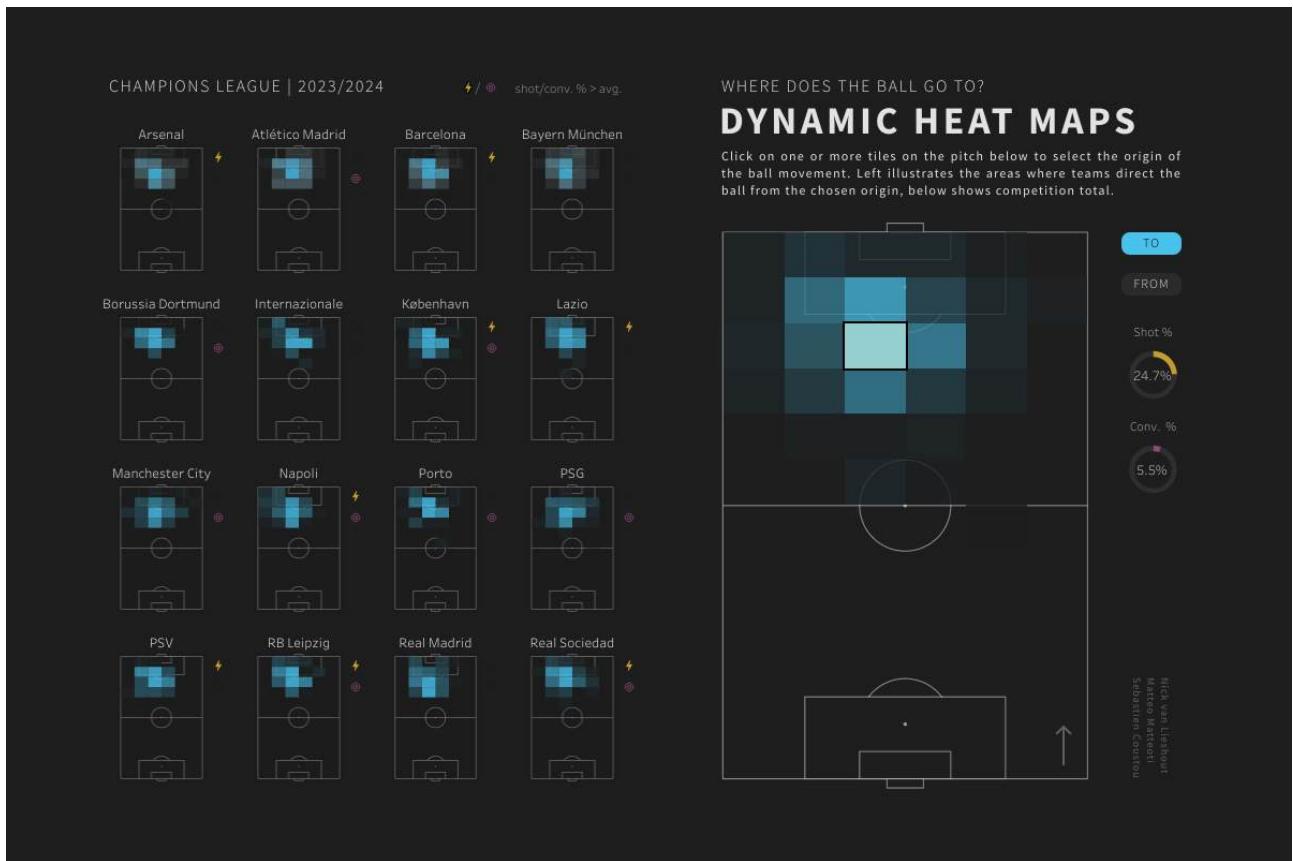


N: For me usually it all starts with pencil and paper. Sketching out how I would ideally like to show something. From there I compare my ideal to what is possible within the realms of my tools, i.e. how far we can reasonably get with the combination of Tableau and Figma while also taking into account the timeframe I have to create it. However, in my day to day work I rarely resort to very advanced chart types, as they are generally a bit more complex to read. But these types of charts do have the ability to attract an audience that would otherwise not be very interested in a certain topic, or to fuel a conversation.

CJ: In your goalkeeper restarts visual and your dynamic heat visual. You add subtle additional information to help enhance the original heatmap, such as in the case of marginal bars, shape icons and pie charts. How do you weigh up providing context vs excessive information, as well as ink ratio and page balance?



N: Using complex charts and bombarding the viewer with excessive information (as well as unnecessary chart elements) risks increasing the cognitive load to a point that a dashboard becomes daunting and off-putting. I always try to be very selective in what I do and don't show, when I show it, and try to avoid anything that does not directly contribute to the understanding of the data. Using shape icons and subtle indicators can help you package quite some contextual information in a limited number of pixels, while making sure you do not completely overpower the main chart.



CJ: Thank you for some of those design tips. Are there individuals in the community that you look to for inspiration, both sports and wider?

N: Andy Kriebel, Sam Parsons, Cj Mayes, Adedamola Ladipo, Ben Norland, Simon Beaumont, Mo Wootten, Gbolahan Adebayo.

CJ Round-Up:

Nick has some serious talent. The way he is able to balance different elements on the page, whilst bringing the perfect amount of context into the visual without it being over bearing is truly fantastic. I would recommend anyone looking to get into sports analytics to use Nick as a source of inspiration. I loved Nicks analogy of booklets, cutting and gluing bits of information together.

LOGGING OFF,

CJ

RUGBY DATA WITH ROB TAYLOR #SPORTSVIZSUNDAY

Hi all,

We have another guest blog this week. Rob Taylor joins to chat all things sports, with a focus on some rugby data. Hope you enjoy the convo!



Rob Taylor

England, United Kingdom

Specialist BI Analyst at JLL | Tableau Pro | Alteryx Core Certified & Innovator | British Data Awards Winner | 2xVOTD | TableauNext2024

Follow

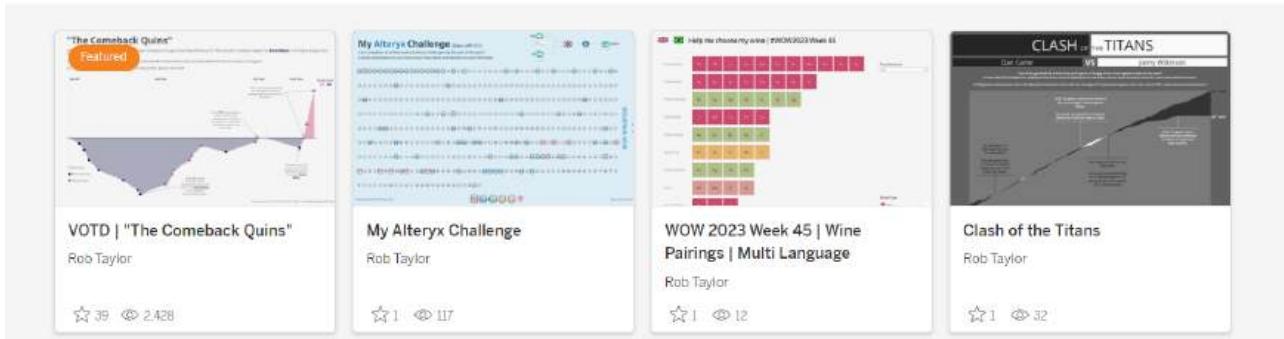
Hire Me

Vizzes

Favorites

Following

Followers



CJ: Rob, thanks for joining. I am so pleased to have a #SportsVizSunday regular join the blog today. For those who don't know, do you want to share a little bit about your journey at LNER & JLL, and where your interest in data started?

RT: Thanks CJ, it's a real pleasure to be invited on to the blog so thanks for having me! I've always been interested in numbers, I'm a very methodical person and love problem solving, numbers just make sense to me. But I also have a real creative passion too. In a former life I used to paint pop art style paintings of high profile rugby players that would be sold off at charity auctions.

I started my analytics career over 16 years ago in excel, automating data processes with macros and VBA. Being able to speed processes up and get to answers quickly was fantastic and I loved doing it.

During my time with London North Eastern Railway, LNER, I was instrumental in how they transformed and used their data. We got Tableau and Alteryx and that's when I knew what I wanted to do and my career path really took off. Building out all sorts of ETL processes to manipulate data so that we could use it in Tableau ticked the methodical part of my brain, then creating really interactive and insightful dashboards the business could use ticked the creative side.

About a year ago a colleague, **Danny Bradley**, introduced me to the datafam and I realised there was this whole bigger world outside of LNER. I was amazed what other people were building and what they could do with the tool. I wanted to be part of that. My recent move to JLL gave me the opportunity to work for a big business and stretch my skills, whilst also getting to work alongside so many of the datafam community I'd come across in the last year. There are so many opportunities for learning and skill growth I can't wait to see where it takes me.

CJ: You are a big rugby fan. I particularly remember your Ironviz entry on the love for Harlequins, as well as a shortform visual on the comeback quins. I particularly love the Comeback quins visual way of showcasing the shortfall of 28 points to then going all square before the final whistle. Can you tell us a little more why you felt this was the best representation of the match data?

Up n Unders- The Harlequins Story

Below we are going to explore the history of the English Rugby Union club that I have supported since I was 8 years old.

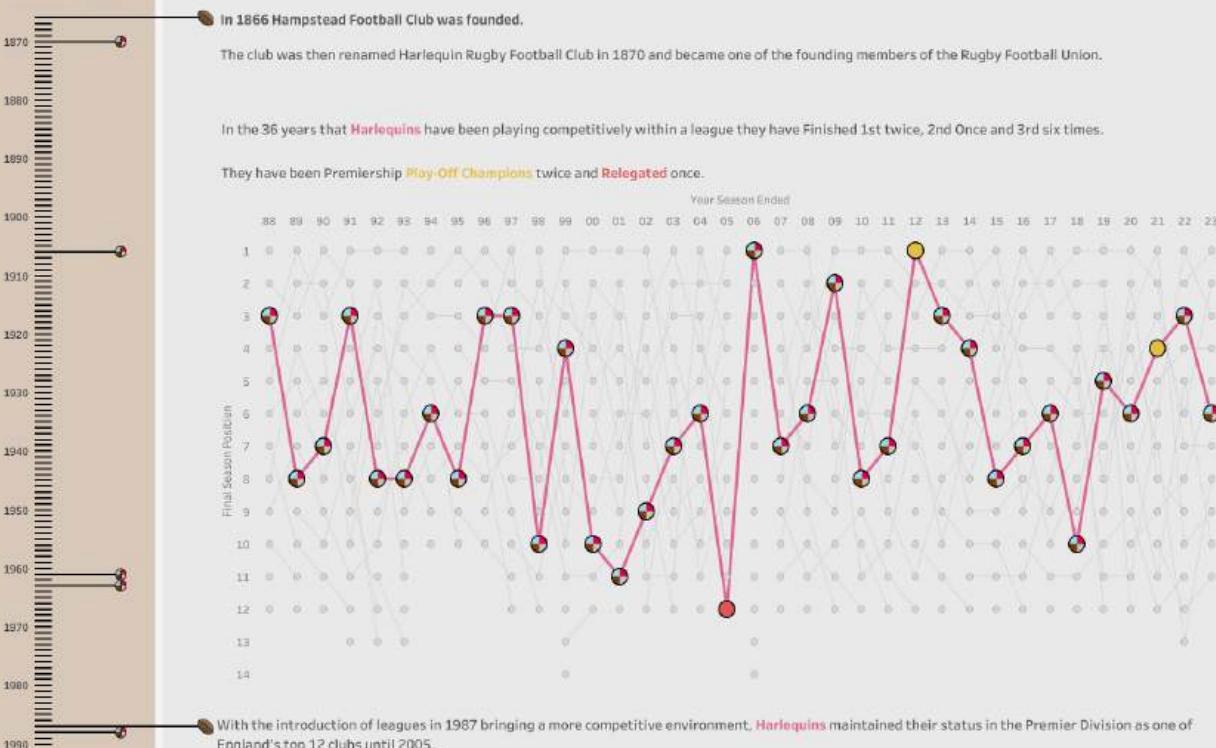
We will follow the time line of the club from where they began in 1866 through some of the key moments in the clubs history. We will see how they performed by seasons, their World Record breaking team and then some high and low points in their performances.

We will even look at the Premiership final I went to watch at Twickenham, what a day!



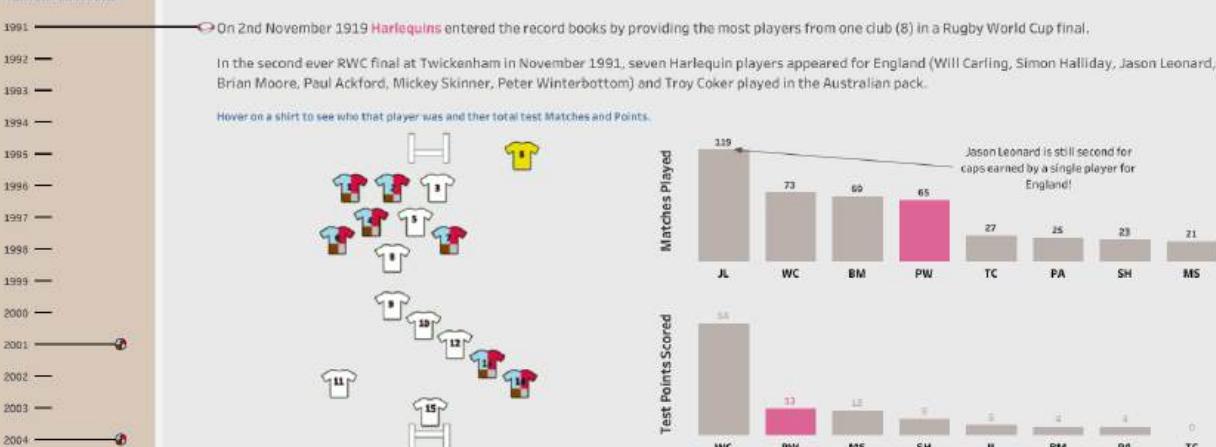
1866-1990

Hover over the Info Dots



1991-2004

Hover over the Info Dots



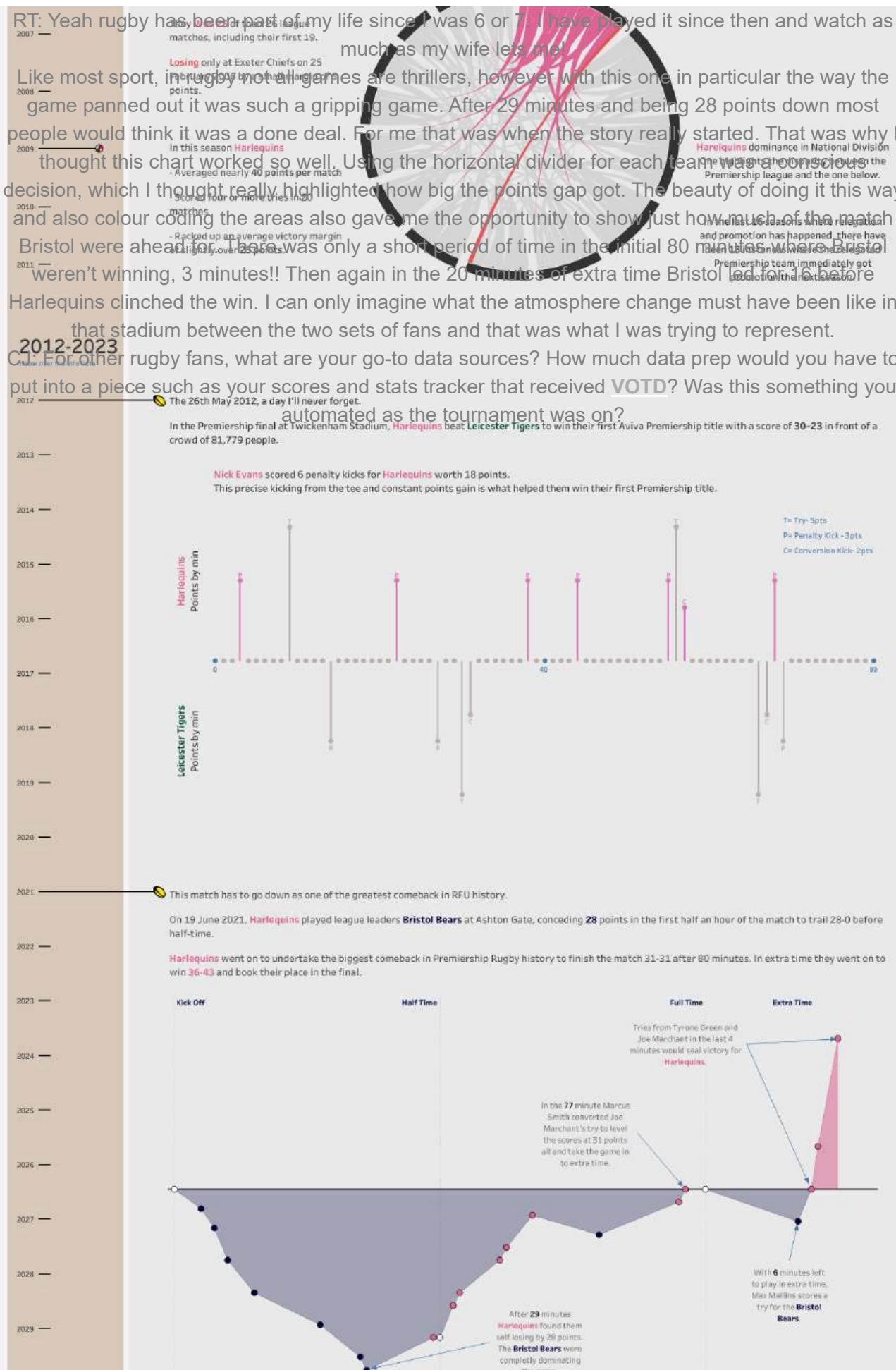
2005-2011

Hover over the Info Dots

Following their relegation from the Premiership in 2005, **Harlequins** utterly dominated National Division One in the next season.

Hover over the Chord chart to see what the results of the different games were and how the other teams fared.







Scores & Stats Tracker

Use the dashboard below to track all the fixtures, scores and key stats for this years rugby world cup. This is the 10th men's rugby world cup, which is taking place in France.

Can the defending champions South Africa win again, or will the home nation lift the William Webb Ellis trophy?

Click to select a match:

- [1](#)
- [2](#)
- [3](#)
- [4](#)
- [5](#)
- [6](#)
- [7](#)
- [8](#)
- [9](#)
- [10](#)
- [11](#)
- [12](#)
- [13](#)
- [14](#)
- [15](#)
- [16](#)
- [17](#)
- [18](#)
- [19](#)
- [20](#)
- [21](#)
- [22](#)
- [23](#)
- [24](#)

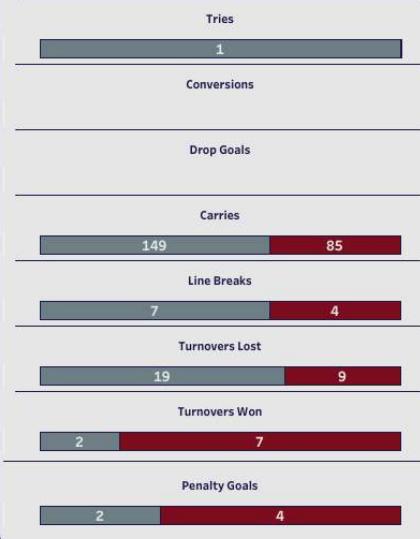
(Hover for Big Game Stats)

- [25](#)
- [26](#)
- [27](#)
- [28](#)
- [29](#)
- [30](#)
- [31](#)
- [32](#)
- [33](#)
- [34](#)
- [35](#)
- [36](#)
- [37](#)
- [38](#)
- [39](#)
- [40](#)
- [41](#)
- [42](#)
- [43](#)
- [44](#)
- [45](#)
- [46](#)
- [47](#)
- [48](#)

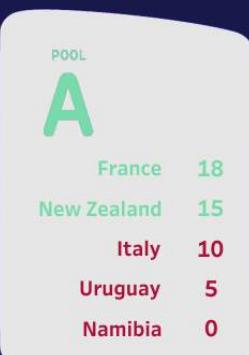
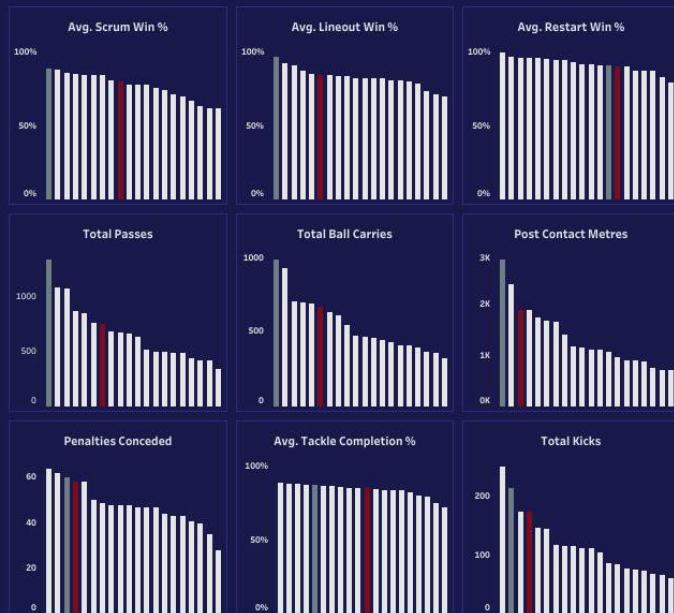
Selected
Played
Not Played

Match 48 Head to Head

New Zealand South Africa

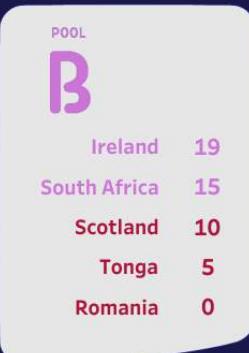


Key Team Stats up to Match 48



The Road to the Final

(After Match 48)



Quarter Final 4

France
28-29
South Africa

Quarter Final 2

Ireland
22-28
New Zealand

Semi Final 1

Argentina
6-44
New Zealand

Final

New Zealand
11-12
South Africa

Semi Final 2

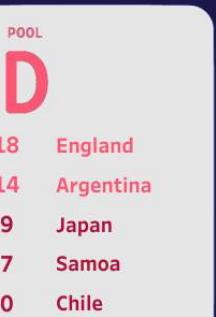
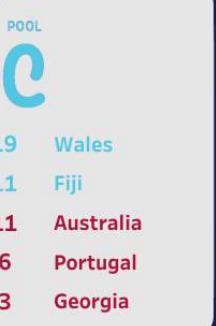
England
15-16
South Africa

Quarter Final 1

Wales
17-29
Argentina

Quarter Final 3

England
30-24
Fiji



RT: For me it can be anything from a properly collated data source that I find like rugby.statbunker.com to just websites that I come across that are based on tournaments or competitions. This was particularly the case with the World Cup one. I came across a website that had all the information that I needed and was almost updated in real time.

This for me is where the data prep comes in and my love of problem solving. My go to tool is Alteryx. From the first time I used it I was blown away with its capabilities and ease of use. Much like Tableau there is a great community out there where much cleverer people than me have done some amazing things. I used it to build out a workflow that would scrape all the different parts of the website I needed and **collate it all together** in a friendly Tableau building format. The initial build took a bit of time but it was on purpose so that it was future proof. The beauty of this was that as soon as each game ended I could run my workflow and update my dataset and refresh my dashboard. I think it only needed a few small tweaks as the tournament went on because of the initial time investment.

I watched a lot of the games and the data geek in me couldn't help running the workflow straight after to see what the stats showed and how it compared to how I saw the game.

Here are a few snippets of the workflow I used. The first part involves me getting the URL's for all the individual games. This was initially scraped from the landing page on the website and combined in to a list so I was future proofed for when new games happened.

Record	URL	Match ID
1	https://www.rugbypass.com/live/france-vs-new-zealand/stats/?g=102301	1
2	https://www.rugbypass.com/live/italy-vs-namibia/stats/?g=102302	2
3	https://www.rugbypass.com/live/ireland-vs-romania/stats/?g=102303	3
4	https://www.rugbypass.com/live/australia-vs-georgia/stats/?g=102304	4
5	https://www.rugbypass.com/live/argentina-vs-england/stats/?g=102305	5
6	https://www.rugbypass.com/live/scotland-vs-south-africa/stats/?g=102307	6
7	https://www.rugbypass.com/live/fiji-vs-wales/stats/?g=102308	7
8	https://www.rugbypass.com/live/chile-vs-japan/stats/?g=102306	8
9	https://www.rugbypass.com/live/france-vs-uruguay/stats/?g=102309	9

Once I had those I could put them through the download tool and start parsing out just the data I wanted to visualise. I did that mainly using the Parse tool and text to columns, then a few different calculations to focus on the number I wanted to get. In this case how many carries each player had in each game.

The screenshot shows a KWF UNL UND Gatherer interface. At the top, there's a configuration panel for a step named "Carries". It includes a preview of the extracted data, which contains fields like "Match ID", "Ascending RecordID", and "Ascending". Below this is a code editor with the following snippet:

```

Carries = IF LENGTH([Carries]>3 [field]
REPLACE([Carries]#&#39; ToNumber
[Row = 1:Carries]
"");
div class="name">"ELSE NU...
", div class="total...

```

Below the configuration is a "Results - Select (138) - Output" table with 144 records displayed. The columns are "Record", "Match ID", "Data Type", "Name", and "Qty". The data includes rows such as:

Record	Match ID	Data Type	Name	Qty
1 1	Carries		Besuden Barrett	24
2 1	Carries		Ardie Savea	20
3 1	Carries		Mark Telea	16
4 2	Carries		Tiaan Swanepoel	16
5 2	Carries		Tommesso Allan	15
6 2	Carries		Richard Hardwick	15
7 3	Carries		Joe McCarthy	20
8 3	Carries		Hugo Keenan	20
9 3	Carries		James Lowe	19
10 4	Carries		Tom Hooper	16
11 4	Carries		Will Siefken	15
12 4	Carries		Mark Nawaqanitawase	15
13 4	Turnovers Conceded		Manu Vatuvei	11

This was repeated for all the different metrics and then unioned back together with a Player details list
I also scraped to combine all the data in one file.

The screenshot shows a KWF UNL UND Gatherer interface. At the top, there's a configuration panel for a step. Below it is a code editor with the following snippet:

```

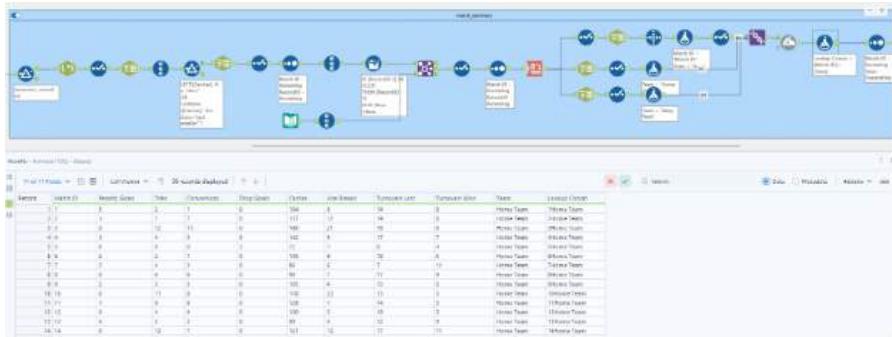
Match ID - Ascending Data Type - Ascending Team - Descending

```

Below the configuration is a "Results - Sort (192) - Output" table with 860 records displayed. The columns are "Record", "Match ID", "Data Type", "Name", "Qty", and "Team". The data includes rows such as:

Record	Match ID	Data Type	Name	Qty	Team
1 1	Carries		Ardie Savea	20	New Zealand
2 1	Carries		Beauden Barrett	24	New Zealand
3 1	Carries		Mark Telea	16	New Zealand
4 1	Clean Breaks		Mark Telea	3	New Zealand
5 1	Clean Breaks		Damian Penaud	2	France
6 1	Clean Breaks		Peato Mauvaka	2	France
7 1	Completed Tackles		Charles Ollivon	15	France
8 1	Completed Tackles		Gael Fickou	14	France
9 1	Completed Tackles		Gregory Alldritt	15	France
10 1	Dominant Tackles		Aaron Smith	3	New Zealand
11 1	Dominant Tackles		Will Jordan	2	New Zealand
12 1	Dominant Tackles		François Cros	2	France
13 1	Turnovers Conceded		Mark Telea	3	New Zealand

Finally I did something similar with the actual Match data to combine it all together in one easy to use file so I could plot the game info on the dashboard.



CJ: Rob has kindly shared the

workflow with us, and is downloadable at the top of the page from the git repository.

CJ: Is there anyone in the community that particularly inspires your work? Where do you look to for sports analytics inspiration?

RT: You mentioned #SportsVizSunday at the start, which I found early on and loved it. There are so many great people contributing there with their own different styles about lots of different sports. **Bo McCready's** Vizzes always looks like a piece of art which I love, **Mo Wootten's World Cup family tree** really made me want to reverse engineer it and work out how he did it and I also love **Ben Norland's** style of Vizzes. I've definitely recreated quite a few of the things he has done. His techniques are so unique.

Women's 2023 World Cup Ball
 coloured by progress through the tournament (darker colours show further progress)
 top country in each group finished first; bottom country finished last
 click any part of the diagram to highlight that country's path through the tournament
 highlight colours are reflective of primary home kit colours.



design | @mo_wooten
 radial family tree template | Brian Moore
 #SportsVizSunday | IronQuest

Outside of sport I love movies too so **Zach Bowders** Vizzes are right up my street. I've also spent a lot of time looking at #IronViz champion **Chris Westlakes** stuff and also **Ant Pulley**, I still don't know how he makes some of his dashboards do what they do!

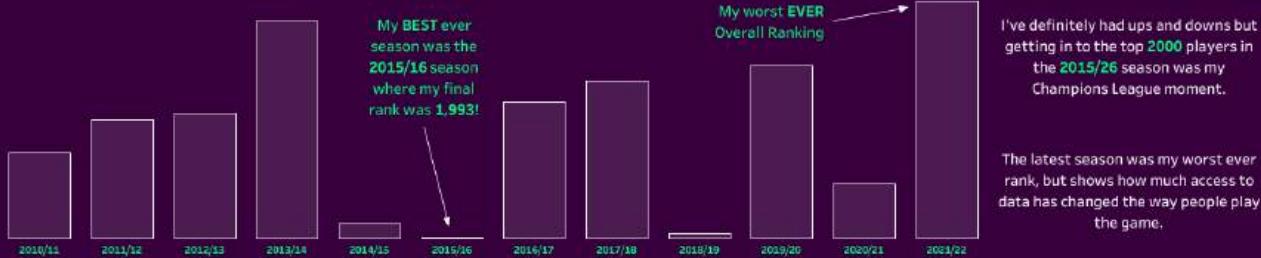
There are so many great people out there though I find it hard to narrow it down. I actually made a Christmas advent calendar last year too. That was a little nod to a lot of the people who had inspired me to get better.

CJ: You have two wonderful **FPL** visuals, one on **draft selections** another based on your own data. For someone looking to do their own analysis what are the best steps to being able to create analysis to the detail level you've gone through?



Fantasy Premier League

I started playing the fantasy premier league game in the 2010/11 season to try and beat my best man at his own game, from that point I was hooked. I loved the data and analysis you needed for it and more recently the joy I have got from using Alteryx and Tableau to dig deeper.



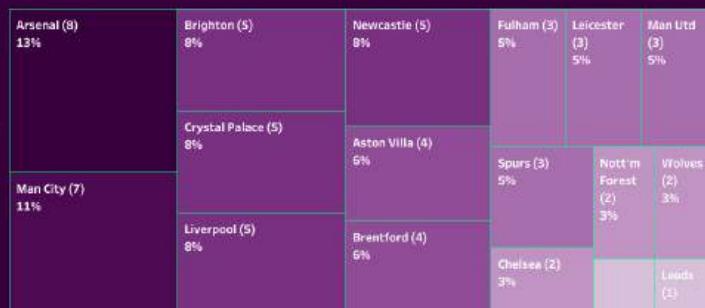
Through all the seasons I have played, I have managed to top my two main mini leagues 9 times.



So how did I get on this season?



I used 17 different teams throughout the season whilst mixing up the Arsenal and Manchester City players regularly.



The top 5 players by appearance



First Wildcard Played

Triple Captain Played

Second Wildcard Played



Here is a breakdown of how I scored my points over the gameweeks.

RT: Yeah FPL is a great competition. I only really got into it to try and beat my really keen football friends. The beauty of that was that I wasn't swayed by any underlying loyalties and was happy picking players, as sometimes wouldn't purely on what the data was telling me. It's a great resource for data. They have huge amounts of data and a load of API's to go with it too. This makes it super easy to pull your own data. On top of that there are Alex's workflows already built on the community for you to use with just a few simple changes.

Once you have all that data the hardest bit is trying to decide what stats you want to use. I also did a viz on how our current draft league was going, but it was a bit biased to be honest and I built it to make myself look better than I was actually doing at the time, mostly to wind up my mates.

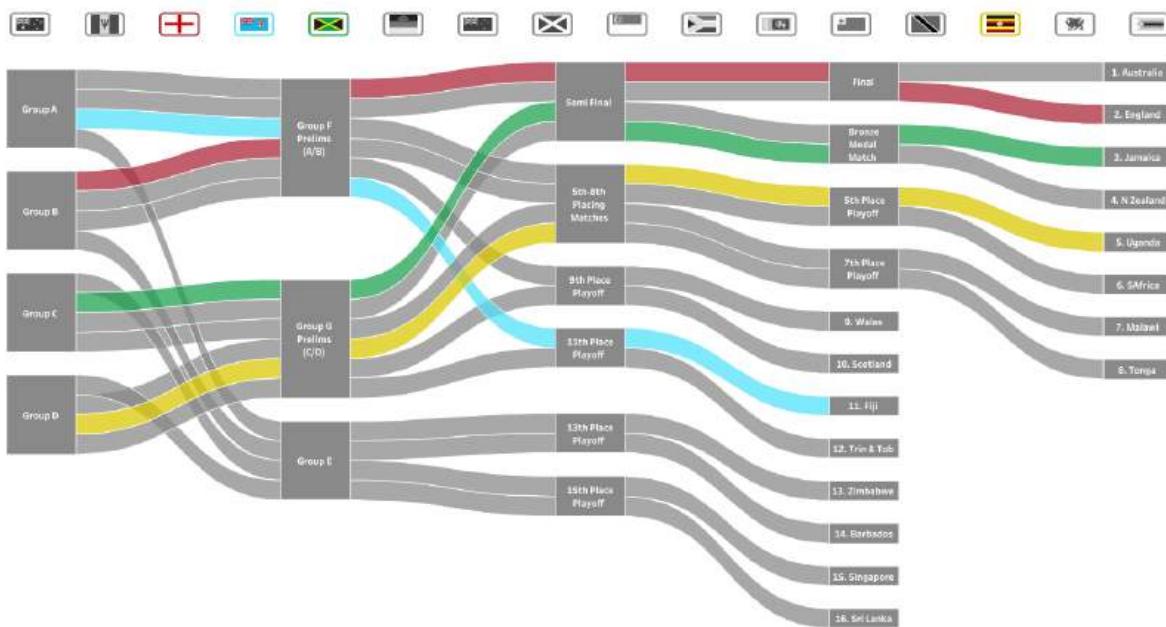


The 2023 Netball World Cup was the sixteenth staging of the Netball World Cup, the premier competition in international netball, contested every four years. The tournament was held from 26 July to 6 August at the International Convention Centre in Cape Town, South Africa. Its first time in Africa. This marked the tournament's 50th anniversary, as it was founded in 1963.

Sixteen nations contested for the title and after two group stage rounds, England, Jamaica, Australia and New Zealand all made it to the semi-finals, with England qualifying for the final for the first time. Australia would take home their 12th title after defeating England 63-45. Jamaica finished in third place, after defeating New Zealand.

