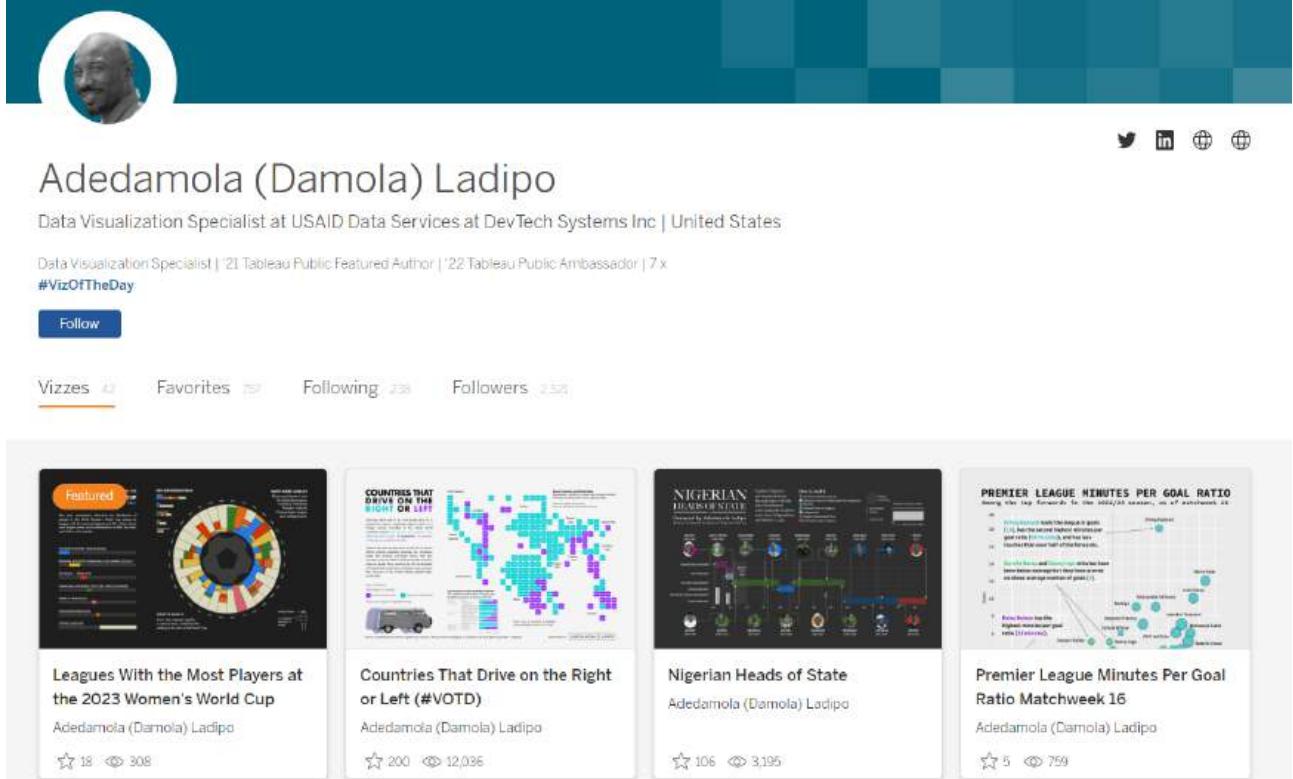


Enjoy this site?  Gift the author a WordPress.com plan.

[Gift](#)

THE INTERSECTION OF ART, DESIGN AND DATA WITH DAMOLA

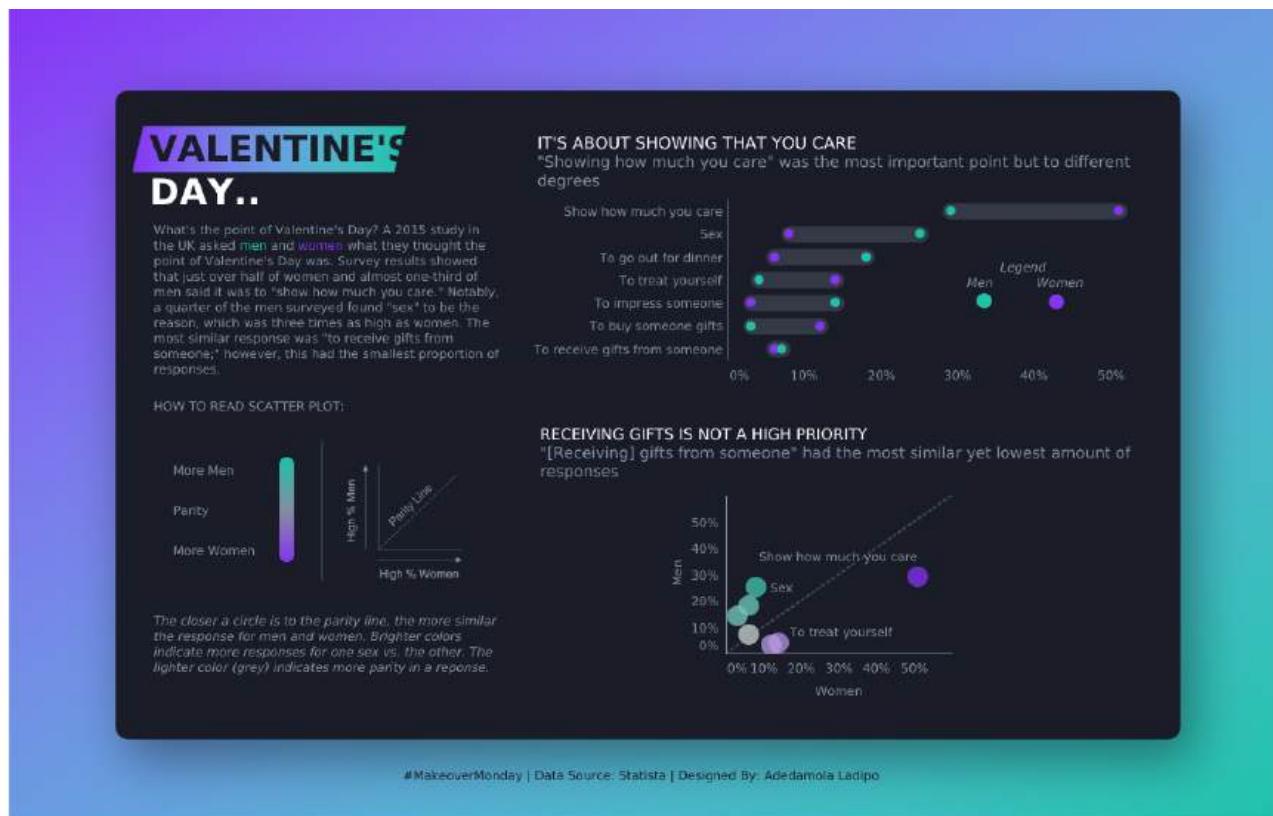
Hi all, I'm really pleased to be able to close out the guest blogs for 2023 with none other than [Damola](#). Damola brings to Tableau Public a unique style, blending both art, design and technical skill to curate what is probably one of my favourite portfolios for Tableau right now. It will come as no surprise that Damola's efforts are therefore recognised through Tableau Public's ambassador program and landing 7 viz of the days at the time of writing. I sat down with Damola to learn more about his designs.



The screenshot shows Damola's Tableau Public profile. At the top is a circular profile picture of him smiling. Below it is his name, "Adedamola (Damola) Ladipo". Underneath his name is his title, "Data Visualization Specialist at USAID Data Services at DevTech Systems Inc | United States". To the right of his name are social media icons for Twitter, LinkedIn, and a globe. Below his title is a small note: "Data Visualization Specialist | '21 Tableau Public Featured Author | '22 Tableau Public Ambassador | 7x #VizOfTheDay". A blue "Follow" button is below this. Below the bio are four tabs: "Vizzes" (42), "Favorites" (757), "Following" (238), and "Followers" (2,521). Below the tabs are four examples of his work:

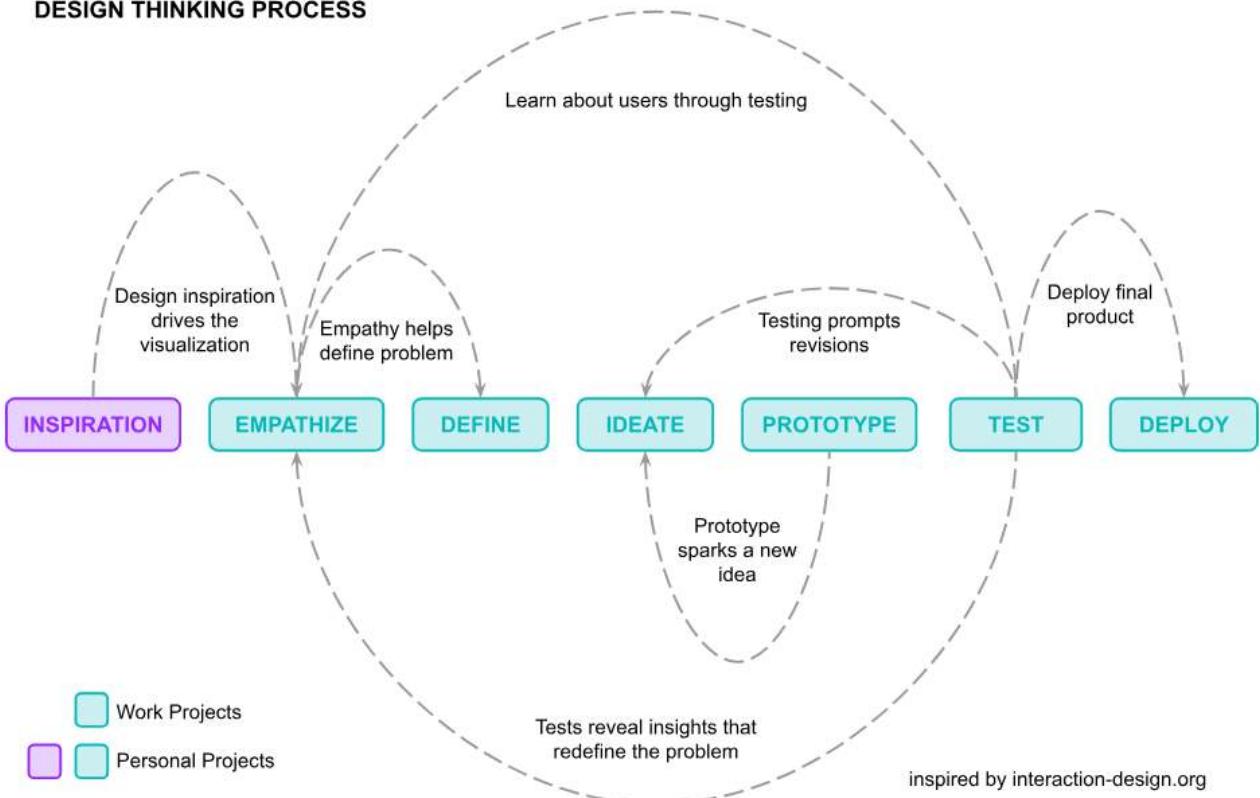
- Leagues With the Most Players at the 2023 Women's World Cup** by Adedamola (Damola) Ladipo. This is a circular sunburst chart showing the distribution of players across different leagues.
- Countries That Drive on the Right or Left (#VOTD)** by Adedamola (Damola) Ladipo. This is a choropleth map of the world where each country is colored based on whether it drives on the left or right side of the road.
- Nigerian Heads of State** by Adedamola (Damola) Ladipo. This is a dashboard featuring a timeline, a bar chart of life expectancy, and a map of Nigeria.
- Premier League Minutes Per Goal Ratio Matchweek 16** by Adedamola (Damola) Ladipo. This is a bubble chart showing the relationship between match week and average minutes per goal.

CJ: Damola, thanks for joining. For those that are unaware, May you tell us a little about how you ended up in the world of data? Does joining the community in 2020 coincide with your move to DevTech systems? D: I'm grateful for the invitation, CJ, and truly honored to be interviewed. Before my role at DevTech Systems, my professional journey led me through positions as a Monitoring and Evaluation (M&E) Specialist for NGOs in Central Asia and West Africa, and as an Assessment Officer for a Think Tank in the Middle East. Engaging in projects spanning climate change adaptation, internal and external displacement monitoring, water, sanitation, hygiene, education, and shelter, I found immense satisfaction in the multifaceted challenges. Yet, what truly captivated me was the world of data visualization. In these roles, I created charts, maps, and tables for factsheets, reports, dashboards, and presentations. While honing my skills on the job, I self-taught Tableau, Power BI, R, and Adobe InDesign, gradually building a portfolio of impactful visualizations. Around early 2020, fueled by a desire to focus my career on data visualization, I sought a role aligning with this passion within the international relations field. My portfolio and professional background ultimately led me to my current position at USAID Data Services with DevTech Systems, where I serve as a Data Visualization Specialist, primarily crafting Tableau dashboards for the agency. Indeed, you're correct in noting that my entry into the Tableau community coincided with my tenure at DevTech Systems. It was around the summer of 2020 that I discovered Tableau visualizations on Twitter, marked with #datafam and #makeovermonday. Witnessing the designs and possibilities in Tableau, I was inspired to create similar visualizations. **CJ: Makeover Monday was your first introduction of visuals to your page. Did you find different data sets helped you both from an analytical point of view but also testing your design skills?** D: Absolutely. Participating in #makeovermonday provided an exhilarating challenge, compelling me to assemble visualizations within tight timeframes while continuously pushing the boundaries of my skills and exploring new techniques. What made it particularly enjoyable was immersing myself in the designs of others, comparing and contrasting to identify areas of improvement in my own work. Drawing inspiration from diverse visualizations became an integral and rewarding aspect of my creative process. I embarked on #makeovermonday visualizations during the peak of the pandemic, and the community aspect played a vital role, offering a sense of connection during an isolated period and providing an expressive outlet in a medium I had grown passionate about.



CJ: I would consider your portfolio one of the most complete and beautiful tableau pages there is. Has design always been a passion of yours? Could you talk us through your process of design? Is it analytically, chart style, or design driven to start? D: I really appreciate that, CJ. I would say the same for your portfolio. Years ago I wanted to be a graphic designer like my father, but life took me down a different path. I think a part of me still wonders what it would have been like to have been one, which is probably one of the reasons design is such a major part of my work. My process for designing business dashboards or stand-alone data visualizations at work primarily follows the often non-linear design thinking process.

DESIGN THINKING PROCESS

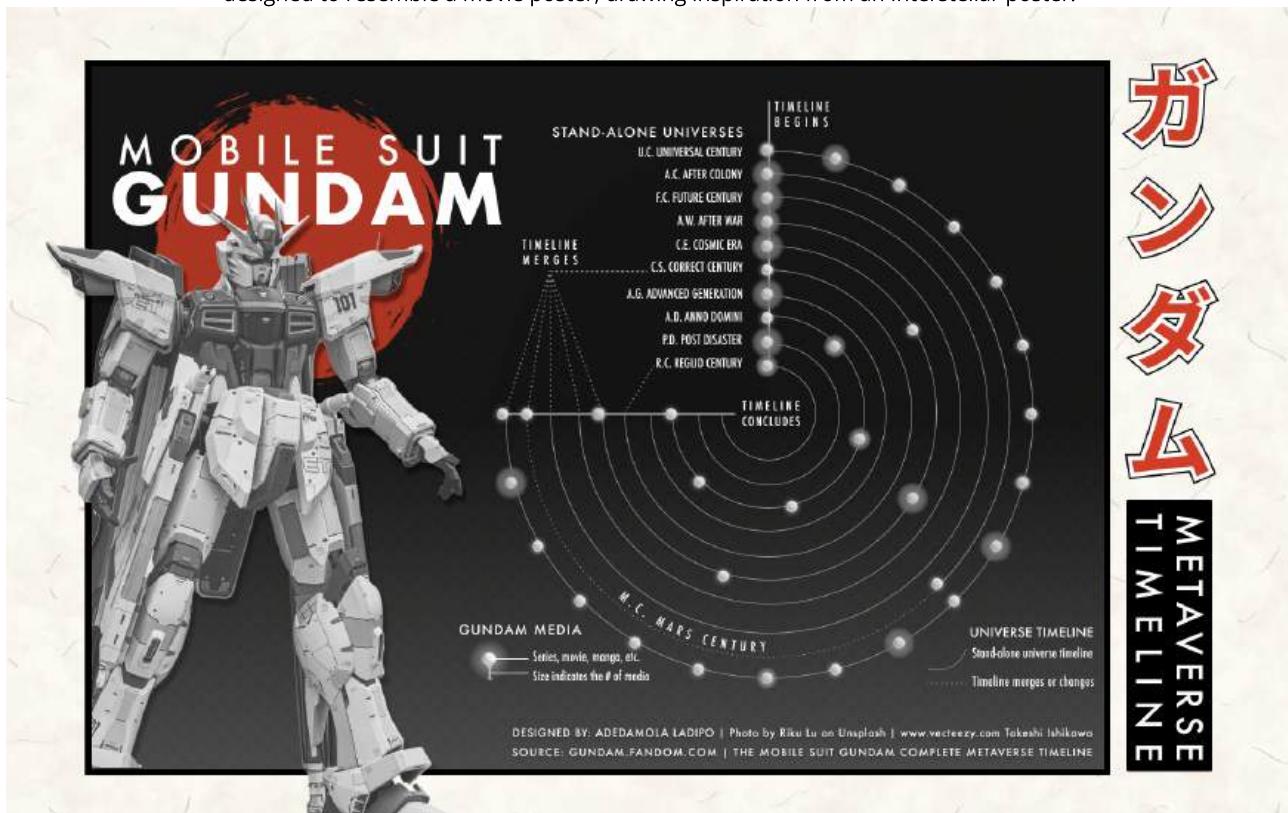


However, when delving into personal projects, inspiration for a specific chart style, layout, or design aesthetic often takes precedence. This is because I draw inspiration not only from other visualizations but also from diverse sources such as articles,

advertisements, scenes from shows, movies, documentaries, art, photography, and even packaging. These elements often become the focal point before diving into the empathy for the audience and ideation stages. This is usually because I gain a lot of inspiration from not just other visualizations, but also things I come across in articles, articles, advertisements, scenes from shows, movies, documentaries, art, packaging, etc. And I make that the first priority, before the empathy for the audience and ideation stage. Several instances illustrate scenarios where the design took precedence:



You can view the Christopher Nolan Films visual, [here](#). A data visualization about Christopher Nolan's movies (I'm a huge fan) designed to resemble a movie poster, drawing inspiration from an Interstellar poster.