Host nation South Africa finished sixth overall.

Below we look at how the 16 countries progressed through the tournament, from the group stages on the left through to their final placing on the right. Click on the country flags to highlight them on the chart.



Credits & Datasources:

LADATAVIZ
"Sankey Template Creator"

SPORTSVIZSUNDAY
"September Challenge & Data"

WIKIPEDIA
"Info and Additional Data"

RT: Thanks, yeah I was particularly pleased with how that one turned out and the challenge was a fun way to use a dataset I probably wouldn't have picked.

Part of the design was me wanting to use two things I hadn't done before; multi select parameters and a Sankey. **Tristan Guillevin's** amazing tool helped me get the basic template without having to do all the painful maths. From there I then added a load of functionality with the aim of allowing the users to select any countries they were interested in. The grey scale/colour selected for each country would then really make their progress through the tournament pop out to the viewer. The choice of flags as the selection tool felt almost natural but also made it easy to get them all above the Sankey in a nice uniformed format, keeping my inner OCD happy.

CJ round-up:

So pleased to have Rob join this week. I feel Rob's portfolio has such breadth in design, sports and also tools to help with the data transformation. I hope that passion has come across in todays conversation and once again thank you to Rob for sharing the finer details around creating an automated solution for the rugby world cup data.

LOGGING OFF,
CJ

BUILDING FLAPPY BIRD THROUGH PROMPTS (PYTHON)

Hi all,

I am the first to admit I am not the best coder out there. Truth be told the only satisfaction I get from coding is when the run button actually works after countless errors. But the process of getting to that stage can sometimes be quite fun.

A colleague at work specifically sent me **this article**.

*"Based on the provided search results, it appears that two new AI chatbot models named "im-a-good-gpt2-chatbot" and "im-also-a-good-gpt2-chatbot" have recently emerged on the LMSYS Chatbot Arena benchmarking site, sparking significant speculation and interest within the AI community."....
"The models excel at coding tasks, with examples of generating code for games like Snake and Flappy Bird with minimal prompting"*

Now chatGPT and the equivalents have been knocking around for a bit now and seemingly are only getting better at knowing what I mean when I frantically ask it to amend the code for me. So, I thought why not set myself a challenge. Can I create my own version of flappy bird simply through asking it to rewrite my script... over and over and over again?

So by simply asking it to “create the flappybird game” in python it gave me some clear indication of width, height, different images that will need loading and how it leverages the pygame package.

```
import pygame
import random

# Global Constants
WIDTH = 288
HEIGHT = 512
FPS = 60
GROUND_HEIGHT = 80

# Colors
WHITE = (255, 255, 255)
BLACK = (0, 0, 0)

pygame.init()
pygame.mixer.init()
screen = pygame.display.set_mode((WIDTH, HEIGHT))
pygame.display.set_caption("Flappy Bird")
clock = pygame.time.Clock()

# Load game assets
background_img = pygame.image.load("background.png").convert()
bird_img = pygame.image.load("bird.png").convert()
pipe_img = pygame.image.load("pipe.png").convert()

# Scale game assets
background_img = pygame.transform.scale(background_img, (WIDTH, HEIGHT))
bird_img = pygame.transform.scale(bird_img, (56, 40))
pipe_img = pygame.transform.scale(pipe_img, (80, 320))
```

Followed by some functions on how to create the pipes, what happens when the bird collides with the object – and generally how things render on the page.

```

def draw_text(surf, text, size, x, y):
    font = pygame.font.Font(None, size)
    text_surface = font.render(text, True, WHITE)
    text_rect = text_surface.get_rect()
    text_rect.midtop = (x, y)
    surf.blit(text_surface, text_rect)

def draw_background():
    screen.blit(background_img, (0, 0))

def draw_bird(bird_y):
    screen.blit(bird_img, (50, bird_y))

def draw_pipes(pipes):
    for pipe in pipes:
        screen.blit(pipe_img, (pipe[0], pipe[1]))
        screen.blit(pygame.transform.flip(pipe_img, False, True), (pipe[0], pipe[1] - 420))

def is_collision(bird_x, bird_y, pipes):
    if bird_y < 0 or bird_y + 40 > HEIGHT - GROUND_HEIGHT:
        return True

    for pipe in pipes:
        if bird_x + 50 > pipe[0] and bird_x < pipe[0] + 80:
            if bird_y < pipe[1] + 320 or bird_y + 40 > pipe[1] + 420:
                return True

```

When I first gave it a go a few things stood out:

- I needed to design the images for the background, “bird” and pipes like in the game
- I would need to make sure the pipes dont overlap one another
- I needed to stop the bird falling through the floor off the screen.
- It would be nice to add some audio to the game too (I ended up cloning these from an online repository I saw.)

So let's jump ahead into the main code and explain what it does and things that needed amending as I went on with the prompts to create my own version of Flappy bird called “Flying Logo”



You can see a gif of it working below.



The code is an implementation of the Flappy Bird game using the Pygame library in Python. You can find it in the repository.

```
import pygame, random, time
from pygame.locals import *

# VARIABLES
SCREEN_WIDTH = 400
SCREEN_HEIGHT = 600
SPEED = 20
GRAVITY = 2.5
GAME_SPEED = 15

GROUND_WIDTH = 2 * SCREEN_WIDTH
GROUND_HEIGHT = 100

PIPE_WIDTH = 80
PIPE_HEIGHT = 500

PIPE_GAP = 150

flying = 'files/sound/flying.wav'
hit = 'files/sound/hit.wav'

pygame.mixer.init()

class logo(pygame.sprite.Sprite):

    def __init__(self):
        pygame.sprite.Sprite.__init__(self)

        self.images = [pygame.image.load('files/images/up.png').convert_alpha(),
                      pygame.image.load('files/images/mid.png').convert_alpha(),
                      pygame.image.load('files/images/down.png').convert_alpha()]

        self.speed = SPEED

        self.current_image = 0
        self.image = pygame.image.load('files/images/up.png').convert_alpha()
        self.mask = pygame.mask.from_surface(self.image)

        self.rect = self.image.get_rect()
        self.rect[0] = SCREEN_WIDTH / 6
        self.rect[1] = SCREEN_HEIGHT / 2

    def update(self):
        self.current_image = (self.current_image + 1) % 3
        self.image = self.images[self.current_image]
        self.speed += GRAVITY

        # UPDATE HEIGHT
        self.rect[1] += self.speed

    def bump(self):
        self.speed = -SPEED

    def begin(self):
        self.current_image = (self.current_image + 1) % 3
        self.image = self.images[self.current_image]

class Pipe(pygame.sprite.Sprite):

    def __init__(self, inverted, xpos, ysize):
        pygame.sprite.Sprite.__init__(self)

        self.image = pygame.image.load('files/images/pipe.png').convert_alpha()
        self.image = pygame.transform.scale(self.image, (PIPE_WIDTH, PIPE_HEIGHT))

        self.rect = self.image.get_rect()
        self.rect[0] = xpos

        if inverted:
            self.image = pygame.transform.flip(self.image, False, True)
            self.rect[1] = - (self.rect[3] - ysize)
        else:
            self.rect[1] = SCREEN_HEIGHT - ysize

        self.mask = pygame.mask.from_surface(self.image)
```

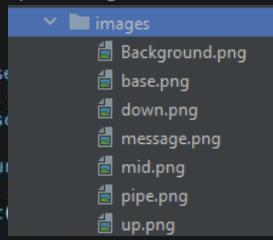
```
1. def update(self):
    self.rect[0] -= GAME_SPEED
```

Here's an explanation:

The code sets up the screen size, game variables, and loads the game assets such as images and sounds.

These needed playing around with depending on the screen size and what you want to include.

```
class Ground(pygame.sprite.Sprite):
```



convert_alpha()
IDTH, GROUND_HEIGHT))

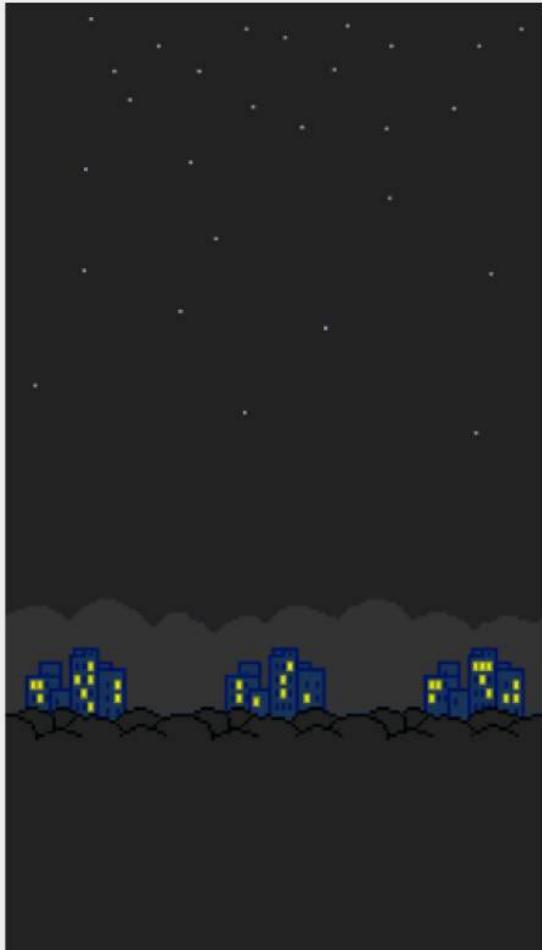
Only 7 images are needed. Even then its overkill as I create a "Wobble" effect to the logo for when it

```
def update(self):
    self.rect[0] -= GAME_SPEED
```

is in flight. (up mid & down)

Below are some screenshots of those files in Figma. You'll find them in the repository. You'll see they mostly look like the original game in terms of the pipe and background. I amended a few cosmetics and the colors to suit my branding above – but architecturally reflect the original game. Most

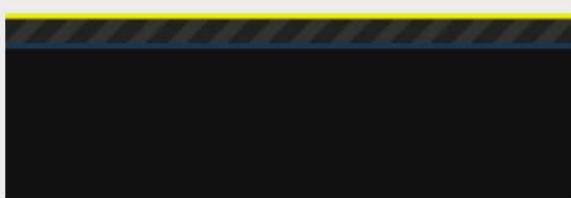
these files you can find on the internet with ease.



Flying



Logo



```
1. ground_group.remove(ground_group.sprites()[-1])
```

The logo class represents the bird (logo) character. It has attributes like speed, current image, and position on the screen. The update method is responsible for updating the position and image of the bird. The bump method is used to make the bird jump. The begin method is used to animate the bird at the beginning of the game.

2. The `Pipe` class represents the pipes that the bird needs to navigate through. It has attributes like position and size. The `update` method is responsible for updating the position of the pipes. This is exactly how it is seen in the original game, except the original game uses green pipes instead.
3. The `Ground` class represents the ground in the game. It has attributes like position and size. The `update` method is responsible for updating the `ground_group.draw(screen)` position of the ground.
4. The `is_off_screen` function checks if a sprite is off the screen.

The ground and screen are important to remember because we want everything to visually be formatted in respect to one another as well as not letting the logo, or bird, fall off the screen!

1. `clock.tick(15) get_random_pipes` function generates a random set of pipes (top and bottom) at a given x position.

When using the prompts originally, I found the pipes the hardest bit to sort out. I was either loading in

- ```
if event.type == KEYDOWN:
 if event.key == K_SPACE or event.key == K_UP:
 pipe_group.update()
 pipe_group.add(pipes)
 pygame.mixer.music.play()
```
1. The first part of `if event.type == KEYDOWN:` is the beginning state of the game where the player needs to press space or up arrow key to start the game. It displays the beginning image and waits for the player's input.
  2. Once the game starts, the `while loop` of the game loop runs. It handles the main game logic. It checks for collision between the bird and the ground or pipes. It updates the `pygame.sprite.Sprite`s `update`. It also checks for off-screen pipes and ground and removes them, creating new ones `pygame.mixer.music.play()`. The game loop continues until a collision occurs.
  3. Finally, the game loop breaks out of the loop once a collision is detected, and a sound effect is played briefly.

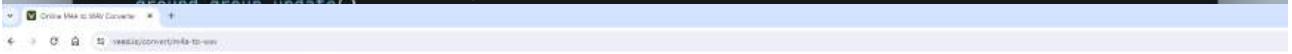
The sound effects I took out a repository I found online – I thought it added a nice element to the game.

I even went as far as to think how can you amend the sound to a custom sound. I was going to make it be my own voice saying "Ha ha You Lose" but thought against it in the end. If you do want to amend

```
new_ground = Ground(GROUND_WIDTH - 20)
ground_group.add(new_ground)

if is_off_screen(ground_group.sprites()[0]):
 ground_group.remove(ground_group.sprites()[0])
```

There are two files. The first is `flying.wav` – so each time you click the up arrow or space bar it makes the sound similar to if you were to click on your phone. The second is the hit sound that plays when you hit into the pipe or floor.



Those audio files you will see there are two in the repository.

```
pipe_group.update()
pipe_group.add(pipes[1])
logo_group.update()
ground_group.update()
```

## M4A to WAV

Convert M4A to WAV and other audio files to WAV online, free

Choose M4A File



If you do want to change the sounds I would recommend the VEED.IO which can help you convert different sound files into WAV. (Especially as iphone recordings come in M4A.)

So how did I find the prompts?

Well if we compare the first set of code when I asked the original question to the final output there are 4 main differences.

1. Using classes instead of functions (new for me I rarely use them)
2. We contain the different files into folders for images and sound, the original had no audio elements.
3. We add in a basic wobble effect for the logo when its going up or down.
4. The first code would start and end too rapidly with no welcome message or ending.

If you'd like to give it a go yourself, just download the files and hit run. Of course change out the logo for your own animal / item / and amend the background as you see fit.

LOGGING OFF,

CJ

## OPEN-HIGH-LOW-CLOSE CHARTS

Hi all,

One for the “Finance Bros” this week. Only joking, but we will look to create different variations on the open high low close charts that are very popular, especially when looking at stock market variations.

### Variations of Open High Low Close Charts

Axis are truncated for visual effect.



These charts are all fairly simple to make – so will leave technical run-throughs for the more complex chart types.

This piece will be mainly a written chat around the different chart types and the design, as with all my workbooks – they workbooks are downloadable\* so they should be easy enough to reverse! (*\*with the exception of paid work examples*) Just click the tableau button under the header to be directed to the free visual.

Before we start please excuse the truncated axis throughout, it's simply so you can see the clear chart style throughout.

First up, if you want to create any type of stock or market type graph, google finance is your best friend. For these examples I quickly just took the example

# GOOGLEFINANCE

Fetches current or historical securities information from Google Finance.

## Sample usage

```
GOOGLEFINANCE("NASDAQ:GOOG", "price", DATE(2014,1,1), DATE(2014,12,31),
"DAILY")

GOOGLEFINANCE("NASDAQ:GOOG","price",TODAY()-30,TODAY())

GOOGLEFINANCE(A2,A3)
```

The documentation for this you can find [here](#).

You will want to amend the calculation where it says price with the open, close, high and low text.

There are four variations we will look to chat about today, as seen above.

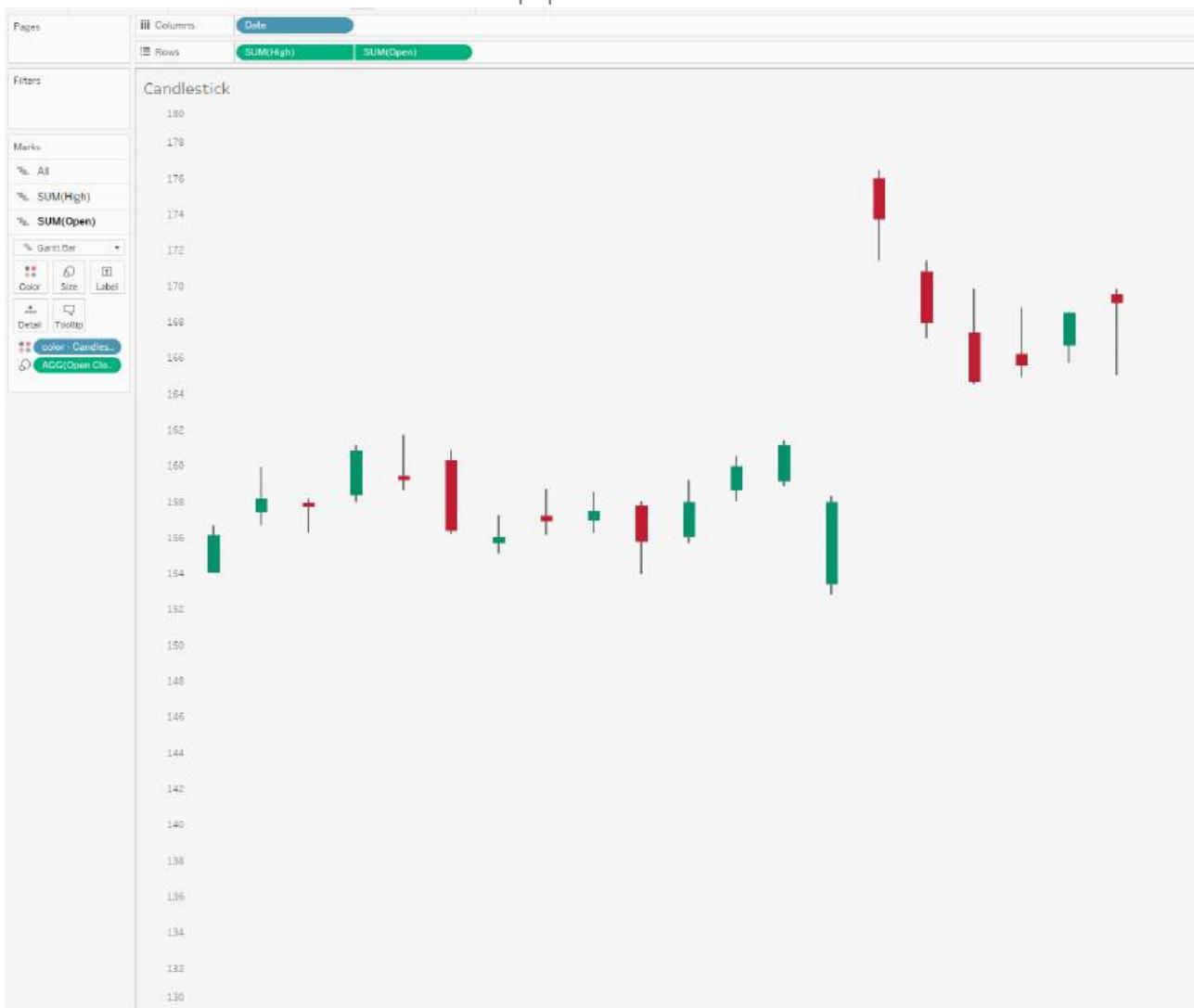
Top Left is the candlestick chart.

Top right is a chart that looks at open and close with a running average line.

Bottom left is a anchor chart.

Bottom right I couldn't find an official name for so I am just going to call it the standard OHLC chart.

Lets start with the most popular of these the candlestick chart.



So I will start by saying this is my favourite variation in Tableau to build. The reason being is, we don't need any "shapes". The chart above will show the high and low for the stock through the use of the

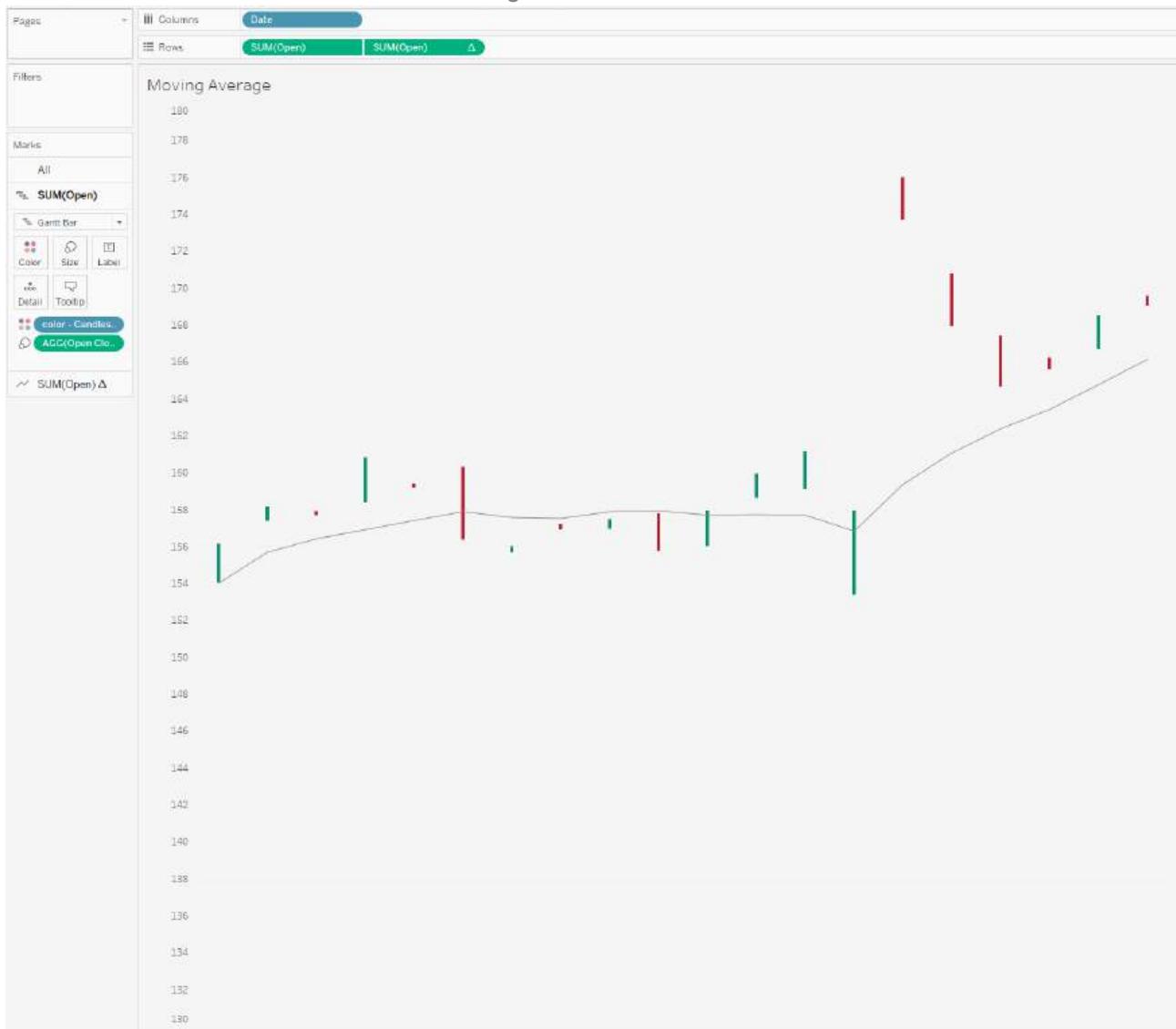
grey line. Just think min and max during a day.

The red or green gantt overlay is looking at whether it is falling or rising between its opening and close price. (A stock will change consistently in price over the course of a day remember)  
For me this is the clearest way of showing fluctuations or swings throughout a day as well as seeing the net positive or negative change each day.

Only a few calculations are needed due to the use of the gantt marks. Two spread calculations one for the difference between the high and low, and another calculation for the spread between open and close. Then it is a simple dual axis with some amendments to size/

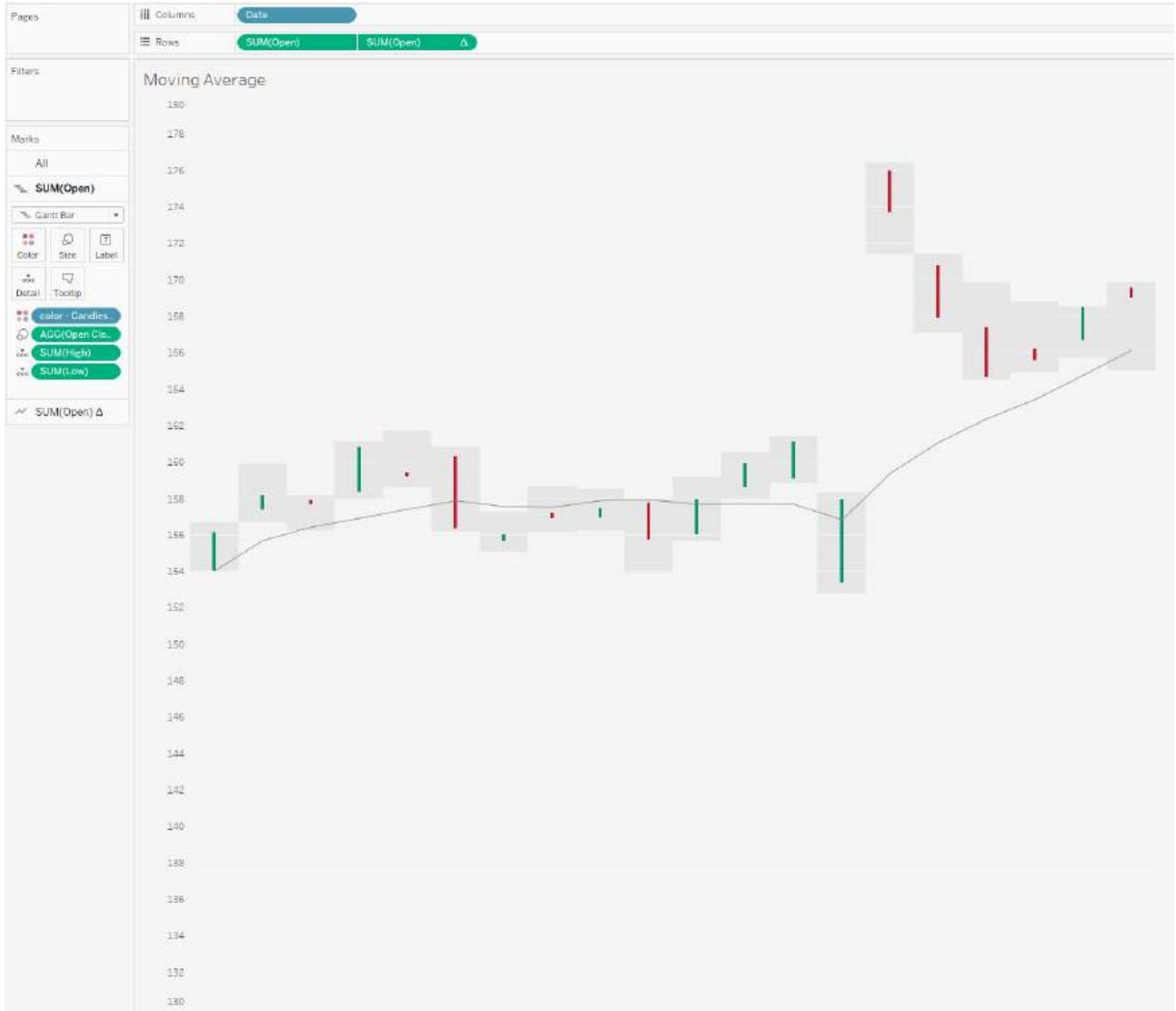
Which brings me onto example 2.

Adding in a reference line.

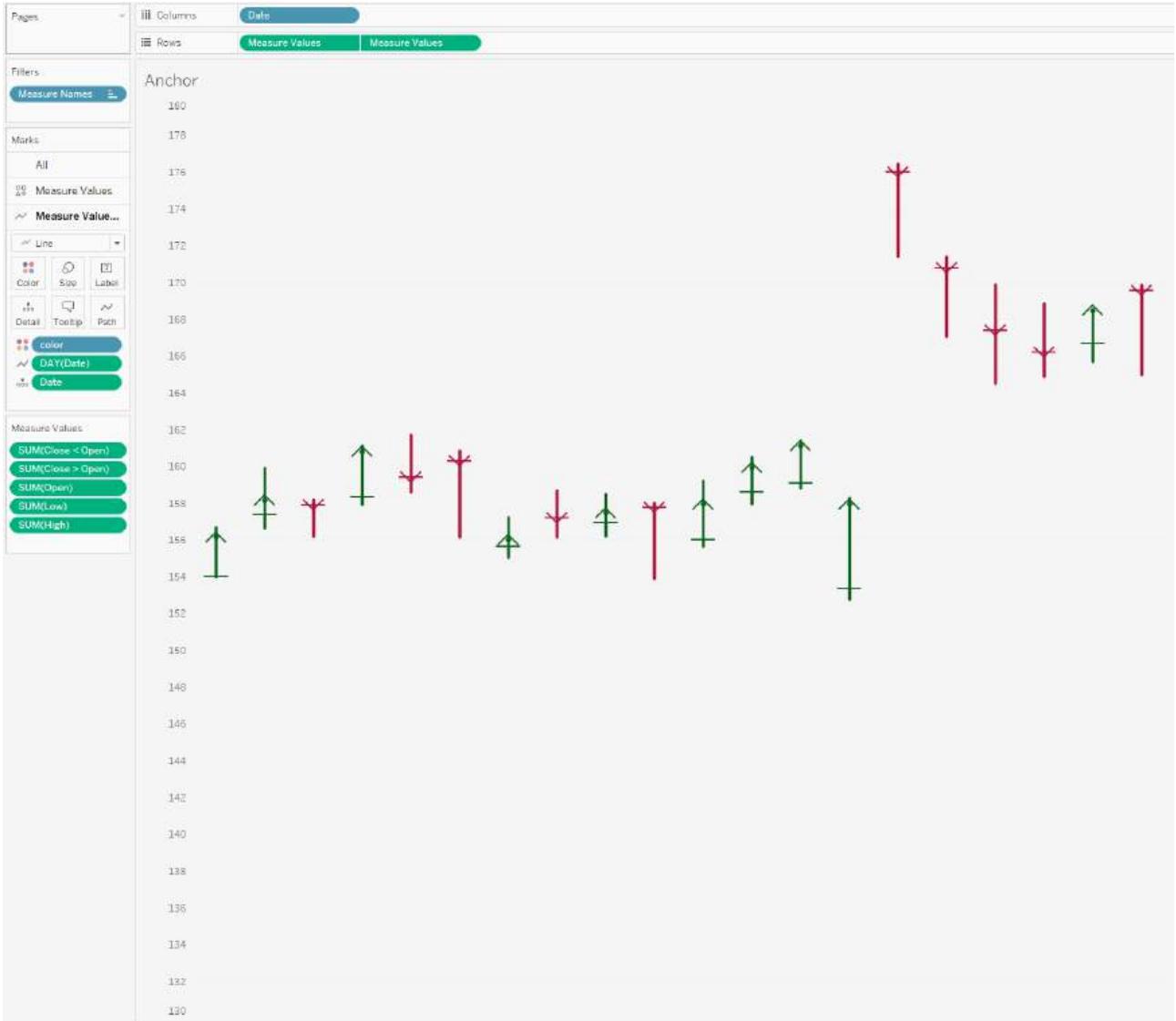


If you download the workbook you will see I put in a rolling average of 7 days (usually I suspect you would do this over a longer period but I only am showing a few days of data as it is)  
I quite like this additional touch to show general trend which you don't get with the previous chart, but to make it as simple as possible it meant that I had to remove one of my dual axis from the previous chart so can't also overlay the low & high of the day. Trying to keep this as simple as possible so map layers weren't an option.

Theoretically you could add this context back in with a reference band but I'm not sure if its overkill or not. I'll let you be the decider.



Moving on in order of favourite ways of designing these charts we have the Anchor chart. I don't see too many of these on Tableau, probably because of the popularity of the candlestick chart.



Each symbol or ‘anchor’ represents a trading session.

The vertical line through the middle represents the highest price from the top (or starting) position, while the bottom (or end) represents the lowest price.

The short, horizontal line marker indicates the open price, while the arrowhead marks the close price. I think the good thing going for it is the love for arrows. Arrows give clear direction and signage, and you theoretically wouldn't need the color to know what is going on (even though it does help!)

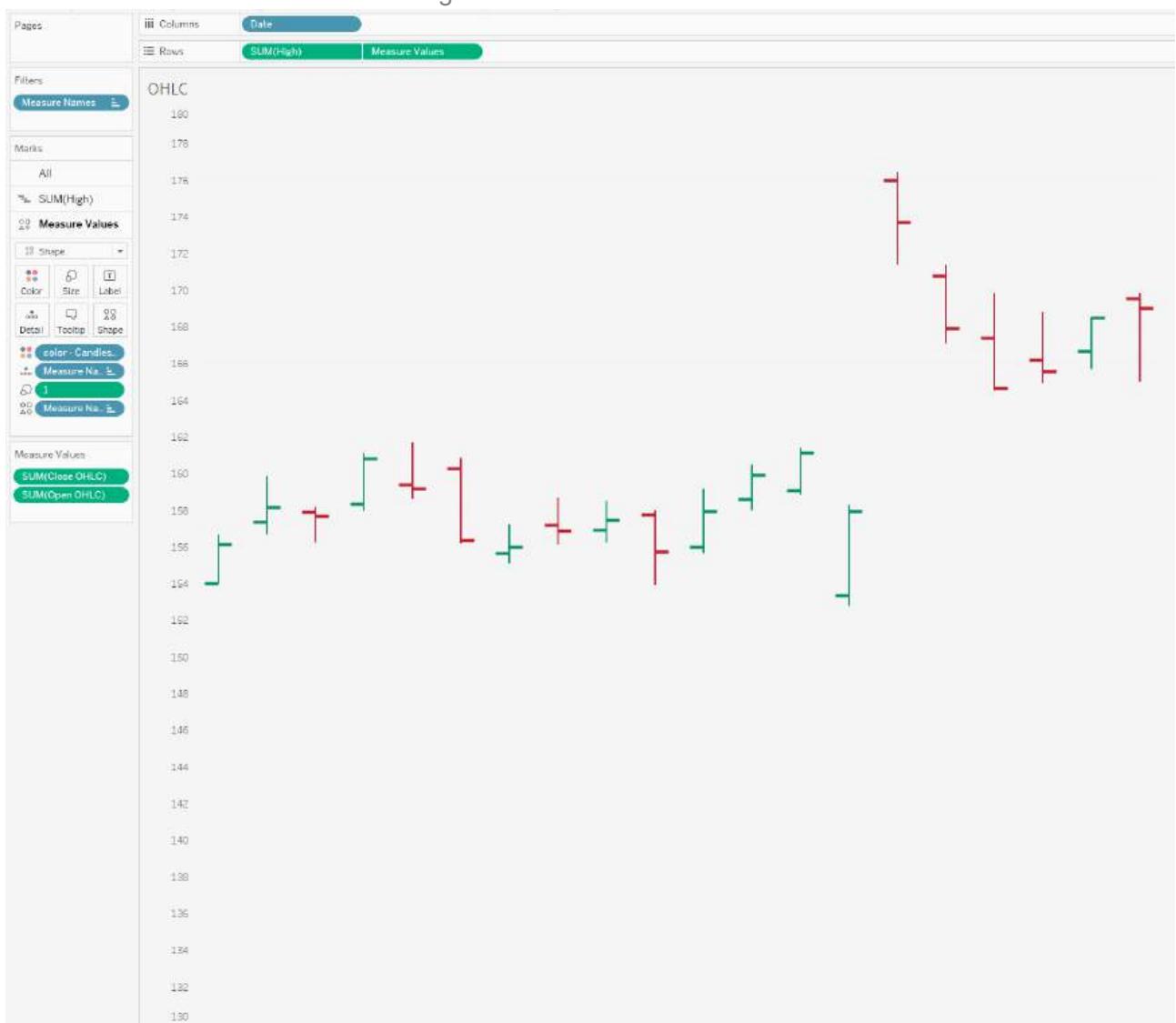
The only real negative with this chart is probably the way I built it for simplicity sake. I think it would be better to have more flexibility in being able to edit the arrow heads (currently they are shapes and I just don't think they come out well, both in size shape and definition), same with the line for the opening price. I think if we were to redesign this in a more technical way we could get some better results using actual lines.

Now the last example I came across from the Data Viz Catalog. You can read about the chart [here](#).

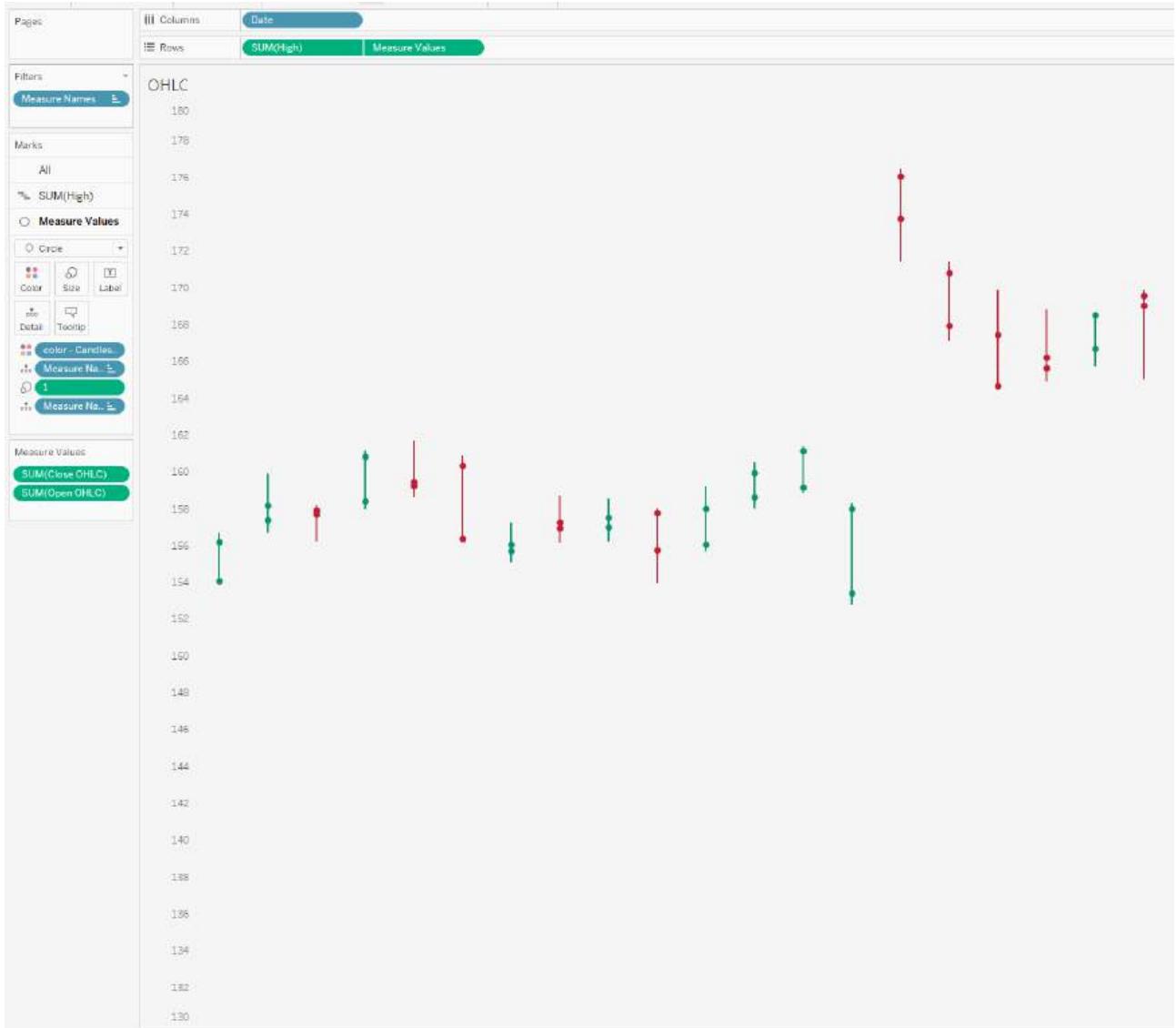
## Open-High-Low-Close Chart



On each single time period, an OHLC Chart plots a symbol that represents two ranges: the highest and lowest prices traded, and also the opening and closing price on that single time period (for example in a day). On the range symbol, the high and low price ranges are represented by the length of the main vertical line. The open and close prices are represented by the vertical positioning of tick-marks that appear on the left (representing the open price) and on right (representing the close price) sides of the high-low vertical line. Here is it in Tableau.



To be able to create this chart, again you need two shape files slightly offset from the middle to be able to appear to the right and the left of the chart based on opening and close. See if I change the mark back to a circle you'll notice that the marks aren't *actually* offset, it is the custom shape file.



Again whilst I like the simplicity to this chart – the idea of having to use a shape file ruin it for me a little.

So there are 4 ways that you can create custom finance charts geared towards stock prices. Let me know what variation you think works best, and whether there are any other alternative chart types important to take into consideration.

Speak soon,

LOGGING OFF,

CJ

TOP 10 ALTERYX TIPS WITH MEDHA KHURANA

Hi all,

The site has grown in recent years to explore different tools in the data world. One of which I have started to enjoy more and more during my time at JLL and that is Alteryx.

This week, Medha Khurana joins the site to talk about some of top Alteryx tips. Medha recently completed her Advanced exam and also has kickstarted her own site

<https://www.medhakhurana.com/>

I love when people make that next jump into sharing more through different means, so to see the site come alive is very exciting.

CJ: Medha, Thanks for joining the blog. For those that are unaware, can you tell us a little about your journey to date and what excites you about the world of data analytics?

The grid displays six blog posts from the Alteryx website:

- Top Left:** **Alteryx Advanced Certification: Preparation and Tips** (Tutorial). Description: If you want to get started in Alteryx and pass the core exam, you can read about it here... Date: Apr 5, 2024 - 3 min read.
- Top Middle:** **Book Review: Storyworthy** (Blog). Description: What is it about? Storyworthy by Matthew Dicks is a book that teaches the art of storytelling. It's filled... Date: Mar 31, 2024 - 3 min read.
- Top Right:** **Find the street crime in your area: Macros and APIs in Alteryx** (Tutorial). Description: Living in London, a city that's a dream destination for many, is an incredible experience. Its beauty am... Date: Mar 31, 2024 - 5 min read.
- Bottom Left:** **From Hypothesis to Conclusions: A Deep Dive into T-Tests for Data-Driven Analysis** (Tutorial). Description: In our journey as analysts, our goal is to harness the power of data to bring intelligence to the decision... Date: Mar 9, 2024 - 5 min read.
- Bottom Middle:** **The Statistical Frontier of the Global AI Race Through Hypothesis Testing** (Tutorial). Description: In today's hyper-connected world, the race for AI dominance is heating up like never before. In thi... Date: Mar 2, 2024 - 9 min read.
- Bottom Right:** **Optimize Your Alteryx Experience: Leveraging Automated Alerts for Workflow Failures** (Tutorial). Description: Welcome to the edition 'Optimize Your Alteryx Experience', crafted for professionals like you who... Date: Feb 18, 2024 - 2 min read.

Hello CJ, thank you for the invitation to your guest blog. It's truly an honour to be featured on your website.

I'm a data professional based in the city of London, originally from Delhi, India. My introduction into the world of data began during my undergraduate studies, where I was introduced to the basic concepts within data analytics and management. Eager to delve deeper, I embarked on a journey of self-learning through online courses, books, and personal projects. However, my career initially led me to a role as a Software Engineer, which I soon discovered wasn't my true passion. To align my career with my interests, I pursued a Master's degree in Business Analytics from Warwick Business School. It exposed me to advanced analytics techniques, business statistics, text analytics and I did several collaborative projects and made me fall in love with data analytics.

Following studies, I joined a financial consulting firm in their Data Analytics division. I worked on a variety of projects—from data visualization to building data pipelines, before taking on a role as a Senior Business Intelligence Analyst at JLL.

What excites me the most about being part of this industry is that it never stands still, innovation and learning are constant companions here. Both of these elements are a perfect blend of what I seek in life.

CJ: You recently started your own site <https://www.medhakhurana.com/>. I love the grayscale theme to it. What prompted you to start and what can we expect of the site in 2024?

Thank you so much. I actually started it on my birthday, as a gift to myself! It took me approximately three weeks to set it up, from selecting the hosting platform to crafting its design and filling its pages with some insightful content.

Teaching is practically in my DNA. Coming from a family of teachers, sharing knowledge brings me immense satisfaction. It's a reciprocal process: it strengthens my understanding while empowering

others to start their own learning journeys. When I learn something new, I like to make it simple and share it so others can learn too. It's a philosophy I live by: to truly master something, one must teach it.

Secondly, I've long drawn inspiration from remarkable members of the data community, such as yourself who have built a brand of their own by sharing what they are passionate about with others. I wanted to do that too – to cultivate my own brand, be a source of inspiration and help others along the way.

You'll find a variety of content there. I'll keep up with my Alteryx Optimization series, sharing helpful tips. There will be more tutorials on how to use advanced Alteryx tools for interesting projects. Plus, I'll dive into topics like statistics, data visualization, or anything else I find fascinating. The goal is to keep sharing knowledge and helping each other learn!

I'm also aiming to cover off personal reflections on book reviews, and insights into student career paths.

CJ: You recently passed the Alteryx advanced exams. What did you find most challenging as part of it? Are there certain tools that you found important to revisit beforehand?

Despite using Alteryx daily in my work and solving weekly challenges regularly, I found the practical questions in the exam quite demanding to complete within the allotted time. Perhaps it was the pressure of the exam itself. Nonetheless, I recommend revisiting advanced data preparation and transformation tools like Multi-Field Formula, Multi-Row Formula, and Regular Expressions. Additionally, it's beneficial to brush up on less commonly used tools such as 'Block Until Done' and 'Run Command' to refresh your understanding of their functionalities.

|                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                              |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <br>Alteryx Designer Core<br><br>2.5 Hours • Free<br>Beginner exam with 80 questions<br><br><a href="#">Prep Guide</a> <a href="#">Sign In</a> | <br>Alteryx Designer Advanced<br><br>2.5 Hours • Free<br>Advanced exam with 51 questions<br><br><a href="#">Prep Guide</a> <a href="#">Sign In</a> | <br>Alteryx Designer Expert<br><br>3 Hours • USD 150 per exam<br>Expert exam with 7 questions<br><br><a href="#">Prep Guide</a> <a href="#">Sign In</a> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

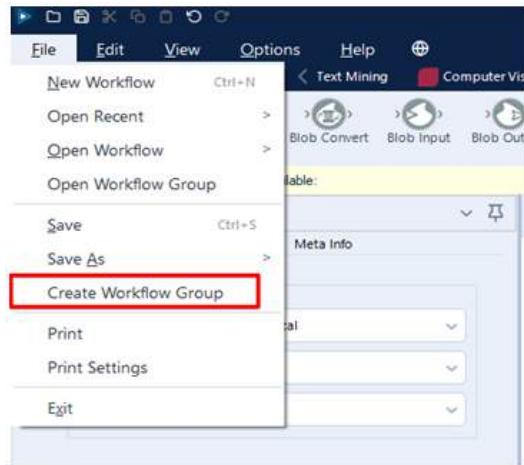
I've dedicated a [blog](#) to Alteryx advanced exam preparation and tips, which you can find on my website for those interested.

CJ: What are your 10 top tips for Alteryx users, from a technical standpoint? Is this something you can share a copy with us?

Absolutely, I love using the cool hacks in Alteryx to work in a more efficient way. Here are top 10 tricks that I use most often:

#### TIP 1: Create Workflow Groups

Creating workflow groups enables you to open multiple workflows at once. For example, if you're working on three separate workflows and want to pick up where you left off the next day, you can create a workflow group. Simply go to File and select 'Create Workflow Group'. This means that instead of opening each workflow individually, you can simply open the workflow group, and all three workflows will open automatically.



### TIP 2: Sort/Filter/Cleanse from Results Window

If you want to preserve the sorting of a column in the results window, simply click the green tick icon at the top right. It's a real time-saver!

| Results - Select (29) - Input |                             |            |              |                          |        |                  |          |          |
|-------------------------------|-----------------------------|------------|--------------|--------------------------|--------|------------------|----------|----------|
|                               | Movie Title                 | Movie Year | Runtime      | Genre                    | Rating | Director         | Votes    | Gross    |
| 1                             | The Cabinet of Dr. Caligari | 1920       | Data Cleanse | Mystery, Thriller        | 8      | Robert Wiene     | 66,749   | 0        |
| 2                             | Nosferatu                   | 1922       |              | Horror                   | 7.9    | F.W. Murnau      | 1,00,942 | 0        |
| 3                             | Frankenstein                | 1931       | Filter       | Horror, Sci-Fi           | 7.8    | James Whale      | 73,779   | 0        |
| 4                             | Orcula                      | 1931       |              | Fantasy, Horror          | 7.4    | Tod Browning     | 55,288   | 0        |
| 5                             | Freaks                      | 1932       | Sort         |                          |        | Zod Browning     | 43,391   | 630000   |
| 6                             | The Mummy                   | 1932       |              |                          |        | Karl Freund      | 28,354   | 0        |
| 7                             | The Invisible Man           | 1933       | 71           | Horror                   |        | James Whale      | 37,549   | 0        |
| 8                             | King Kong                   | 1933       | 100          | Adventure                |        | Merian C. Cooper | 80,019   | 10000000 |
| 9                             | The Bride of Frankenstein   | 1935       | 75           | Drama                    |        | James Whale      | 50,611   | 4360000  |
| 10                            | The Wolf Man                | 1941       | 70           | Horror, Mystery, Romance | 7.2    | George Wagener   | 28,724   | 0        |

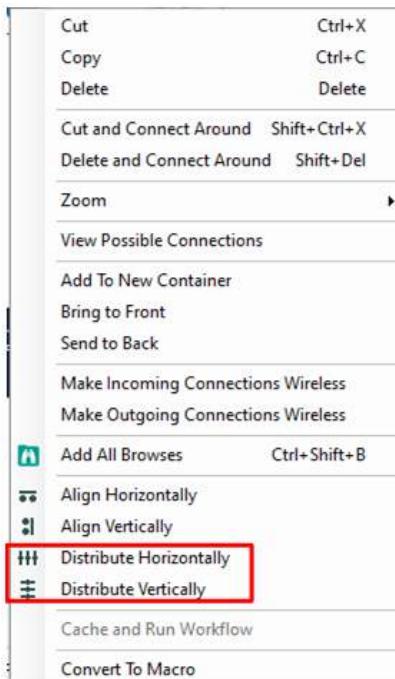
### TIP 3: Arrange the tools in workflow

When working on a chunky workflow, organizing tools is key. Instead of manually aligning each one, save time with two handy shortcuts:

- Align tools horizontally by clicking Ctrl Shift –
- Align tools vertically by clicking Ctrl Shift +

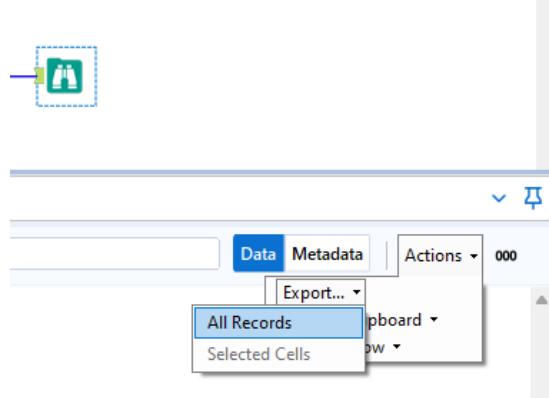
### TIP 4: Distribute Tools Equally

If you love clean and well-organized workflows, another useful tip for arranging tools in your workflow is to select all the tools in a row, then right-click and choose 'Distribute Tools Horizontally/Vertically'.



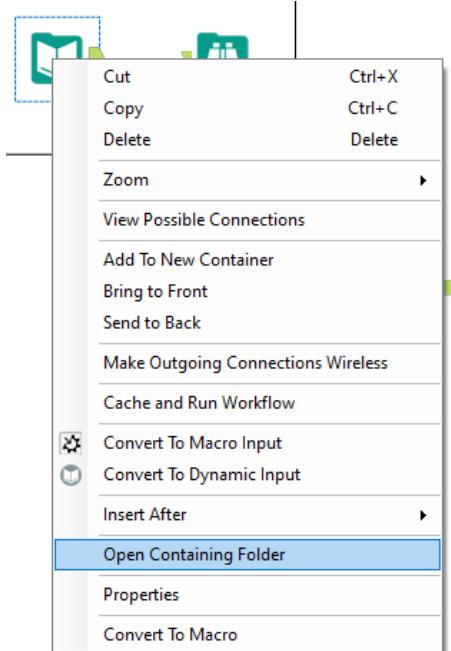
### TIP 5: Output Data Using Browse Tool

Ever since I have discovered about this tip, I've completely replaced my use of output tool. Now I simply use a browse tool instead – navigate to the results pane, click ‘Actions’, select ‘Export,’ then ‘All Records,’ and quickly download data in any desired format.



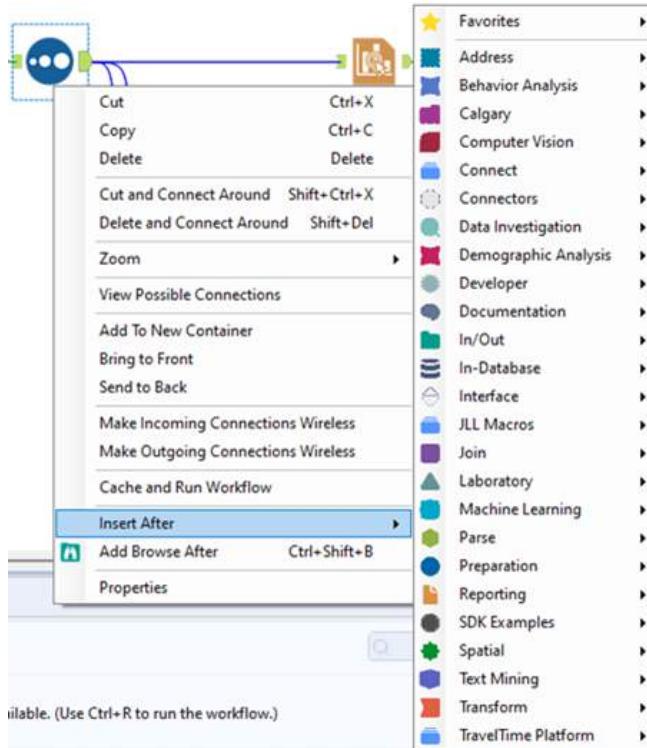
#### TIP 6: Open Containing Folder

One of my favourite tips in Alteryx. Right click on the tool or workflow to select ‘Open containing folder’ and voila! You’ll be instantly directed to the folder where your file/workflow is stored. It’s a game-changer for saving time—no more endless digging through nested folders!



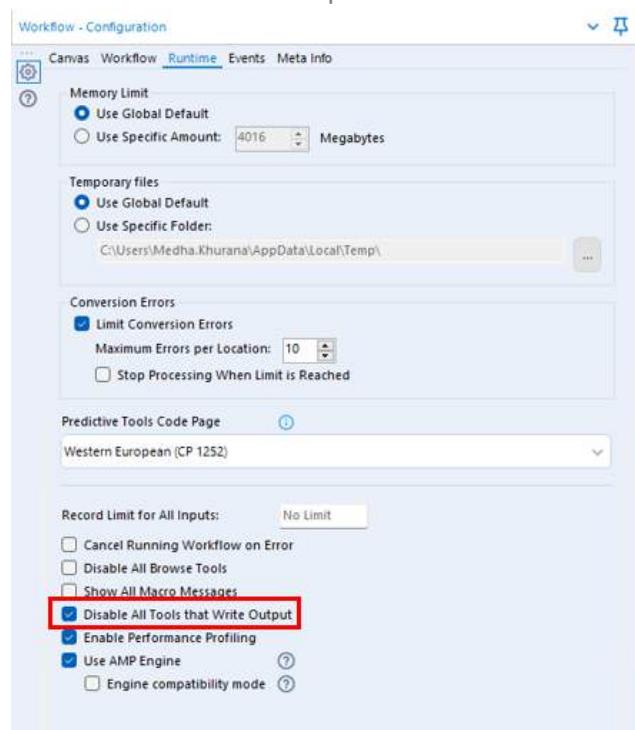
#### TIP 7: Use Insert After to include a new tool

Imagine you have an input tool connected to five other tools, and you need to add a select tool after it. Instead of reconfiguring each connection individually, you can simply right-click on the input tool and choose ‘Insert After’ to add the select tool. This will seamlessly insert the select tool between the input tool and its five connected tools.



#### TIP 8: Disable all tools that write output

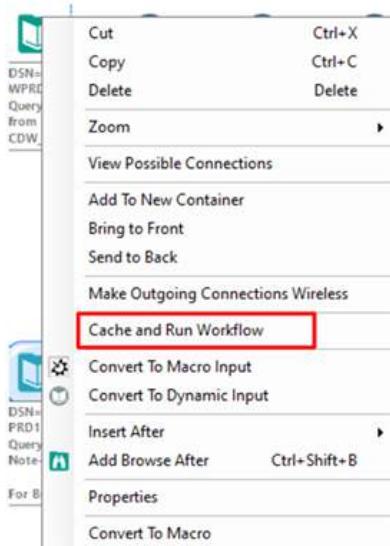
Imagine a workflow that writes data into database or sends an email to a list of recipients with every run. However, during development, you don't want this to happen and end up putting a container for each output tool and then disable it. Instead, you can simply click on the blank space in your workflow, go to Runtime tab in Configuration Window on right and select 'Disable All Tools that Write Output'.



#### Tip 9: Cache and run workflow

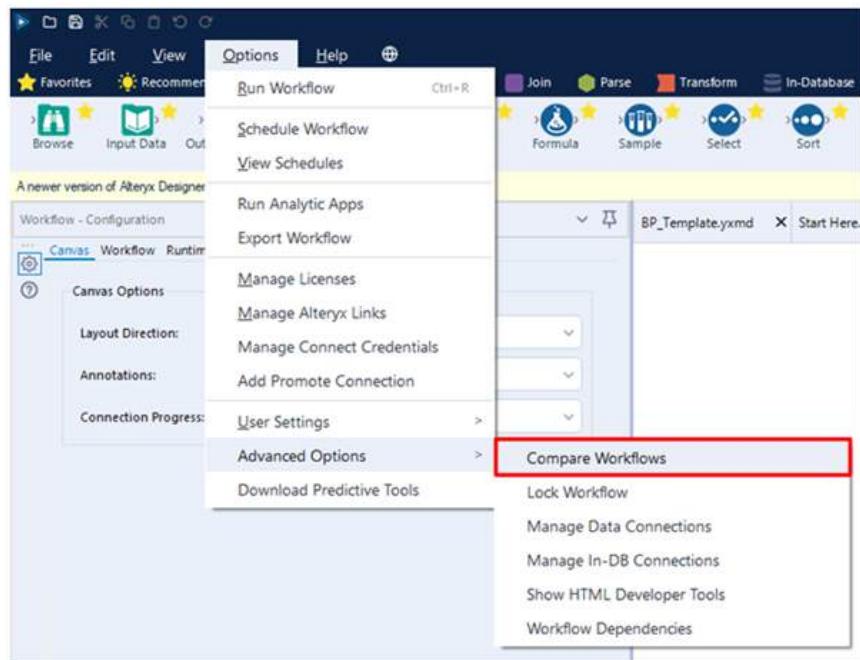
Another tip when your workflow is in the development phase. Imagine you're fetching data from a SQL server, and the lengthy query takes a couple of minutes to execute. Instead of waiting for the query to run each time you run the workflow, there's a handy shortcut. Right-click on the input tool

and choose 'Cache and Run Workflow'. This action loads the data from the SQL server once, storing it in memory and drastically cutting down the workflow's run time.



#### TIP 10: Compare Workflows

Alteryx gives you the ability to compare one workflow from the other. All you need to do is go on Options, select Advanced Options and then select 'Compare Workflows'. A new window pops up where you can select the two workflows you want to compare.



Compare Workflows

Select the workflows to compare.

Workflow 1

Select Workflow

Workflow 2

Select Workflow

That's it! Those are my top 10 Alteryx tips. Hope you enjoy them!

CJ: You've written previously about workflow optimisation and the need for templates. Building robust solutions is often a struggle in many workplaces. How can a developer keep this in mind when building an Alteryx workflow?

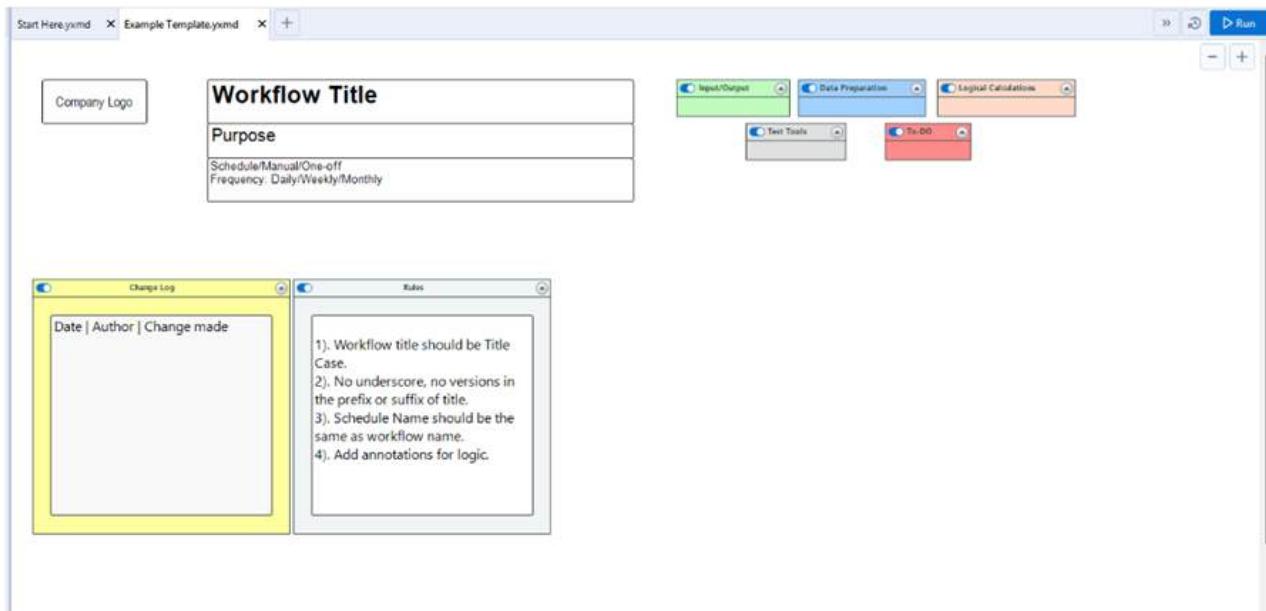
Here's how I like to think about every workflow/solution I build: Will my future self understand this workflow a year from now? And if someone else must use this workflow later on, will they know what's going on?

To ensure you are building sustainable and robust workflows, it is beneficial to stick to a few best practices. One of them is incorporating workflow templates. Templates provide a standardized structure that maintains uniformity across all your workflows. A typical workflow template consists of the following elements:

- Workflow title and purpose
- Workflow schedule
- Change Log
- Legend

By adhering to a set template, you're making sure that your workflow is not just easy to understand, but also simple to update and fix if needed.

Here's a simple template you can use as a starting point. Feel free to customize it to fit your needs and your organization's style.



Another good practice is to annotate the business logic on specific tools wherever necessary. This would ensure you are not leaving your co-worker in a pickle wondering why you adopted a particular approach.

Annotating is simple, click on the tool you want to annotate, go to the configuration panel on the right, select “Annotation” and simply input the supporting logic/documentation.

CJ: When it comes to automation, what are the most important things to consider when creating a workflow?

That's a great question. We often create workflows and schedule them to perform their expected job such as refresh data in tableau server or sending emails etc. However, there are a few things to consider when creating workflows for automation.

1.

#### Prepare for Unforeseen Errors

Leverage tools like the ‘Error’ and ‘Message’ tools in Alteryx to customize error messages for unexpected occurrences in your workflow. For instance, when joining two datasets and expecting no output in the left anchor, configure these tools to send tailored error messages. Harry Osborne from

Data School has done a great job explaining this – <https://www.thedataschool.co.uk/harry-osborne/message-test-tools/>

2.

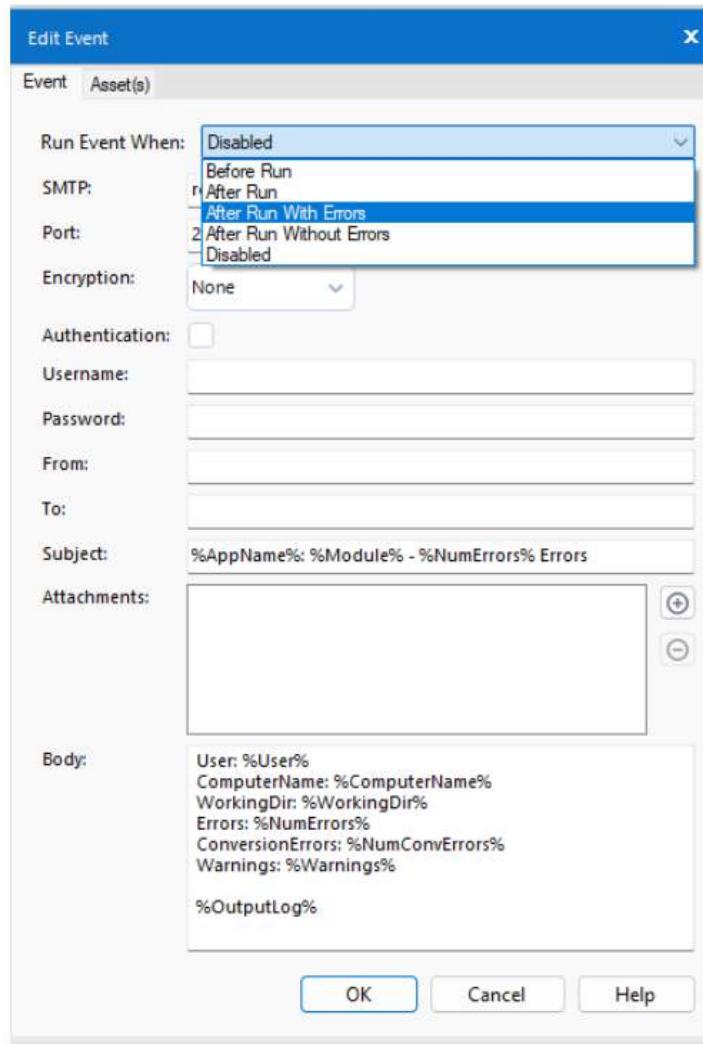
#### Identify and Remove Bottlenecks from your Workflow

Remove all the browse tools before publishing your workflow on the server besides any other tools that are not required. Enable performance profiling to identify and resolve bottlenecks that might be slowing down your workflow.

The screenshot shows the Alteryx Workflow Configuration interface. On the left, the 'Runtime' tab is selected, displaying settings for conversion errors (Limit Conversion Errors, Maximum Errors per Location: 10, Stop Processing When Limit is Reached), predictive tools code page (Western European (CP 1252)), and performance profiling options (Record Limit for All Inputs: No Limit, Cancel Running Workflow on Error, Disable All Browse Tools, Show All Macro Messages, Disable All Tools that Write Output, **Enable Performance Profiling** (which is checked), Use AMP Engine, Engine compatibility mode). On the right, the 'Street Crimes in UK.yxmd' workflow is visible, showing a series of tools connected by arrows. Below the canvas is the 'Results - Workflow - Messages' pane, which lists workflow statistics. A red box highlights the 'Designer x64' section, which contains a note about memory management: 'The Designer x64 reported: 28.78ms have been spent on memory management. 0.62% of the total workflow execution time.' The 'All' tab is selected in the results pane.

#### 3. Configure Alerts on Workflow Failure

Errors in the workflow hosted on the server can disrupt operations without immediate detection. To mitigate such risks, leverage events in Alteryx to trigger alerts that proactively notify you every time a workflow fails on server. You can find out the detailed process up setting up alerts on my [website](#).



Simple measures like these enable swift action to prevent/resolve issues on workflows that are built for automation before they escalate. By investing a bit of time upfront, you can prevent costly oversights and maintain the reliability of your workflows.

#### CJ Round-Up:

A blog packed full of tips to help you elevate your Alteryx game. I particularly love the tips around making sure the solutions are robust and conform to a minimum level of standard in way of creating a workflow. Any time we can save letting the workflows run and not have to mitigate bugs and errors is time well saved.

Please join me in following **Medha's journey**, she put it so poetically the impact community can make. "I've long drawn inspiration from remarkable members of the data community ... who have built a brand of their own by sharing what they are passionate about with others. I wanted to do that too – to cultivate my own brand, be a source of inspiration and help others along the way."

Lovely sentiment to close out this week. Take care.

LOGGING OFF,

CJ

A code share piece this week. Recently I have become quite the fan of the darts and spend Thursday evenings in front of the TV watching the Premier League. It is part of what has been coined "The Littler Effect" since the Christmas period. So whether you're into darts or just want to know the basics behind Selenium web scraping this ones for you.

The data we will look at is from Darts24. Here is an example of that.

The screenshot shows a Darts24 match summary for a game between Littler L. (World Ranking: 26) and Aspinall N. (World Ranking: 4). The match ended with a score of 6-4 in favor of Littler L. The summary includes a list of throws for each set, with some marked as '140+' or '180'. The detailed history shows the breakdown of each set's throws.

| Set | Score | Throws                                                                                       |
|-----|-------|----------------------------------------------------------------------------------------------|
| 1   | 1 - 0 | 416:501, 416:321 180, 334:321, 334:264, 236:264, 236:124 140+, 136:124, 136:68               |
| 2   | 1 - 1 | 501:401, 402:401, 402:315, 304:315, 304:175 140+, 170:175, 170:70, 25:70 140+                |
| 3   | 1 - 2 | 402:501, 402:407, 306:407, 306:347, 225:347, 225:207 140+, 131:207, 131:36 140+, 60:36       |
| 4   | 1 - 3 | 501:321 180, 401:321, 401:281, 344:281, 344:221, 254:221, 254:81 140+, 194:81, 194:16, 92:16 |
| 5   | 2 - 3 | 361:501 140+, 361:367, 261:367, 261:227 140+, 121:227 140+, 121:170, 40:170, 40:114          |
| 6   | 3 - 3 | 501:361 140+, 364:361, 364:261, 244:261, 244:206, 125:206, 125:133                           |
| 7   | 4 - 3 | 404:501, 404:439, 264:439 140+, 264:394, 164:394, 164:312, 70:312, 70:177                    |
| 8   | 5 - 3 | 501:420, 401:420, 401:323, 310:323, 310:223, 185:223, 185:88, 51:88, 51:14                   |
| 9   | 5 - 4 | 401:501, 401:404, 341:404, 341:264 140+, 207:264, 207:164, 157:164, 157:70, 60:70            |
| 10  | 6 - 4 | 501:378, 361:378 140+, 361:318, 224:318, 224:198, 140:198, 140:113, 20:113, 20:36            |

You will notice a few things.

- That we get a marker for 3 darts worth 140+ or 180.