[View Mobile Suite Gundam Metaverse Timeline](#)



A visualization inspired by a Mobile Suite Gundam (one of my favorite animes) [Gunpla](#) (Gundam Model) box, capturing the artistic packaging.

The Faces of BATMAN

Batman is one of the most well-known comic book characters of all time, and several actors have donned the cowl of the Dark Knight on film. This data visualization shows the actors who played the Caped Crusader in feature-length live-action films from 1966 to 2022, where Batman was either the lead or co-lead. The width of each actor's image indicates the total run time of their respective Batman movie(s).

Film	Year	Runtime	Change*
ADAM WEST	Batman: The Movie	1h 45m	
MICHAEL KEATON	Batman	2h 6m	▲
	Batman Returns	2h 6m	
VAL KILMER	Batman Forever	2h 2m	▼
GEORGE CLOONEY	Batman and Robin	2h 5m	▲
CHRISTIAN BAILEY	Batman Begins	2h 20m	▲
	The Dark Knight	2h 32m	▲
	The Dark Knight Rises	2h 45m	▲
BEN AFFLECK	Batman vs Superman	2h 33m	▼
ROBERT PATTINSON	The Batman	2h 56m	▲

*▲ = increase or ▼ = decrease in a film's runtime compared to the previous. Each image was captured from the respective movie promotional material.

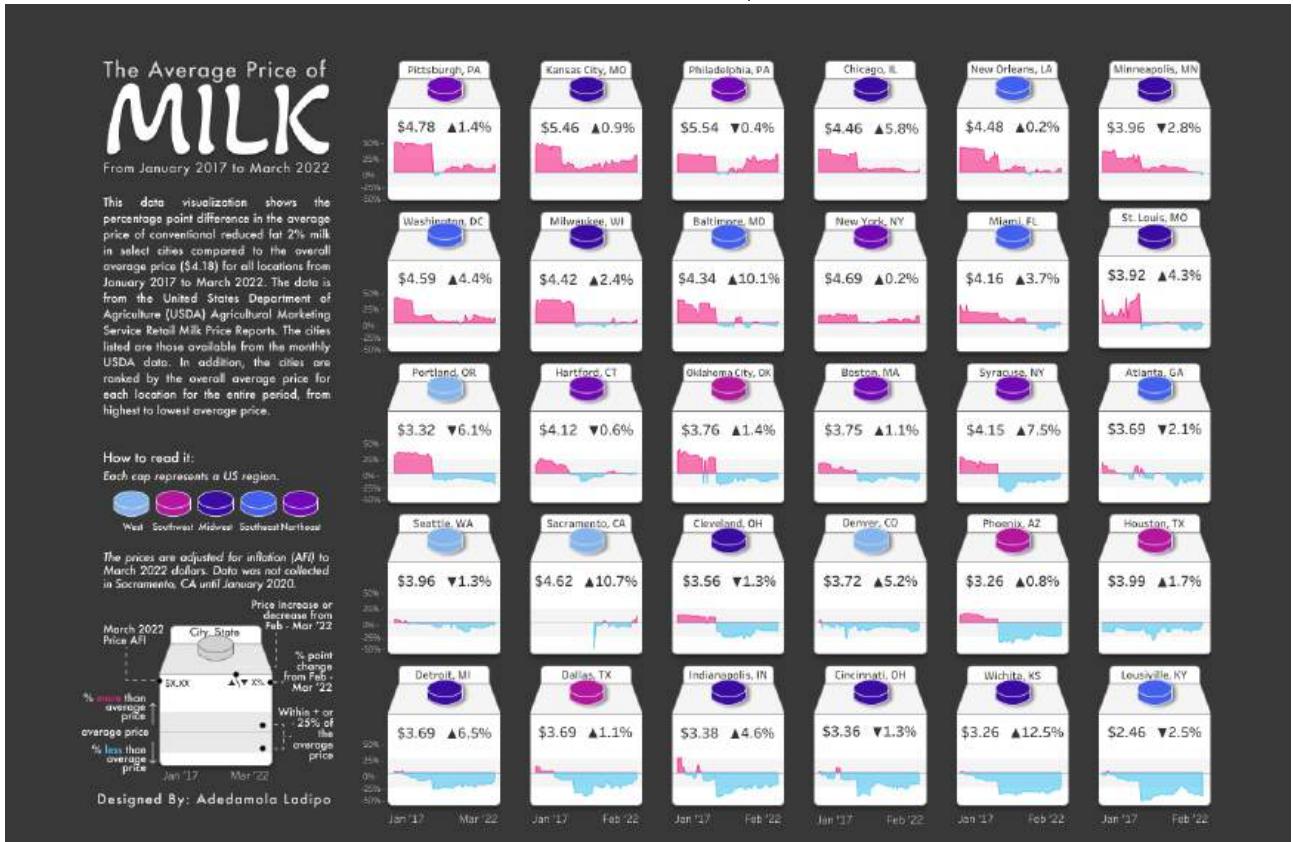


Designed by [ADEDAMOLA LADIPO](#)

[View The Faces of Batman](#)



The [Face of a Nation visualization by Guney Soykan](#) sparked my desire to create a similar visual of one of the most well-known character faces in film, Batman.



The Average Price of Milk The inspiration for this is probably the most obvious. Inspired by the idea of using milk cartons for small multiple charts, the visualization utilizes aspects of a typical milk carton to convey data for various locations. **CJ: Your profile raises awareness of some key events around racial bias, integration, and various diversity topics. I've noticed you steer more towards clear, more 'standard' chart types in these visuals – was this a storytelling choice?**

FOOD APARTHEID IN WASHINGTON, D.C.



This data visualization shows an analysis of food apartheid in Washington, D.C., by neighborhood cluster. Unlike the more commonly used term 'food desert,' food apartheid looks at the whole food system by taking into account income, race, and geography and how it impacts social and racial inequalities. The term food desert leaves out the critical social disparities that exist in food access. Overall, food apartheid shows that the systems in place make it difficult for people of color and low-income areas to access fresh and healthy food.

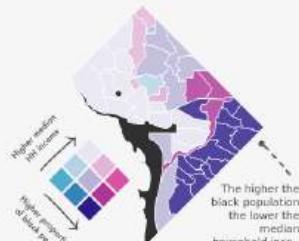
Specifically, the visualization includes four analysis areas of the proportion of black people and grocery stores and farmers markets per capita, the proportion of black people to median household (HH) income, grocery stores, and farmers markets per capita compared to median HH income, the location of every grocery store and farmers market with a 0.5-mile radius (standard distance for 100+ HHs in an urban area to access fresh food).



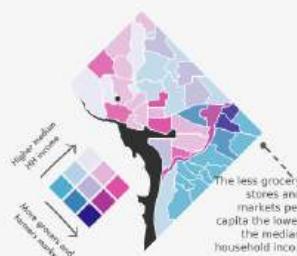
Proportion of Black Population:
& Farmers Markets and Grocery Stores Per Capita



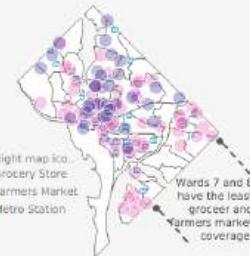
Median Household Income
& Proportion of Black Population



Farmers Markets and Grocer Stores Per Capita
& Median Household Income



Farmers Markets, Grocery Stores,
and Metro Station Locations



Source: Open Data DC, ArcGIS, Greater DC, DC Fiscal Policy Institute, DC School Reform, Investopia, The Green Urban Lunchbox
Designed By: Adedamola Ladipo

D: I'm deeply passionate about social issues, particularly those aligned with my personal interests in sports and film. However, my day job, focused on International Development operations dashboards, doesn't naturally always cover these topics. While I find the work enjoyable, personal projects offer the freedom to explore different data visualization approaches beyond creating standard business dashboard charts, maps, and KPIs. One of my early social issues data visualizations focused on "Food Apartheid in Washington, D.C.," shedding light on a prevalent yet insufficiently discussed issue in the city. The response from Tableau, the Data Visualization Society, Black In Data, the DC Department of Health, and the GitHub news blog exceeded my expectations. This experience reinforced my commitment to creating data stories on diverse social issues, aiming for general awareness and advocacy. **CJ: The breadth in chart styles on show in your portfolio is incredible, is showcasing data in different ways important to you? Was there one that you found particularly challenging to create, if so what technical elements played a part in that?** D: Thanks again for the compliment. Diversifying chart styles is indeed crucial to me. I thrive on the challenge of creating something novel. Reflecting on it, my visualization of "The Success and Failure of West Ham Managers" posed a few challenges. It marked one of the first instances where I incorporated custom shapes and map layers, coupled with my first ever attempt at a radial chart in Tableau.

WEST HAM UNITED

THE SUCCESS AND FAILURE OF MANAGERS

The 2000/01 to 2020/21 Premier League Seasons

How have West Ham United managers (gaffers) performed over the past 20 Premier League seasons? The manager's win, draw, and loss percentage per season are in the central radial chart in this visualization. In addition, the badge icons around the chart display summary statistics of each gaffer. These badges show the average stats of each manager compared to the average of all West Ham managers since the '00/01 season.

The season:

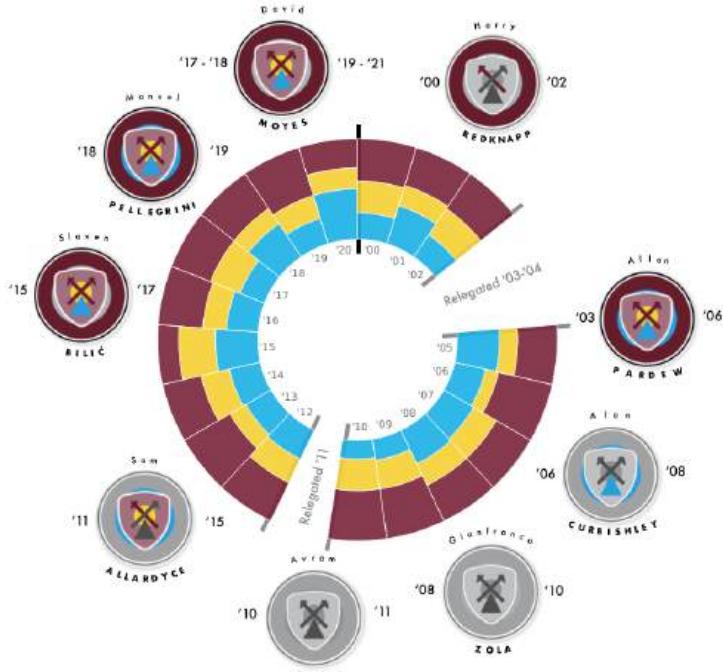
Each piece of the radial chart represents one season, and gaps indicate relegation. Relegation seasons are excluded.

Year season started → '00

The gaffer:

Colored badge icon = manager performed better than the average of WHU managers.

Grey badge icon = manager performed worse than the average of WHU managers.

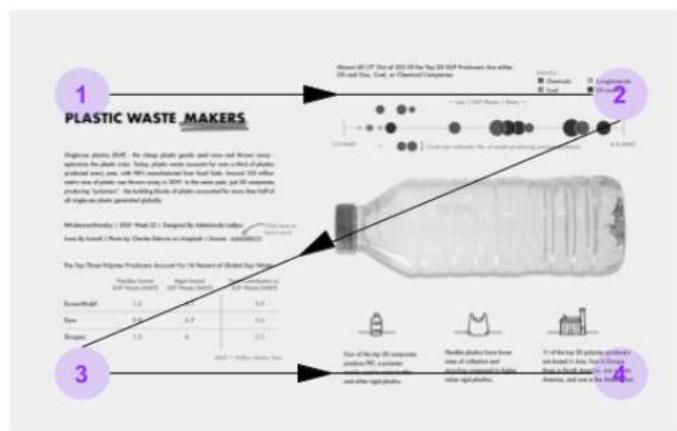
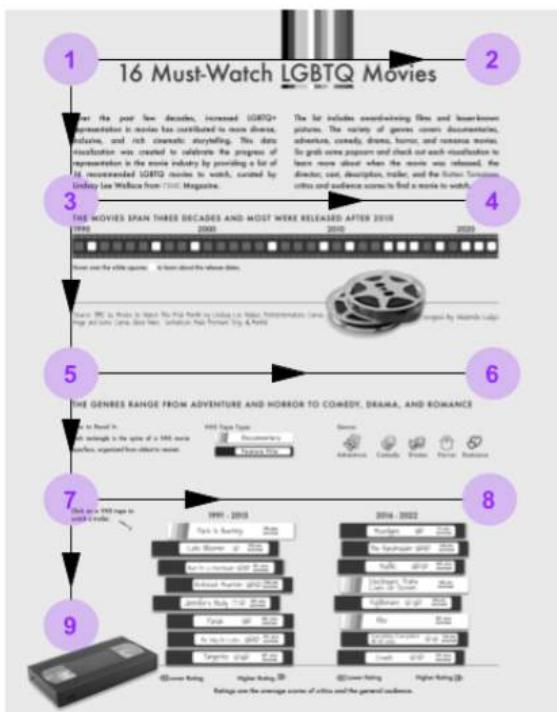


Source: worldfootball.net | Photo by: Photo by Samuel Regan-Asante on Unsplash | Designed by: Adedamola Ladipo

The Success and Failure of West Ham Managers The noteworthy technical challenge layer in deciding to display information only during West Ham's time in the Premier League. For periods of relegation, I wanted the visual to remain blank, emphasizing their time out of the Premier League. Achieving this involved utilizing calculations in Tableau and elements from Figma, such as grey lines to underscore when West Ham dropped down. I also owe you a lot of thanks for your "Game, Shape, Match" tutorial, which was instrumental in providing significant inspiration for the custom West Ham logos I created. **CJ: With much of your recent work, your workbooks consist of only one or two charts on the page. How do you approach the number of charts needed to solidify your story? Does positioning play an important role when thinking about legends and additional text context?**

D: My focus on a low number of charts is intentional because I like to center my visuals around one focal point and technique. I occasionally include complementary visuals to provide added context to the data story. The choice of where to place a chart is typically on a case-by-case basis, with an emphasis on visual hierarchy as the deciding factor. Hierarchy in a dashboard guides the user's attention to key elements. The average person's attention span today is 8 seconds, so I try to follow the reading pattern and rule of thirds hierarchy to create engaging content. **Reading pattern** Positioning is a crucial factor for the reader. I design my charts with the viewer's eye in mind, ensuring a natural flow that builds an

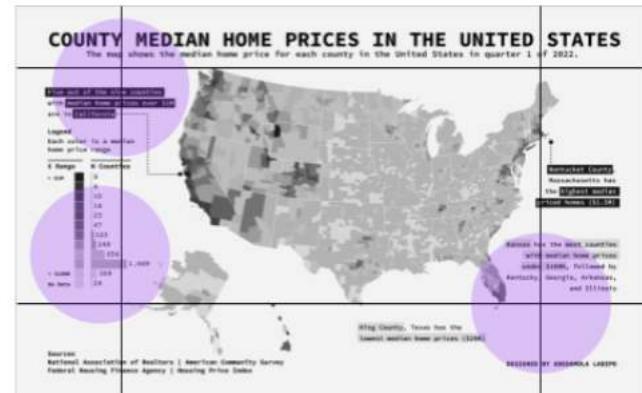
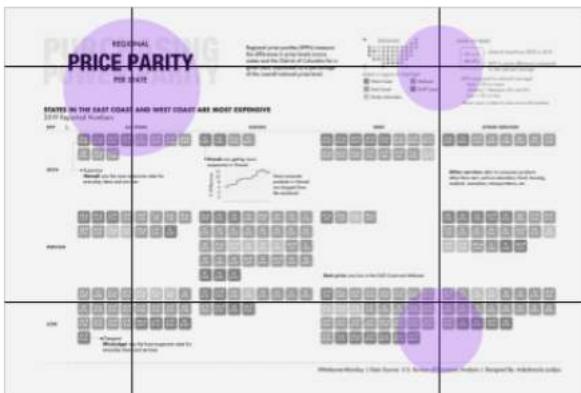
understanding of the portrayed story. Organizing visuals to follow an "F" or "Z" pattern.



16 must watch movies

Plastic waste makers

As the reader follows the flow of the dashboard, the arrangement ensures a logical sequence. **Rule of thirds** The elements in my visuals typically follow the rule of thirds when it comes to placement. This rule involves two vertical and horizontal lines crossing on a page. The points where they intersect direct the reader's eyes. I usually use these points to position titles, legends, source information, context information, or certain visualizations.



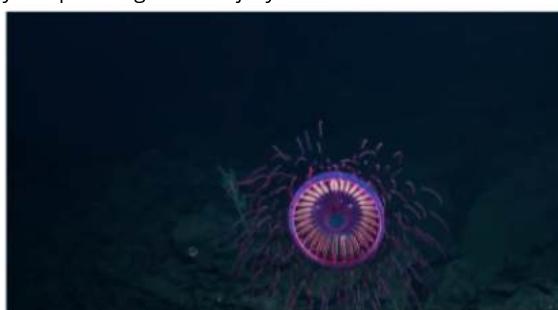
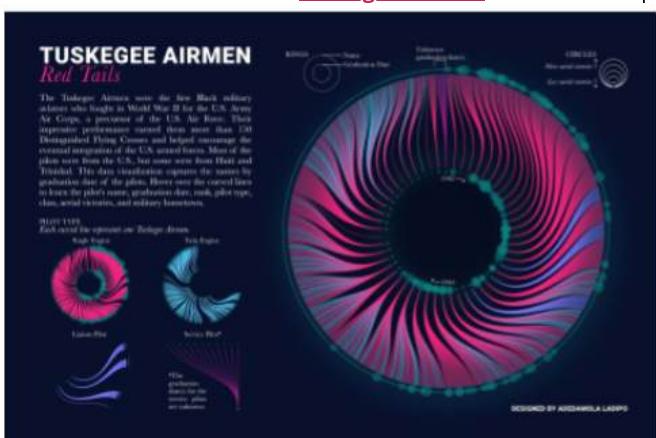
Regional Price Parity Per State

County Median Home Prices in the United States

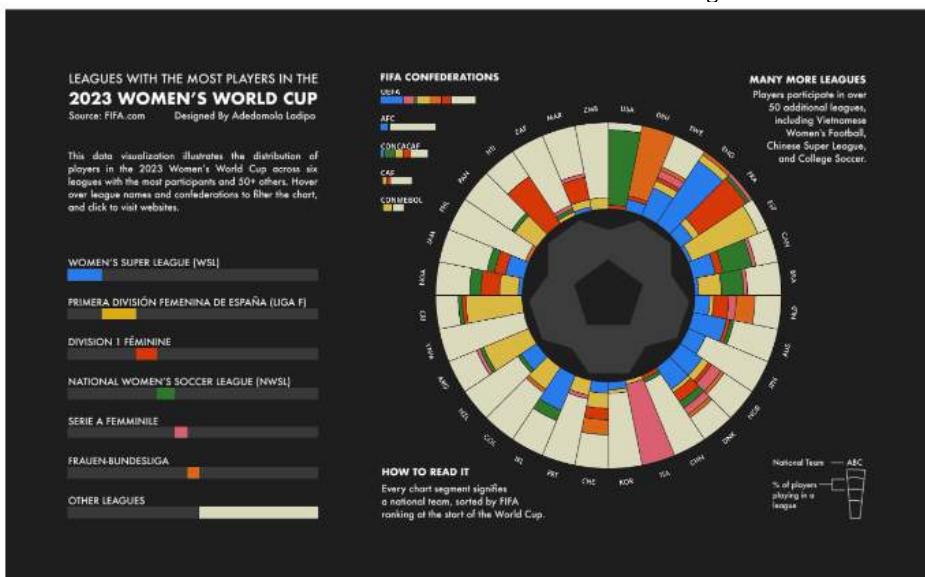
For a more in-depth explanation of design hierarchy also covering size, scale, and typography I recommend checking out [What is Visual Hierarchy and Why is it important](#) CJ: I am a huge fan of your SportsVizSunday content, obviously! Within your All-Time fastest 200-meters visual and your NFL record visual you tend to use a color palette of just 2-3 colors. How do you approach the logic of applying color to visuals? D: Thanks again! I'm also a big fan of yours. Your soccer/football visualizations are truly impressive and in my opinion, match the standards set by Opta Analyst, and Tifo for the Athletic as well as other sports analysts. Regarding my use of color, I am a firm believer in the "less is more" philosophy, a mantra I have even displayed on a poster next to my workspace as a constant reminder. I approach color selection with a deliberate mindset, considering its application when there's a clear need for specific colors, such as those from a logo. Additionally, I assess whether certain colors can evoke specific emotions or enhance visual appeal and engagement from the audience. The [All-Time Fastest 200 Meters](#) colors were inspired by the concept of speed and Nike.



The [Tuskegee Airmen](#) colors were inspired by a captivating video of jellyfish.



The [Leagues with the Most Players in the Women's 2023 World Cup](#) colors derived from the logo to maintain consistency with the branding.



CJ: What design tools do you currently use for your public work? Is there any data or design products you'd like to add to your toolkit? D: In my public work, I leverage Figma to prototype ideas and craft backgrounds, titles, and custom shapes for my Tableau visualizations. Recently, I've delved back into learning Tableau Prep to refine my data cleaning and preparation skills.

I aspire to rekindle my visualization journey with R and Python, both of which I intermittently use in my professional capacity.

Learning Blender and D3.js is also on my radar, though the challenge lies in carving out the time to do so. Lately, I've been exploring the tools available on [Ladataviz](#) by Tristan Guillemin. Tristan offers a suite of tools facilitating the creation of technically advanced charts—network, Sankey, bump, beeswarm, Voronoi, and chord—in Tableau with remarkable simplicity.

For instance, my [Agricultural Land Use](#) chart employed the Voronoi generator tool, and I look forward to incorporating additional tools from Ladataviz in future projects. **CJ: Are there any community members either within data visualization or**

the wider design community that have inspired your gallery of work? What was it that caught your attention? D: So many people in Tableau public and the wider data visualization community inspire me. The list could go on for a long time, but I'll to name a few (This is specific to Tableau Public): The following people inspire me on Tableau Public:

- [**Abhinav KR**](#) – Impeccably clean and captivating designs.
- [**Adam Green**](#) – Demonstrates high-level techniques in sports visualizations.
- [**Agata Ketterick**](#) – Crafts clean, well-designed visuals on informative topics.
- [**Alexander Varlamov**](#) – Utilizes advanced techniques in a remarkable manner.
- [**Ali Tehrani**](#) – Consistently produces clean, clear, and creative designs.
- [**Anastasia Komissarova**](#) – Showcases clean and eye-catching designs.
- [**Andy Kriebel**](#) – A vast library of visualizations, providing valuable learning resources.
- [**Ant Pully**](#) – Pushes the boundaries with techniques that seem almost impossible.
- [**Arshad Ejaz**](#) – Creates stunning technical visuals.
- [**Ben Norland**](#) – Each sports viz is a must-see.
- [**Autumn Battani**](#) – Combines cleanliness with creativity, offering helpful tips and guides.
- [**Chimdi Nwosu**](#) – Maintains a portfolio that sets the highest standard in design.
- [**Christian Felix**](#) – Demonstrates purposeful use of color and techniques.
- [**CJ Mayes**](#) – Strikes an excellent balance between technique and unique design.
- [**Dennis Kao**](#) – GIS wizard creating maps
- [**Ellen Blackburn**](#) – Sets the standard for well-designed business dashboards.
- [**Gbolahan Adebayo**](#) – Crafts clean and informative dashboards.
- [**Ghafar Shah**](#) – Covers interesting topics, telling visually captivating data stories.
- [**Idris Akilapa**](#) – Presents analytics in an informative way.
- [**Ivet Alexa**](#) – Possesses a portfolio filled with clean, creative designs and techniques.
- [**Joti Gautam**](#) – Creates visualizations that reach a level of data art.
- [**Judit Bekker**](#) – Each viz in the portfolio could be a viz of the day.
- [**Ken Flerlage**](#) – Presents a wide array of visualizations with great tips and tricks.
- [**Kevin Flerlage**](#) – A huge portfolio of excellent visualizations and example templates.
- [**Kevin Wee**](#) – Crafts eye-catching visuals with innovative techniques.
- [**Kimly Scott**](#) – Tells excellent data stories with captivating visualizations.
- [**Lindsay Betzendahl**](#) – Presents technical, unique, clean and innovative visualizations.
- [**Luke Abraham**](#) – A brilliant data storyteller with a captivating and consistent aesthetic.
- [**Pradeep Kumar G**](#) – A business dashboard extraordinaire.
- [**Priya Padham**](#) – Maintains one of the cleanest and aesthetically appealing portfolios.
- [**Ryan Soares**](#) – Tells data stories that captivate the audience with clear and clean visuals.
- [**Sam Epley**](#) – Creates a massive number of technical sports visualizations.
- [**Samuel Parsons**](#) – Presents a portfolio of awe-inspiring visualizations
- [**Simon Beaumont**](#) – Delivers brilliant sports visuals with advanced techniques.
- [**Tristan Guillevin**](#) – Creates spectacular visual tools using advanced techniques
- [**Wendy Shijia**](#) – Possesses a portfolio of stunningly creative visualizations
- [**Will Sutton**](#) – Provokes thought with data stories and eye-catching visuals.
- [**Zainab Ayodimeji**](#) – An amazing storyteller with an eye for design and unique visuals.

Outside of Tableau I draw inspiration from data visualizations created by;

- Ali Torban
- Edit Gyenge
- Frederica Fragapane
- Gabrielle Merite
- Krisztina Szucs

Their data visualizations are beyond brilliant and many of them transcend into not just being data visualizations but also. Their work is often awe-inspiring and well worth taking a look. **CJ Round-up:** Wowee what a list for individuals to refer to if they are just getting started with Tableau Public content. I'm so pleased that I had the opportunity to sit down with Damola, I find it fascinating when you mirror the end visuals with the source inspiration, be it through color palettes, posters, or objects further a field. I hope individuals loved this guest blog as much as me, some great takeaways around reading patterns and the rule of third. That's a wrap for guest blogs in 2023. Let me know who you think should be in the 2024 editions. **LOGGING OFF, CJ**

IRONVIZ FEEDER RESOURCES 2024

Hi all,

My 3rd time entering Iron Viz....

The past two times entering IronViz I've tended to write a mini blog on some of the aspects. They've never really covered anything to do with the mark scheme but solely looked at a few techniques that were used within the viz that may inspire your own work. This will be the same.

If you'd like to view the visual you can find it at the top of the page. There are 8 different views to the visual.

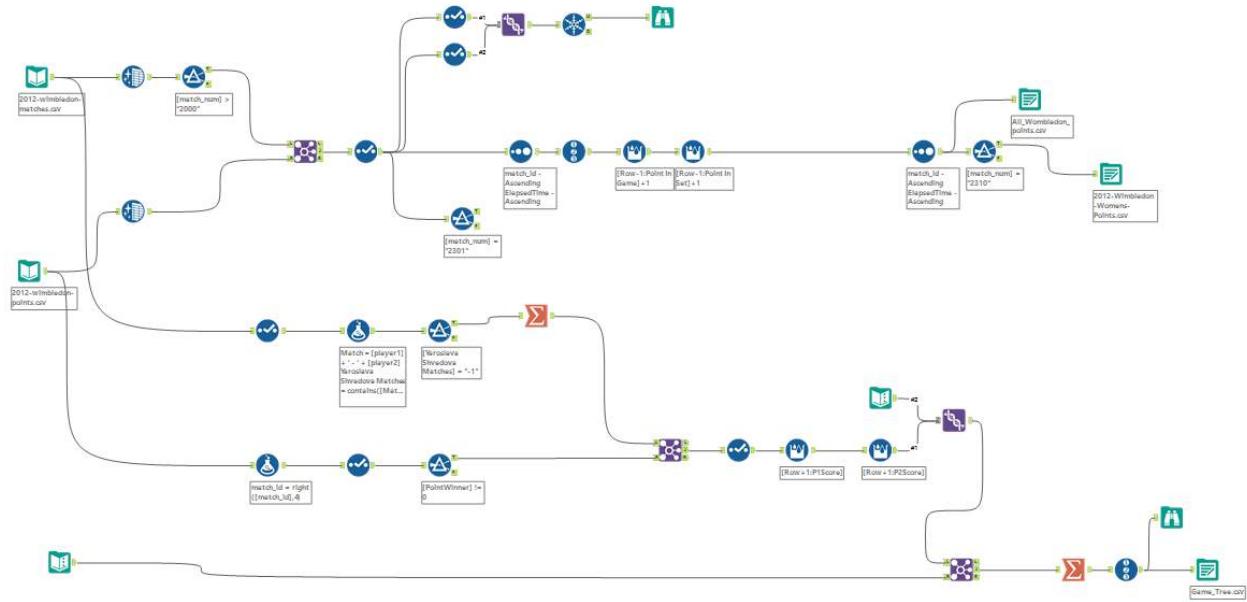


1.

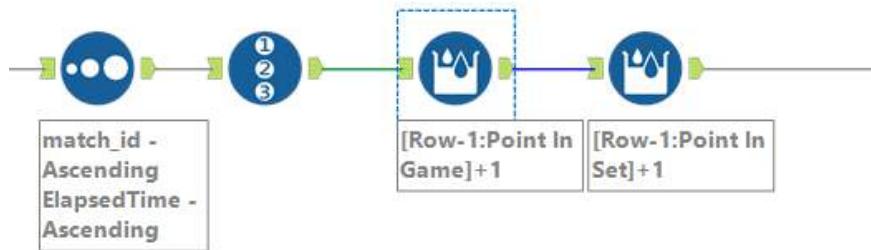
DATA PREP

In previous years, I've tended to use a variety of python scripts to manipulate a lot of the data. That or just sucked it up and amended data in excel. After 2 years of playing around with Alteryx I built what was quite a scrappy workflow. (I'm too embarrassed to upload it to Github but you can see from the screenshot the various outputs) The original data that I use ended up coming from Jeff Sackmanns public repository on Github. You can see it credited within the visual.

Some stand out transformations include:



- Taking a lot of Game Tree logic from a previous blog I wrote and adapting it to not look at server but to look at game and set combinations. You can read about the [game tree's here](#).
- For row point data, it was often the case I needed to know who the winner of the next point was.



- Adding in information such as current overall score, or changing a "Game" or "Set" flag to be a running count.
- Small subsets of data where there was a need to glue the names together to find specific match details.

Output Column	Data Preview
Match	Novak Djokovic - Juan Carlos Ferrero
fx [player1] + ' - ' + [player2]	

Output Column	Data Preview
Yaroslava Shvedova Match	0
fx contains([Match], 'Yaroslava Shvedova')	

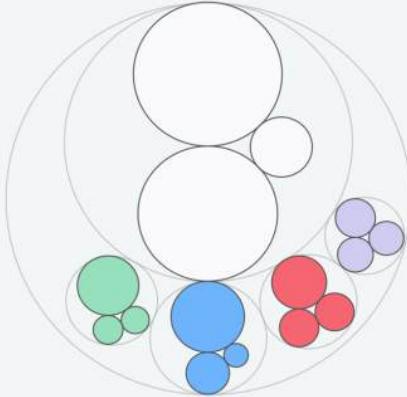
I think generally to speak of data prep over the years, this year was comparatively easier than my 2022 feeder (which had a lot of sentiment analysis, text cleaning and needing to use different python packages). Something to consider if you do enter and

what tools you do feel comfortable using, it meant more time in Tableau visualising relative to cleaning.

The only time I used actual code was a copy and paste job from my old tutorial on **circular packing**.

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

2. FONTS & DESIGN

This year IronViz was a lot more transparent around font sizing and accessible design.