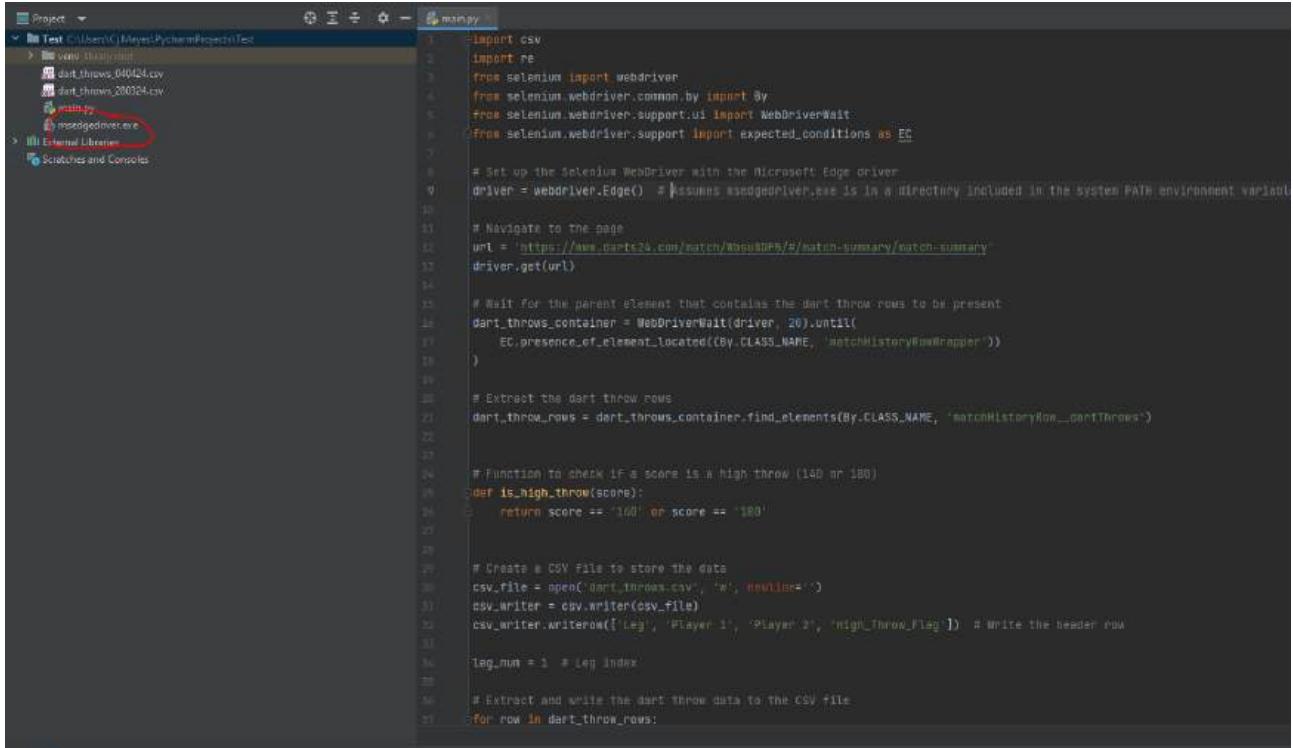
- The players scores are split by a colon.

These will be two things we will need to keep in mind when referring to our code.

So how do I get set up?

Well firstly, I am using the Edge Web Browser (Others are available, depending what you use.) For the Selenium package to work you have to install the correct driver. This will allow the code to open up the designated web page. For my version of Edge, I downloaded it from [here](#).

You will see in my python file I host this driver locally for quick access.



```

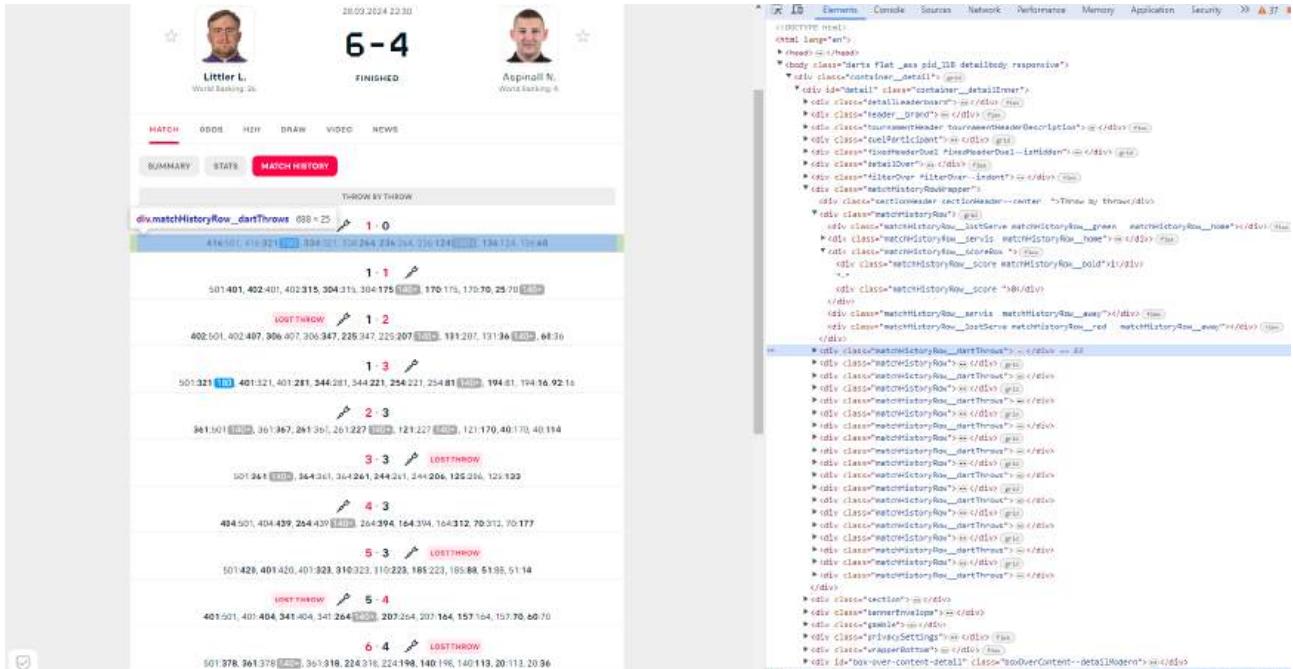
1 import csv
2 import re
3 from selenium import webdriver
4 from selenium.webdriver.common.by import By
5 from selenium.webdriver.support.ui import WebDriverWait
6 from selenium.webdriver.support import expected_conditions as EC
7
8 # Set up the Selenium WebDriver with the Microsoft Edge driver
9 driver = webdriver.Edge() # Assumes edgewebdriver.exe is in a directory included in the system PATH environment variable
10
11 # Navigate to the page
12 url = 'https://www.darts24.com/api/v1/matchHistory?#match-summary'
13 driver.get(url)
14
15 # Wait for the parent element that contains the dart throw rows to be present
16 dart_throws_container = WebDriverWait(driver, 20).until(
17 EC.presence_of_element_located((By.CLASS_NAME, 'matchHistoryRow__contThrows'))
18)
19
20 # Extract the dart throw rows
21 dart_throw_rows = dart_throws_container.find_elements(By.CLASS_NAME, 'matchHistoryRow__contThrows')
22
23
24 # Function to check if a score is a high throw (140 or 180)
25 def is_high_throw(score):
26 return score == '140' or score == '180'
27
28
29 # Create a CSV file to store the data
30 csv_file = open('dart_throws.csv', 'w', newline='')
31 csv_writer = csv.writer(csv_file)
32 csv_writer.writerow(['Leg', 'Player 1', 'Player 2', 'High_Throw_Flag'])
33
34 leg_num = 1 # Leg index
35
36 # Extract and write the dart throw data to the CSV file
37 for row in dart_throw_rows:
38

```

## How does Selenium work?

Selenium is a package that enables browser automation, allowing developers to control and interact with web browsers programmatically. It supports multiple browsers and provides tools for simulating user interactions, extracting data from websites, and performing cross-browser testing. With Selenium, users can automate repetitive tasks, scrape web content, and test web applications across different browser platforms. When you click run on the code it will open the web page and start the web scraping.

It uses the html paths to do this.



Let's look at the code in greater detail. (You can find a copy at the top of the page under the header)



1. Import the required modules:
  - csv for reading and writing CSV files.
  - re for regular expression operations.
  - webdriver from the selenium package for browser automation.
  - By and WebDriverWait from selenium.webdriver.common.by and selenium.webdriver.support.ui respectively for locating elements on the webpage.
  - expected\_conditions from selenium.webdriver.support for specifying the expected conditions for waiting.
2. Set up the Selenium WebDriver:
  - Initialize the WebDriver with the Microsoft Edge driver. Make sure the driver executable is in your system PATH or provide the path to the driver executable.
  - In this example, webdriver.Edge() is used to start an Edge browser instance. You can replace it with a different WebDriver (e.g., webdriver.Chrome()) if desired.
3. Navigate to the webpage:
  - Set the url variable to the desired webpage URL for the desired darts game on Darts 24.
  - Use driver.get(url) to navigate to the webpage.
4. Wait for the dart throw rows to be present:
  - Use WebDriverWait to wait until the parent element that contains the dart throw rows with class name 'matchHistoryRowWrapper' is present on the webpage.
  - The EC.presence\_of\_element\_located expected condition is specified.
  - The maximum wait time is set to 20 seconds (you can adjust this if needed).
5. Extract the dart throw rows:
  - Use find\_elements to locate all the child elements of the dart throw container element with class name 'matchHistoryRow\_\_dartThrows'.
6. Define the is\_high\_throw function:
  - This function takes a score as input and checks if it is a high throw (140 or 180) by comparing it against the values '140' and '180'.
  - It returns True if the score is a high throw or False otherwise.
7. Create a CSV file to store the data:
  - Open a new CSV file named 'dart\_throws.csv' in write mode using open.
  - Create a csv.writer object to write the CSV rows.
  - Write the header row containing the column names 'Leg', 'Player 1', 'Player 2', and 'High\_Throw\_Flag'.
8. Extract and write the dart throw data to the CSV file:
  - Iterate over each dart throw row.
  - Split the row text using ',' as the delimiter to separate the player scores.
  - Iterate over each dart throw in the row.
  - Use map and str.strip to remove leading/trailing whitespaces and split each dart throw into player 1 and player 2 scores using ':' as the delimiter.
  - Use regular expressions (re) to remove any non-digit characters from the player scores by using re.sub(r'', '', score).
  - Check if player 2's score is a high throw by checking if it ends with '140' or '180'.
  - If it is a high throw, extract the high throw flag and remove it from the player 2 score.
  - Write a new row to the CSV file containing the leg number, player 1 score, player 2 score, and high throw flag.
- 9.
10. Close the CSV file using csv\_file.close().  
Close the browser using driver.quit().

```

main.py × dart_throws_040424.csv × dart_throws_280324.csv ×
1 Leg,Player 1,Player 2,High_Throw_Flag
2 1,416,501,
3 1,416,321,180
4 1,334,321,
5 1,334,264,
6 1,236,264,
7 1,236,124,140
8 1,136,124,
9 1,136,68,

```

You will see by the output we get the associated leg, scoring of the visit and the high throw flag of 140+ or 180.

Two examples of the finals from the last few premier league weeks can be found in the git repo.

Want to take it further? Why not try:

- Add in who was on throw to start the leg.
- Add in a running score timeline.
- Add in whether the player won or lost the throw.
- Add in a row record before the next leg begins to show the players score reaching zero.
- Visualise the scoring through a visualisation platform of your choosing

\*\*With thumbnail thanks including Adam Greens Tableau Visual, which uses some radial logic from previous blogs\*\*

LOGGING OFF,

## BUILDING A MARGINAL HISTOGRAM

Hi all,

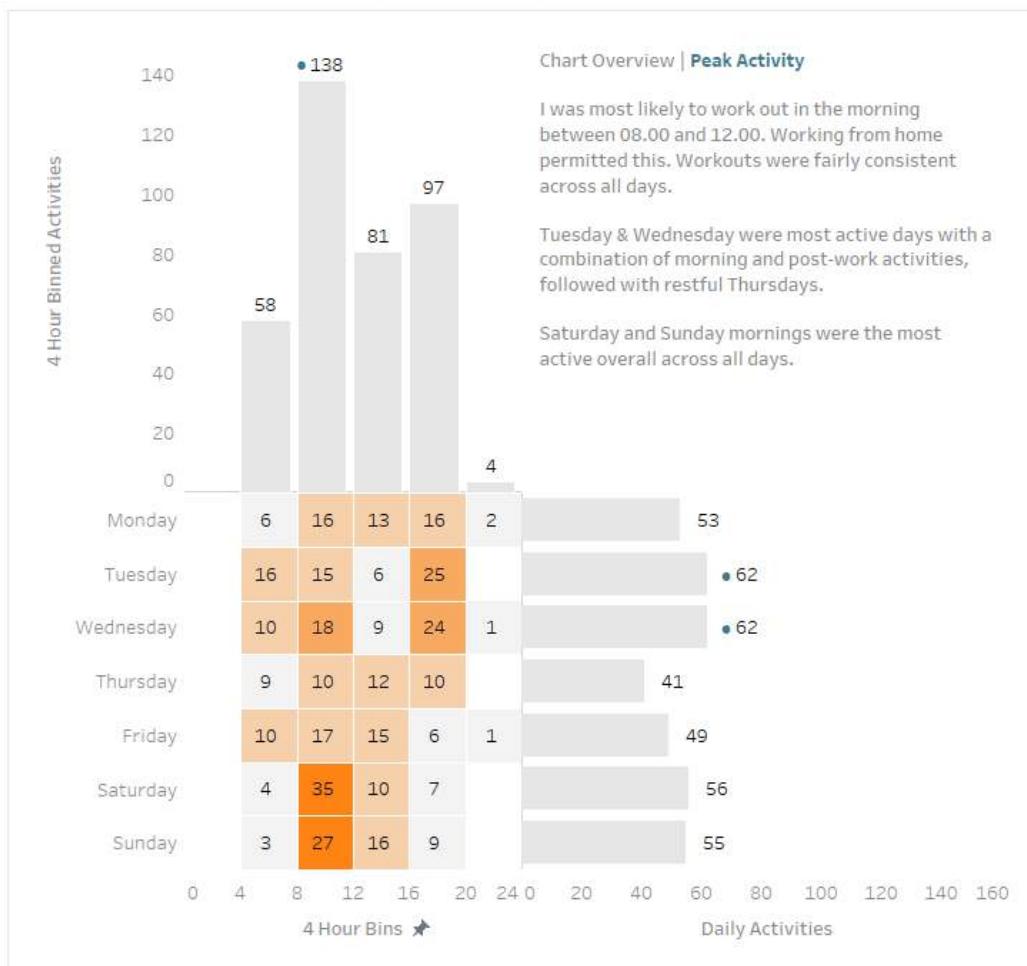
Hope everyone is doing well. We are coming up to the end of Q1 which officially means..... the sun almost starts to shine in the UK!

I have been away from action a little these last months, which is a combination focussing on training, work, my friendships and I welcomed a new puppy into my life. They often talk about the different “burners” in life, and you can’t have all of them turned all the way up. Fortunately, between puppy and friends it has been a welcome down period in terms of producing website content. Enough rambling, we are back this week with a chart specific blog, something geared towards more beginner level when it comes to build, but a good opportunity to talk through some design choices.

We will look at both the build and design aspects of a marginal histogram.

### Strava | Marginal Histogram

This visual Looks At Which Weekdays & Time of Day I Was **Most Active** In 2023.



This is what we are going to move towards using Strava data. You will find my export in the repository. We wont amend any of the data before hand so the tableau calculations will work for your own data.

Ideally we want to split our chart by hours and day.

`DATEPART('hour',)`

`DATEPART('weekday',)`



For this I'm just looking at a specific day. Instantly something looks a little off to me here. As I see a 10 for the number 1 at around 10am.



I'm sure you can spot it now, how my date part for the weekday aligns 1 as Sunday. Something to be cautious of and the type of calendar you use. So we should make sure when we use

`DATEPART('weekday', )`

that we are putting the days in the order we want. I.e we want to show Monday-Sunday.

Design wise, we have added the count distinct onto the color pane. A few things from me here.

- I find gradual color changes harder to interpret than a color change of palettes using 2-5 colors.
- We have some hours of the day missing.
- We may want to play around with some formatting for the labels but also the cells.
- This is a lot of information to consume all at once. It is good for very specific hours on days, but hard to judge across all days, or brackets of hours.



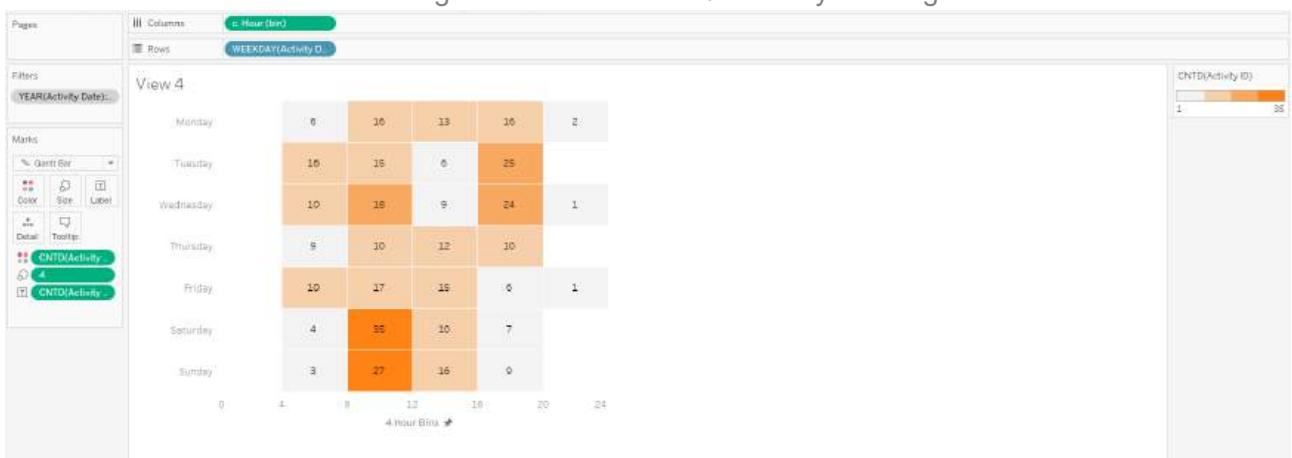
Above we add in the full day range. Notice how I don't work out before 5am or after 21.00. Shocking right? By adjusting the timeframe this becomes increasingly obvious.

Here I add what % each hour on a day is as a proportion of total. For me a couple % change isn't huge in impact, so I thought maybe this needs additional bins...

Edit Bins [c. Hour]

|                                   |               |                                       |    |
|-----------------------------------|---------------|---------------------------------------|----|
| New field name:                   | c. Hour (bin) |                                       |    |
| Size of bins:                     | 4             | Suggest Bin Size                      |    |
| Range of Values:                  |               |                                       |    |
| Min:                              | 4             | Diff:                                 | 18 |
| Max:                              | 22            | CntD:                                 | 19 |
| <input type="button" value="OK"/> |               | <input type="button" value="Cancel"/> |    |

Using this we separate our day into 4 hour chunks. (of course you can use other options but I felt like these were good brackets and 24/4 is easy to categorise.



I much prefer this design here because of its emphasis on the Saturday and Sunday morning. We also see some numbers worth digging into for the volume on Tuesdays and Wednesdays and the impact it has on Thursdays.

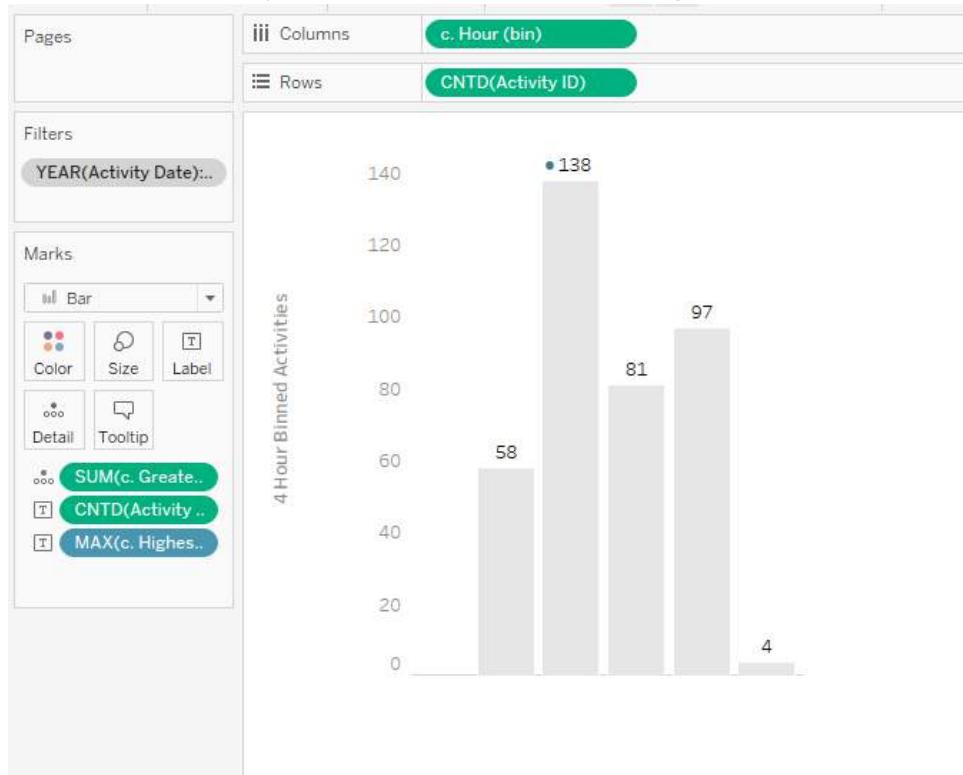
Stylistically, I use 4 colors to emphasize volume, but also keep the range of bins across the full 24 hours, notice 0-4 hours empty.

Some individuals will prefer no labels, I've actually gone with a punchy all black color. People in the comments will have to let me know whether they are against this because of "ink on the page". For me... it helps reinforce the value here, but of course we centre align it.

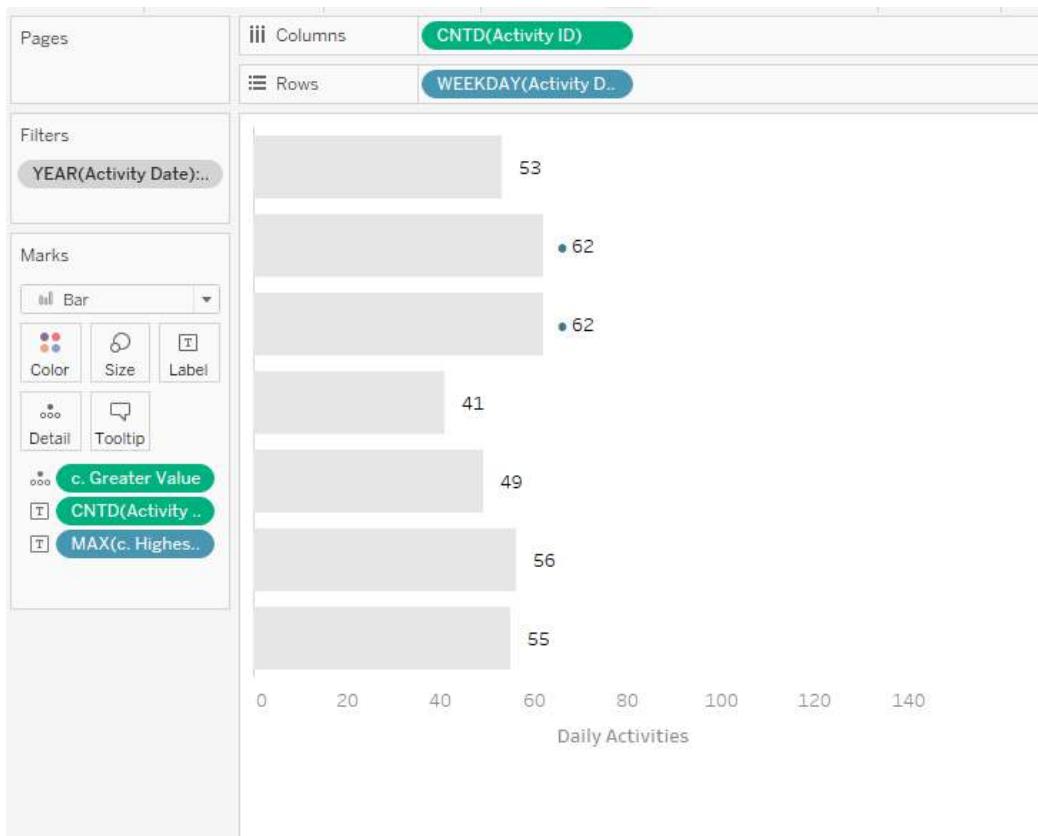
Another tip is to either add borders, (or dividers if you want a thicker color) – White usually works well against colors on the page. Sometimes a light grey works well too.

So the view above is great for specific hours on a page, but what if i want to easily see volume across a day or a block of hours. Well this is where the full chart comes into play.

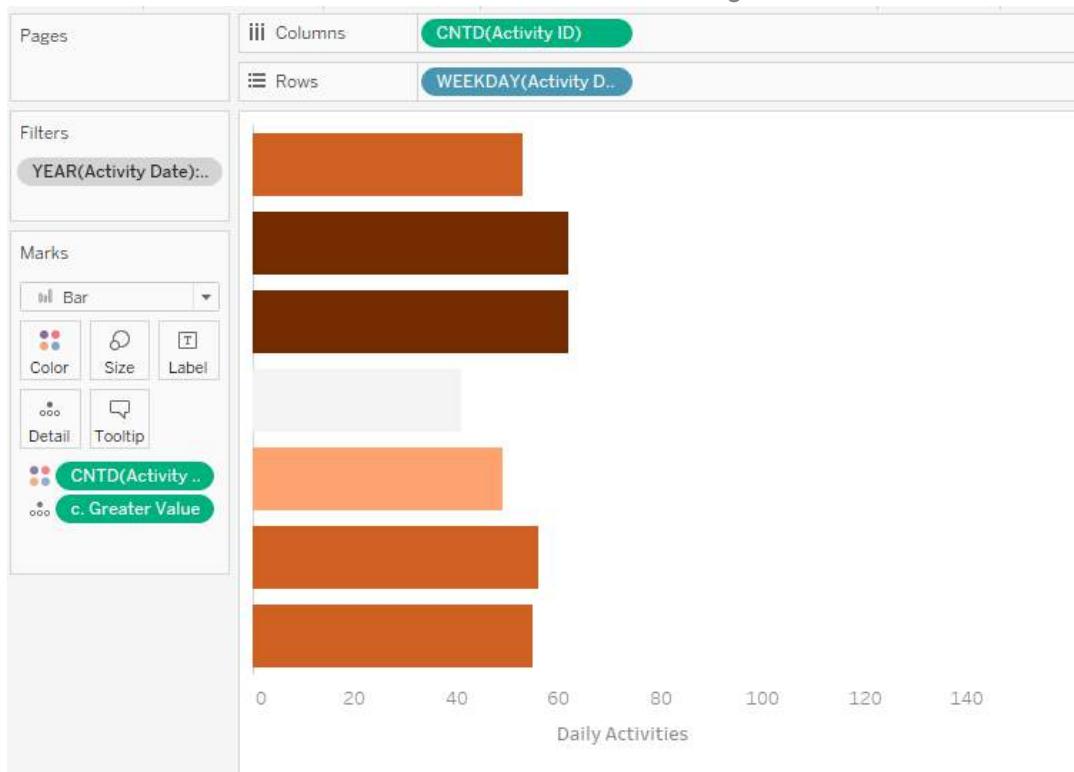
We simply need to build the corresponding bar charts.



followed by the daily bar,



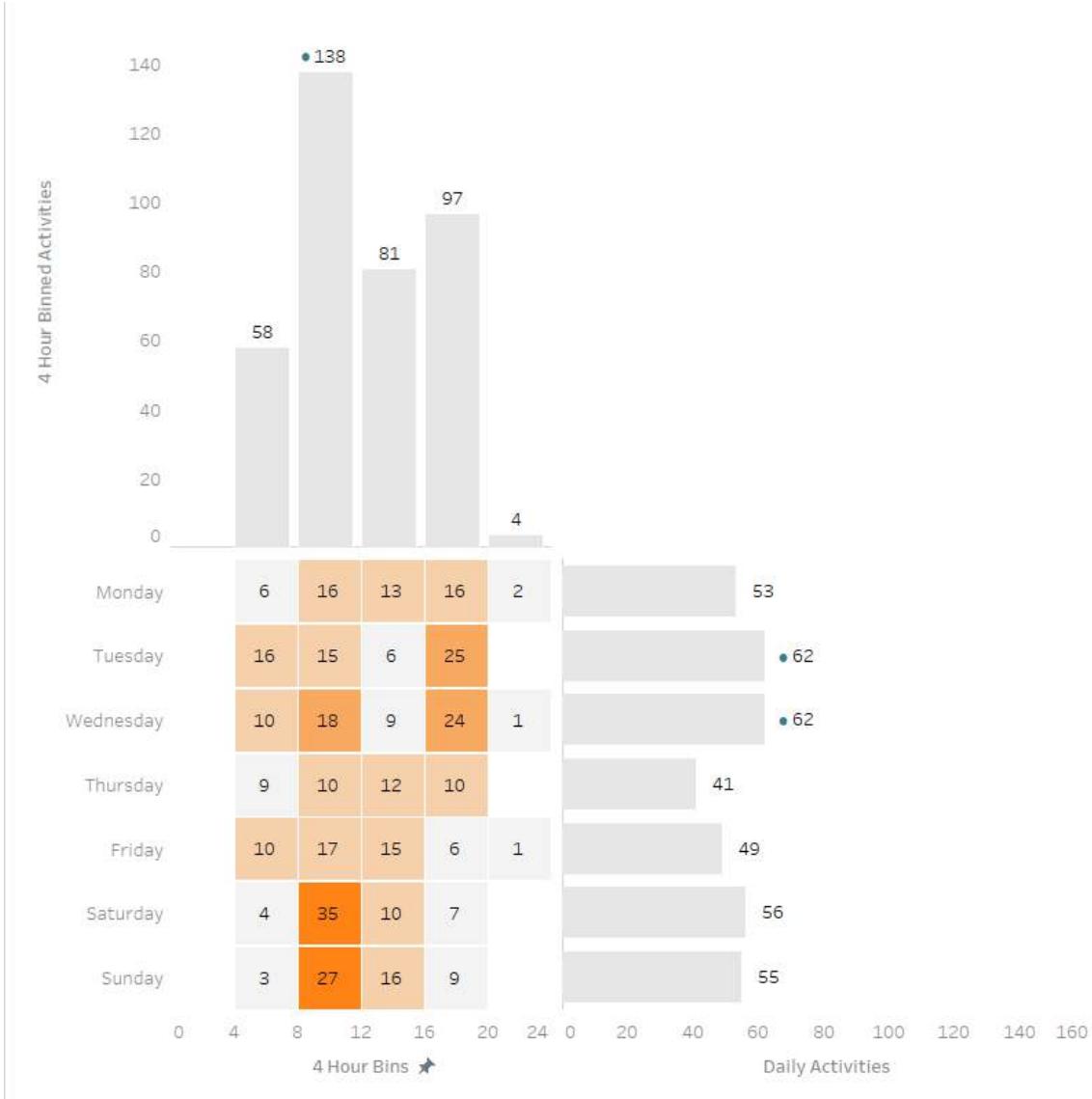
Thing to notice with these bars is i put them in a separate color.  
I see a lot of bars with the color encoding on it still.



For me it raises two things:

- Of course my longer bar is going to be darker in color. That feels like double encoding.
- The second thing is, when you drag color onto the bar here, the color gradient is working to a different length of shade compared to our grid, as they are working across different aggregation blocks. For me that is confusing as the same shade is not worth the same amount of activities across the whole chart.

Final thing then, is getting it onto a page and aligned as one chart.



A few cosmetic things to note:

I really like having the padding and alignment feel fairly square. This allows for the bars and the heatmap to have similar shape. The bars are equal across the two bar visuals.

You may consider removing some of the axis. I kept them for alignment purposes but greyed them slightly as I have the labels on the page. Totally get why you may want one or the other when it comes to labels and axis.

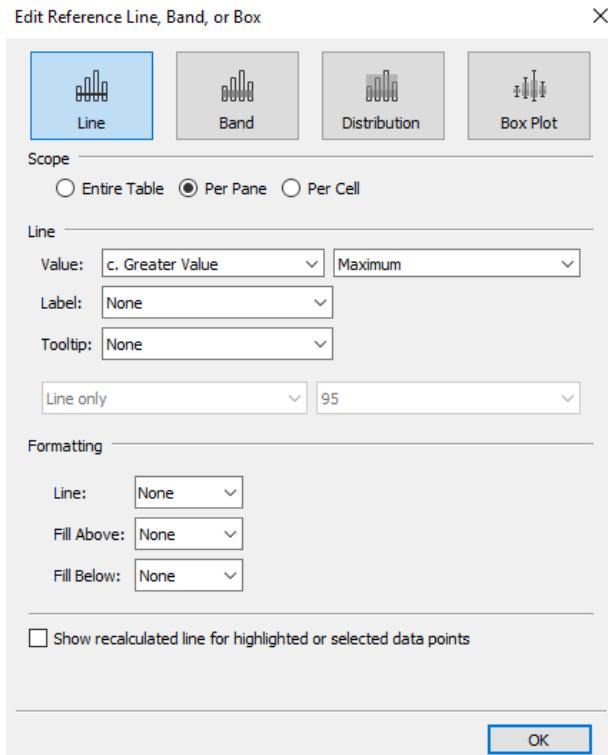
I use a calculation to add a reference line to allow that my two bar axis to be of similar scale. Having the bars in proportion I feel is a subtle but useful touch here.

Here is the calculation for that:

```

if
{ fixed: max({ Fixed : COUNTD()})}
>
{ fixed: max({ Fixed DATEPART('weekday',) : COUNTD()})}
then { fixed: max({ Fixed : COUNTD()})}
ELSE
{ fixed: max({ Fixed DATEPART('weekday',) : COUNTD()})}
END

```

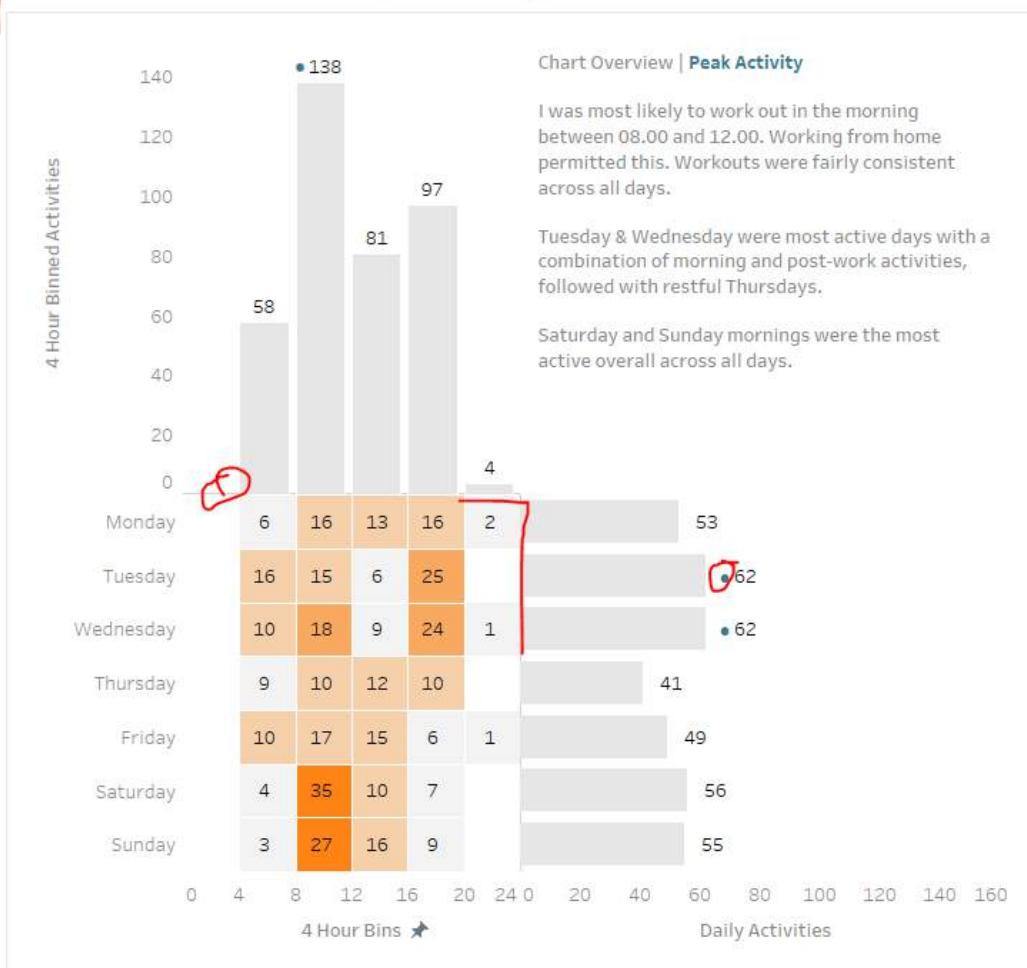


What you could consider here is using the bars as % instead that way you could keep consistency between 0-100.