Judging Criteria

Design

Visual elements add to the overall understanding of the visualization rather than distract

Interactivity and layout are user-friendly, instructed/specified, 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

In some ways this made life easier.... given only a number of fonts render well on Tableau public and desktop. But it also gave the opportunity to try build something completely native in Tableau.

The size 12 font minimum limit also made me much more consider the amount of text I was putting on the page. I often find with Iron Viz we add huge chunks of text elements into our entries, so this year I wanted to go a little more off piste and try utilise information buttons and let the user interpret a sequence of charts to showcase the main points of my story, without me having to explicitly say it.

In terms of resources, the safe fonts blog by the **Flerlage twins** is always a good one to go revisit.

The screenshot shows a dashboard titled "THE (ONLY) 7 SAFE FONTS FOR TABLEAU PUBLIC VIZZING". It includes sections for Arial (28 pt), Courier New (28 pt), Georgia (28 pt), Times New Roman (28 pt), Trebuchet MS (28 pt), and Verdana (28 pt). Each section displays sample text in various styles (normal, italic, bold) and accessibility samples. A note at the bottom states: "If you choose to incorporate a non-Tableau safe font by using an image of the text, be sure to include Alt text of the image to maintain accessibility. The Accessibility Sample text reads: 'eye-all-one-size-been-zero-share-an-em'."

Font	Style	Sample Text	Accessibility Sample
Arial (28 pt)	Normal	The quick brown fox jumps over the lazy dog. 1234567890 ?! (11 pt)	Accessibility Sample: II1 db C00 rn m
Courier New (28 pt)	Normal	The quick brown fox jumps over the lazy dog. 1234567890 ?! (11 pt)	Accessibility Sample: III db C00 rn m
Georgia (28 pt)	Normal	The quick brown fox jumps over the lazy dog. 1234567890 ?! (11 pt)	Accessibility Sample: III db C00 rn m
Times New Roman (28 pt)	Normal	The quick brown fox jumps over the lazy dog. 1234567890 ?! (11 pt)	Accessibility Sample: III db C00 rn m
Trebuchet MS (28 pt)	Normal	The quick brown fox jumps over the lazy dog. 1234567890 ?! (11 pt)	Accessibility Sample: III db C00 rn m
Verdana (28 pt)	Normal	The quick brown fox jumps over the lazy dog. 1234567890 ?! (11 pt)	Accessibility Sample: III db C00 rn m

The above screenshot is from one of **Elisa Davis** dashboards, which highlights the rendering of safe fonts on Tableau.

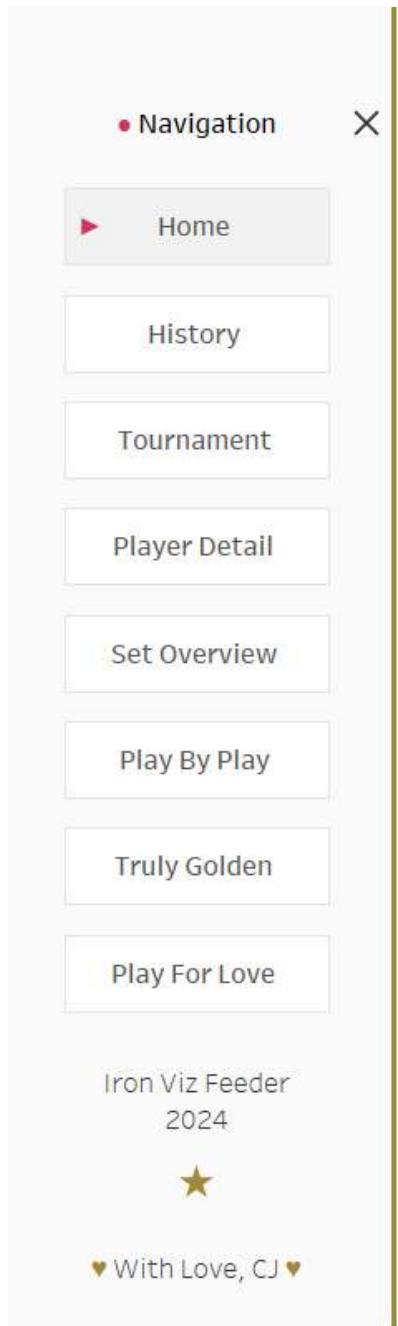
I ended up going for Tableau light (Which is of course a safe font given its Tableau) as well as a few titles in Cambria. Safe when rendering on tableau public on google chrome, but given I used it sparingly, and I quite liked it's effect it seemed like a reasonable choice.

3. LEVERAGE WORKBOOKS AND BLOGS

I think with everyone's entry, you tend to pick things up from the community over time. I know personally, I've picked up tips and tricks that have helped improve my work.

I wanted to point out a few more subtle influences of the dashboard.

I first learnt how to use dynamic zone visibility from work of Sam Parsons, in addition to using a show hide container to create collapsible containers. He has some wonderful advanced tutorials on his **Youtube**.

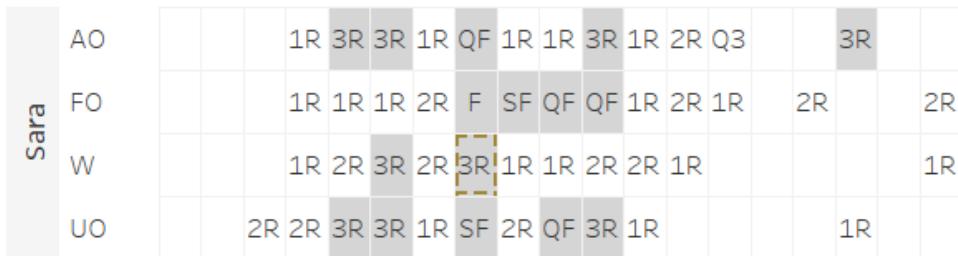


If you fancy something simpler, by all means I have some tutorials, [here](#). Next up was a bit more of a creative chart. Of course I can't enter Iron Viz without building something not seen before.



I would say the main inspiration behind this was Kevin Flerlage and Brian Moore. It is a bit of a mix between circular packing and rank charts. You can view my tutorial on circular packing [here](#).

I love the idea of combining different tutorials together to create something new.



This chart above is inspired actually by **WorkOut Wednesday**. Remember that really tough one set by Luke Stanke. I ended up writing about reference borders [here](#).

Requirements



[Click to open in Tableau Public](#)

- Dashboard Size: 700 x 700px
- 1 sheet
- Create a heat map that shows sales by month and sub-category – you will probably have one cell that shows no value, that's okay.
- For each cell of the heat map show the cumulative percent of total sales for each sub-category (the values should only count up–unless there are returns)
- Use the following calc to group your color coding: **ROUND([pct_of_total]*50, -1)**
- For every adjacent cell, if the colors are different, add a line that separates the colors of the two cells.
- Turn off tooltips
- Match the formatting
- ADVANCED: Make sure values return for all cells.

With the H2H chart there are a whole bunch of resources out there in terms of creating an arc chart. My favourite go to's for this are usually Brian Moore, or Toan Hoang.

Given I didn't need any "actual" data / information in, I could easily add this into the container and account for padding and adjustments to the Y axis to get it to align to the text.

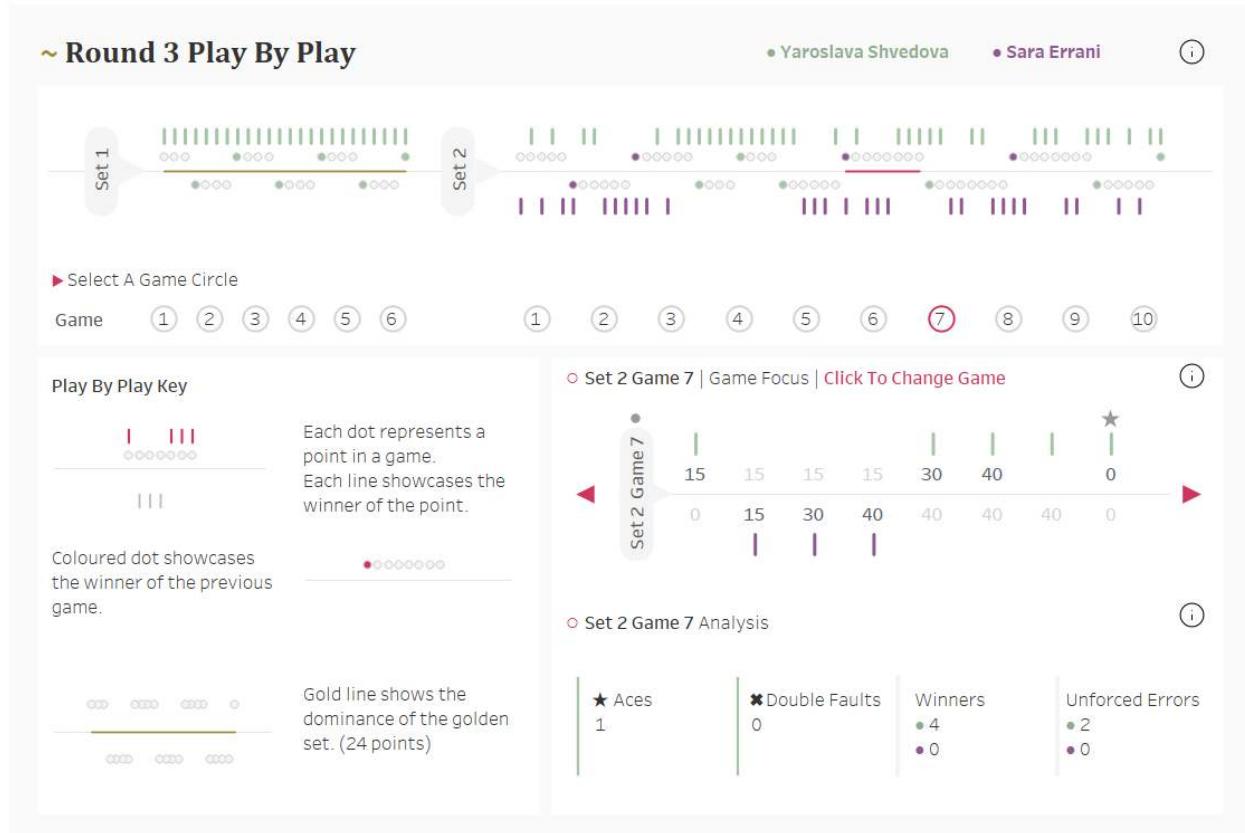
~ Head To Head



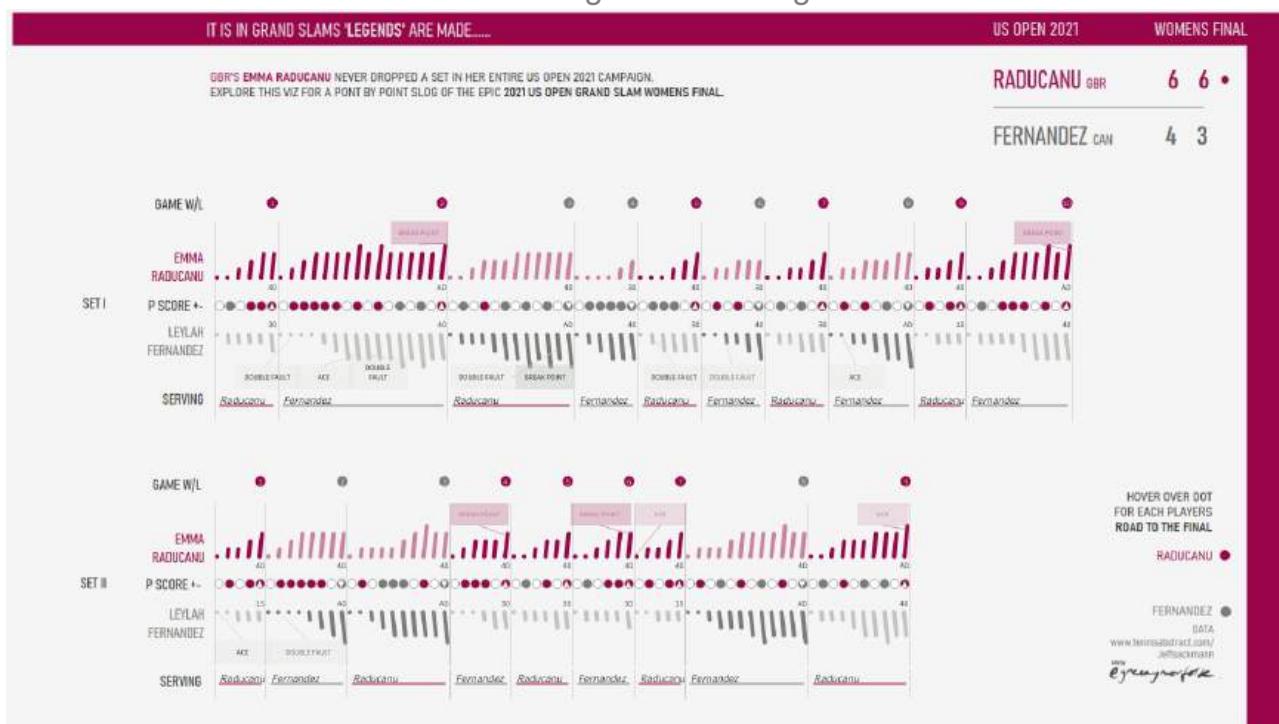
Yaroslava Shvedova and Sara Errani have played each other 6 times, both winning 3 matches.

2016 Charleston Round of 32	Y.S	3	3	-
	S.E	✓	6	6
2016 Dubai ...	Y.S	3	0	-
	S.E	✓	6	6
2012 Wimbledon Round of 32	Y.S	✓	6	6
	S.E	0	4	-
2011 Miami Round of 128	Y.S	✓	6	5
	S.E	3	7	6
2010 French Open Round of 128	Y.S	✓	6	7
	S.E	1	5	-
2009 Palermo Quarter Finals	Y.S	✓	2	5
	S.E	✓	6	7

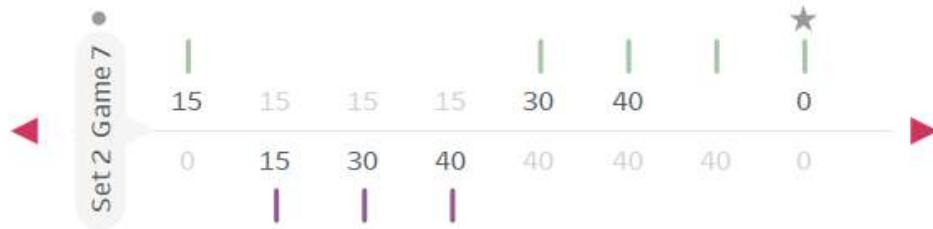
The final two parts of inspiration can be seen on the play by play stage



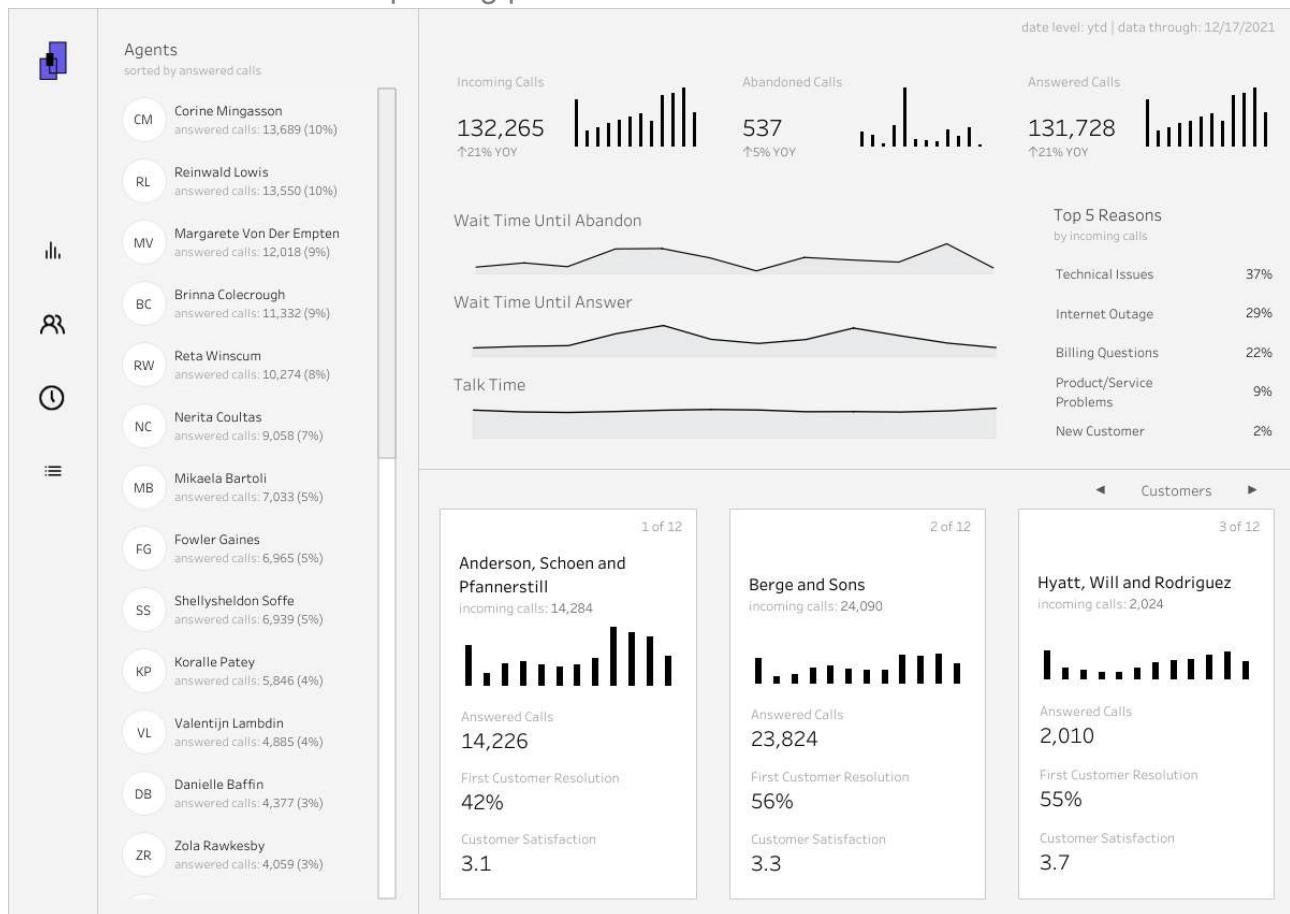
I am of course an avid Tennis fan and one of my favourite visuals comes from Adam Green who built this amazing visual looking at Emma Radacanu.