Final things are a few cosmetic amendments.

## Strava | Marginal Histogram

This visual Looks At Which Weekday & Time of Day I Was **Most Active** In 2023.



- Title alignment through padding. Adding a grey box around the whole chart, and adding in a small amount of text to help give some analysis to the chart.
- The circle shows where the zero axis has been added. I feel this helps frame the bars and the heatmap. It also re-emphasizes the fact 0-4 hours there were no activities, as opposed to just a big gap design wise!
- Finally adding some peak activity circles to demonstrate the longest bars for specific hour blocks or days.

Anyway, that's my two cent for today. A fairly easy one to complete in terms of technical skills but a lot to consider design wise. Let me know if you agree or disagree with some of my thoughts, love talking about these things in greater detail.

LOGGING OFF,

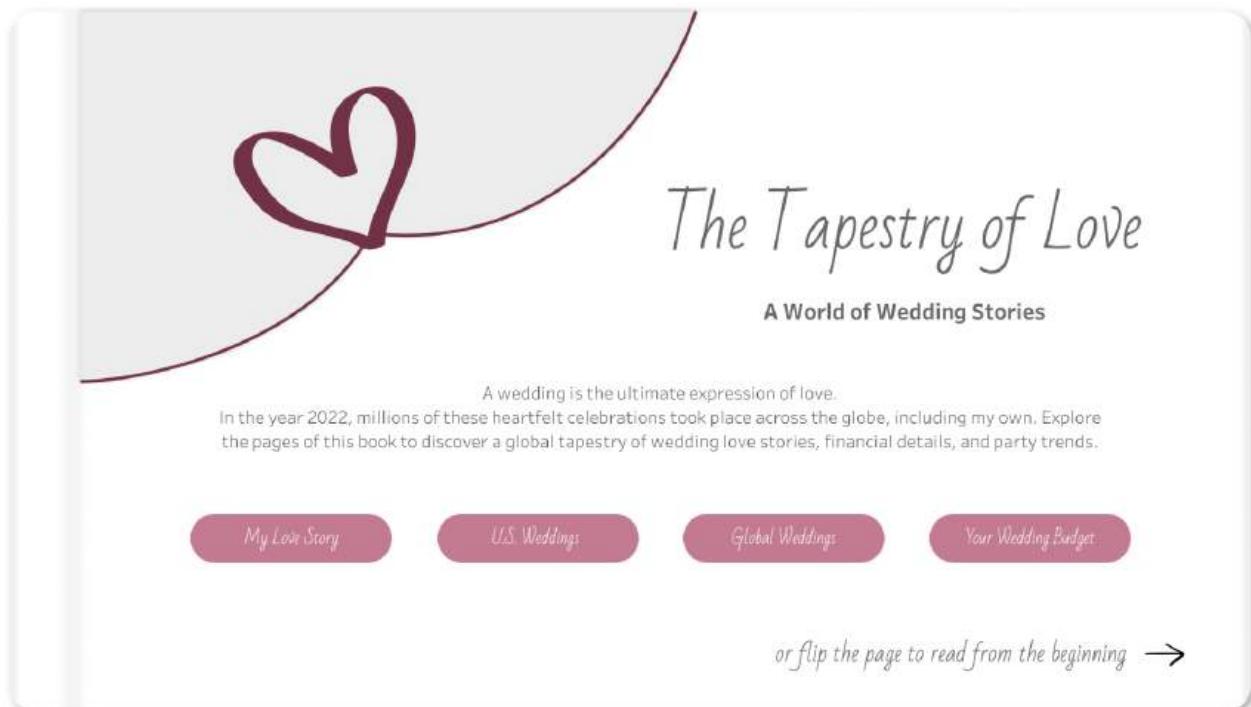
CJ

## IRON VIZ FEEDER REFLECTIONS – JADE CORLEY

Hi All,

Welcome back to another guest blog. I am so pleased to be joined by Jade this week to talk to us about her Iron Viz journey over the past few years. Jade's Iron Viz entry was one of my favorites this year and so it's a real honour to be able to dive into some of the details with her today. The 2024

feeder was the theme of love, so what better topic to talk about than Jade's personal story of finding love.



CJ: Jade, thanks for joining. For those that are unaware, May you tell us a little about how you ended up in the world of data? When did the passion for Tableau ignite, was it when you joined Analytic Vizion or earlier?



[Twitter](#) [Facebook](#) [LinkedIn](#)

## Jade Corley

Solutions Architect at Analytic Vizion | Greenville, South Carolina, United States

Tableau Social Ambassador

[Follow](#)

Vizzes 28

Favorites 25

Following 118

Followers 701

| Visualization Title                                            | Author      | Rating | Views  |
|----------------------------------------------------------------|-------------|--------|--------|
| Lost but Not Forgotten - IronViz 2022 Top #4 Qualifier   #VOTD | Jade Corley | ☆ 99   | 6,267  |
| The IronViz Qualifer is like a Marathon                        | Jade Corley | ☆ 0    | 28     |
| The Tapestry of Love   IronViz Submission 2024                 | Jade Corley | ☆ 10   | 1,494  |
| #TC23 Presentation - Superstore                                | Jade Corley | ☆ 1    | 174    |
| Chiari I Malformation                                          | Jade Corley | ☆ 117  | 5,055  |
| Est. in '93 - #TC23 Tip Battle                                 | Jade Corley | ☆ 5    | 197    |
| Tableau Tips for Tomorrow's Leaders                            | Jade Corley | ☆ 3    | 110    |
| Example Professional/Business Dashboard                        | Jade Corley | ☆ 16   | 10,086 |

J:Hi CJ, it's so cool being a part of this! Thanks for having me on your blog. My journey with data began in 2017 during an internship I had in grad school. I was working alongside and learning from former Tableau Visionary Matt Chambers. He saw an infographic I made, said I had an eye for design, and introduced me to Tableau. I was instantly hooked. To this day I still enjoy data visualization because I believe it's the perfect balance of creativity and analytical problem solving. It bridges together two different ways of thinking to tell a story that will hopefully positively impact someone else. After grad school, I worked for a few years at Clemson University, and today I'm a Solutions Architect and consultant with Analytic Vizion. It's been a fun journey full of learning and growth.

CJ: You've entered Iron Viz last time, earning spots in Top 10 and most recently built one of my favorite visuals this year. What is it about IronViz that makes you want to enter?

# Lost but not forgotten

*a look at famous art heists around the world*

There it is, hanging right there against the wall. A priceless work of art worth millions. It's only a few pounds of weight at most. An easy item to snatch under the right conditions. Maybe after museum hours under the cover of darkness. The thrill of the thought gets your adrenaline pumping. A relatively easy, non-violent crime to commit. Is the risk worth the reward? ...What would you do?

It's a crime that's been around since the dawn of time. **Art theft**, sometimes called artnapping, is the stealing of paintings, sculptures, or other forms of visual art from galleries, museums, or other public and private locations. And it still happens today all over the world.

There are  
**50,000**  
Est. Art Thefts  
per year

worth an est.  
**\$7.0M**  
in total value per year

and only **10%** of these stolen works  
are actually recovered

**1.5%**  
of cases see the  
art recovered &  
criminal prosecuted

Although the odds of getting caught are very low, making a career as an art thief is a lot harder than you would think. Most thieves don't realize how "hot" a particular artwork can be, especially those that are extremely valuable. Unless they have a buyer ahead of time, it can be very difficult, if not impossible to sell even on the black market. Most criminals will instead attempt to use the art as collateral or hold it for ransom. It's a federal crime that could land a thief up to 10 years in prison or heavily fined. So why is art napping still so popular? Are they trying to re-create a scene from the movies or is it just for the thrill?

Here's a look at some of the most famous art heists from around the globe.

J: Thank you that means a lot CJ! I'll start by saying that while using Tableau for 7 years now, I've only entered IronViz twice.

My first entry was: **Lost but not forgotten**

My second entry was: **The Tapestry of Love**

That's because it truly is an intimidating, time consuming, pour your heart into it, challenge. So for me, I have to really be inspired by the topic. I find that I get more inspired when I talk about the challenge with others, even if they have no idea what Tableau is. For example, for the Art themed IronViz Qualifier, I wouldn't consider myself an artist, but I was inspired by conversations with my husband after we watched a documentary about stolen art. I found some supporting data and the story unveiled in my mind from there. The #datafam coming together for the IronViz challenge is also an encouraging and fun season to be a part of. Not going to lie, the prize money is also a nice incentive. Each time I've entered, I've truly wanted to make it on that stage; to push myself and just go for it!

## My Love Story

The Planning

The Numbers

The Party



Let's begin our journey with my personal love story. It began on October 21, 2018, where a mutual friend set me up on a blind double date at an escape room. What a way to dive in! I remember immediately thinking my date was so handsome. We solved puzzles, laughed, and escaped successfully. The next day I received a Facebook friend request from him. I was so excited. This sparked the sweetest and best relationship I've ever experienced.

That next week we went on our first date to the typical dinner, movie, and drinks afterwards. We talked for hours about life and all the taboo things you're not supposed to talk about on a first date like politics and religion. Somehow it all still worked. To this day our heartfelt, deep conversations are one of my favorite things about our relationship.

With both of us being in our mid-twenties, we were looking for something serious but didn't want to rush anything. However, by 10 months in I knew this relationship was something special - I knew I loved him. I can remember being so nervous, overthinking how to tell him, and wondering if he felt the same way. Finally, one night we were having dinner at his house, salmon and (my absolute favorite side) sautéed mushrooms. We were both sitting on the couch when I offered him some mushrooms. He goes "Whoa! You're going to give me the rest of your mushrooms?" And I responded, "I mean well yeah, because I love you." My eyes grew wide, his smile grew big, and he said "I love you too!"

### Our Timeline



The next 3 years were filled with sweet memories as we learned more about each other. From adventures like snowboarding, beach trips, and watching rocket launches we talked, shared, laughed, and had the hard conversations, I'm thankful for every moment.

In October 2021, we flew to Nova Scotia to visit my dad and family. Little did I know he also used that opportunity to ask my dad for his marriage blessing. It warmed my heart that he wanted to uphold this sweet tradition. 4 months later, our love story came full circle. A typical Saturday, we met some friends to do an escape room. At the very end there was a key with my name on it. When I op-

I waited a whole 4 days to start planning! I was so ready to marry him. I planned for about 10 months, and we got married on October 1, 2022 in our home state of South Carolina. It was a day full of so much love, family, friends, and fun. The day was such a blessing to me. This was the man and the relationship I had been praying for. I can't wait to see where the next years take us!



CJ: With the topic of love, you landed on such a sentimental topic. What made this topic so special for you?

## My Love Story

The Planning

The Numbers

The Party



It's finally time for the big day! It went by so fast, but it was amazing to see all of the planning come to life through the beautiful venue, vibrant colors, and all our friends and family gathered.

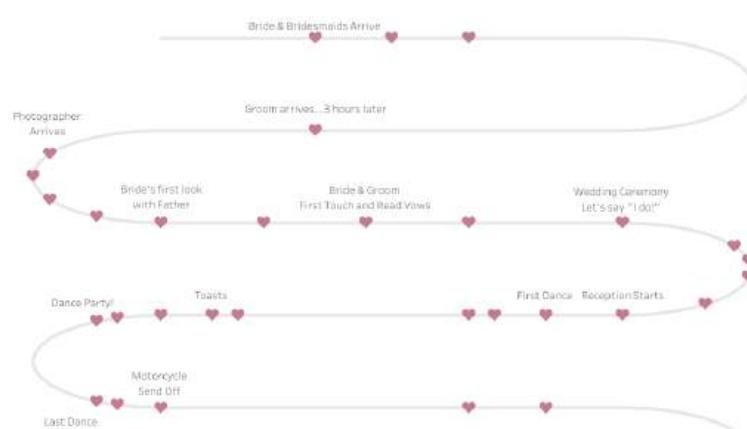
For my venue, I opted for a Modern House. It had everything I envisioned with an outdoor ceremony location, bridal suite, and coordinator.

I wanted classic, modern, and sweet. So my overall wedding theme was Romantic. The venue, florals, and colors made this come to life.

I chose shades of Purple for my color scheme, with lilac, lavender, and plum displayed in the bridesmaid dresses.

### Our Wedding Day Timeline

Did the day go exactly like I planned? Absolutely not. But it was still the most incredible, love-filled day that I will never forget!



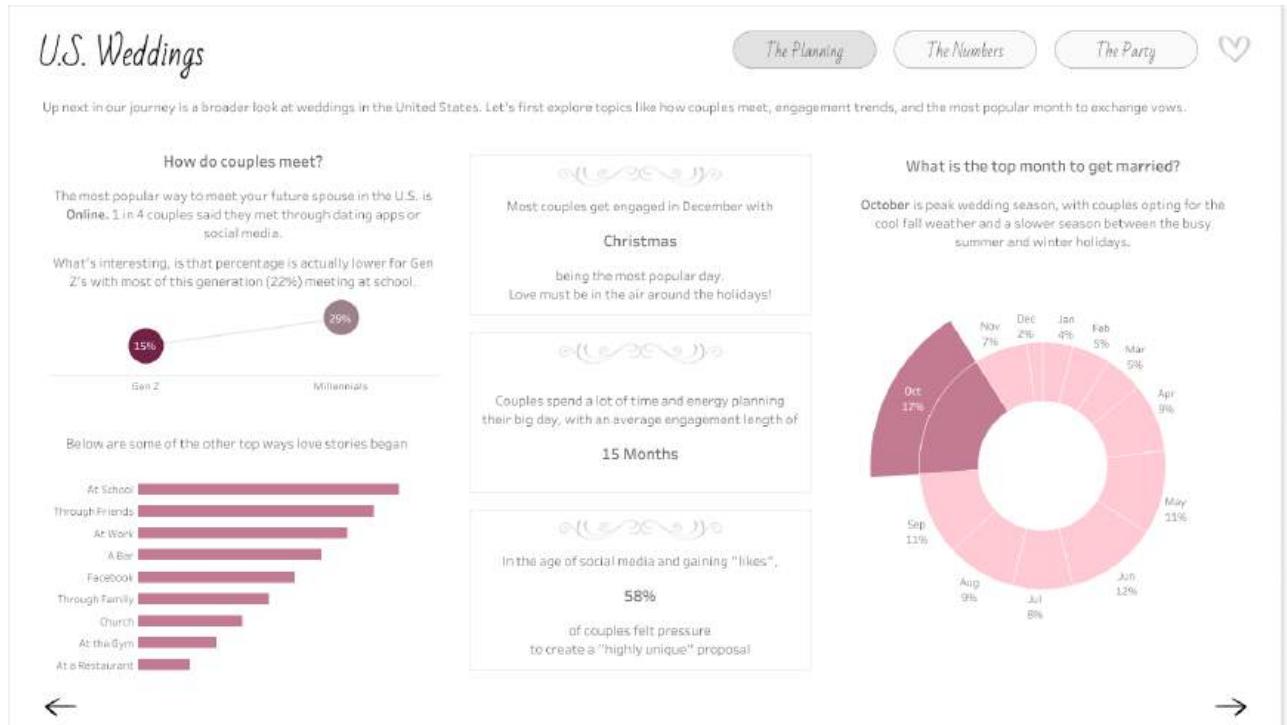
flip the page for U.S. Weddings →

J: Love was in the air! I had recently had my own wedding back in October 2022 and had been wanting to visualize some of the wedding data I collected through the wedding planning process. The topic of love was the perfect reason for me to set aside time to build a viz that had already been on my mind. My wedding day was such a beautiful day and a blessing. It was more than the planning and the data. It was a day full of love, family, and friends. I wanted to express and relive some of that through my IronViz visualization.

CJ: From a design perspective, I loved the consistency in button style and heart icons, light color palette, and framing of each visual with the "table cards", charts and labels. They really help elevate

the visual and reduce clutter between text and chart. Were there any stand out design decisions you think played well into your entry?

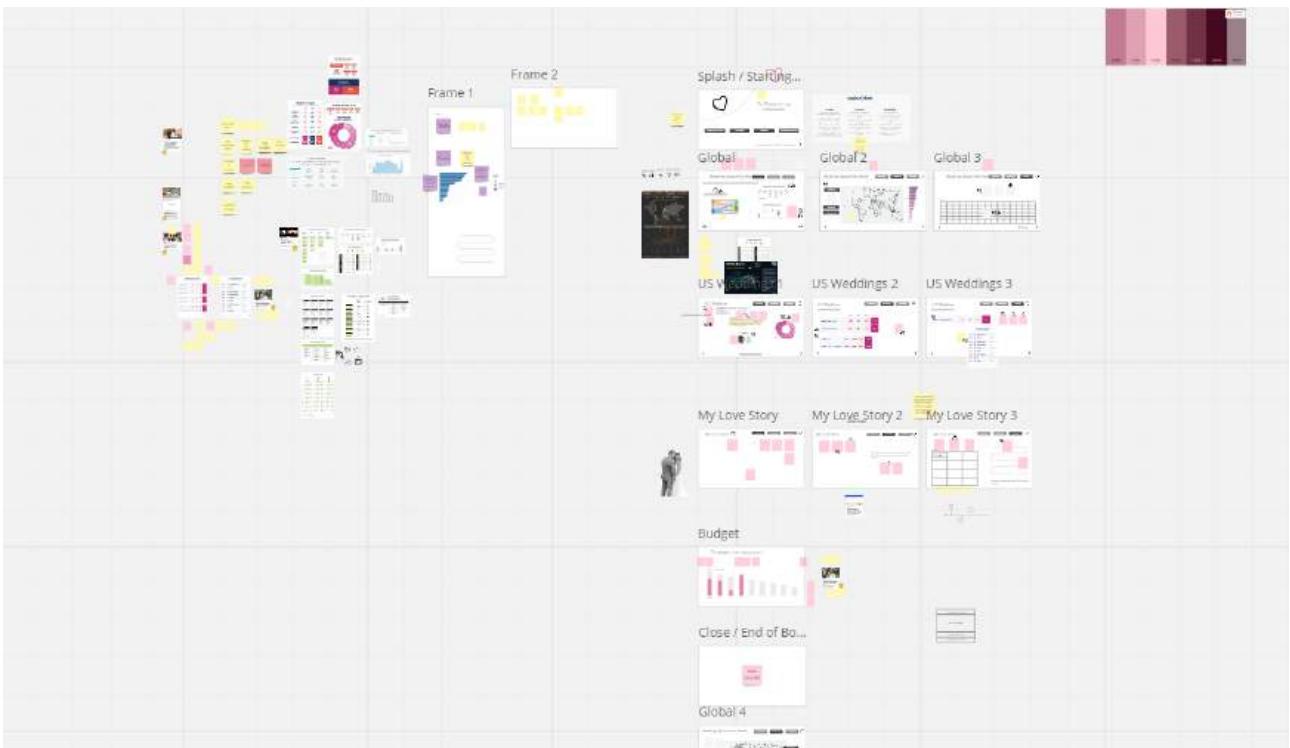
J: From the beginning, I strived to use data viz and dashboarding best practices. I've learned a lot through building dashboards in consulting and what is helpful to end users. However, I still wanted the design to be beautiful. It was a balance I tried to achieve through consistency, a minimalist design, and clear visuals.



- Consistency –
  - Same flow of topics at each level. Going from micro to macro (my wedding, U.S. Weddings, to Global Weddings) I kept the story flow the same so viewers could follow along and not get bogged down with too much information. Each level hit The Planning (how couples met & engagement), The Numbers (the costs), and The Party (the wedding day theme, reception, and vendors).
  - Titles, subtitles, buttons, and KPIs are in the same exact place with a consistent design on every page so end users become familiar with reading and navigation
- Minimalist Design –
  - Color – I wanted to use color wisely, choosing a love inspired mauve pink as the pop of color against the white and grays to draw the viewer's eye.
  - White space – I added lots of padding in Tableau to create balance of text and whitespace
  - Accessible text – while it's more popular to use a smaller text, I opted for size 12 font along with size 14 font in the tooltips to be more accessible and available to the broader community and #datafam.
- Clear Visuals – while I wanted to have a few sophisticated charts (like the sankey), most of my charts were basic (bars, lines, maps). I've learned that there's a reason basic is so popular! They are intuitive, simple, tell the story, and can still be beautiful.

CJ: In the [herdata video](#) you mention the wedding guest book style to your design. I personally love that it segments the story into chapters as well as choosing a select few appropriate charts for each "page". Can you talk to us a little more around the wireframing you did in Miro?

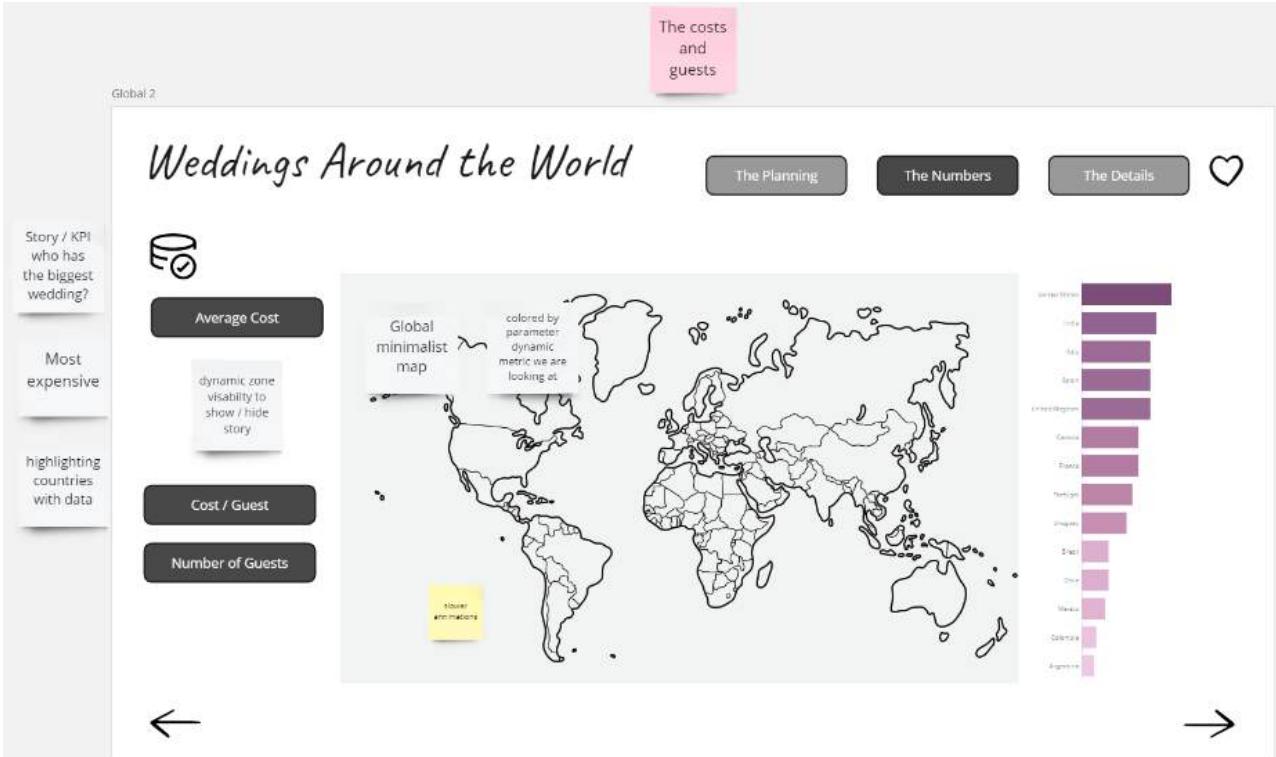
J: I love when I can design my data viz in a way that is symbolic of the topic I'm sharing. I did this a few years ago with my gardening viz; I wanted the dumbbell charts to look like rows of plants growing out of the dirt on the dark brown background. So I tried to do this again with my IronViz submission, through making it look like an interactive wedding guest book.



Using a whiteboarding tool like Miro really helped me with this. I'll try to keep it short because similar to Tableau, I could talk about the benefits of Miro all day long lol. It gave me a space to:

- Save links, images, and data from my research
- Highlight and narrow down the data and topics I wanted to focus on for my viz
- Test out a vertical vs horizontal design flow
- Test color palettes and fonts
- Create the story flow and wireframes with example visuals for every single page
- Keep up with next steps to do (or change) in Tableau Desktop

Here's an example of the "Global Weddings – The Numbers" tab wireframe. I used Miro to make a plan and low fidelity wireframe to help guide what the actual dashboard would look like and function.



CJ: From an analytical perspective I really enjoy the ability to highlight the weddings around the world page as well as to drill down supportive information on the side. How do you go about designing a

specific page and the balance between what is vital information to highlight, compared to complimentary information that may be useful to a user.

J: This is just a balance I've learned over the years. I remember the first Makeover Monday I did, I tried to get all of the data into one viz, not focusing on highlighting the story or what I wanted my end users to take away from it. Now I know that just because I have all of the data doesn't mean I need to viz it all. For this IronViz submission I started by exploring the data myself. Drag and dropping it into Tableau, analyzing it, asking my own questions of it. And then when I found the bright spots, I designed to best tell that story – with visuals and supporting text to help guide the users.

## Weddings Around the World

The Planning

The Numbers

The Party



The cost to tie the knot can range across the world. Click through the cards on the left to see how each country ranks when it comes to overall expense, number of guests, and the cost per guest.



CJ: From a storytelling perspective you chose to go through your personal story, and then open it up to US and global analytics before reflecting on the budget. What determined this navigation of the story? How did you find balancing the emotional story of your wedding with the financial context behind weddings?

J: I actually originally designed it to go from macro to micro, thinking starting from a global lens would be more engaging and create a better splash experience. However, after showing it to some of my colleagues for feedback, they suggested that I start with my own story. The theme was love and that's such a personal topic. Why not start with why my topic was so important to me? So I swapped it to go from micro to macro instead. I also wanted an element where viewers could personally interact with the dashboard. After sharing my viz with a friend, she suggested creating a page where you could see costs based on your own wedding budget, and thus the "Budget: Let's help you Plan!" page was born.

This is why I believe feedback from people you trust is so important.

CJ: The budget planning page is a wonderful way to add interactivity and user curiosity to the visual to really drive the users own personal budget. What were some of the complexities behind the ability to remove and add vendor finances into the page?

## Budget: Let's help you plan!



There are so many big and small decisions to make when it comes to planning a wedding. Knowing what to expect cost wise can relieve a lot of stress. If you're a future bride or groom, use this tool below to see approximately how much to spend on each vendor based on your own budget.

1) Enter your total budget

2) Select your average cost distribution source

3) Check or uncheck any vendors you want to include or exclude

4) Use the final budget breakdown to compare vendors as you plan your dream wedding day!



close the story →

J: This view was fun to make! For the viz, I imagined having an interactive list where a bride or groom could check off which vendors they do or don't want to hire for their wedding, and therefore adjust the budget breakdown. I could have done this with a typical field dropdown filter, but I wanted it to look prettier than that. Getting that checklist is easier said than done because you can't apply both the "Add" and "Remove" value from Set action to the same sheet. Other people in the Tableau community have found ways around it, but the examples I saw require at least 2 clicks or multiple Sets (Add and Remove Options in the tooltip menu, Add and Remove Buttons, Using multiple sets)

I found a Tableau hack way to make it happen with containers, one Set, two Set Actions, and hidden sheets. I hope to write a more thorough How-To blog on this one day soon!

CJ: Whilst technical expertise isn't sadly within the remit of the IronViz, I love that you have included some great advanced features within your visual, around dynamic zone visibility and some non standard "out of the box" chart types. Do you use IronViz to showcase some technical flair or were they a conscious decision for the flexibility in analysis and navigation?

J: I like to use IronViz as an outlet for the culmination of all of the things I've learned about Tableau throughout the year. The qualifier is a big challenge so I like to use it to practice Jordan Peterson's idea of having one foot in comfort and the other foot in chaos. To me, IronViz is a balance of getting into flow state, practicing what I know/do well like design, those basic/intuitive visuals, and story flow; but also introducing chaos, like integrating new Tableau features or creating visuals that I don't get to do often for business dashboards (like my wedding timeline and vendor bump chart). I like that bit of challenge; it's like solving a new puzzle.

CJ: I really liked your follow up visual comparing Iron Viz to a marathon effort. A wonderful analogy of time, effort and commitment. What did you want individuals to take away most from the visual?

# Competing in the IronViz Qualifier is like Running a Marathon

Every year, Tableau unveils the IronViz Qualifier contest, and the thrill kicks in. If you decide to take the plunge, the next 4 weeks are a whirlwind of brainstorming your topic, scouring the internet for data, and immersing yourself in building this viz you've imagined in Tableau. All of this effort is geared towards securing a coveted Top 10, or better yet, a Top 3 spot for a chance to shine on the stage at the Tableau Conference IronViz showdown. You've invested your time, blood, sweat, and tears into this viz. So, when your creation doesn't grace that Top 10 list, emotions can run high. It's disheartening to see something you poured so much into, fall short.

Well I'm here to tell you that just submitting a viz, pushing your limits, and daring to go for it is an achievement in itself!

You are a part of the 0.01% who take on this challenge!

Out of the 8 Billion people in the world, only 1.1 Million run a marathon each year

Out of the 3 Million authors on Tableau Public, only 196 submitted a 2024 IronViz Qualifier entry



Remember, win or learn you can't lose!

J: In 2019 when I got #4 in the Art themed Qualifier, I remember just feeling so crushed and defeated. I literally cried and even wrote a blog post about my experience **Not Quite an IronViz Finalist**. I've learned now that defeat and disappointment isn't what I should have taken away from the experience, putting my viz out there and getting Top 10 is truly something to be proud of! But that doesn't mean the sad and negative feelings aren't a reality for those who participate and didn't make the cut. So I wanted to send some data-backed, words of encouragement out into the world of IronViz.

My husband and his family are big runners. They do marathons all the time. I think it's crazy and seems way too hard. So I asked my husband why on earth he chooses to do something like that? He said it's because of the challenge. Running a marathon is so challenging that only .01% of the world takes it on! He wants to be a part of that .01%.

After another year of competing in the IronViz Qualifier, and this time not even getting in Top 10, the data nerd in me wanted to do the math... am I also in that .01%?

Yes! Out of the 3 million authors on Tableau Public, only 196 submitted a 2024 IronViz Qualifier entry. If you are one of those 196, you are the .01%! Be proud and encouraged by that!

CJ: Are you happy to share your feedback? What did you think of it and is there anything on reflection you would change about your visual?

J: Yes I'm happy to share my feedback. At first, reading it was hard. It was easy to only focus on the negative, but there were some positives in there too. Now I know of some things that are judged that weren't as obvious from the criteria. For example, I got points off because "some unused fields were not hidden." Based on the feedback, I also think I tried to cram too much into the viz, with the global, U.S., and my story. Next time, I'll narrow my scope a bit.

I really thought my viz would go farther in the Qualifier, I gave it my all. However, I can't imagine how hard judging this competition is. I know it's all volunteer based so I'm thankful for the people who give up their time for that.

Overall I wouldn't change my topic or story. Feedback is hard, but there's always growth in challenge!

CJ: What are you looking forward to most in the community this year?

J: Honestly I'm looking so forward to Tableau Conference! It's nice to chat here and there and see people on X and LinkedIn, but it's just so much better hanging out with everyone in person!

CJ Round-up:

Grateful for Jade making time to be able to share with us the highs and lows of Iron Viz, the growth moments, and of course, her passion, both in creating such a stunning visual but also her love for

Tableau. Honestly, this visual from Jade is one of my favourites, it takes real attention to detail to get padding, alignment, white space and color so perfectly balanced.

Well done to all those who entered, cherish the feeling of being part of the .01%.

LOGGING OFF,

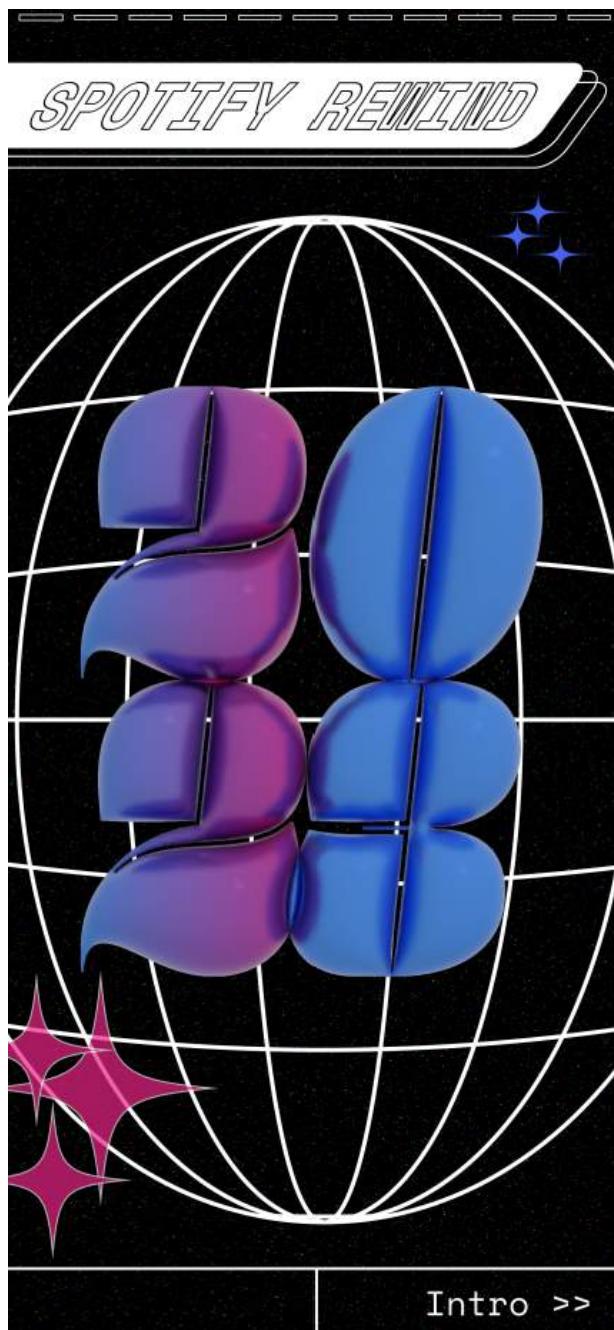
CJ

## SPOTIFY WRAPPED DATA WITH BEN MANGEL

Hi all, It's great to be able to have another guest blog today. Calling all music fans with this one. I am delighted to have Ben Mangel join the site today to be able to share his love for music, data prep and data visualisation through his take on spotify wrapped. I hate to spoil the surprise but he offers the solution both in Alteryx and Python, so pick the tool of your choosing! You can follow along using Ben's repo [here](#) Ben has a [creative portfolio](#) with a sprinkle of community initiatives and personal projects that are well worth checking out.

The screenshot shows Ben Mangel's Tableau Public profile. At the top, there is a circular profile picture of Ben Mangel. Below it, his name "Ben Mangel" and title "Data School Consultant at The Information Lab | Hamburg, Germany" are displayed. A "Follow" button is present. Below this, there are navigation links for "Vizzes 24", "Favorites 50", "Following 17", and "Followers 62". The main area displays eight data visualization examples:

- Spotify Rewind 2023** by Ben Mangel: A dashboard featuring a chart titled "MoM 2023 | Week 10 | Decline in Meat Production".
- MoM 2023 | Week 10 | Decline in Meat Production** by Ben Mangel: A line chart showing trends over time.
- ARTIST PROFILE | MAC MILLER** by Ben Mangel: An artist profile card for Mac Miller with statistics: 8,699,339 plays, 86 songs, and 12 albums.
- KPI Cards Map Layer Practice** by Ben Mangel: A dashboard titled "SUPERSTONE SALES PER SUB-CATEGORY IN 2022" showing multiple KPI cards.
- Journey through the World Wide Web** by Ben Mangel: A dashboard showing a Google search interface.
- Bloody Breaking Bad (Dashboard Week Day 4)** by Ben Mangel: A dashboard titled "Bloody Breaking Bad" featuring a dark-themed image.
- British Optimism Survey (Dashboard Week Day 3)** by Ben Mangel: A dashboard titled "GLASS HALF FULL?" showing survey results.
- Crimes in GU10 4JU UK (Dashboard Week Day 1)** by Ben Mangel: A dashboard titled "CRIMES IN GU10 4JU UK" showing crime data.



## Intro >>

Ben, over to you.

Spotify Wrapped is an exciting phenomenon, my friends and I were talking about it, showing us our musical besties and guilty pleasures. Also at least my Instagram and tiktok were full of people showing who and what and also important how much they have listened to their favourite artists in the last year. You can compare, sideeye other persons wrapped, admire and enjoy this throwback of the past year and live a little bit in the memories you've made in these months. But what if you can go further, learn identify some interesting patterns and, most important, tell more stories about your life. This blog aims to give you an instruction to create your own Spotify Wrapped in Tableau. I'll give you a walkthrough through the resources I used and how we build everything together. What do we need?

- Data from Spotify
- Data preparation
- Data Visualisation

Data from Spotify:

# Download your data

By using our Download your data tool, you can request a copy of your personal data. You may download three different packages of data, either separately or all at once. Please see below what the packages include (if applicable to you) and choose what you want to download.

For more details about the data categories in the packages, please see [Understanding My Data](#) or [contact us](#).

## Account data

- Playlists
- Streaming history for the past year
- A list of items saved in your library
- Search queries
- No. of followers, accounts you follow, and blocked accounts
- Payment and subscription data
- User data
- Customer Service History
- Family Plan data
- Inferences
- Voice input
- Podcast interactivity
- Spotify for Artists data

 Preparation time 5 days

Select Account data

## Extended streaming history

Extended streaming history for the lifetime of your account, including track information, and when and how you streamed content.

 Preparation time 30 days

Select Extended streaming history

## Technical log information

Technical log information that we have collected about your account to provide and troubleshoot the Spotify service.

 Preparation time 30 days

Select Technical log information

**Request data**

Spotify collects a whole bunch of data about you. And for GDPR reasons they have to offer you a way to get those personal data about you they've collected. The process is quite simple:

1. You log in with your Spotify account and go to the [privacy center](#)
1. Select the second option to get your full streaming history
1. Confirm the link in the email Spotify sends you
1. Wait up to 30 days -> probably the hardest part in the whole process

Spotify needs some time to prepare the data for you, this is a little bit frustrating, but the way it goes. For this you will want the extended streaming history for an annual summary so make sure to check the additional boxes as part of the request.

## Data Preparation

I have done the whole data preparation in Alteryx + Python and fully in Python. Depending on which tool you have access to or more comfortable to use you can go on. We can divide the whole process in 4 steps:

1. Load the raw data
1. Prepare the data initially and create new features
1. Get the metadata from the Spotify API
1. Bring everything together and output to Tableau

## Load the raw data

The data you will get from the request to Spotify are in a .zip file containing a folder with a ReadMe.pdf and a lot of .json files. To parse the data I use Python. The function is basically going to a directory you pass into the function and writing every filepath + filename into a list, that starts with "Streaming\_" and is a .json file. So I make sure to get just the correct files. Then I iterate over these filepaths and load the json as pandas dataframes. Every file is its own dataframe and those dataframes are stored in a list. The concat function in pandas allows us to "union" those dataframes in the list, this works because we assume that every file has the same structure.

```

def union_json_files(directory):
 """
 Union multiple JSON files from a directory into a single dataframe.

 Parameters:
 - directory: Path to the directory containing the JSON files

 Returns:
 - DataFrame containing unioned data
 """

 # List files in the directory that start with "Streaming_" and end with ".json"
 file_paths = [os.path.join(directory, file) for file in os.listdir(directory)
 if file.startswith("Streaming_") and file.endswith(".json")]

 # Initialize an empty list to store dataframes
 dfs = []

 # Iterate over each file path
 for path in file_paths:

 with open(path, "r") as file:
 data = json.load(file)

 # Convert the JSON data into a dataframe
 df = pd.DataFrame(data)

 # Append the dataframe to the list
 dfs.append(df)

 # Concatenate all the dataframes together
 result_df = pd.concat(dfs, ignore_index=True)

 return result_df

```

The result is written back as a .csv to use in the next steps.

### Prepare the data initially

We have to do some basic transformations before we can use the data. At the end of this chapter datatypes are correctly set, new features are calculated and the data are filtered. Both can be done in Alteryx and Python and lead to the same result. The tasks we are facing are:

1. Convert the ISO date timestamp to a date type we can use
1. Calculate based on the date a starting timestamp
1. The timestamps are in UTC time, so we have to handle our own timezone
1. Filter NULL values
1. Filter rows that are not necessary for this analysis
1. Flag if a row contains a Song or Podcast

## Alteryx

1. Input the .csv file, you've created with the python function
1. Use a DateTime Tool to convert the ts column into a date The format is an ISO Date and can be parsed like this: **yyyy-MM-ddThh:mm:ss**
1. Add, depending on your timezone, hour(s) to the new timestamp column This column is your end\_date, so the timestamp where you've finished the track or episode
1. I prefer working with an start date, that seems more logically to me, so we have to subtract the ms\_played from the end\_date. Alteryx does not allow to use milliseconds, so we use this formula: `DateTimeAdd(, / 1000 * -1, "seconds")`
1. Filter every row, where episode\_name AND master\_metadata\_track\_name are NULL, this can happen, I guess those are tracks or episodes which are deleted from Spotify
1. To get the information if it's a podcast or not, we can check, if master\_metadata\_track\_name is NULL, then we can be sure it is a podcast, otherwise it is a track
1. Currently all data Spotify has provided are used, but for our purpose of a Wrapped, we just need the last 2 years, and because the next step is to call the Spotify API, we'll reduce the processed data here by filtering on these last 2 years (for the sake of ease, I have filtered everything from 2022 onwards, depending on your timeframe, this has to be adjusted)



This can also be done as a function in Python

```
def prefilterData(df: pd.DataFrame) -> pd.DataFrame:
 """

 Goal:

 - Creates timestamp
 - add or subtract hours depending on the users location
 - Prefilters data to the current and last year
 - add start date feature
 - rename some columns
 - create feature to indicate podcasts

 Params: dataframe with spotify data from the union_json_files function

 Returns: modified df
 """

 # Convert ts in UTC timestamp
 df['end_date'] = pd.to_datetime(df['ts'])
 df = df.drop('ts', axis = 1)

 def convert_UTC_toLocal(df):
 """
 Gets the local time of the user and converts the UTC time into the timezone of the user
 """

 local = pytz.timezone(str(get_localzone())) # Convert timezone to string
 df['end_date'] = df['end_date'].dt.tz_convert(local) # Use tz_convert instead of tz.convert
 return df

 df = convert_UTC_toLocal(df)

 # filters everything, where col episode_name AND master_metadata_track_name are null
 df = df.dropna(subset=['episode_name', 'master_metadata_track_name'], how='all')

 # Filter everything which is older than current year and last year
 lastYear = df['end_date'].max().year -1

 df = df[df['end_date'].dt.year >= lastYear]

 # Add start_date as a feature
 df['start_date'] = df['end_date'] - pd.to_timedelta(df['ms_played'], unit= 'ms')

 # Get the Track ID from the URI
 df['uri'] = df['spotify_track_uri'].str.split(':').str[-1]

 # Rename columns
 df.rename(columns={
 'master_metadata_track_name' : 'track_name',
```

The time zone is automatically recognised by the local time zone of the machine and processed in the function. Moreover it takes care of the last 2 years of data depending on the maximum date in the dataset.

## Get Data from the Spotify API

To enrich our analysis we will include additional data which come from the Spotify API. APIs offer us a way to ask for data in a structured way and get a response from the server we've asked. The API Spotify offers is quite good documented and has various endpoints (questions we can ask). For this case we'll use the tracks endpoint, which gives us information about a single track.

## Authentication

But first of all, to use the API, we have to authenticate. We need a Token, which we get from the API itself. Therefore we use the token endpoint. To do this, we need to POST a couple of information's, otherwise the API can't help us. We need:

```
1. return df
1.
```

Those information come from a Spotify App. Not the one on your smartphone, we are talking about a developing app, you can create on the developers page from Spotify, when you navigate to your dashboard: <https://developer.spotify.com/dashboard> (you have to be logged in). There



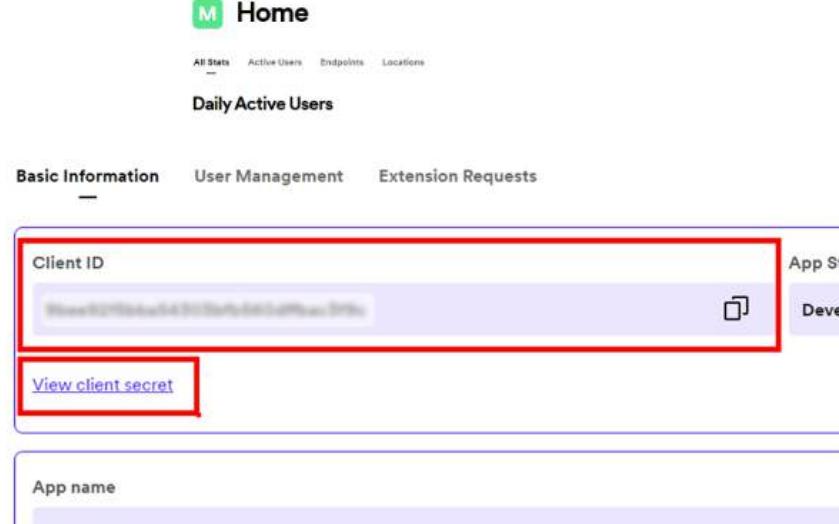
## Dashboard

you can create an app:

need to go to Settings, then you'll come to a page where you can copy your client secret and client ID



In the created app you



Great, now we are good to go to request

To automatically generate the API token, I've built a macro, which uses the gathered information and gives back the token:



You just need to change the client secret and the client id with yours. The information are appended and then fed into a download tool, which uses the following data with this configuration:

| Record | client_id      | client_secret      | Token URL                                                                                   | grant_type         |
|--------|----------------|--------------------|---------------------------------------------------------------------------------------------|--------------------|
| 1      | your client id | your client secret | <a href="https://accounts.spotify.com/api/token">https://accounts.spotify.com/api/token</a> | client_credentials |

What comes back is parsed and cleaned, so the output of the macro is just the token, ready to use for new API adventures.

### Python

Now let's look at how to do the same using python,

Same input and same approach, we make use of the request library and use this function, with a small safety net, if the call did not work, because of not working inputs.

```

def getAuthentication(clientID: str, clientSecret: str, tokenURL: str, grantType: str) -> str:

 """
 Goal: Gets the OAuth token from the Spotify API
 Params: come from the Spotify App created here: https://developer.spotify.com/dashboard
 Returns: OAuth Token in access_token variable
 """

 data = {
 "client_id": clientID,
 "client_secret": clientSecret,
 "grant_type": grantType
 }

 response = requests.post(tokenURL, data=data)

 if response.status_code == 200:
 access_token = response.json().get('access_token')
 return access_token
 else:
 print("Failed to obtain access token")
 print("Status Code:", response.status_code)
 print("Response:", response.text)

```

### Data Prep for API call

Before we are ready to make the API call, we need to figure out, what the API is expecting from us and prepare it. The endpoint we are using expects us to give per API call max 50 comma separated Track IDs. There is also the option to request every track in a single request, but this would take ages and it is possible, that the access token isn't valid anymore (expires after 1h). So we need to build batches of 50 track ids.

1. The track ids are hidden in the Spotify URI, so we need to parse them out there
1. We need to know how many batches we will have if 1 batch contains 50 ids
1. Build these batches together



1. We filter every NULL value (the podcasts)
1. Use the Text to Columns and the Select tool to split the Spotify URI column and get the track id
1. Then we count the number of rows, divide this by 50 (our batch size) and round this up with the ceil() function
1. Now we know how many batches we need, so we append this number to the original data and use a Multi Row Formula tool to assign groups to every row. From the previous step we know how many groups we have for our data, when every group should not have more than 50 members
1. Last step is to concatenate the track ids based on this groups

Output is 1 column with looooong strings containing up to 50 track ids. **Python** We do exactly the same as in Alteryx with this function:

```
def prepareSpotifyURIGroupsForAPI(df: pd.DataFrame) -> pd.DataFrame:
 """
 Goal: Create batches of 50 track ids that can be fed into the API
 Params: main df, output from preFilterData()
 Returns: df with 1 column containing comma seperated track ids
 """

 # get unique Spotify URI and remove NULL values
 uri = pd.DataFrame(df['uri'].dropna().unique()).rename(columns={0: 'uri'}, inplace=False)

 # Create URI groups
 numberOfRowsRecords = len(uri)

 count = math.ceil(numberOfRecords / 50) # divide by 50 to get the number of groups which have the
 length of 50

 uri['group'] = 0
 groupCounter = 0

 for i in range(numberOfRecords):
 groupCounter += 1

 if groupCounter > count:
 groupCounter = 1

 uri.at[i, 'group'] = groupCounter

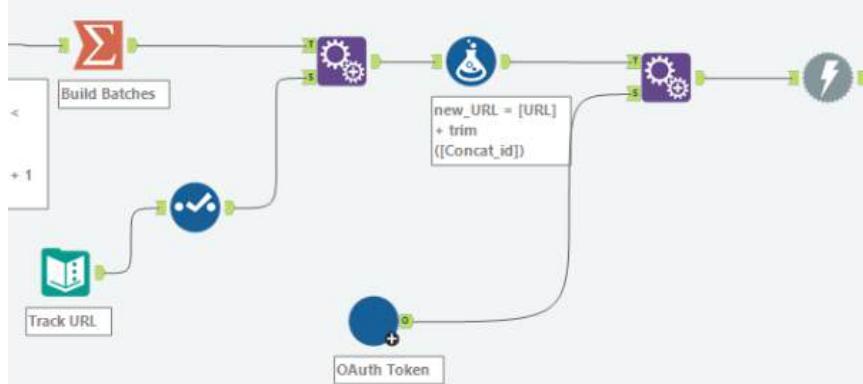
 # concatenate groups together
 uri = uri.groupby('group')['uri'].apply(lambda x: ','.join(x))

 return uri
```

## Make API call

### Alteryx

We have everything together, the API token, the requested data, now this needs to be assembled in 1 table, therefore we append all of these information together and feed it into the download tool:



The configuration for the download tool looks like this:

The answer comes as JSON and can be parsed with the JSON Parse tool. In the following steps I reshape the data to get a table containing the features I am interested in, in a table format with 1 row per track ID and columns for every feature. The goal is to join these information at the end back to the main table.



I also built a test, which throws an error, when the status of the response is not ok, so I am aware of those. You may think, why is it outputted as a csv, when he says he wants to join the data back to the main table? It is just for the sake of ease during developing, the API call takes some time and I do not want to run this every time, it's like manual caching. I can just deactivate the container with the API call and work with the downloaded data.

Python

To make the API call, we use again the request library, and assemble the variables we need together:

```
def trackRequest(prepared_uri: pd.DataFrame) -> list:
 """
 Goal: Access the Spotify API to get metainformation about the tracks
 Params: df of track ids, comma seperated, 50 ids per row
 Returns: API response as a list containing json with desired information
 """

 baseURL = 'https://api.spotify.com/v1/tracks?ids='
 urls = [baseURL + group_uris for group_uris in prepared_uri]
 bearer_token = getAuthentication(clientID, clientSecret, tokenURL, grantType)

 headers = {
 "Authorization": f"Bearer {bearer_token}"
 }
 responses = []
 for url in urls:
 response = requests.get(url, headers= headers)
 if response.status_code == 200:
 responses.append(response.json())
 else:
 print("Failed to access tracks endpoint")
 print("Status Code:", response.status_code)
 print("Response:", response.text)
 print(f"URL = {url}")

 return responses
```

The output comes as JSON, which we have to parse by ourselves. This function is iterating to the JSON and parsing the information I am interested in

and put everything into a dataframe:

```
def parseAPIResponse(list: list) -> pd.DataFrame:
 """
 Goal: Parse the json from the API response and get only the track_id, duration and popularity.
 Other features can be extracted as well

 Params: list, that is the output from the getTrackMetadata function

 Returns: df with parsed features, the df is joinable to the main df (output from the
 prefilterData function)

 """
 track_info_list = []

 for i in range(len(list)):
 current_batch = list[i]['tracks']

 for j in range(len(current_batch)):
 track = current_batch[j]
 track_id = track.get('id', 'unknown')
 duration_ms = track.get('duration_ms', 0)
 popularity = track.get('popularity', 'unknown')

 # store results in dict
 track_info = {
 'track_id': track_id,
 'duration_ms': duration_ms,
 'popularity': popularity
 }

 # Append track infos to list
 track_info_list.append(track_info)

 track_df = pd.DataFrame(track_info_list)

 return track_df
```

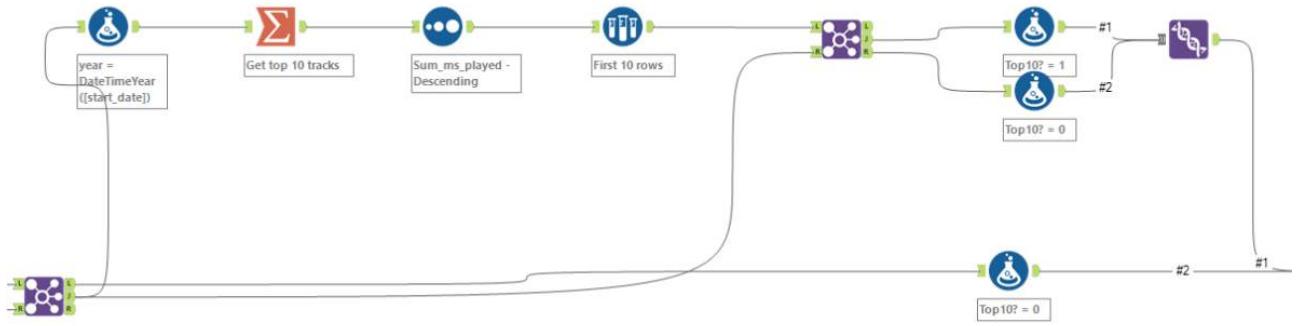
## Get Top 10 tracks

In my analysis, I want to compare the Top 10 tracks from the previous year, these can be calculated in Tableau, but it is easier to do this already in the data preparation. Therefore I identify the top 10 tracks per year, I have in my prefiltered table and group them per year, sort them by ms\_played and get the first 10 entries per group. The result is then joined back to the main table.

### Alteryx

In Alteryx we use first parse the year as a column to be able to group by this column, and then we use the Summarize Tool to group by track and year, sort the results and then take the sample tool to get the first 10 rows per year. The results get a Top10=True flag, every other track and the

podcast episodes get a false.



## Python

This function does exactly the same thing, of picking the top 10 rows per year. Because I do a left join, I am able to flag the non-matching results which have a NULL in the Top10 column with no matching columns.

```
def getTop10Tracks (df: pd.DataFrame) -> pd.DataFrame :
 """
 Goal: Get the top 10 tracks per year
 Params: pre-filtered df
 Returns: df with top 10 tracks per year
 """

 data = df

 df['start_date'] = pd.to_datetime(df['start_date'])
 # Extract the year into a new column
 df['year'] = df['start_date'].dt.year

 grouped_df = data.groupby(['uri', 'year'])['ms_played'].sum().reset_index()
 filtered_df = grouped_df.dropna()

 top_entries = (filtered_df.sort_values(['year', 'ms_played'], ascending= False)
 .groupby('year')
 .head(10))

 top_entries['top10?'] = True
 top_entries = top_entries.drop(['year', 'ms_played'], axis = 1)

 return top_entries
```

```
def prepareFinalDf(preFiltered_data: pd.DataFrame,
 metadata: pd.DataFrame,
 top10_tracks: pd.DataFrame) -> pd.DataFrame :

 """
 Goal: Join API response, Top 10 tracks and prefiltered data together
 Params: all the created df
 Returns: final df

 """
 # Join with metadata
 data_metadata = preFiltered_data.merge(
 metadata,
 left_on='uri',
 right_on='track_id',
 how = 'inner'
)

 # Join with top10 tracks
 final_df = data_metadata.merge(
 top10_tracks,
 how = 'left',
 on = 'uri'
)

 # Fill null values in top10? column for songs that are not top 10
 final_df['top10?'] = final_df['top10?'].fillna(False)
 print('Success! Your table has been created!')

 return final_df
```

## Export Result

The result of the Alteryx workflow can be outputted to a .csv file, in Python we need to call all the functions we have created with our parameters like this, and output the result of this dataframe.

```

Directory containing the JSON files
directory = r"directory_with_your_spotify_json_files"

Union the files
unioned_df = union_json_files(directory)
pre_filtered_data = prefilterData(unioned_df)

clientID = 'yourClientID'
clientSecret = 'yourClientSecret'
tokenURL = 'https://accounts.spotify.com/api/token'
grantType = 'client_credentials'

track_metadata = getTrackMetadata(df=pre_filtered_data,
 clientID=clientID,
 clientSecret=clientSecret,
 tokenURL=tokenURL,
 grantType=grantType)

top_10_tracks = getTop10Tracks(pre_filtered_data)

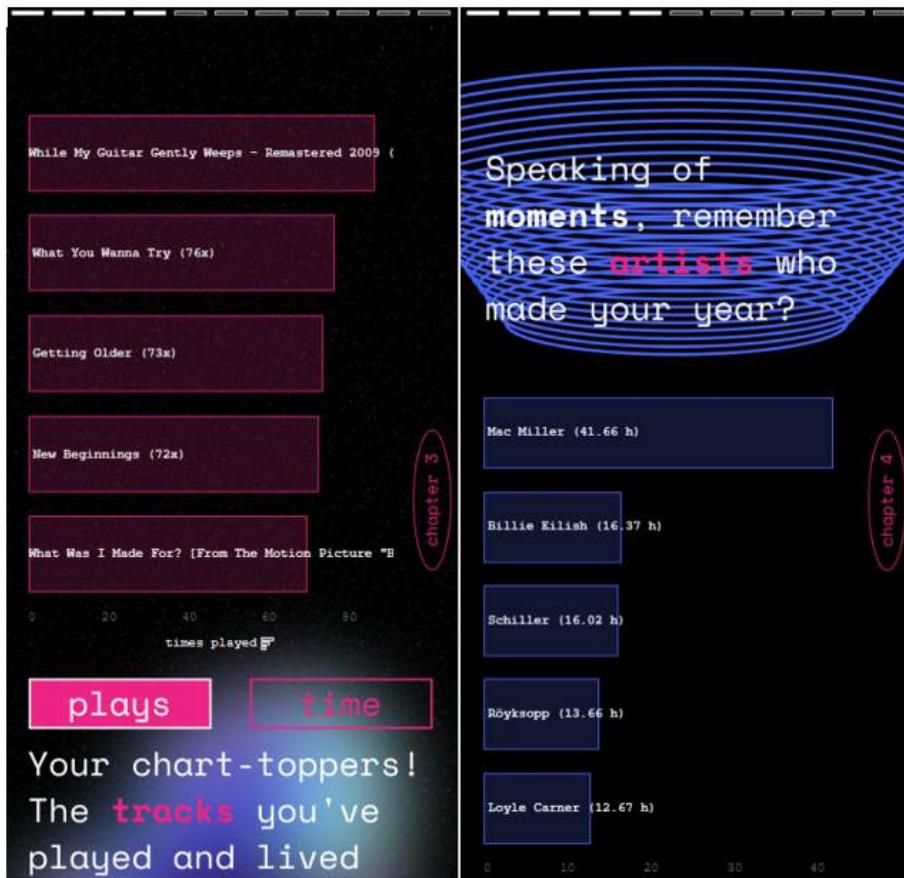
final_df = prepareFinalDf(preFiltered_data= pre_filtered_data,
 metadata=track_metadata,
 top10_tracks=top_10_tracks)

```

Now you are ready to plug these data into the Tableau workbook and take a look at your own data! Therefore you just need to direct the path of the datasource in the Tableau workbook in the repository to the output of your workflow and you should be good to go!

### Data Visualization

Goal of this section is to describe which decisions led me to the visualisations I've built, not describing the actual visualisations and techniques itself. It was very important to me, to enhance what Spotify offers, but not to overcomplicate it visually. Because of the size limitations and assuming the user not to have any knowledge about what he / she is expecting, I went for basic charts, which are as self-explanatory as possible and no filtering in or other interaction elements from Tableau. Moreover I wanted to keep the charts as visual light as possible, keep it simple regarding colours and keep those colours consistent. Therefore I stucked to the colour palette that I've used to create the backgrounds, and distinguish between podcasts and music with colour. I used navigation buttons to allow the user to navigate through the chapters or change charts. The position and size of the charts is mostly the same, depending on where the text in the chapter is. For the backgrounds I've used Adobe Illustrator and tried to stick to a Y2K look, with spacey elements, a little bit of grain on the visuals in a mix of 2D and 3D elements. To give the user an idea where he / she is in the story I used elements like the Instagram Story like bars at the top, and a stamp indicating the chapter the user is in. Generally I tried to keep it in a design language, that fits into the current fashion of internet culture, connected with elements that are known from social media. I like the design of the original Spotify Wrapped, but wanted to make my own version of it. To adjust the backgrounds you can use the AI or SVG files. From my experience, Figma is a great tool to work with SVGs and is also free to use!



There are many different features that can be extracted out of the data, also connected with the Spotify API, or enriched by data from the whole music industry. If you, dear reader, have other ideas I am happy to hear about them, moreover, if there are any things I missed, or can be done better, please let me know! **CJ Round-Up:** Wow, what a detailed write up of how to transform the spotify export into something useable for building your own spotify wrapped. Once again thank you so much Ben for sharing the content with us, you can find a copy in the git repository at the top of the page or following along to [Bens repo](#). **LOGGING OFF, CJ**

### INTRO TO VORONOI VISIBLE FRAMES (PYTHON)

Hi All,

Just a small code piece to share this week.

I've previously written about shot freeze frames and visibility in Tableau [here](#).

That version access' the freeze frames in a similar way but preps the data fit for Tableau by creating a path element and all the points of the polygon.

This version of code looks at how we can plot a series of freeze frames (not necessarily shot data) and then plotting the chart using python packages instead of the need for Tableau.

You can find a copy of the code in the **Github** link at the top of the page.

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from statsbombpy import sb
from mlsoccer import Pitch, Sbopen
import os

Get competitions
comp = sb.competitions()

Find a match_id required
match_id = 3906390
match_events = sb.events(match_id=match_id)

Filter frames to include only frames when a shot was taken
shot_frame_ids = match_events[match_events['type'] == 'Shot']['id'].unique()
frames, visible = Sbopen().frame(match_id)
print(frames)
frames_details = pd.DataFrame()

Specify the absolute path to the folder where the images will be saved
output_folder = os.path.abspath("output_frames")

Create the output folder if it does not exist
if not os.path.exists(output_folder):
 os.makedirs(output_folder)

Dataframe to store frame details
frames_details = pd.DataFrame()

Draw plot for each shot frame and save as an image
for idx, frame_id in enumerate(shot_frame_ids):
 frame_data = frames[frames['id'] == frame_id].copy() # Make a copy to avoid
SettingWithCopyWarning
 teammate_locs = frame_data[frame_data.teammate]
 opponent_locs = frame_data[~frame_data.teammate]

 # Check if 'visible_area' is available for the current frame
 if frame_id in visible['id'].unique():
 visible_area_data = visible[visible['id'] == frame_id].visible_area.values
 if len(visible_area_data) > 0:
 visible_area = np.array(visible_area_data[0]).reshape(-1, 2)

 # Draw plot
 p = Pitch(pitch_type='statsbomb')
 fig, ax = p.draw(figsize=(12, 8))

 # Plot Voronoi
 team1, team2 = p.voronoi(frame_data.x, frame_data.y, frame_data.teammate)
 t1 = p.polygon(team1, ax=ax, fc='red', ec='white', lw=3, alpha=0.4)
 t2 = p.polygon(team2, ax=ax, fc='grey', ec='white', lw=3, alpha=0.4)

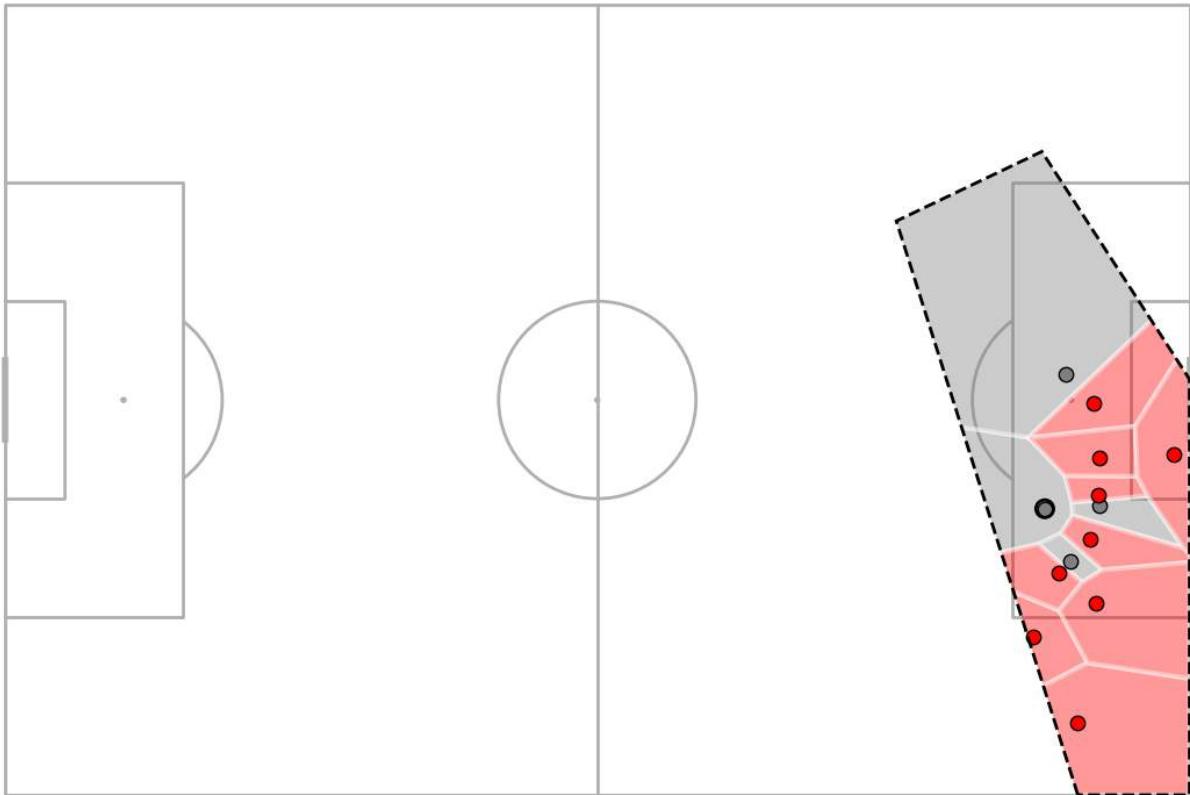
 # Plot players
 sc1 = p.scatter(teammate_locs.x.iloc[0], teammate_locs.y.iloc[0], c='grey', s=80,
ec='k', ax=ax)
 sc2 = p.scatter(opponent_locs.x, opponent_locs.y, c='red', s=80, ec='k', ax=ax)

 # Plot the first player with a slightly bigger size
 sc3 = p.scatter(teammate_locs.x.iloc[0], teammate_locs.y.iloc[0], c='None',
edgecolors='black', s=120, linewidths=2, ax=ax)

 # Plot the visible area with the same colors
 visible_polygon = p.polygon([visible_area], color='None', ec='k', linestyle='--', lw=2,
ax=ax)
 for pl in t1:
```

- ```
p1.set_clip_path(A brief explanation of the code,
p1.set_facecolor('grey') # Team color
p1.set_edgecolor('white') The necessary libraries are
imported:[numpy, pandas, matplotlib.pyplot, statsbombpy, mplsoccer, and os].
for p2,i
p2.set_clip_path(vtstate_polygon[i])
p2.set_facecolor('white')
p2.set_edgecolor('white')
The variable comp is assigned the result of calling sb.competitions(). This retrieves a
Dataframe containing information about available competitions.
3. The variable match_id is set to a specific match ID. This match ID is used to retrieve the events
for that specific match using sb.events(match_id=match_id).
4. The shot_frame_ids variable is assigned the unique frame IDs where shots occurred during the
match. This is done by filtering the events DataFrame to include only events of type 'Shot' and
plt.savefig(output_path,bbox_inches='tight')
else:
extracting the unique frame IDs.
5. The variable frames is assigned the result of calling Sbopen().frame(match_id). This retrieves
else:
print("No visibility information for Frame ID", frame_id)
the frames for the specified match.
6. The variable visible is assigned the result of calling Sbopen().visible(match_id). This
retrieves the visibility information for the specified match.
7. The output_folder variable is set to the absolute path of the "output_frames" folder. This is the
folder where the generated images will be saved.
8. If the output folder does not exist, it is created using os.makedirs(output_folder).
9. The current frame's data is retrieved from the frames DataFrame using frames == frame_id].
10. The code checks if the 'visible_area' information is available for the current frame
using frame_id in visible.unique().
11. If the 'visible_area' information is available, a plot is created using the Pitch class from
the mplsoccer library.
12. The Voronoi polygons for the two teams are plotted on the pitch using the teammate and
opponent locations.
13. The figure is saved as an image in the output folder with a filename containing the index and
frame ID.
14. The loop continues until all shot frames have been processed.
```

Overall, this code generates individual frames for each shot event in a match, plots the players' positions, visible area, and Voronoi polygons, and saves the frames as images in a specified output folder.



Going Further:

- Why not try plot this in Tableau? or data viz tool of your choosing.
- Build a small multiples to show a variety of different actions.
- Try create an [animated voronoi like Will Suttons](#).

As always, Let me know how you get on with this one. I can be reached on Twitter (X), @_CJMayes.

LOGGING OFF,

CJ

IT'S TRANSFER SEASON (PYTHON)

Hi All,

With the end of the transfer season, I thought it would be good to do a quick code share on how to look at transfer data.

For this tutorial we look at a previous season, 2022, and solely the winter transfers within the Premier league.

Filter by season:

Date of transfer:

Loans:

Transfers within the club

Show

ARSENAL FC

In	Age	Nat.	Position	Market values	Left	Fee
Jakub Kiwior	22	PL	Centre-Back	€25.00m	ITA Spezia Calcio	€25.00m
Leandro Trossard	28	BEL	Left Winger	€38.00m	ENG Brighton	€24.00m
Jorginho	31	ITA BRA	Defensive Midfield	€15.00m	ENG Chelsea	€11.30m
Arthur Okonkwo	21	ENG GHA	Goalkeeper	€2.00m	ENG Crewe Alexandra	End of loan Jan 15, 2023

Average age of arrivals: 25.5 Total market value of arrivals: €80.00m Expenditure: €60.30m

Here is the website screen shot of Transfermarkt and the filters we are looking at of Winter and year 22/23.

Notice that this is a little more challenging to previous weeks because we have the club name now in a banner, and the associated details below.

Out	Age	Nat.	Position	Market values	Joined	Fee
Arthur Okonkwo	21	 	Goalkeeper	€2.00m	  Sturm Graz	loan transfer
Marquinhos	19		Right Winger	€7.00m	  Norwich	loan transfer
Albert Sambi Lokonga	23	 	Central Midfield	€12.00m	  Crystal Palace	loan transfer
Cédric Soares	31	 	Right-Back	€2.00m	  Fulham	loan transfer

Average age of departures: 23.5

Total market value of departures: €23.00m

Income: 0

Transfer record: **€-60.30m**



ASTON VILLA

In	Age	Nat.	Position	Market values	Left	Fee
Jhon Durán	19		Centre-Forward	€17.00m	  Chicago	€16.64m
Álex Moreno	29		Left-Back	€10.00m	  Real Betis	€13.50m
Viljami Sinisalo	21		Goalkeeper	€500k	  Burton Albion	End of loan Jan 2, 2023
Bertrand Traoré	27		Right Winger	€10.00m	  Basaksehir	End of loan Jan 31, 2023

Average age of arrivals: 24.0

Total market value of arrivals: €37.50m

Expenditure: **€30.14m**

We also have sections that are whether the play was “in” or “out”, i.e arriving or departing the club.

What does the code do in general?

This Python script is a web scraper designed to extract information about Premier League football transfers from the Transfermarkt website. The script utilizes the requests library to send HTTP requests, the BeautifulSoup library to parse HTML content, and the pandas library to organize the collected data into a DataFrame. The target information includes details about incoming and outgoing transfers for each Premier League team, such as player names, ages, nationalities, positions, market values, and transfer fees.

The script begins by sending a request to the specified URL, which contains data on Premier League transfers for the 2022 season. It then extracts the list of teams from the page and filters out any empty strings. After that, it defines a function (extract_player_data) to extract player information for a given team, considering both incoming and outgoing transfers. The script then iterates through each team, finds the relevant HTML elements for incoming and outgoing transfers, and uses the function to extract and store the player data in a list (all_player_data).

Finally, the script creates a pandas DataFrame from the accumulated player data, defines column names, and prints the DataFrame to the console. Additionally, it exports the DataFrame to a CSV file named ‘transfer_data.csv’ in the current working directory, providing a structured and tabular representation of the Premier League transfer information.