I wanted to add my own stamp to the idea so built it in one sheet, as well as added in a zoomed in interactivity element.



I ended up actually going and digging through some of **Autumn Battani's work**. I remembered she had built an amazing card switch in one of her call center dashboards. Of course, so different in design, but helped tremendously when thinking about logic of updating parameters and calculations.



4. A SMALL FEEDBACK GROUPS ARE BETTER

You don't need to necessarily hear too many opinions, just a select few who will take the time to give you honest feedback. I sometimes think it's a bit of a trap asking too many people.



Acknowledgements & Resources

- > Thanks to Michelle, Nicole and Brian in the Tableau Community for giving early feedback
- > This visual wouldn't be possible without the open data from Jeff Sackmann
- > All elements are built solely within Tableau
- > Dashboard was created as part of the Iron Viz Feeder Competition 2024

For me, this year I grabbed 15 minutes with Nicole, Michelle and Brian. I would say they saw it at a stage where I was 70% complete so was able to give some good thoughts around the flow of the visual in terms of ordering of the pages (Rarity, history, tournament, player, match etc). Once you've been in the detail for so long, it felt really beneficial to be pulled back out to look at how the different sections sync together. This was particularly important to me given I was trying to *make my charts do the talking*. Another piece of feedback I thought was particularly handy was the idea to put the how to read on the page for more complex visuals. Understandably people don't want to have to click / potentially miss hover material and not understand the more custom chart types.

The final big hitter (whey, punny right) was Brian mentioned that my Set Overview page would be nice to have both the non aggregated (each game) as well as aggregated view (set).

My instant reaction was ... “I wish you didn’t tell me that because I’m not sure I have enough time to figure it out!” but funnily enough a few tweaks to my Alteryx flow allowed me to group by game too and bring that in as a separate layer – thanks for the idea Brian!

~ Set Overview | A Closer Look



2012 Wimbledon | 3rd Round

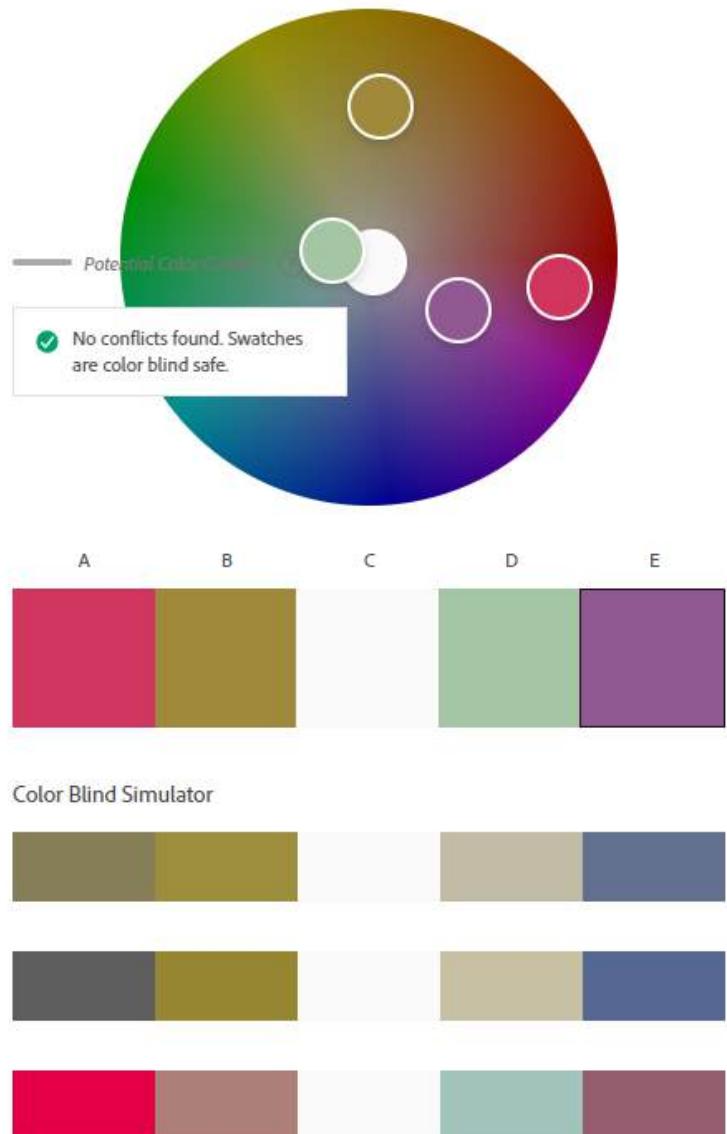


5. ADOBE COLOUR

I played around with a colour tool on **Adobe**, where I also created my palette to try and make sure it was suitable for all eyesight's. It is always my go to since coming across it years ago.



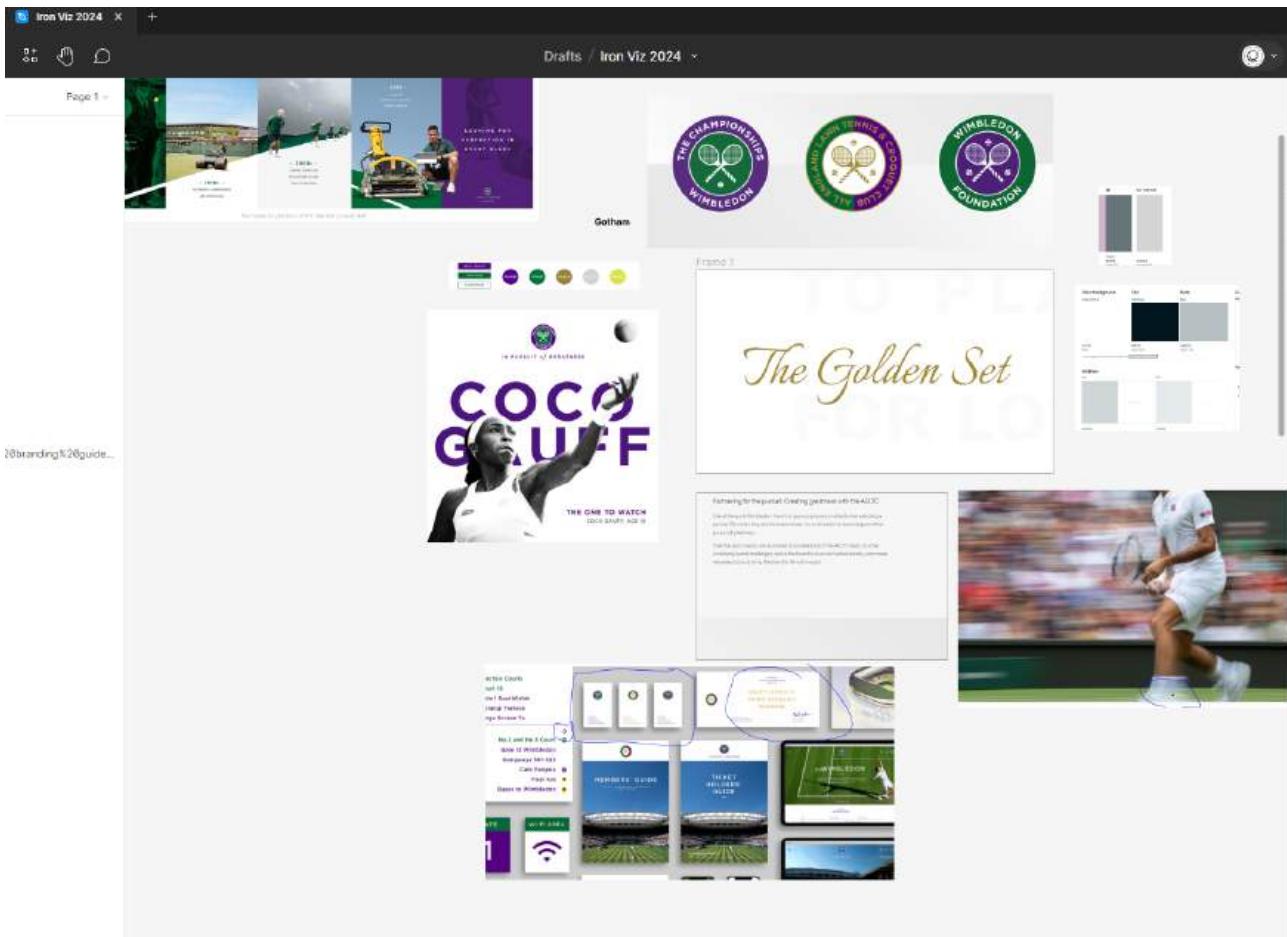
You can put it through a colour blind simulator. The colours I aimed for was to amplify the idea of the two players through a Wimbledon-esque colour scheme, as well as the gold as a highlight ability for the specific game and set. Other than that you'll see more neutral colours to support the visual in terms of the whites and greys on the page.



6. WIDER INSPIRATION

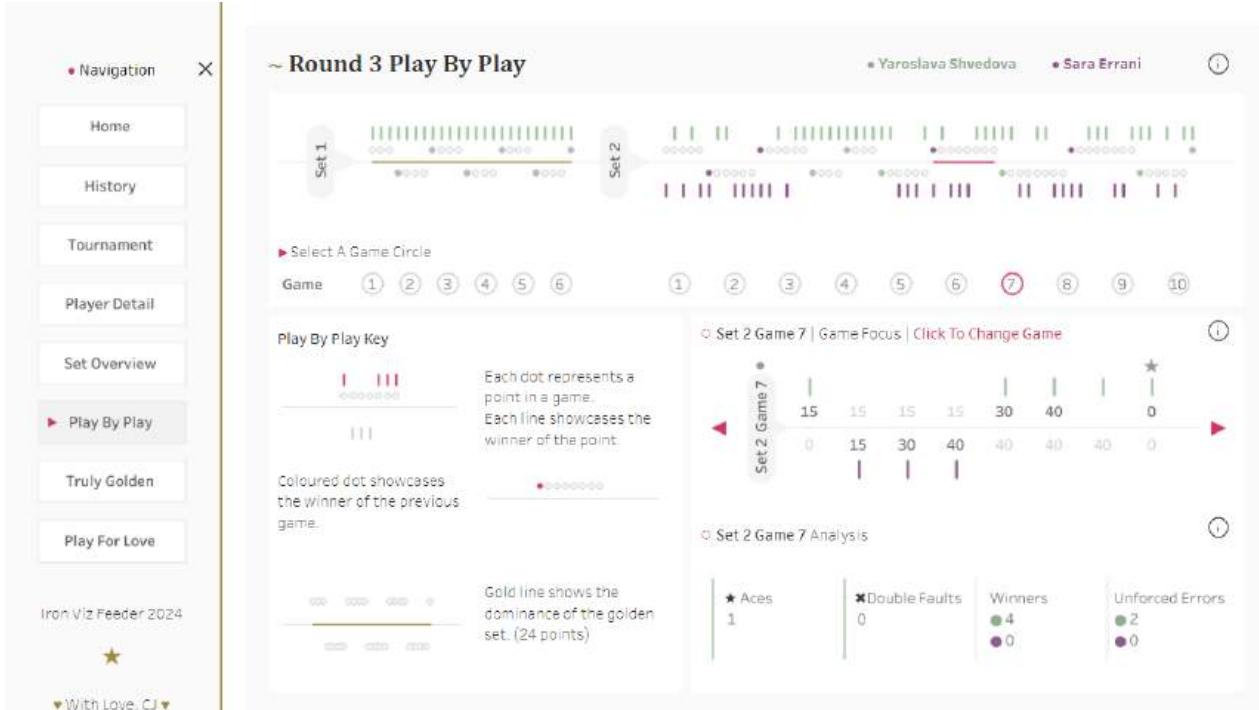
Nothing too new here. It goes without saying a board always helps when planning out visuals. Where else to go other than **Pinterest** and **Behance**.

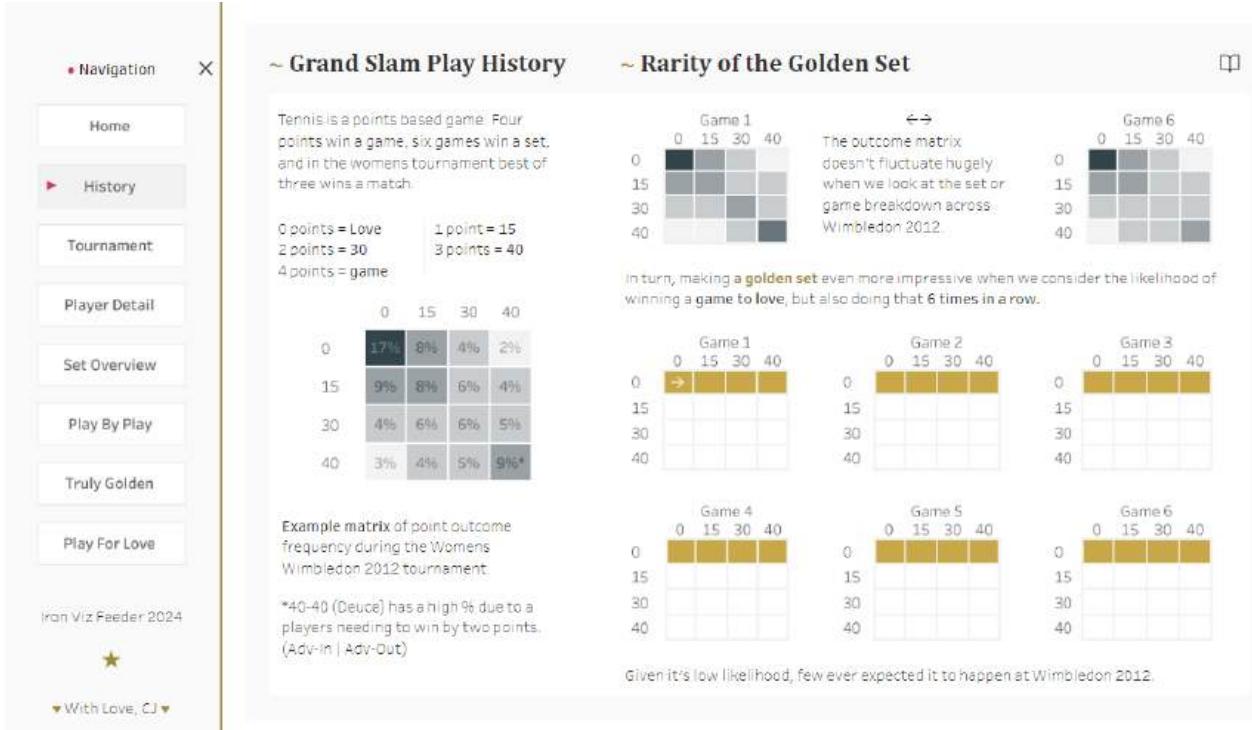
This year, given everything was built in Tableau I didn't have too many external resources, but I ended up using Figma as a bit of a scrap book and resource centre. So for example any websites referencing the golden set I would make a log of, as well as supportive design materials that may inspire the piece. I think you can see how the color palette slowly formed over time.



CJ Round-Up:

Below are a few of my favourite views from the workbook as a whole.





Looking forward to celebrating next years contestants. Iron Viz has a very special place

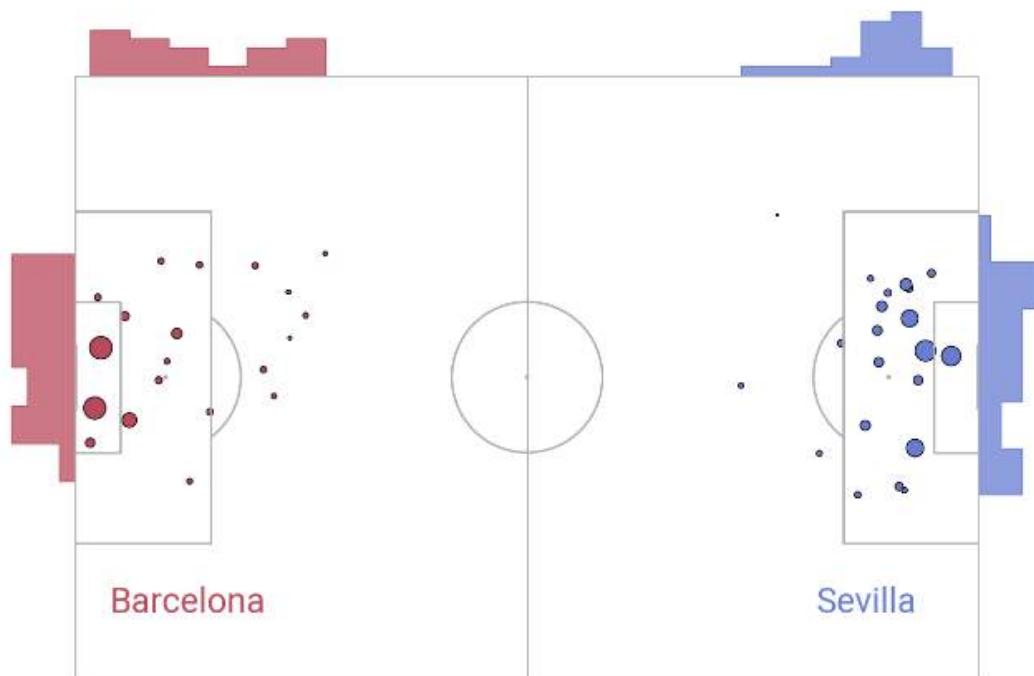
in my heart.

LOGGING OFF,

CJ

SOCCER STEP CHARTS IN TABLEAU

Hi all, Back to some basics this week. I've previously written extensively on soccer shot maps, looking at hex bins, heat maps, zones. This week we go back to our roots with the most simple shot map there is possible. Simple X and Y co-ordinates plotted on the pitch. If you're new round here, by all means check out my [very first tutorial](#) on this topic. Today we will be making a small addition to our standard shot maps, by adding in step charts. If you would prefer to do this in python instead of Tableau, I've taken the inspiration from the mplsoccer site, [read more here](#).

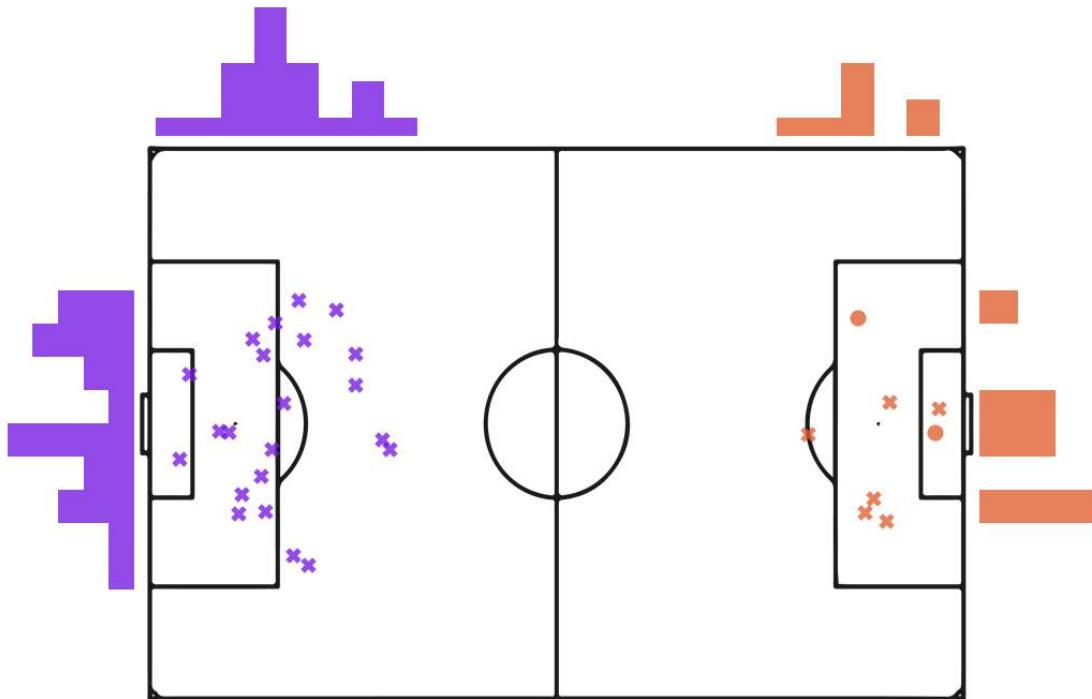


Todays shot

map data looks back to this match between Brentford and Arsenal from 2021. (No reason other than it just happens to be the data I had to hand at the time – of course use any match of your choosing – You will find a copy in the repo at the top of the page)

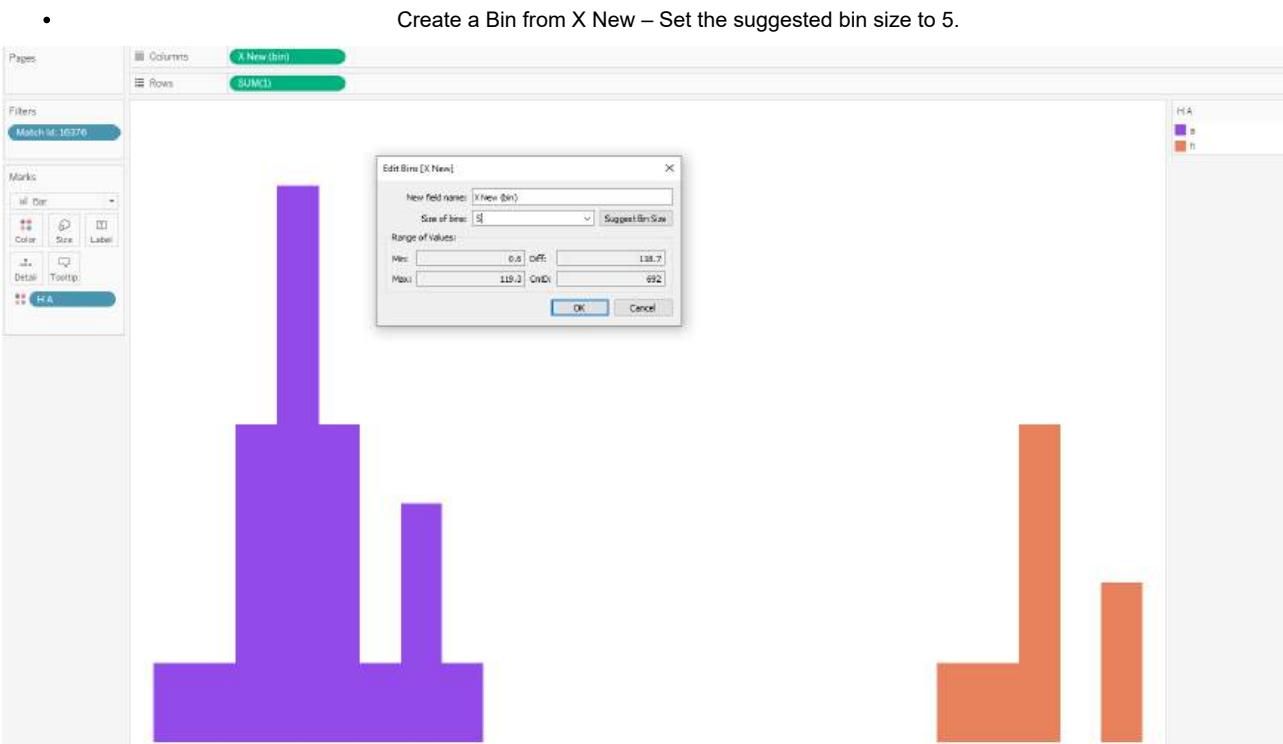


Step charts are useful to see the overall frequency of shots from designated zones. Of course we would both want and expect better chances to occur if created more front on to the goal and not wider as well as being able to determine shot frequency from distance from the goal, though this would hugely depend on the type of shot (header, foot etc)



Hopefully, by reading my original Understat blog post, I can assume you are comfortable with mapping on a background image, as well as plotting X and Y co-ordinates, though do message me if you get stuck and download the workbook. As a reminder, we split the x and y co-ordinates based on team (mirroring them each side) using the following calculations: X New if = "a" then -(*120)+120 ELSEIF ="h" then *120 END Y New if = "a" then -(*80)+80 ELSEIF ="h" then *80 END Now, Lets look at the Step Charts.

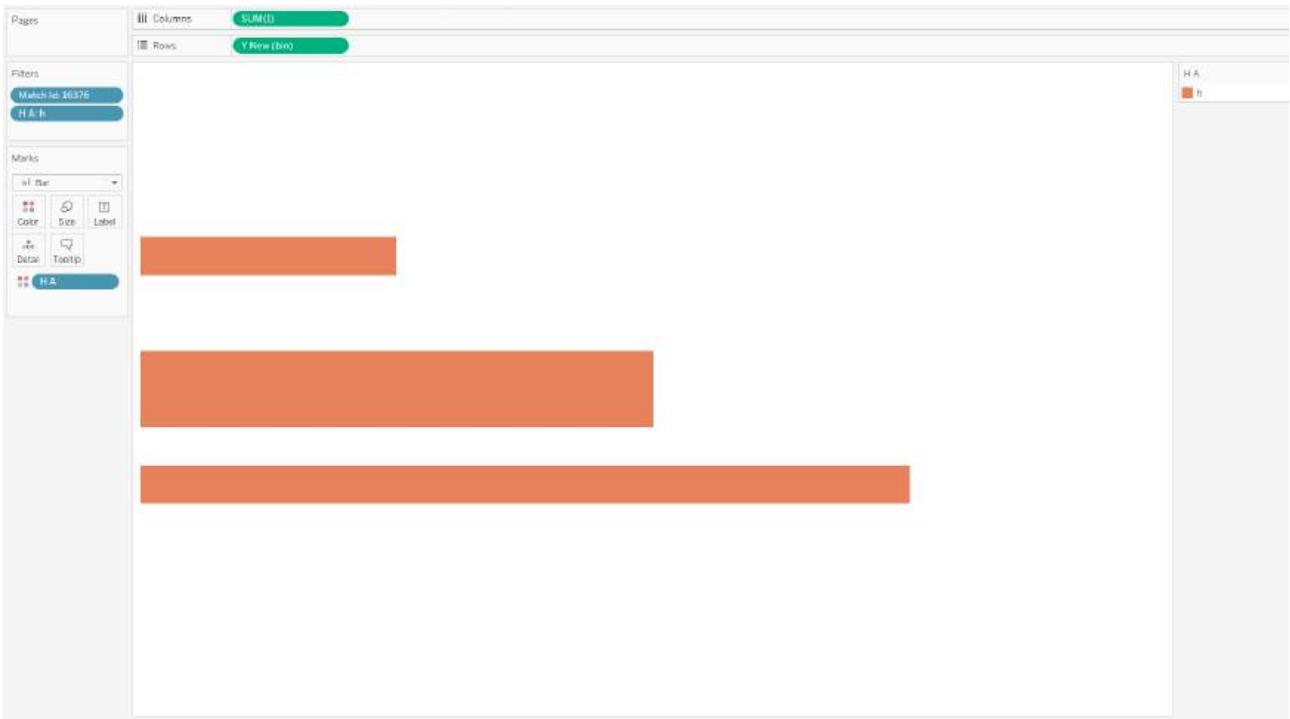
First we want to create a sheet to place above the top of the pitch.



When you first build the chart, the bin will be discrete. This is okay, but what you will find is that we need the axis to stretch the length of our pitch co-ordinates.

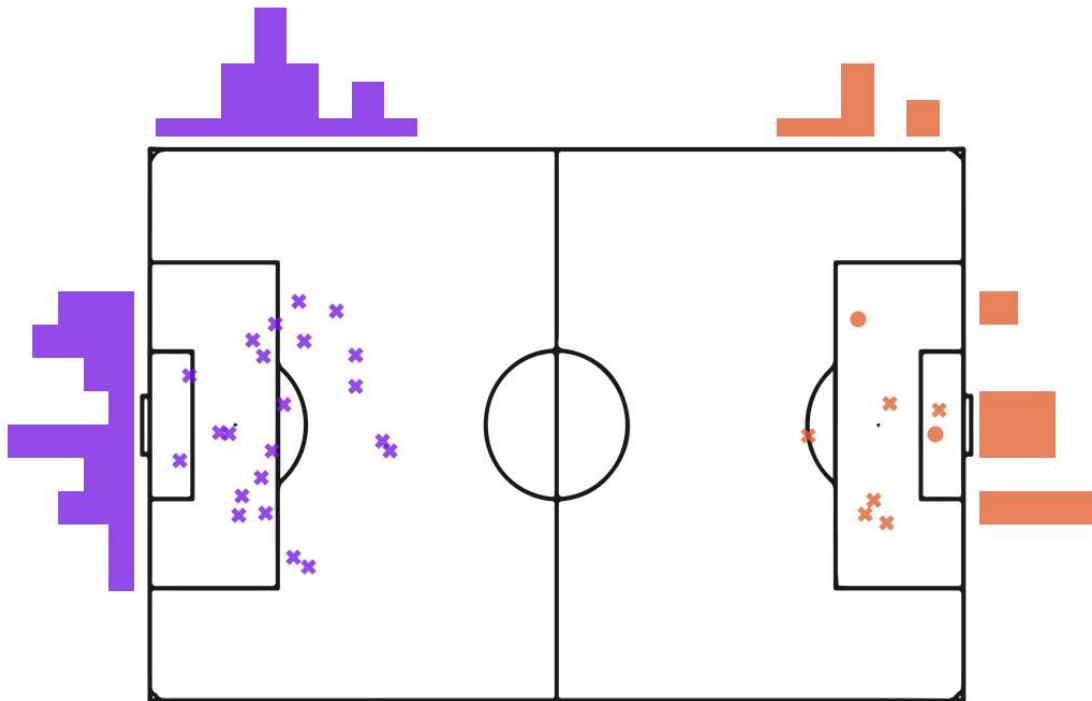
- Convert the bin to continuous.
- Make the appropriate cosmetic changes, I like to make my bars fixed width with no border.

- Adjust the axis, fixing it to the pitch length. (-2 , 122)
- Next we want to create the right and left hand bars.
- Create a Bin from Y New – Set the suggested bin size to 5.



We can copy the sheet, replacing the filter to the other team. Adjust the columns pill from sum(1), to sum(-1) to invert the axis. Again, edit the axis to accommodate the full length of the pitch, making it fixed. From there it is a matter of mastering your containers to place the 4 sheets on the dashboard!

Basic Shot Map with Step Charts | ● Goal | X Shot



Should be a nice easy one to re-create. Message me if you have any questions. Speak soon, **LOGGING OFF, CJ**

THE BEST VERSION OF YOU (QUANTIFIED SELF)

Hi all, A thought blog this week mixing a few of my passions together. Over the years, I've slowly been collecting more data about myself to track my progress, mostly relating to physical and mental wellbeing. Some individuals might steer away from the concept of this, after all it can be fairly personal thing but for me it really helps me stay on track. I have found my physical and mental have wider impacts on work, friendships and relationships. This blog will be about three of the main applications I use currently, as well as how we can start to link those datasets up together. If you use other apps too, then let me know!

1. Strava – (Free) Many will of heard of it, I use this to track all exercise I do, This includes workouts, weightlifting, running, as well as playing hockey. (I wear an apple watch)

You can follow me on [Strava](#) here, and join my [fitness club](#) here.

1. Hevy – (Free) This app I got at the start of the year, I use it specifically only for weightlifting sessions. It helps me track repetitions, sets and types of workouts I've been doing. (No tech required, manual app input)
2. WHOOP – (Paid) More recently I've been using whoop to track my sleep schedules, mental well being as well as log my activities.... they push straight to strava which is a life sender, The only time I don't use it is for GPS activities. (I wear the whoop 4.0 band)

You can find some snippets of my personal data in the repo. I've removed a huge chunk for protection reasons, but if its of interest what kind of information you can find and whether there are metrics in there you'd like to track. Let's take a little look at each application before looking to join the data up. **Strava App** I've written at length about strava, and how you can download your data in the past. They export it in a huge zip file and it contains everything from activities, to gpx file, through to your social elements of events, followers and logins. It's actually slightly overwhelming the amount of data Strava stores! the main one we will look at is the **activities**.