```
import requests
from bs4 import BeautifulSoup
import pandas as pd

headers = {
    'User-Agent': 'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537.36'
}

# Initialize empty lists to store player information and teams
all_player_data = []

# Initial page
base_url = "https://www.transfermarkt.co.uk/premier-league/transfers/wettbewerb/GB1/plus/?saison_id=2022&s_w=w&leihe=1&intern=0&intern=1"

pageTree = requests.get(base_url, headers=headers)
soup = BeautifulSoup(pageTree.content, 'html.parser')

# Extract teams from the page and filter out empty strings
teams = soup.select('h2.content-box-headline a')
teams_list = [team.text.strip() for team in teams if team.text.strip()]

# Function to extract player information for a given team
def extract_player_data(team_name, team_div, transfer_type):
    transfer_table = team_div.find_all('table')

    # Use the first table for incoming transfers and the second for outgoing transfers
    if transfer_table and len(transfer_table) >= 2:
        headers = [th.text.strip() for th in transfer_table[transfer_type].find('thead').find_all('th')]
        rows = transfer_table[transfer_type].find('tbody').find_all('tr')

        for row in rows:
            data = [td.text.strip() for td in row.find_all('td')]
            player_data = [team_name, 'In' if transfer_type == 0 else 'Out'] + data
            all_player_data.append(player_data)

# Iterate through each team and extract player information for both incoming and outgoing transfers
for team_name in teams_list:
    team_anchor = soup.find('a', text=team_name)
    team_div = team_anchor.find_parent('div', class_='box') if team_anchor else None

    # Check if the team_div is found
    if team_div:
        # Extract incoming transfers
        extract_player_data(team_name, team_div, transfer_type=0)

        # Extract outgoing transfers
        extract_player_data(team_name, team_div, transfer_type=1)

# Create a DataFrame from the collected player data
columns = ['Team', 'Transfer Direction', 'Player', 'Age', 'Nationality', 'Position', 'Short Position', 'Market Value', 'Left Team', 'Left Team Flag', 'Fee']
```

```

I've added this example file transfer_data for the 22/23 winter transfer season to the repo
# Display the DataFrame
print(df)

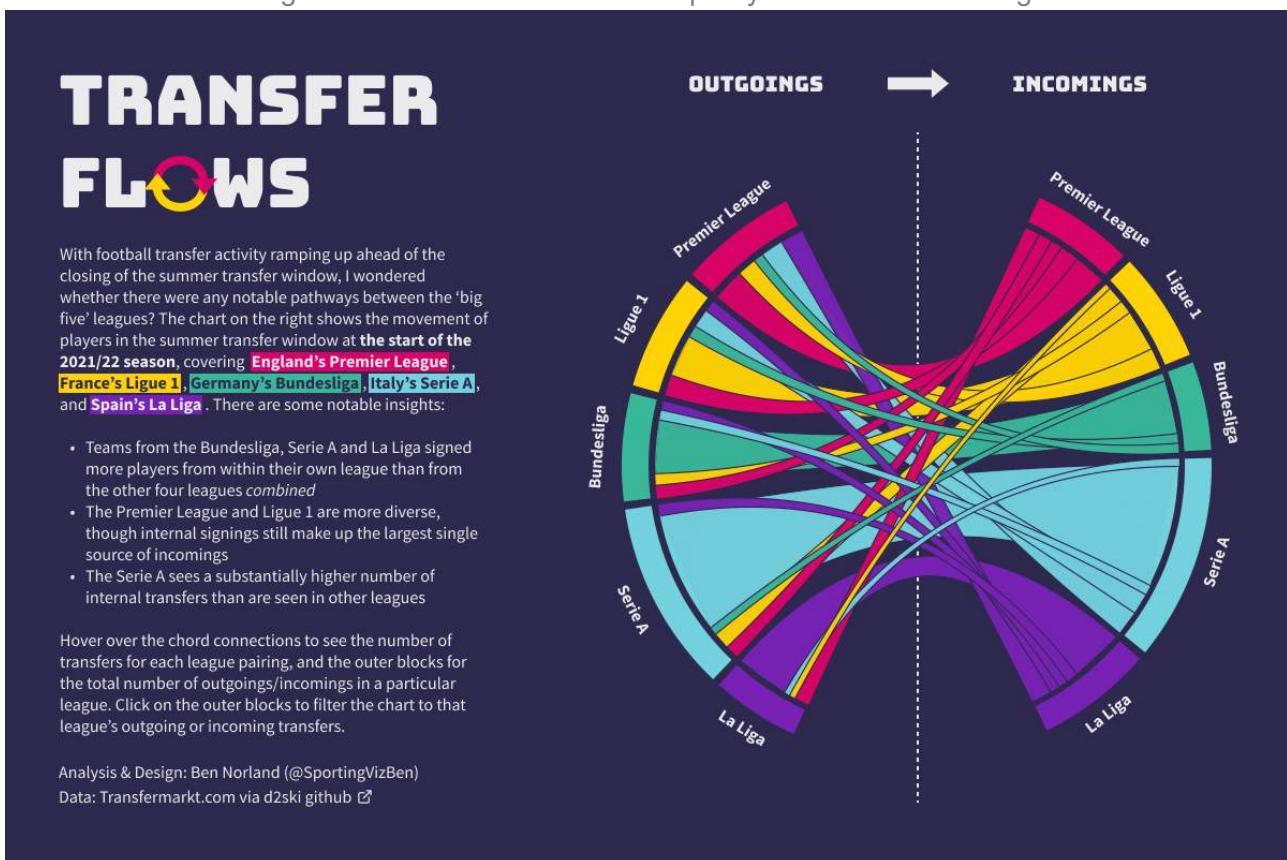
Bonus:
I've added some additional code that loops through multiple years, (both summer and winter)
transfers. I've added this to the repository for years from 2000- before the transfer season started in
January. This will be the "TwentyYearsPLTransfers.py" & corresponding 20Y_PL_transfer_data.csv
dataset.

Going Further:

```

- Why not try collate data for another league?
- Why not try visualise flow of individuals between clubs?

If you are looking for the art of what is possible, why not check out this dashboard by Ben Norland looking at football transfer flows. It's a pretty awesome chord diagram!



As always, Let me know how you get on with this one. I can be reached on Twitter, @_CJMayes.

LOGGING OFF,

CJ

INTRO TO TRANSFERMARKT PT 2 (TABLEAU)

Hi All,

Previously I've written about beautiful soup, as well as Transfermarkt. I want to build a little bit on that code, and within this tutorial share how we can replicate the logic within Tableau.

If you missed the intro, you can find the code, [here](#).

In todays blog we will look to expand on this code and add in some new details on finding what the most a player has been worth in their career, and who sits at the top of that list.

The new code looks at the Name, Club, Position, Age and Value of the player currently, as well as the maximum value of the player. We also bring back the date of the update for awareness.

You can find a copy of the new code and data output in the repository, [here](#).

As a brief recap.

```
import requests
from bs4 import BeautifulSoup
import pandas as pd

headers = {
    'User-Agent': 'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537.36'
}

# Initialize empty lists to store player information
all_player_data = []

# Initial page
base_url = "https://www.transfermarkt.co.uk/premier-league/marktwerte/wettbewerb/GB1/pos//detailpos/0/altersklasse/alle/plus/1"
pageTree = requests.get(base_url, headers=headers)
soup = BeautifulSoup(pageTree.content, 'html.parser')

# Number of pages to iterate through (adjust this based on the actual number of pages)
num_pages = 5

for page_number in range(1, num_pages + 1):
    # Find all player rows in the table
    player_rows = soup.find_all('tr', {'class': ['odd', 'even']})

    # Iterate through each player row
    for player_row in player_rows:
        # Extract player information
        player_name_tag = player_row.find('td', {'class': 'hauptlink'})
        player_position_tag = player_row.find_all('td', {'class': 'None'})[2]
        player_age_tag = player_row.find_all('td', {'class': 'zentriert'})[2]

        player_club_tag = player_row.find_all('td', {'class': 'zentriert'})[3].find('a')
        player_club = player_club_tag['title'] if player_club_tag else None

        # Extract player position
        player_position = player_position_tag.text.strip() if player_position_tag else None

        # Extract player age
        player_age = player_age_tag.text.strip() if player_age_tag else None

        # Extract current market value from the player's profile page
        player_link_tag = player_name_tag.find('a')
        if player_link_tag:
            player_link = player_link_tag['href']
            player_page_url = f"https://www.transfermarkt.co.uk{player_link}"
            player_page_tree = requests.get(player_page_url, headers=headers)
            player_page_soup = BeautifulSoup(player_page_tree.content, 'html.parser')

            # Extract current market value
            current_value_wrapper = player_page_soup.find('a', {'class': 'data-header__market-value-wrapper'})
            if current_value_wrapper:
                current_value_text = current_value_wrapper.text.replace('€', '').replace('m', '').strip()
                # Split value and date
```

The provided Python code is a web scraping script designed to extract information about football players in the English Premier League from the Transfermarkt website. The script utilizes the requests library to send HTTP requests and retrieve the HTML content of the Web pages, and the BeautifulSoup library for parsing and navigating the HTML structure. Additionally, it uses the pandas library to organize and export the scraped data into a CSV file.

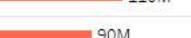
The script iterates through multiple pages of player data, where each page represents a subset of players. For each player, it extracts relevant information such as the player's name, club, position, age, current market value, last update date, and maximum market value. The current market value is obtained by navigating to the player's individual profile page and extracting the value from a specific HTML element. The extracted data is then stored in a list of dictionaries (`all_player_data`). After scraping all the pages, the script creates a DataFrame using pandas and exports the data to a CSV file named `player_data.csv`. The provided User-Agent header in the HTTP requests is used to mimic a web browser and avoid being blocked by anti-scraping mechanisms on the website.

```
player_data = {  
    'Name': player_name_tag.find('a').text.strip() if player_name_tag  
    else None,  
    'Club': player_position,  
    'Position': player_position,  
    'Age': player_age,  
    'Current_Value': player_current_value,  
    'Last_Update': player_last_update_date,  
    'Max_Value': player_max_value  
}
```

The current value was a little more tricky than in the previous tutorial, namely because I end up going to the players page to take the value alongside the update date. If anyone was able to withdraw it from the main table, let me know.

I thought it would be nice to be able to showcase how we can re-create the simple table and add in some pagination style to the visual.

Top 100 Transfermarkt Valued Players | Index Pagination Blog

index()	Name	Age	Club	Position	Max Value	Current Value
1	Erling Haaland	23	Manchester City	Centre-Forward	€180.00m	 180M
2	Bukayo Saka	22	Arsenal FC	Right Winger	€120.00m	 120M
3	Declan Rice	24	Arsenal FC	Defensive Midfield	€110.00m	 110M
4	Phil Foden	23	Manchester City	Right Winger	€110.00m	 110M
5	Rödri	27	Manchester City	Defensive Midfield	€110.00m	 110M
6	Julián Álvarez	23	Manchester City	Second Striker	€90.00m	 90M
7	Martin Ødegaard	25	Arsenal FC	Attacking Midfield	€90.00m	 90M
8	Moisés Caicedo	22	Chelsea FC	Defensive Midfield	€90.00m	 90M
9	Bruno Guimarães	26	Newcastle United	Defensive Midfield	€85.00m	 85M
10	Gabriel Martinelli	22	Arsenal FC	Left Winger	€85.00m	 85M
11	Bernardo Silva	29	Manchester City	Attacking Midfield	€100.00m	 100M
12	Enzo Fernández	22	Chelsea FC	Central Midfield	€85.00m	 85M
13	Josko Gvardiol	21	Manchester City	Centre-Back	€80.00m	 80M
14	Rüben Dias	26	Manchester City	Centre-Back	€80.00m	 80M
15	Christopher Nkunku	26	Chelsea FC	Attacking Midfield	€80.00m	 80M
16	Dominik Szoboszlai	23	Liverpool FC	Central Midfield	€75.00m	 75M
17	Luis Díaz	26	Liverpool FC	Left Winger	€75.00m	 75M
18	William Saliba	22	Arsenal FC	Centre-Back	€75.00m	 75M
19	Alexander Isak	24	Newcastle United	Centre-Forward	€70.00m	 70M
20	Bruno Fernandes	29	Manchester United	Attacking Midfield	€90.00m	 90M

You can access the visual using the link at the top of the page.



The build is fairly generic in terms of being a bar chart. A bunch of things on rows (with them sorted on current value), followed by adding in an index value to know where they are ranked.

The buttons are where the slightly more difficult element comes into practice.

First we need the Pager calculation

```

if index() <= 20 then 1
elseif index() <= 40 then 2
elseif index() <= 60 then 3
elseif index() <= 80 then 4
elseif index() <= 100 then 5
END
  
```

This will help us separate our players.

Then we need a parameter that will help us with a click action to navigate through the pages.

Edit Parameter [P. Page]

Name: P. Page

Properties

Data type: Integer	Display format: 1														
Current value: 1	Value when workbook opens: Current value														
Allowable values:															
<input type="radio"/> All <input checked="" type="radio"/> List <input type="radio"/> Range															
<table border="1"> <thead> <tr> <th>Value</th> <th>Display As</th> </tr> </thead> <tbody> <tr><td>1</td><td>1</td></tr> <tr><td>2</td><td>2</td></tr> <tr><td>3</td><td>3</td></tr> <tr><td>4</td><td>4</td></tr> <tr><td>5</td><td>5</td></tr> <tr><td colspan="2">Click to add</td></tr> </tbody> </table>		Value	Display As	1	1	2	2	3	3	4	4	5	5	Click to add	
Value	Display As														
1	1														
2	2														
3	3														
4	4														
5	5														
Click to add															
<input checked="" type="radio"/> Fixed <input type="radio"/> When workbook opens <input type="button" value="Add values from ▾"/>															
<input type="button" value="Remove Selected"/>															
<input type="button" value="Cancel"/> <input type="button" value="OK"/>															

A true false calculation is needed for the filters, to only show the relevant index values.

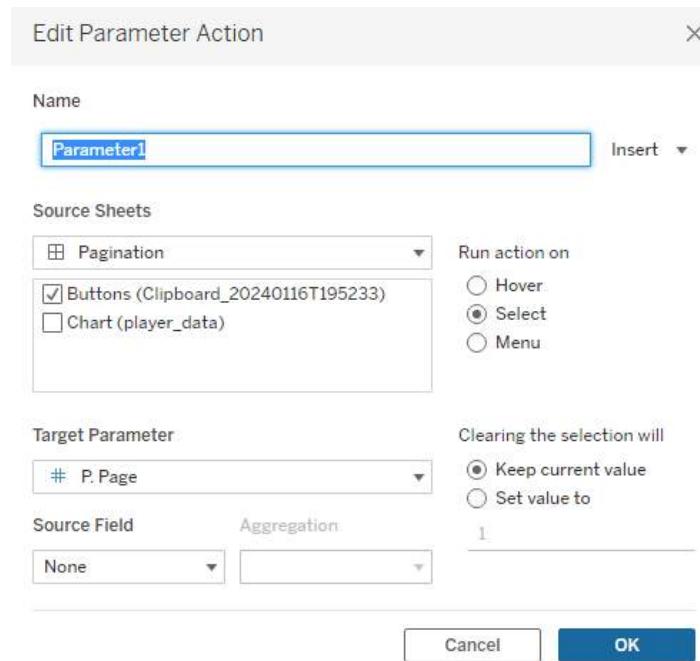
=

Finally,

I bring in a new dataset of values 1 through 5 to act as the separate page buttons.

The screenshot shows the Tableau Data Source interface. On the left, under 'Tables', there is a new dataset named 'Buttons'. This dataset has three columns: 'Page' (containing values 1, 2, 3, 4, 5), 'Title' (containing 'Buttons'), and 'Rows' (containing an empty list). In the bottom right corner of the interface, there are five circular buttons numbered 1 through 5. The first button (labeled 1) is highlighted with a red border, while the others are grey.

Then, all we need is an action on the dashboard and we are good to go.



Going Further:

- Why not try and find the market value of top players in a different league?
- Build the chart into a full dashboard on player metrics.
- Create a chart to show the difference between player market price, and max price the player has been valued at. Group pricing by club.

As always, Let me know how you get on with this one. I can be reached on Twitter, @_CJMayes.

LOGGING OFF,

CJ

INTRO TO TRANSFERMARKT (PYTHON)

Hi All,

Welcome back,

I hope everyone has had a lovely festive period and closed out 2023 in style. Looking ahead to 2024 I am excited to share more with the community, this may be through a variety of different channels compared to previous, but we will see how things play out.

Last year I really found good flow to the tutorial content normally posting live on a Monday morning – and hit a milestone of a blog a week on average! It was great to see the site grow in views a further 33% YoY – excited people haven't got fed up yet of the efforts.

Anyway, wishing everyone all the best. Let's dive into some content.

Many sports fans will already be aware of transfermarkt and its existence. It is a wonderful resource to use when looking at... well transfers.

TRANSFERS	NATIONAL TRANSFERS	INTERNATIONAL TRANSFERS	AVAILABLE
> Latest Transfers <small>TOP</small>	> Premier League	> La Liga	> Free agents
> Top transfers	> Latest transfers	> Bundesliga	> Contracts ending <small>TOP</small>
> Current loan players	> Summer Transfers	> Serie A	> Available coaches
> Transfers by day	> Winter transfers	> Ligue 1	> Available managers
> Transfer fees	> Transfer records	> Premier Liga	
> Transfer records	> Championship	> Primeira Liga	
> Transfer Records	> Latest transfers	> Calculate solidarity contribution	<small>NEW</small>
> Most valuable transfers	> Scottish Premiership	> Latest transfers	
> Transfer profits	> Latest transfers		
> Income / Expenditure			
> Transfer balance			
> Purchase value			
> Transfer window			
> Contract extensions			

It holds everything from rumours, fees, to contracts that are ending soon amongst a whole variety of clubs. It's not just popular amongst your twitter footy fans but also a lot of clubs and sports companies / agencies.

Today we will take a first peak at how we can start to extract some of that information through python. This will be for those who are really just starting out their journey, but will look to build upon it in future weeks.

The code can be found in the repository through the [github link](#) at the top of the page.

```
import requests
from bs4 import BeautifulSoup
import pandas as pd

headers = {
    'User-Agent': 'Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537.36'
}

# Initialize empty lists to store player names and values
all_player_names = []
all_player_values = []

# Initial page
base_url = "https://www.transfermarkt.co.uk/premier-league/marktwerte/wettbewerb/GB1/pos//detailpos/0/altersklasse/alle/plus/1"
pageTree = requests.get(base_url, headers=headers)
soup = BeautifulSoup(pageTree.content, 'html.parser')

# Number of pages to iterate through (adjust this based on the actual number of pages)
num_pages = 5

for page_number in range(1, num_pages + 1):
    # Find all player rows in the table
    player_rows = soup.find_all('tr', {'class': ['odd', 'even']})

    # Iterate through each player row
    for player_row in player_rows:
        # Extract player name
        player_name_tag = player_row.find('td', {'class': 'hauptlink'})
        player_name = player_name_tag.a.text.strip() if player_name_tag
        else None
        all_player_names.append(player_name)

        # Extract player value
        player_value_tag = player_row.find('td', {'class': 'rechts'})
        player_value = player_value_tag.span.text.strip() if player_value_tag
        else None
        all_player_values.append(player_value)

    # Find the link to the next page
    next_page_link = soup.find('a', {'class': 'tm-pagination__link', 'title': f'Page {page_number + 1}'})

    # Break the loop if there is no next page link
    if not next_page_link:
        break

    # Construct the URL for the next page
    next_page_url = base_url + next_page_link['href']
    pageTree = requests.get(next_page_url, headers=headers)
    soup = BeautifulSoup(pageTree.content, 'html.parser')

# Create a DataFrame to store the data
df = pd.DataFrame({'Player Name': all_player_names, 'Player Value': all_player_values})

# Export the DataFrame to a CSV file
df.to_csv('player_data.csv', index=False)
```

What does the script do?

This Python script is a web scraper that extracts information about player names and their corresponding values from the Transfermarkt website, specifically focusing on the Premier League.

The script utilizes the requests library to send HTTP requests and the BeautifulSoup library to parse and navigate HTML content. The extracted data is then organised into a pandas DataFrame and exported to a CSV file.

#	Player	Nat.	Age	Club	Highest value in career	Last update	Market values
1	Erling Haaland Centre-Forward	🇳🇴	23	ACL	€180.00m	Dec 19, 2023	€180.00m
2	Bukayo Saka Right Winger	🇬🇧	22	ARS	€120.00m	Dec 19, 2023	€120.00m
3	Declan Rice Defensive Midfield	🇮🇪	24	ARS	€110.00m	Dec 19, 2023	€110.00m
4	Rodri Defensive Midfield	🇸🇵	27	CLB	€110.00m	Dec 19, 2023	€110.00m
5	Phil Foden Right Winger	🇬🇧	23	CLB	€110.00m	Dec 19, 2023	€110.00m
6	Julián Álvarez Second Striker	🇦🇷	23	CLB	€90.00m	Dec 19, 2023	€90.00m
7	Martin Ødegaard Attacking Midfield	🇳🇴	25	ARS	€90.00m	Dec 19, 2023	€90.00m
8	Moisés Caicedo Defensive Midfield	🇨🇴	22	CFC	€90.00m	Dec 19, 2023	€90.00m
9	Bruno Guimarães Defensive Midfield	🇧🇷	26	MAN	€85.00m	Dec 19, 2023	€85.00m
10	Gabriel Martinelli Left Winger	🇧🇷	22	ARS	€85.00m	Dec 19, 2023	€85.00m

Let's take a step by step look...

1.

First we import the necessary libraries.

2. Then we set the user agent headers up. These headers are used to mimic a web browser when making requests. Some websites may block requests without a proper user-agent.

3. I create empty lists because I know I want to end up adding the players names and current value to the list.

4. The base url is the easy bit... just copy and paste the page you are on. You will see for this tutorial we are looking at premier league players.

5. The page tree and soup send a request to the initial page and parse the content.

What do we mean by parse the content?

Well, right click on a web page and go to inspect.

OVERVIEW TABLES & TRANSFERS - MARKET VALUES - PLAYERS - CLUBS - INFORMATION & FACTS - HISTORY - NEWS						
Competitor	Player	Nat.	Age	Out	Market value	Highest value
1 Erling Haaland	Erling Haaland	Norway	23	---	€100.000k	€100.000k
2 Dani Olmo	Dani Olmo	Spain	23	---	€50.000k	€50.000k
3 Lukas Kral	Lukas Kral	Czech Republic	23	---	€50.000k	€50.000k
4 Bernd Leno	Bernd Leno	Germany	24	---	€25.000k	€25.000k
5 Rui	Rui	Portugal	23	---	€20.000k	€20.000k
6 Phil Foden	Phil Foden	England	20	---	€20.000k	€20.000k
7 Raphinha	Raphinha	Brazil	23	---	€20.000k	€20.000k
8 Julian Nagelsmann	Julian Nagelsmann	Germany	40	---	€20.000k	€20.000k
9 Max Wergert	Max Wergert	Austria	23	---	€20.000k	€20.000k
10 Weston Coyle	Weston Coyle	Defender (midfield)	23	---	€20.000k	€20.000k
11 Bruno Soares	Bruno Soares	Portugal	20	---	€20.000k	€20.000k
12 Lukas Hradecky	Lukas Hradecky	Czech Republic	22	---	€20.000k	€20.000k
13 Edson Perneta	Edson Perneta	Central Midfield	22	---	€20.000k	€20.000k
14 Renato Silveira	Renato Silveira	Attacking Midfield	29	---	€20.000k	€20.000k
15 Nicanor Usta	Nicanor Usta	Central Midfield	26	---	€20.000k	€20.000k
16 Charles Aranguiz	Charles Aranguiz	Chile	21	---	€20.000k	€20.000k
17 Gheorghe Marin	Gheorghe Marin	Middlefield	20	---	€20.000k	€20.000k
18 Renzo Soriano	Renzo Soriano	Central Midfield	23	---	€20.000k	€20.000k
19 William Saliba	William Saliba	Defender	20	---	€20.000k	€20.000k
20 Luis Muriel	Luis Muriel	Colombia	28	---	€20.000k	€20.000k
21 Forest Whosoever-Arnold	Forest Whosoever-Arnold	England	20	---	€20.000k	€20.000k
22 Gabriel Jesus	Gabriel Jesus	Central Forward	23	---	€20.000k	€20.000k
23 Bruno Fernandes	Bruno Fernandes	Portugal	29	---	€20.000k	€20.000k
24 Alexander Isak	Alexander Isak	Central Forward	23	---	€20.000k	€20.000k

```

<td><img alt="Norway flag" data-bbox="108 100 125 115"/> Erling Haaland</td>
<td>Norway</td>
<td>23</td>
<td>---</td>
<td>€100.000k</td>
<td>€100.000k</td>
<td>€100.000k</td>

```

The screenshot shows a table of football players from Transfermarkt. The first player listed is Erling Haaland, with his name highlighted in yellow. The developer tools are open, showing the DOM structure for the first row. The 'highlighted' class is applied to the first 'td' element containing 'Erling Haaland'. The 'rechts' class is applied to the 'td' element containing '€100.000k'.

You will see the different tagging available in the web page. Notice the highlighted title of Erling Haaland. Essentially we have to navigate through these different trees to find the different attributes we want.

This is what the following parts are doing
`player_name_tag = player_row.find('td', {'class': 'hauptlink'})`

```
<td class="hauptlink">
  <a title="Erling Haaland" href="/erling-haaland/profil/spieler/418560">Erling Haaland</a>
  ...

```

`player_value_tag = player_row.find('td', {'class': 'rechts'})`

One thing to note here is i am taking the highest value the player has reached instead of the current market value.

Sometimes we will have to do some additional search or cleansing depending on how the page is structured.

So generally that's how we retrieved specific players details.

The For loop will help us iterate over all the different players on the page.

But what if we want the full 100 players, and their highest value they have reached in their career?

Compact	Detailed	Gallery					
#	Player	Nat.	Age ↗	Club	Highest value in career ↗	Last update ↗	Market values ↗
1	Erling Haaland Centre-Forward	🇳🇴	23	Manchester City	€180.00m	Dec 19, 2023	€180.00m ⬜
2	Bukayo Saka Right Winger	🇬🇧 / 🇩🇪	22	Arsenal	€120.00m	Dec 19, 2023	€120.00m ⬜
3	Declan Rice Defensive Midfield	🇬🇧 / 🇮🇪	24	Arsenal	€110.00m	Dec 19, 2023	€110.00m ⬈
4	Rodri Defensive Midfield	🇸🇵	27	Manchester City	€110.00m	Dec 19, 2023	€110.00m ⬈
5	Phil Foden Right Winger	🇬🇧	23	Manchester City	€110.00m	Dec 19, 2023	€110.00m ⬜
6	Julián Álvarez Second Striker	🇦🇷	23	Manchester City	€90.00m	Dec 19, 2023	€90.00m ⬈
7	Martin Ødegaard Attacking Midfield	🇳🇴	25	Arsenal	€90.00m	Dec 19, 2023	€90.00m ⬜
8	Moisés Caicedo Defensive Midfield	🇨🇴	22	Chelsea	€90.00m	Dec 19, 2023	€90.00m ⬜
9	Bruno Guimarães Defensive Midfield	🇧🇷 / 🇸ປ	26	Newcastle United	€85.00m	Dec 19, 2023	€85.00m ⬜
10	Gabriel Martinelli Left Winger	🇧🇷 / 🇮🇹	22	Arsenal	€85.00m	Dec 19, 2023	€85.00m ⬈
11	Enzo Fernández Central Midfield	🇦🇷	22	Chelsea	€85.00m	Dec 19, 2023	€80.00m ⬜
12	Bernardo Silva Attacking Midfield	🇵🇹	29	Manchester City	€100.00m	Dec 19, 2023	€80.00m ⬜
13	Rúben Dias Centre-Back	🇵🇹	26	Manchester City	€80.00m	Dec 19, 2023	€80.00m ⬜
14	Josko Gvardiol Centre-Back	🇭🇷	21	Man City	€80.00m	Dec 19, 2023	€80.00m ⬜
15	Christopher Nkunku Attacking Midfield	🇫🇷 / 🇩🇪	26	Chelsea	€80.00m	Dec 19, 2023	€75.00m ⬛
16	Dominik Szoboszlai Central Midfield	🇭🇺	23	Liverpool	€75.00m	Dec 19, 2023	€75.00m ⬈
17	William Saliba Centre-Back	🇫🇷 / 🇩🇪	22	Arsenal	€75.00m	Dec 19, 2023	€75.00m ⬈
18	Luis Díaz Left Winger	🇨🇴	26	Liverpool	€75.00m	Dec 19, 2023	€75.00m ⬜
19	Trent Alexander-Arnold Right-Back	🇬🇧	25	Liverpool	€110.00m	Dec 19, 2023	€70.00m ⬈
20	Gabriel Jesus Centre-Forward	🇧🇷	26	Arsenal	€80.00m	Dec 19, 2023	€70.00m ⬛
21	Bruno Fernandes Attacking Midfield	🇵🇹	29	Manchester United	€90.00m	Dec 19, 2023	€70.00m ⬛
22	Alexander Isak Centre-Forward	🇸🇪 / 🇩🇪	24	Newcastle United	€70.00m	Dec 19, 2023	€70.00m ⬜
23	Marcus Rashford Left Winger	🇬🇧 / 🇲🇼	26	Manchester United	€85.00m	Dec 19, 2023	€70.00m ⬛
24	James Maddison Attacking Midfield	🇬🇧	27	Leeds United	€70.00m	Dec 19, 2023	€70.00m ⬜
25	Jack Grealish Left Winger	🇬🇧 / 🇮🇪	28	Man City	€100.00m	Dec 19, 2023	€70.00m ⬛

1 2 3 4 ► ⏪ ⏵

Well normally you'd think that the url changes.

In this case it's actually the bottom banner that controls the view on the page. (1,2,3,4..)

So we actually want a way to click through these buttons, that's why we have added in a for loop containing that page, as well as the signal to locate the number box to action the page clicks.