Name	Date modified	Type	Size
activities	29/09/2023 16:01	File folder	
clubs	29/09/2023 16:01	File folder	
media	29/09/2023 16:02	File folder	
routes	29/09/2023 13:42	File folder	
activities.csv	29/09/2023 16:01	Microsoft Excel C...	341 KB
applications.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
bikes.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
blocks.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
categories_of_personal_information_we_... ...	29/09/2023 16:01	Adobe Acrobat D...	52 KB
clubs.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
comments.csv	29/09/2023 16:01	Microsoft Excel C...	32 KB
components.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
connected_apps.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
contacts.csv	29/09/2023 16:01	Microsoft Excel C...	34 KB
email_preferences.csv	29/09/2023 16:01	Microsoft Excel C...	2 KB
events.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
favorites.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
flags.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
followers.csv	29/09/2023 16:01	Microsoft Excel C...	5 KB
following.csv	29/09/2023 16:01	Microsoft Excel C...	5 KB
general_preferences.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
global_challenges.csv	29/09/2023 16:01	Microsoft Excel C...	21 KB
goals.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
group_challenges.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
information_we_disclose_for_a_business_... ...	29/09/2023 16:01	Adobe Acrobat D...	73 KB
local_legend_segments.csv	29/09/2023 16:01	Microsoft Excel C...	1 KB
logins.csv	29/09/2023 16:01	Microsoft Excel C...	3 KB
media.csv	29/09/2023 16:02	Microsoft Excel C...	13 KB
memberships.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
metering.csv	29/09/2023 16:02	Microsoft Excel C...	178 KB
mobile_device_identifiers.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
orders.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
partner_opt_outs.csv	29/09/2023 16:02	Microsoft Excel C...	2 KB
posts.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
privacy_zones.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
profile.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
profile.jpg	29/09/2023 16:02	JPG File	24 KB
reactions.csv	29/09/2023 16:02	Microsoft Excel C...	164 KB
routes.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
segment_feedback.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
segments.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
shoes.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
social_settings.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
starred_routes.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
starred_segments.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
support_tickets.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB
visibility_settings.csv	29/09/2023 16:02	Microsoft Excel C...	1 KB

The activities folder contains GPX files, [this blog](#) shows how to prep a Strava map in Tableau.

The activities

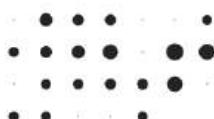
Cj Mayes

Last 4 Weeks

23

Total Activities

M T W T F S S



• Walking • Cycling • Swimming • Hiking



The free version of strava is very useful for:

- Map coordinates
- Looking at pace per mile or km
- Heart rate data
- General overview of number of activities.
- Gamification (badges for whether you've run 5/10/21.1km runs or done over 100km that month are some of my favourite ones to go for)
- Hours spent exercising a month

16:49

You

Progress Activities Profile

Cj Mayes

Following | Followers
112 | 104

Edit

Run | Workout | Weight Training | Walk

This week

Distance Time Elevation
0.00 km 1h 4m 0 m

5h 20m 2h 40m 0h 1h 4m

AUG SEP

Activities 6 hours ago

Statistics This year: 1,101.9 km

Home Maps Record Groups You

Hevy App The Hevy App export is just one csv file – it is at the exercise set level so there are multiple rows per workout. You can check the input file in the github repo to get a flavour for what's included.

Afternoon	20 Sep 2023	20 Sep 2023, 15:08	Lateral Raise (Dumbbell)	1	normal	10
Afternoon	20 Sep 2023	20 Sep 2023, 15:08	Lateral Raise (Dumbbell)	2	normal	10
Afternoon	20 Sep 2023	20 Sep 2023, 15:08	Lateral Raise (Dumbbell)	3	normal	10
Afternoon	20 Sep 2023	20 Sep 2023, 15:08	Lateral Raise (Dumbbell)	4	normal	10
Afternoon	20 Sep 2023	20 Sep 2023, 15:08	Lateral Raise (Dumbbell)	5	normal	10
Morning v	19 Sep 2023	19 Sep 2023, 08:30	Lat Pulldown (Cable)	0	warmup	27.5
Morning v	19 Sep 2023	19 Sep 2023, 08:30	Lat Pulldown (Cable)	1	normal	35
Morning v	19 Sep 2023	19 Sep 2023, 08:30	Lat Pulldown (Cable)	2	normal	40
Morning v	19 Sep 2023	19 Sep 2023, 08:30	Lat Pulldown (Cable)	3	normal	40
Morning v	19 Sep 2023	19 Sep 2023, 08:30	Lat Pulldown (Cable)	4	normal	40
Morning v	19 Sep 2023	19 Sep 2023, 08:30	Dumbbell Row	0	normal	25
Morning v	19 Sep 2023	19 Sep 2023, 08:30	Dumbbell Row	1	normal	25
						10

My favourite things about the app:

- You pick the exercises from a logbook so it has consistency in naming conventions
- There are some summary stats around muscle groups and training frequency
- It reminds you of your previous weights so acts as a great reminder of progress.

16:46

Workout Detail

cjmayes Monday, Sep 25, 2023 - 4:26pm

Afternoon workout 💪

Time	Volume	Sets
28min	2,145 kg	15

0 likes 0 comments

Like Comment Share

Muscle Split

Shoulders 49%

Chest 34%

Arms 17%

Workout Edit Workout

Logged with Apple Watch

Lateral Raise (Dumbbell)

SET WEIGHT & REPS

Home Workout Profile

Here are a few snapshots from the app that help explain.

16:46

Edit Profile cjmAYES  



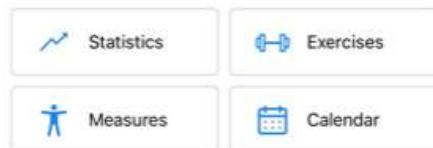
cj-mayes.com

28 mins this week Last 12 weeks ▾

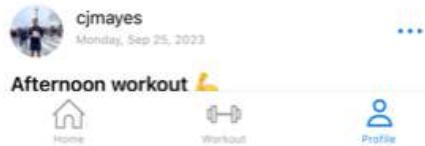


Duration Volume Reps

Dashboard



Workouts



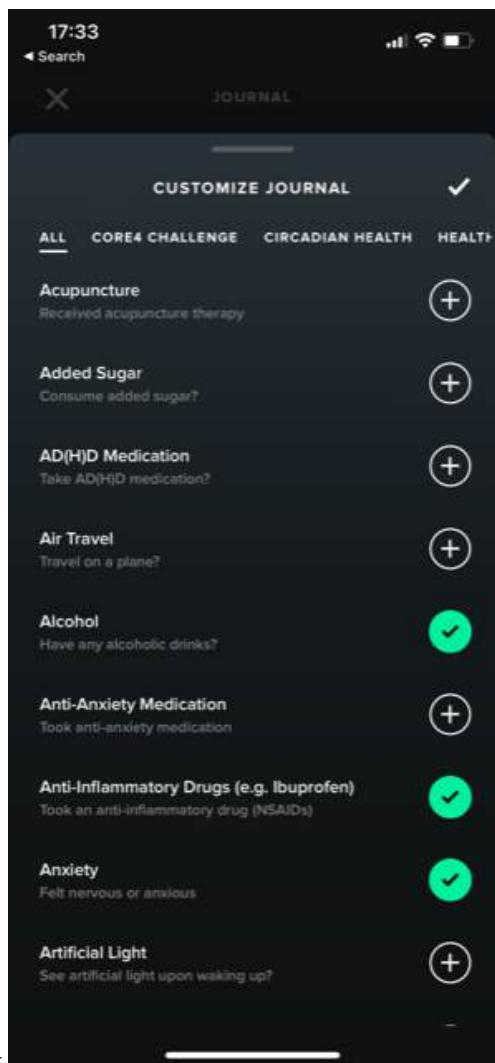
WHOOP Whoop has a whole bunch of metrics that are useful from tracking workouts, through to sleep, and journalling. You can read a little more around [WHOOP here](#).

 journal_entries.csv	29/09/2023 14:46	Microsoft Excel C...	10 KB
 physiological_cycles.csv	29/09/2023 14:46	Microsoft Excel C...	4 KB
 sleeps.csv	29/09/2023 14:46	Microsoft Excel C...	3 KB
 workouts.csv	29/09/2023 14:46	Microsoft Excel C...	3 KB

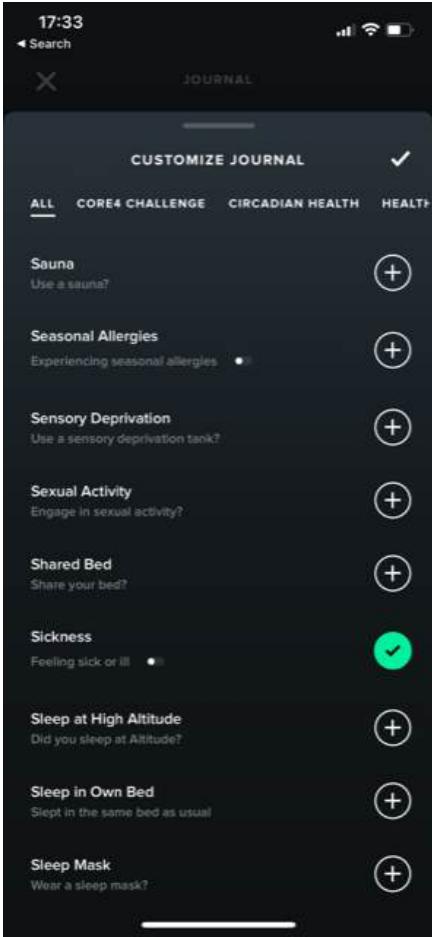
There are 4 main files –

I have uploaded the workout file into the repo. The journal entries file is a bit too personal to upload, but the screen shot below shows some of the items you can add to your journal entry. **Journal**

The screenshot shows a Tableau dashboard titled "Whoop Journal". The interface includes a sidebar with "Pages", "Filters", and "Marks" settings. The main area displays a grid of data points for each day from September 10 to September 28. Each row corresponds to a question, and each column corresponds to a date. The data is represented by colored dots: grey for "False" and red for "True". A legend on the right side of the grid defines these colors. The questions listed are: Hydrated sufficiently, Feeling sick or ill, Felt depressed or down, Felt nervous or anxious, Have any alcoholic drinks?, and Took an anti-inflammatory drug (NSAID).



The journal is customisable to track elements that you may find important



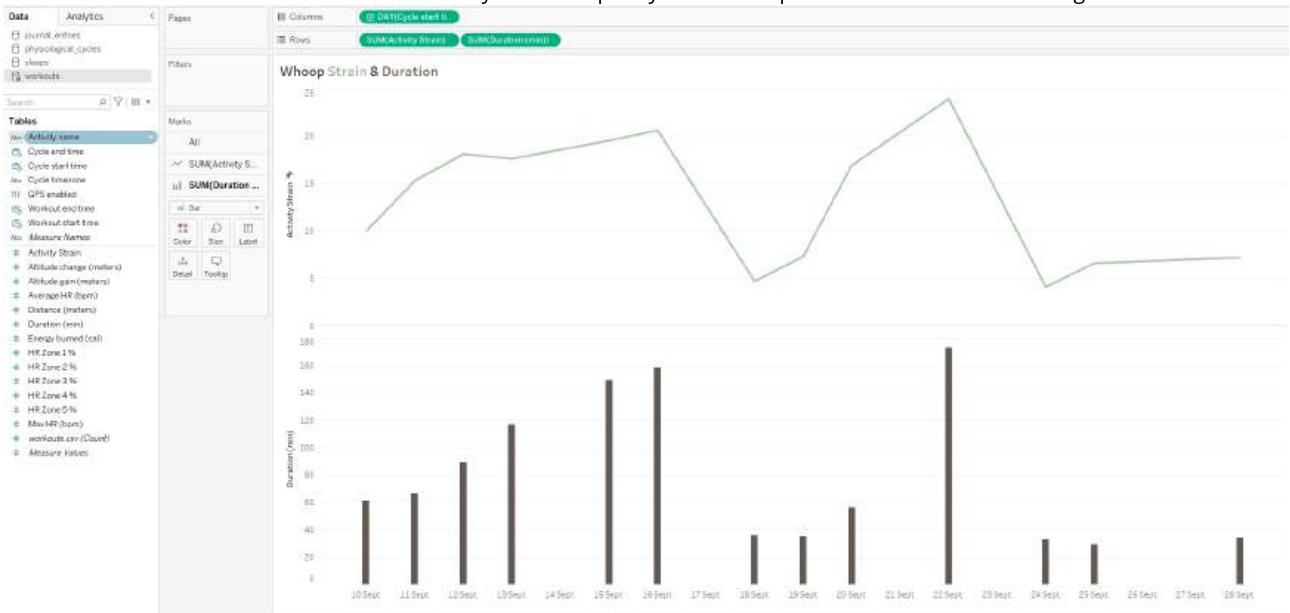
Sleep & Recovery – With the help of WHOOP we can measure the impact of sleep on our recovery.



Sleep Breakdown (Yes I get very little REM sleep and it hurts)



Strain and Duration – We can look at longer runs the impact it has on the body in terms of strain. This lets us optimise future workouts. The ideal is the body recovers quickly even when put under more strain for longer.





Check out what this looks like in the app:

Workout Data

Cycle start time	Cycle end time	Cycle time	Workout start time	Workout end time	Duration [min]	Activity name	Activity Strain	Energy burned (cal)	Max HR (bpm)	Average HRR Zone 1	HRR Zone 2	HRR Zone 3	HRR Zone 4	HRR Zone 5	GPS enabled	Distance [m]	Altitude [m]	Altitude change [meters]
28/09/2023 22:25		UTC+01:00	29/09/2023 08:10	29/09/2023 08:44	34	Running	7.2	223	177	122	7	75	3	8	0	FALSE		
28/09/2023 22:30	26/09/2023 22:24	UTC+01:00	26/09/2023 07:19	26/09/2023 07:49	20	Running	6.6	213	140	136	2	92	3	0	0	FALSE		
24/09/2023 22:38	25/09/2023 22:20	UTC+01:00	25/09/2023 16:21	25/09/2023 16:54	33	Weightlifting	4.1	43	124	84	20	0	0	0	0	FALSE		
22/09/2023 23:05	23/09/2023 23:35	UTC+01:00	23/09/2023 12:24	23/09/2023 14:32	128	Field Hockey	17.6	1440	192	146	6	15	47	27	5	FALSE		
22/09/2023 00:06	22/09/2023 23:05	UTC+01:00	22/09/2023 11:02	22/09/2023 11:47	45	Weightlifting	6.3	192	106	98	7	25	14	0	0	FALSE		
20/09/2023 22:11	22/09/2023 00:06	UTC+01:00	21/09/2023 13:28	21/09/2023 14:24	56	Running	16.9	877	188	169	0	0	5	80	15	FALSE		
19/09/2023 22:29	20/09/2023 22:11	UTC+01:00	20/09/2023 14:33	20/09/2023 15:08	35	Weightlifting	7.3	247	161	124	8	51	30	1	0	FALSE		

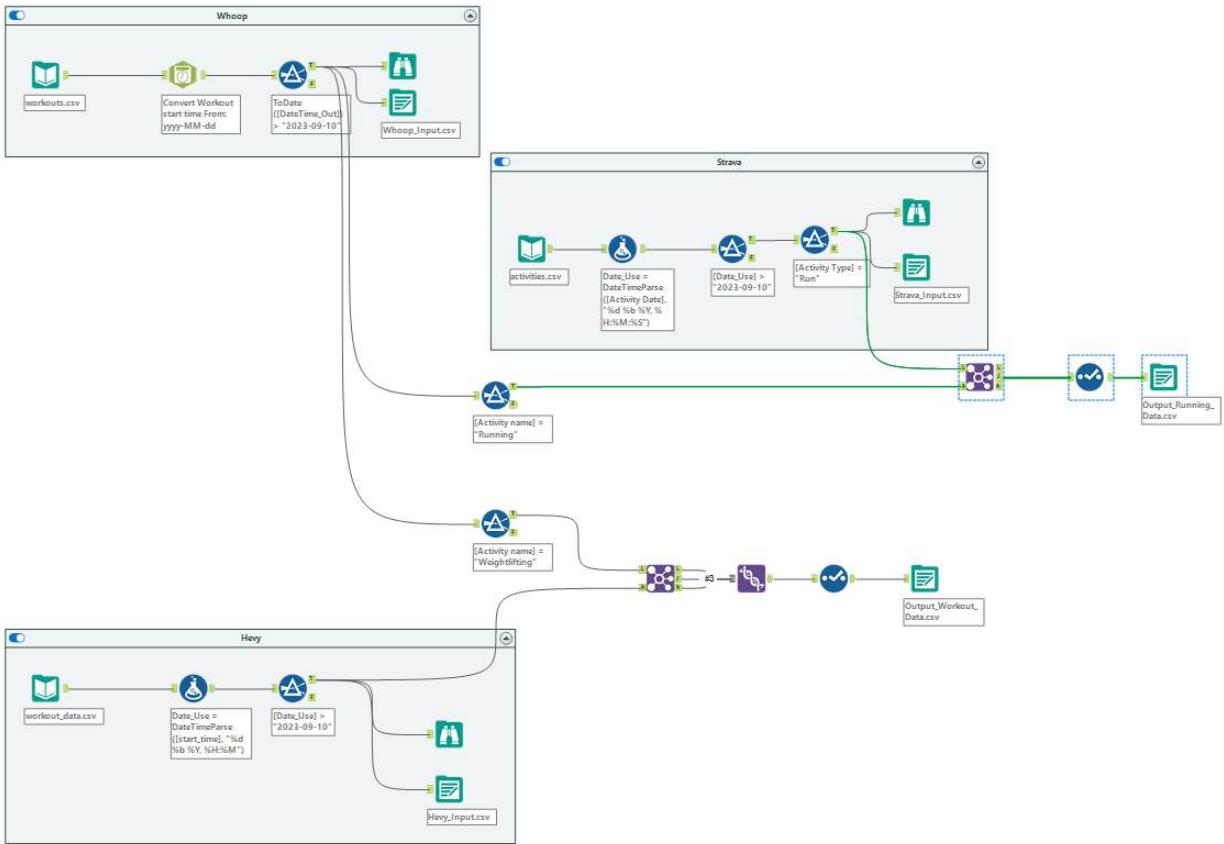
So why collect all this data?