```
next_page_link = soup.find('a', {'class': 'tm-pagination__link', 'title': f'Page {page_number + 1}'})
```

```
if not next_page_link:
```

```
    break
```

```
next_page_url = base_url + next_page_link
```

```
pageTree = requests.get(next_page_url, headers=headers)
```

```
soup = BeautifulSoup(pageTree.content, 'html.parser')
```

The script writes the player information to a CSV file named 'player_data.csv' in the current working directory.

Player Name	Player Value
Erling Haaland	€180.00m
Bukayo Saka	€120.00m
Declan Rice	€110.00m
Rodri	€110.00m
Phil Foden	€110.00m
Julián Álvarez	€90.00m
Martin Ødegaard	€90.00m
Moisés Caicedo	€90.00m
Bruno Guimarães	€85.00m
Gabriel Martinelli	€85.00m
Enzo Fernández	€85.00m
Bernardo Silva	€100.00m
Rúben Dias	€80.00m
Josko Gvardiol	€80.00m
Christopher Nkunku	€80.00m
Dominik Szoboszlai	€75.00m
William Saliba	€75.00m
Luis Diaz	€75.00m
Trent Alexander-Arnold	€110.00m
Gabriel Jesus	€80.00m
Bruno Fernandes	€90.00m
Alexander Isak	€70.00m
Marcus Rashford	€85.00m
James Maddison	€70.00m
Jack Grealish	€100.00m
Lucas Paquetá	€65.00m
Evan Ferguson	€65.00m
Cristian Romero	€65.00m
Mohamed Salah	€150.00m
Rasmus Højlund	€65.00m
Alexis Mac Allister	€65.00m
Jérémy Doku	€65.00m
Darwin Núñez	€70.00m
Gabriel Magalhães	€60.00m
Douglas Luiz	€60.00m
Kevin De Bruyne	€150.00m
Moussa Diaby	€60.00m
João Palhinha	€60.00m
Ollie Watkins	€60.00m

You can find a copy of what this output looks like in the GitHub repo.

Now that I've given you the power to, It's time to add the disclaimer:

Keep in mind that web scraping might be subject to the terms of service of a website, and scraping too aggressively may lead to IP bans or other restrictions. Always make sure you comply with the website's policies and use web scraping responsibly.

Going Further:

- Why not try and find the market value of top players in a different league?
- Try and add additional information in, such as country, age or club.
- Create a chart to show the difference between player market price.

As always, Let me know how you get on with this one. I can be reached on Twitter, @_CJMayes.

LOGGING OFF,

CJ

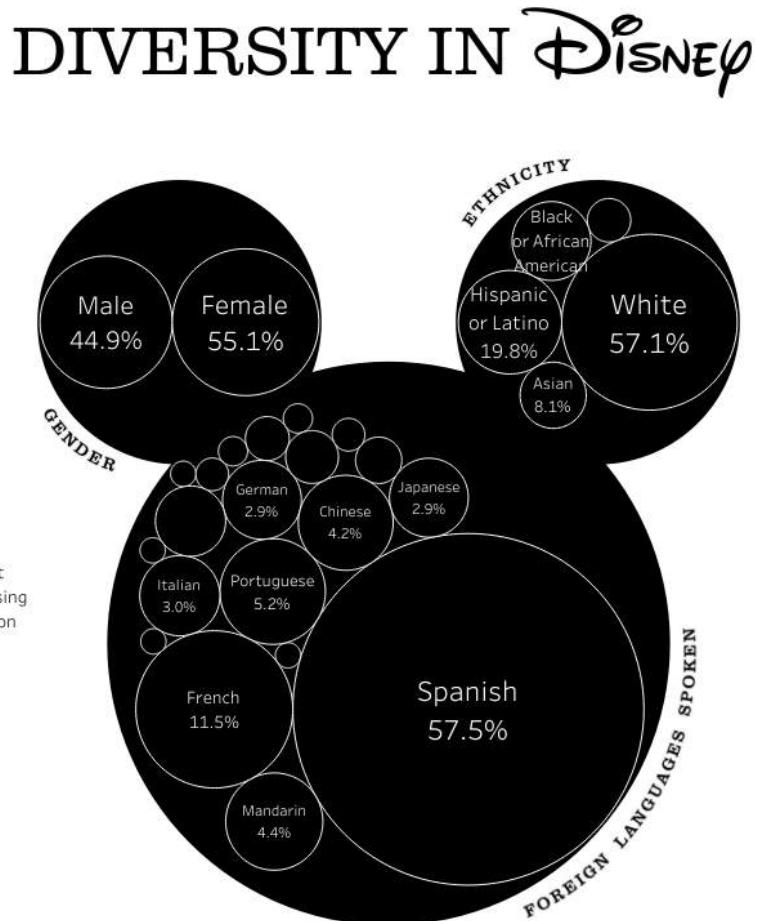
CIRCULAR PACKING IN TABLEAU PART 2

Hi All,

Back in 2022 I wrote about **circular packing**, and created a tutorial on how you could use the code to then export the details to build it within Tableau.

A few people reached out in coming years telling me they had difficulties with being able to frame the initial JSON input.

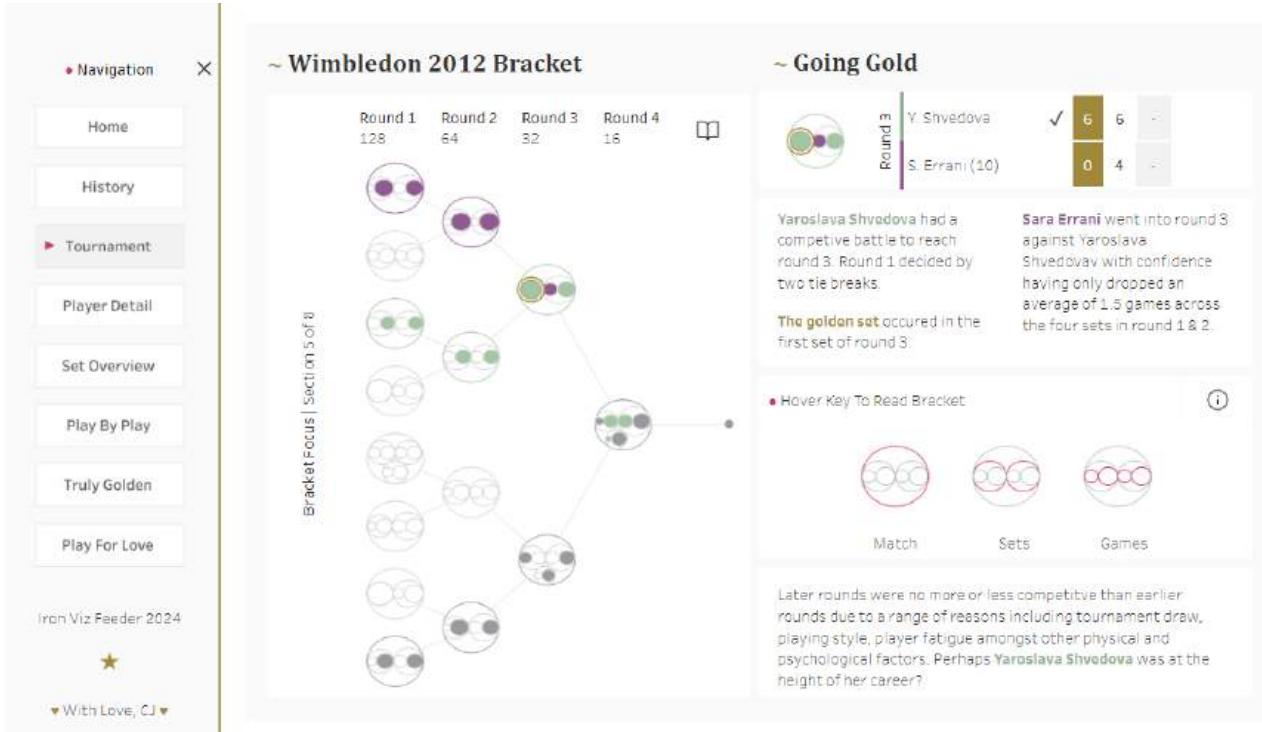
Having had a flurry of wonderful dashboards at the end of 2023 that cite the circular packing tutorial as means of creation such as the likes of this visual by Jess



and [this Simpsons visual by Brian](#),

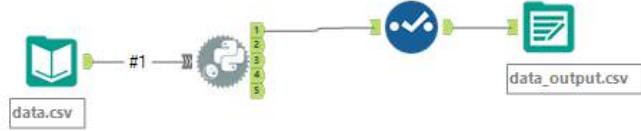


and even using it in my own Iron Viz Feeder last month,



I thought I would re-write some of the code to be able to take an input csv, and create an export file. Doing the leg work in the middle. Rather than just add the code, I've also created it in Alteryx, for those that don't have a python IDE.

- If you only want to use Alteryx, then download the Alteryx workflow from the [Git Repo](#).



The Alteryx code is fairly similar to that stored in main.py

The only big differences will be

- Where the original dataset is stored.
- How Data frames are managed and fed through the code. The main.py version is made to be run locally using an IDE of your choosing.
- Both the workflow, or the python code will create the same output of data, ready to follow the rest of the blog in terms of creating the chart in Tableau.

Let's take a look at the code, and our starting csv file.

This is what the csv file looks like. You will find a copy in the Git Repo.

```

1 id,datum,parent
2 World,6964195249,None
3 North America,450448697,World
4 United States,308865000,North America
5 Mexico,107550097,North America
6 Canada,34033000,North America
7 South America,278095425,World
8 Brazil,192612000,South America
9 Colombia,45349000,South America
10 Argentina,40134425,South America
11 Europe,209246682,World
12 Germany,81757600,Europe
13 France,65447374,Europe
14 United Kingdom,62041708,Europe
15 Africa,311929000,World
16 Nigeria,154729000,Africa
17 Ethiopia,79221000,Africa
18 Egypt,77979000,Africa
19 Asia,2745929500,World
20 China,1336335000,Asia
21 India,1178225000,Asia
22 Indonesia,231369500,Asia
23

```

3 Columns are needed, an id, a value called

datum and a column to identify what the parent is (in effect figuring out what the “children” nodes are)

If you were to run this section of the code, it ends up reading from the csv, and creating the json ready for the circular packing tutorial.

```
[  
  {  
    "id": "World",  
    "datum": 6964195249,  
    "children": [  
      {  
        "id": "North America",  
        "datum": 450448697,  
        "children": [  
          {  
            "id": "United States",  
            "datum": 308865000,  
            "children": []  
          },  
          {  
            "id": "Mexico",  
            "datum": 107550697,  
            "children": []  
          },  
          {  
            "id": "Canada",  
            "datum": 34033000,  
            "children": []  
          }  
        ]  
      },  
      {  
        "id": "South America",  
        "datum": 278095425,  
        "children": [  
          {  
            "id": "Brazil",  
            "datum": 192612000,  
            "children": [  
              {  
                "id": "Argentina",  
                "datum": 192612000,  
                "children": []  
              }  
            ]  
          }  
        ]  
      }  
    ]  
  }]
```

The second part of the code is similar to the 2021 blog, and is all about prepping our data ready for Tableau. **The code is 150 lines long.**

```
  This first bit of code is new, it takes a data.csv file and converts the columns
  into the json we need
  """

import pandas as pd
import os
import json
import circlify
import matplotlib.pyplot as plt

def find_children(df, parent_id):
    children_list = df[df["parent"] == parent_id].to_dict('records')
    for child in children_list:
        child['id'] = child.pop('id')
        child['datum'] = child.pop('datum')
        child['children'] = find_children(df, child["id"])
        del child['parent'] # delete parent field from child
    return children_list

# Get current working directory
cwd = os.getcwd()

# Read the dataframe from csv
df = pd.read_csv(os.path.join(cwd, 'data.csv'))

# Now manually set the first level elements (where parent is NaN or None)
df.fillna('None', inplace=True)
df['parent'] = df['parent'].astype(str)

output = []
for _, row in df[df["parent"] == 'None'].iterrows():
    node = dict(row)
    node.pop('parent') # delete parent field from node
    output.append({
        'id': node['id'],
        'datum': node['datum'],
        'children': find_children(df, node['id'])
    })

# Print final json output
print(json.dumps(output, indent=2))
data = output

"""

The next bit of code is all about translating this Json into the format we want to
use,
Let's make sure our datum values are considered actual values
"""

def convert_datum_to_integer(node):
    node['datum'] = int(node['datum'])
    for child in node.get('children', []):
        convert_datum_to_integer(child)

# Convert "datum" field to integers
for item in data:
    convert_datum_to_integer(item)

# Print the updated data
print(json.dumps(data, indent=4))

"""

```

The last bit of code is about converting the circles into x,y and radius coordinates ready for Tableau

This piece of code performs several tasks related to data processing and visualization:

```
1. It reads data from a CSV file called 'data.csv' using the pandas library.
2. # Compute circle positions thanks to the circlify() function
3. # The maximum radius is set to
4. #           maxradius library to compute circle positions based on the JSON data.
5. circles = circlify.circlify(
6.     data,
7.     show_enclosure=False,
8.     target_enclosure=circlify.Circle(x=0, y=0, r=1)
```

Overall, this code takes data from a CSV file, converts it into a nested JSON structure, calculates circle positions, and visualizes the circles using matplotlib. It also saves the processed data to a CSV

```
# To ensure in Tableau the sizing works the fig size is the same x*y.
```

```
fig, ax = plt.subplots(figsize=(10, 10))
for _, row in df.iterrows():
    print(f'{row["id"]},{row["X co-ord"]},{row["Y-coord"]},{row["Radius"]},{row["Rank"]}')
    # 0,World,0.0,0.0,1.0,1
    # 1,Europe,-0.1891573044970616,0.7725949609994359,0.1964724487306323,2
    # 2,South America,-0.5193811141243917,-0.4774793174718824,0.22650056519090414,3
    # 3,Africa,-0.5250482991363239,0.4940564718994228,0.2398342689140008,4
    # 4,North America,-0.7117329289789401,0.0,0.28826707102105975,5
    # 5,Asia,0.28826707102105975,0.0,0.7117329289789401,6
    # 6,Canada,-0.8015572298502232,0.13991165332617728,0.06017798041665636,7
    # 7,Argentina,-0.6218965087862706,-0.35827194898537407,0.06927524011838612,8
    # 8,Colombia,-0.6715240632168605,-0.49229197511777817,0.07363823567480635,9
    # 9,United Kingdom,-0.20484950837730978,0.8820585650518235,0.08590977893113161,10
    # 10,France,-0.2883116431566897,0.7291956444670085,0.08823620918716854,11
    # 11,Egypt,-0.5807524341097545,0.6266527390697123,0.09606159016055799,12
    # 12,Ethiopia,-0.6616477217438194,0.4515509451194898,0.09682357206293761,13
    # 13,Germany,-0.10145547990466931,0.7291956444670085,0.09861995406485186,14
    # 14,Mexico,-0.8930220906231182,0.0,0.1069779093768817,15
    # 15,Nigeria,-0.42950888228212775,0.4515509451194898,0.13531526739875396,16
    # 16,Brazil,-0.44612447532083954,-0.49229197511777817,0.15176135222121462,17
    # 17,Indonesia,0.2610622289123354,0.5651857971321717,0.15273517341003195,18
    # 18,United States,-0.6047550196020585,0.0,0.18128916164417805,19
    # 19,India,-0.07879852383709784,0.0,0.34466733412078254,20
    # 20,China,0.6329344051418423,0.0,0.3670655948581576,21
```

```
    if circle.level != 1:
        continue
```

TABLEAU

I've appended the Tableau steps from the previous blog, in case you want to finalise your visual.

```
You can download the dashboard at the top of the page.
```

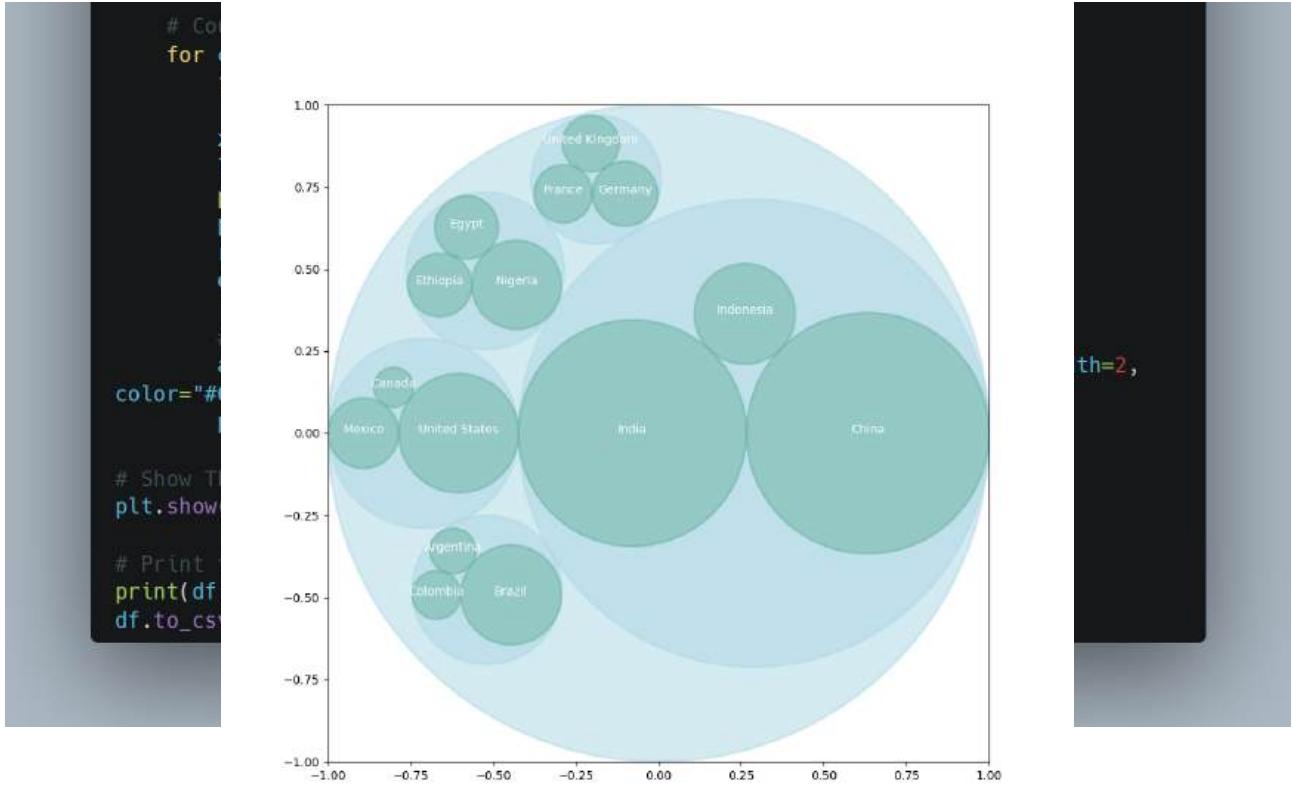
```
If you'd like to give the Tableau part a go, only download the data from the Github repository.
```

```
rank = rank + 1
df.loc[len(df)] = printdata
```

```
# Not needed for Tableau
ax.add_patch(plt.Circle((x, y), r*padding_outer, alpha=0.5, linewidth=2,
color="lightblue"))
```

```
# Continent Level
for circle in circles:
    if circle.level != 2:
        continue
    x, y, r = circle
    label = circle.ex["id"]
    print(label, x, y, r)
    printdata = [label, x, y, r, rank]
    rank = rank + 1
    df.loc[len(df)] = printdata
```

```
# Not needed for Tableau
ax.add_patch(plt.Circle((x, y), r*padding_outer, alpha=0.5, linewidth=2,
color="lightblue"))
```



Once we have run our code we will get the csv export countries, that we will want to add in [another sheet of T values between 1 and 100](#). I haven't included this in the repo but you can find an example in the [2021 repo](#).

We will want to join the sheets with a custom calculation of $1 = 1$. This duplicates our dataset with 100 rows for each country as we will use these points to create circles.



Next we create 6 calculations.

```
1a. Angle
(360/(100-1))
// closes angle gap by 1 (to make a full circle)

// The value is 50 because that is the number
// we have in the T sheet
```

The calculation is valid. 11 Dependencies. Apply OK

Angle – We will want to plot 100 points in a circle. This calculation finds what the angle between each point will be. You will see $1 - 1$ in order for the lines to overlap allowing for me to use a polygon and line tool effectively.

*The T value is 100 not 50 as shown in the comment as I wanted the circle to be a bit more rounded.

1b. Rank Angle

```
([T]) * ([1a. Angle])
// Multiply the number of points in our circle by the angle
```

The calculation is valid. 10 Dependencies

Rank Angle – Find the angle for each point. If you've followed my blogs you will see I tend to build most my radial vizzes using the same logic.

2a. X

```
COS(RADIANS([1b. Rank Angle])) * [Radius]
// We can find the size of each circle plotted
// based on the radius, next we need to offset them
```

The calculation is valid. 7 Dependencies

X – Now we use trigonometry to make our circle.... a circle. We multiply by the radius as otherwise without this all our radius' would be equal to 1. This allows for the circles to be the correct proportion.

2b. Y

```
SIN(RADIANS([1b. Rank Angle])) * [Radius]
// We can find the size of each circle plotted
// based on the radius, next we need to offset them
```

The calculation is valid. 7 Dependencies

Y – Same as above but wrapped in a sin function.

3a. X

```
[2a. X] + [X co-ord]
// Aligning X to the position in the chart
```

The calculation is valid. 6 Dependencies

3a X – So what are we doing here? As you can imagine we have created our sizes circles but at the moment they all sit on top of one another. We add the X co ordinate from the original data in order to shift it to where the centre point of the specific country should be. (I.e Transpose)

3b.Y

[2b. Y] + [Y-cord]
//Aligning Y to the position in the chart

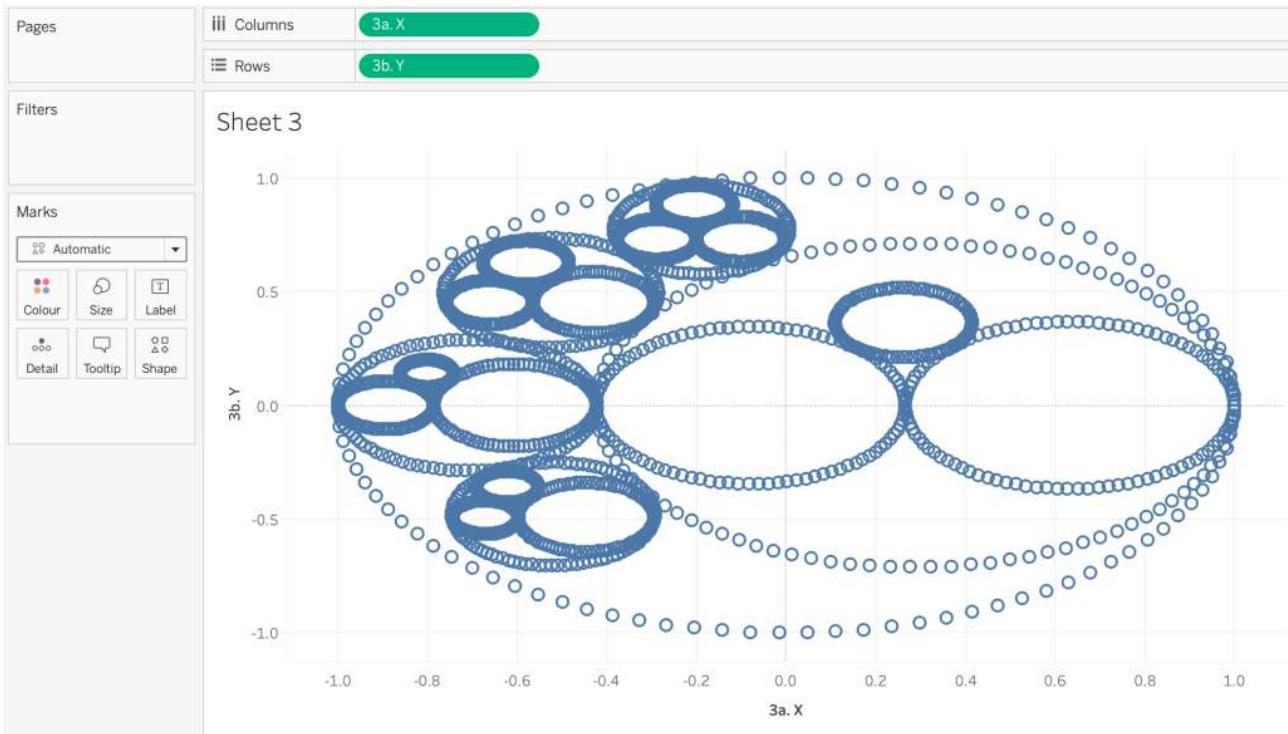
The calculation is valid. 6 Dependencies Apply OK

Same for 3b Y, we transpose the circle adding on the centre point to the calculation to move the circle upwards by Y cord.

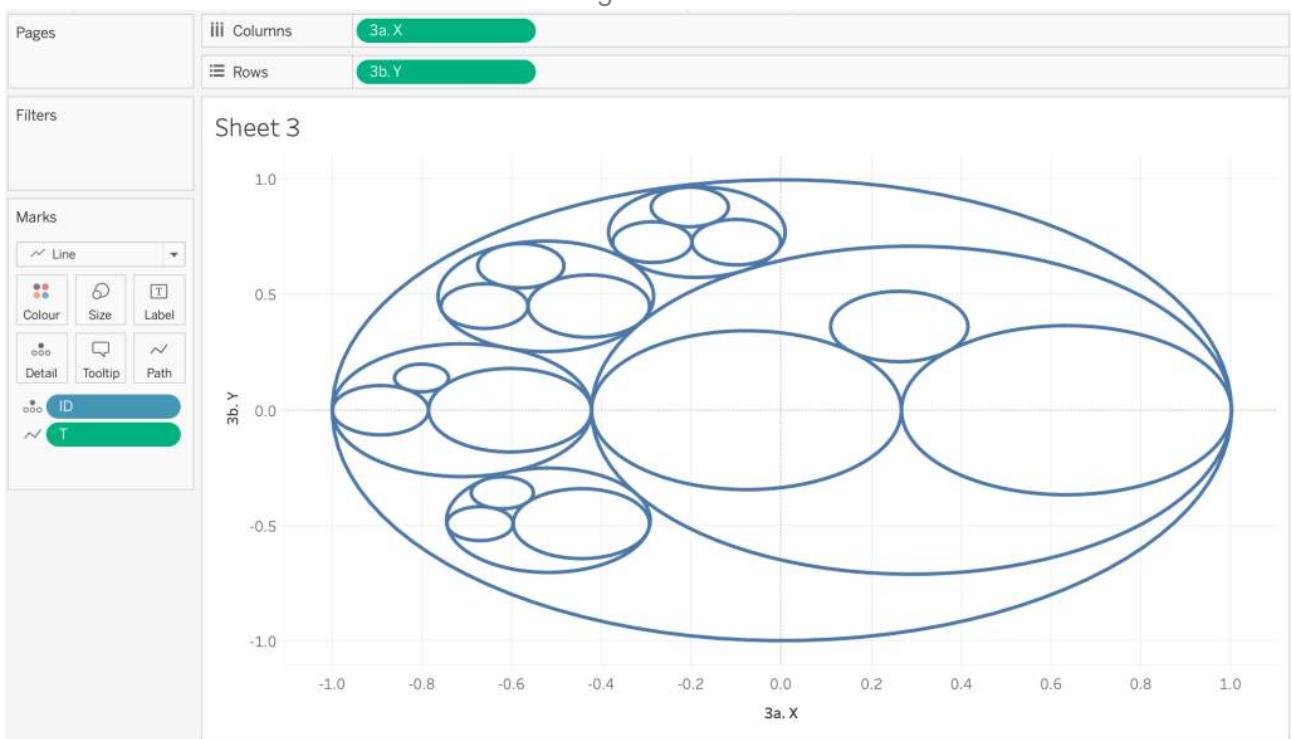
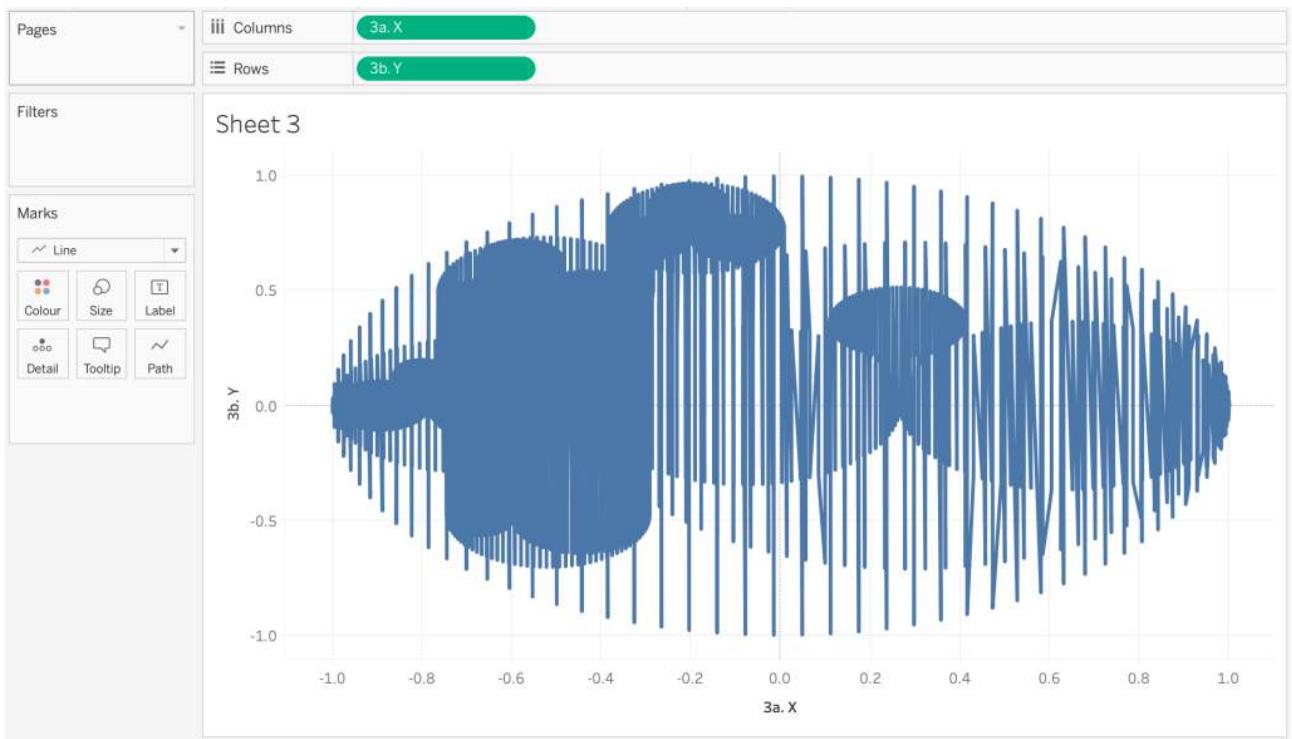
Now you can plot your 3a. X against your 3b. Y

(Side note: I apologise for the naming conventions the use of numbers and lettering is so that you can see in which order the calculations are made)

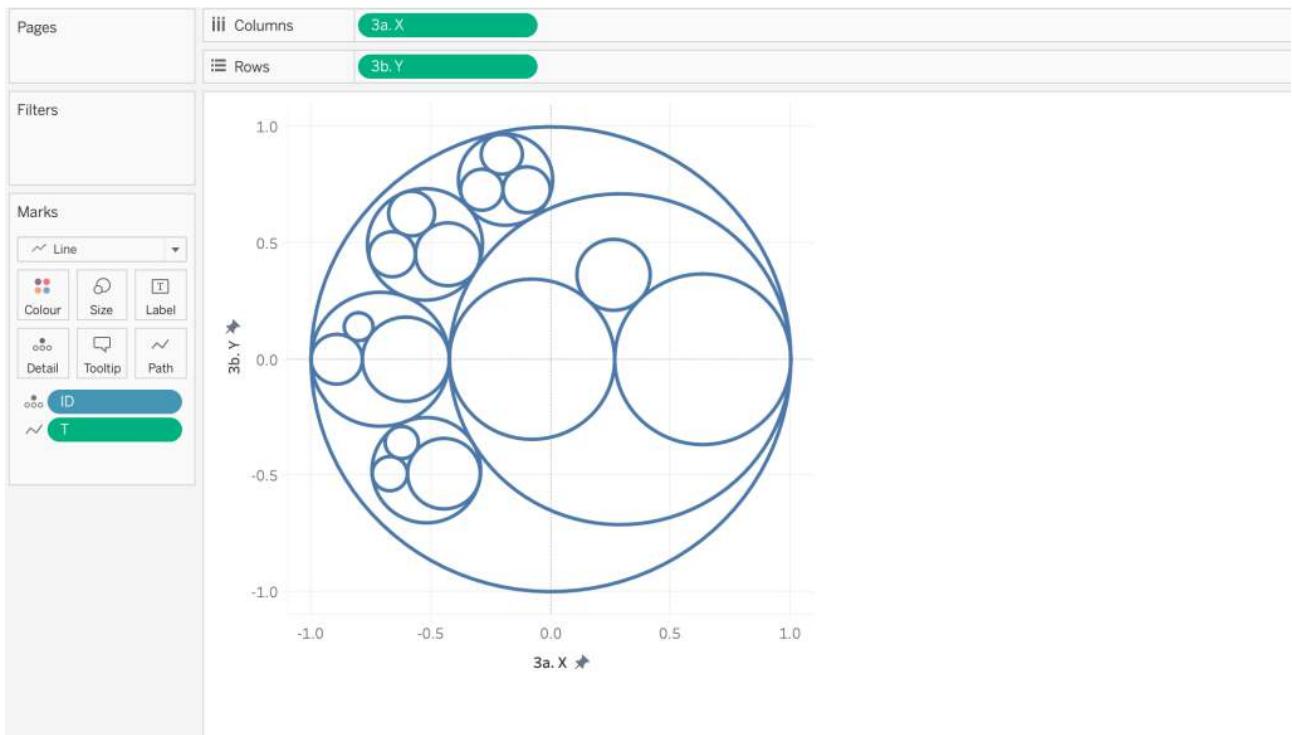
Make both X and Y dimensions.



Add ID to detail and change Marks to Line.

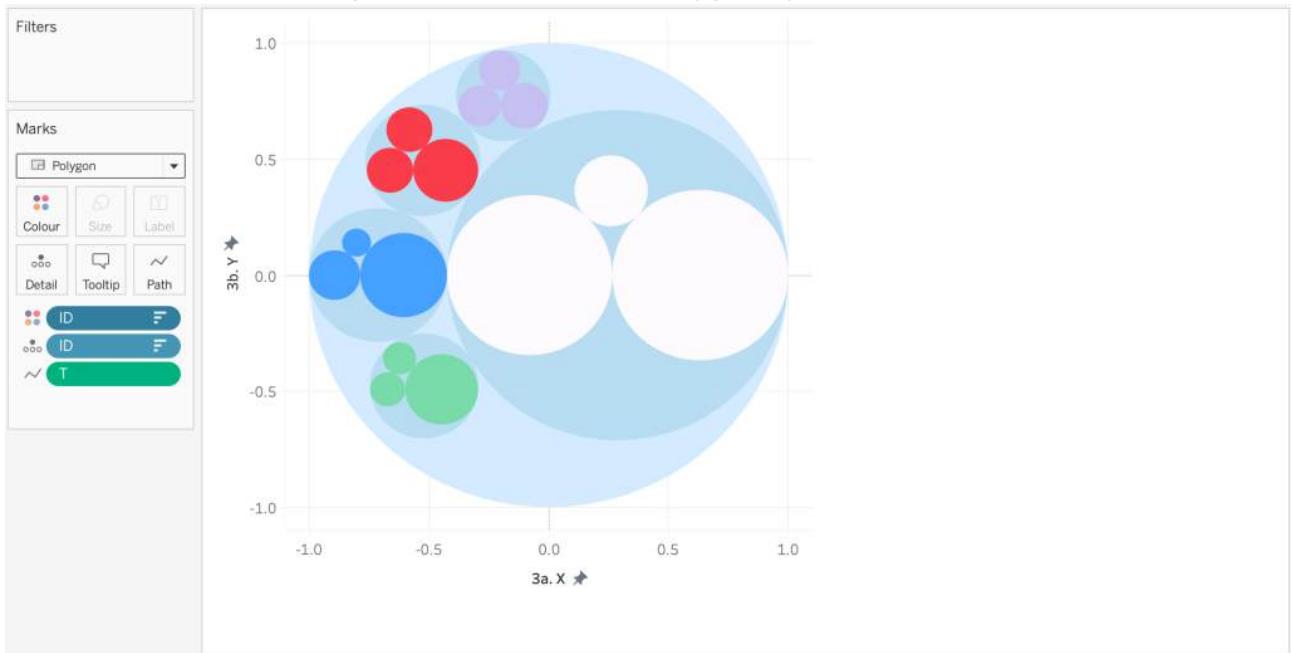


Fix the axis between -1.1 and 1.1, remember our python code made the circles max size 1.



Add ID to colour, Sort the colour descending on Rank. (This is because we want Level 1 at the back, i.e the world. Level 2 of continents to sit in front, and then Level 3 of countries to be on top.)

You can then change the marks card to be polygon if you want the circles to be filled!

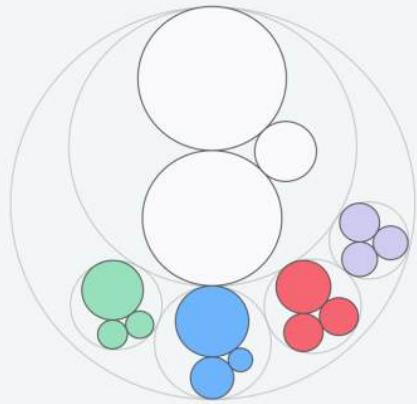


As further steps, I added a few more calculations 4a,5a,5b to the workbook to show how you can make the most of layering functionality to create some pretty cool alternative versions of the chart. They cover off how to make some of the circles lines and some as polygons. Download the workbook to see.

There we have it our finished packed circle!

CIRCULAR PACKING TUTORIAL

CREATED USING PYTHON AND EXPORTING CO-ORDINATES
TABLEAU: USING MAP LAYERS, LINE, POLYGON TOOLS



CIRCULAR PACKING OR CIRCULAR TREEMAP ALLOWS THE INDIVIDUAL TO VISUALISE A HIERARCHIC ORGANIZATION. IT IS AN EQUIVILANT OF A TREEMAP OR A DENDOGRAM, WHERE EACH NODE OF THE TREE IS REPRESENTED AS A CIRCLE AND ITS SUB-NODES ARE REPRESENTED AS CIRCLES INSIDE OF IT

PLEASE NOTE THE USE OF EXAMPLE DATA - THE DATA DOES NOT REPRESENT TRUE VALUES IN RELATION TO A COUNTRY METRICS OR REPRESENT ALL COUNTRIES

As always, Let me know how you get on with this one. I can be reached on Twitter, @_CJMayes.

LOGGING OFF,

CJ