1.

Well interestingly enough, it can be joined together for a more detailed view.

I tend to do this in two ways. I post in the comment of all my strava workouts the url for the hevy link. Mainly though we can connect on date and activity. (It is rare I would go for two runs in the same day) 2) Trends over time – I think the above three apps are great when tracking metrics day to day. But to really improve at anything, It is about doing them over months and years to have an impact. That's why the export data is so useful! 3) Discrepancies in apps – I notice my apple watch and whoop measure max heart rate slightly differently for each exercise, measuring it twice can help me take into account fluctuations. 4)

Building a more well-rounded overview – I no longer have to look at individual events but can now look at the relationship between different elements. For example, how does drinking alcohol negatively impact mood and sport performance the next day. What impact does hydration have on recovery and runs of different distances? No one wants to be running a 13 miler dehydrated from the day before. Does my rolling average mood impact the effort I put into weights sessions? Anyway, hopefully this blog gives some thoughts around ways you may want to track your own health. Maybe I will look to do a larger project that visualises all the data from these apps.... Lots to explore and we have barely touched the surface so far of some of the insights we could get from this data. **Alteryx Joins** Final note from me – I've posted an example of the workflow in the GitRepo – all it has is some date amendment fields, and filters to split out workouts and runs and an example of how to join the datasets from the raw data you would export from the various apps.



LOGGING OFF, CJ

STATSBOMB SUBSTITUTIONS & TACTICS (ALTERYX)

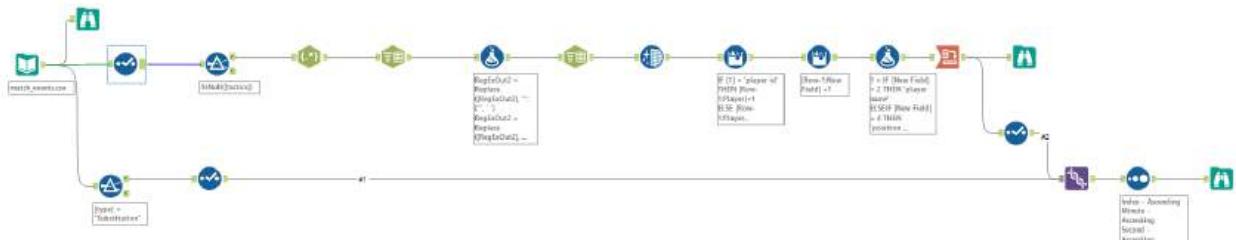
Hi all, An Alteryx flow blog this week looking at preparing data ready for use. As many of you are aware, Statsbomb have made some data free for use within the public domain. Todays tutorial we will revisit the Women's World Cup and look at recording substitutions and changes in tactics as part of the match events data that can be pulled directly from their Git Repo. If you'd like to follow along, by all means the workflow, starting and end dataset are all found in the repository. It is worth mentioning there is an easier way of pulling back some of the records, such as the line up and that is through python. You can find how to convert the json into a readable format on the [mplsoccer website](#), it consists of the following line of code

```
df_lineup = parser.lineup(7478)
```

Anyway. Here I want to help those without a coding background to make their own transformations. Lets look at the workflow and data. We want 3 things from today:

1. To log every substitution
 2. To log whether that substitution then impacts a change in tactics
 3. To create a condensed output file to be able to showcase those 2 things occurring.

Connect to the **match events** data.



First thing we do in the top of the flow is filter out some of the noise. Always important as generally we don't want to bring fields we don't need.

6 of 6 Fields | Cell Viewer | 3,581 records displayed | ↑ ↓

Record	index	minute	second	tactics	team	type
1	1	0	0	{'formation': 4231, 'lineup': [...]}	Spain Women's	Starting XI
2	2	0	0	{'formation': 3412, 'lineup': [...]}	England Women's	Starting XI
3	3	0	0	[Null]	England Women's	Half Start
4	4	0	0	[Null]	Spain Women's	Half Start
5	1913	45	0	[Null]	Spain Women's	Half Start
6	1914	45	0	[Null]	England Women's	Half Start
7	5	0	0	[Null]	England Women's	Pass
8	8	0	2	[Null]	England Women's	Pass
9	11	0	4	[Null]	England Women's	Pass
10	14	0	9	[Null]	England Women's	Pass
11	16	0	11	[Null]	Spain Women's	Pass
12	18	0	12	[Null]	Spain Women's	Pass
13	20	0	14	[Null]	England Women's	Pass
14	32	0	33	[Null]	Spain Women's	Pass
15	36	0	35	[Null]	Spain Women's	Pass
16	45	0	38	[Null]	England Women's	Pass
17	52	0	40	[Null]	England Women's	Pass
18	56	0	48	[Null]	England Women's	Pass
19	60	0	55	[Null]	England Women's	Pass
20	63	1	3	[Null]	England Women's	Pass
21	66	1	9	[Null]	England Women's	Pass
22	69	1	12	[Null]	England Women's	Pass

First thing to notice is the tactics field is stored in a json format. The cell is also truncated meaning we are at risk of the string cutting off some of the details. In this case we click back on our match_events input and change the field length to be longer than the json would be (in this case i set it to 4000) We then look to filter our data to only the tactics.

index	minute	second	tactics	team	type
1	0	0	{'formation': 4231, 'lineup': [...]}	Spain Women's	Starting XI
2	0	0	{'formation': 3412, 'lineup': [...]}	England Women's	Starting XI
1917	45	0	{'formation': 4231, 'lineup': [...]}	England Women's	Tactical Shift
3072	87	31	{'formation': 3412, 'lineup': [...]}	England Women's	Tactical Shift
3081	87	42	{'formation': 4231, 'lineup': [...]}	Spain Women's	Tactical Shift
3140	89	42	{'formation': 433, 'lineup': [...]}	Spain Women's	Tactical Shift

Here is a sample example of the tactics field {formation: 4231, lineup: } Now whilst I would love to use the Json parse too to solve this challenge, it doesn't always work as expected depending on the format of the json (not to mention id have to do a string replace for the quotations to be double quotes), so we have to turn to some regex to be able to split the different parts of the data we need. At this point a major thank you goes to my colleague Filip Rutkowski. I don't have the coding patience I use to... ha. The regular expression needed is `(.*\)(.*\)` – This part matches zero or more occurrences of any character (except newline) and captures it within a group. The . matches any character, and * quantifier means "zero or more occurrences". So, `(.*\)` captures any sequence of characters. \ – This part matches the opening square bracket matches the closing square bracket]. What this will do is trim off the start to the formation and line up part of the string.

Output Column		Data Preview
	RegExOut2	{player id': 45295
	<code>Replace([RegExOut2], "': '", ', ')</code>	
	RegExOut2	{player id: 45295
	<code>Replace([RegExOut2], "'", '')</code>	
	RegExOut2	player id: 45295
	<code>Trim(Replace(Replace([RegExOut2], ')', ''), '{', ''))</code>	
	RegExOut2	player id: 45295
	<code>Trim(Replace(Replace([RegExOut2], ',', ''), '{', ''))</code>	
	RegExOut2	player id: 45295

By

following the above formula tools and a further regex column to split on the : delimiter, we can start to split out field names and values like below

Record	index	minute	second	tactics	team	type	RegExOut1	RegExOut2	1	2
1	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	player id: 45295	player id	45295	
2	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	name: Catalina Thomas Coll Lluch	name	Catalina Thomas Coll Lluch	
3	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	position id: 1	position id	1	
4	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	name: Goalkeeper	name	Goalkeeper	
5	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	jersey number: 23	jersey number	23	
6	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	player id: 46639	player id	46639	
7	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	name: Ona Battle Pascual	name	Ona Battle Pascual	
8	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	position id: 2	position id	2	
9	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	name: Right Back	name	Right Back	
10	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	jersey number: 2	jersey number	2	
11	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	player id: 10165	player id	10165	
12	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	name: Irene Paredes Hernandez	name	Irene Paredes Hernandez	
13	1	0	0	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	{formation: 4231, lineup: [{player: {id: 45295, ...}, Spain Women's Starting XI	position id: 7	position id	7	

It's then a case of using the cross tab to transform the data.

Record	index	minute	second	team	Player	jersey_number	player_id	player_name	position_id	position_name
1	1	0	0	Spain Women's	1	23	45295	Catalina Thomas Coll Lluch	1	Goalkeeper
2	1	0	0	Spain Women's	2	2	46639	Ona Battle Pascual	2	Right Back
3	1	0	0	Spain Women's	3	4	10165	Irene Paredes Hernandez	3	Right Center Back
4	1	0	0	Spain Women's	4	14	32694	Lala Codina Panedas	5	Left Center Back
5	1	0	0	Spain Women's	5	19	49836	Olga Carmona GarcÃ¡a	6	Left Back
6	1	0	0	Spain Women's	6	3	49835	Teresa Albeira DueÃ±as	11	Left Defensive Midfield
7	1	0	0	Spain Women's	7	6	15284	Alitana Bonmati Conca	13	Right Center Midfield
8	1	0	0	Spain Women's	8	17	50151	Alba MarÃ¡ Redondo Ferrer	17	Right Wing
9	1	0	0	Spain Women's	9	10	10151	Jennifer Hermoso Fuentes	19	Center Attacking Midfield
10	1	0	0	Spain Women's	10	8	10161	MarÃ¡ Francesca Caldentey Oliver	21	Left Wing
11	1	0	0	Spain Women's	11	18	222682	Salma Paralluelo Ayingono	23	Center Forward
12	1917	45	0	England Women's	1	1	31538	Mary Alexandra Earps	1	Goalkeeper
13	1917	45	0	England Women's	2	2	10178	Lucy Bronze	2	Right Back
14	1917	45	0	England Women's	3	6	4642	Millie Bright	3	Right Center Back
15	1917	45	0	England Women's	4	5	10252	Alex Greenwood	5	Left Center Back
16	1917	45	0	England Women's	5	16	19422	Jessica Carter	6	Left Back
17	1917	45	0	England Women's	6	8	4643	Georgia Stanway	9	Right Defensive Midfield
18	1917	45	0	England Women's	7	4	4658	Keira Walsh	11	Left Defensive Midfield
19	1917	45	0	England Women's	8	18	15570	Chloe Kelly	17	Right Wing
20	1917	45	0	England Women's	9	10	31534	Ella Toone	19	Center Attacking Midfield
21	1917	45	0	England Women's	10	7	31531	Lauren James	21	Left Wing
22	1917	45	0	England Women's	11	11	15555	Lauren Hemp	23	Center Forward
23	2	0	0	England Women's	1	1	31538	Mary Alexandra Earps	1	Goalkeeper
24	2	0	0	England Women's	2	16	19422	Jessica Carter	3	Right Center Back
25	2	0	0	England Women's	3	6	4642	Millie Bright	4	Center Back
26	2	0	0	England Women's	4	5	10252	Alex Greenwood	5	Left Center Back
27	2	0	0	England Women's	5	2	10178	Lucy Bronze	7	Right Wing Back
28	2	0	0	England Women's	6	9	5058	Rachel Daly	8	Left Wing Back
29	2	0	0	England Women's	7	8	4643	Georgia Stanway	9	Right Defensive Midfield
30	2	0	0	England Women's	8	4	4658	Keira Walsh	11	Left Defensive Midfield

See how we now have the starting 11 for each team. Scrolling further we see that it also has changes in tactics too and what the 11 players for that would look like. The final thing we want to do is add in the substitutions, to add context to the tactics that take place. You will see the bottom half of the flow looks at where type = Substitution. We then do some renaming conventions to both our parts of the flow before doing a union to glue them together

Index	I	Minute	Second	Player_Name	Player_ID	Position_Name	Substitution_Outcome	Player_Replacement	Team	Type	11_Pitch	Position_ID
2	0	0	0	Mary Alexandra Carps	31530	Goalkeeper	[null]	[null]	England Women's	1	1	
2	0	0	0	Jessica Carter	19422	Right Center Back	[null]	[null]	England Women's	2	3	
2	0	0	0	Mille Bright	4642	Center Back	[null]	[null]	England Women's	3	4	
2	0	0	0	Alys Greenwood	10252	Left Center Back	[null]	[null]	England Women's	4	5	
2	0	0	0	Lucy Bronze	10178	Right Wing Back	[null]	[null]	England Women's	5	7	
2	0	0	0	Rachel Daly	5038	Left Wing Back	[null]	[null]	England Women's	6	8	
2	0	0	0	Georgia Stanway	4643	Right Defensive Midfield	[null]	[null]	England Women's	7	9	
2	0	0	0	Kiera Walsh	4658	Left Defensive Midfield	[null]	[null]	England Women's	8	11	
2	0	0	0	Ella Toone	31534	Center Attacking Midfield	[null]	[null]	England Women's	9	19	
2	0	0	0	Alessia Russo	47521	Right Center Forward	[null]	[null]	England Women's	10	22	
2	0	0	0	Lauren Hemp	15555	Left Center Forward	[null]	[null]	England Women's	11	24	
1915	45	0	0	Rachel Daly	50380	Left Wing Back	Tactical	<u>Chloe Kelly</u>	England Women's	Substitution	[null]	[null]
1916	45	0	0	<u>Alys Greenwood</u>	475210	Right Center Forward	Tactical	<u>Lauren James</u>	England Women's	Substitution	[null]	[null]
1917	45	0	0	Mary Alexandra Carps	31530	Goalkeeper	[null]	[null]	England Women's	1	1	
1917	45	0	0	Lucy Bronze	10178	Right Back	[null]	[null]	England Women's	2	2	
1917	45	0	0	Mille Bright	4642	Right Center Back	[null]	[null]	England Women's	3	3	
1917	45	0	0	Alys Greenwood	10252	Left Center Back	[null]	[null]	England Women's	4	5	
1917	45	0	0	Jessica Carter	19422	Left Back	[null]	[null]	England Women's	5	6	
1917	45	0	0	Georgia Stanway	4643	Right Defensive Midfield	[null]	[null]	England Women's	6	9	
1917	45	0	0	Kiera Walsh	4658	Left Defensive Midfield	[null]	[null]	England Women's	7	11	
1917	45	0	0	<u>Chloe Kelly</u>	15570	Right Wing	[null]	[null]	England Women's	8	17	
1917	45	0	0	Ella Toone	31534	Center Attacking Midfield	[null]	[null]	England Women's	9	19	
1917	45	0	0	Lauren James	31531	Left Wing	[null]	[null]	England Women's	10	21	
1917	45	0	0	Lauren Hemp	15555	Center Forward	[null]	[null]	England Women's	11	23	

Take the above example, we see Chloe Kelly and Lauren James sub on at half time and the tactic / line up change and we log both those events in our final dataset. Hopefully you enjoyed some of the complexities Taking It Further:

- Create a minute by minute dataset for which players are on the pitch
- Consider metrics on how you could measure the impact of a substitution
- Try convert the alteryx workflow into a code of your choosing

LOGGING OFF, CJ

SHAPE MAKING IN TABLEAU

Hi all, I wrote the following blog back in 2020 for Amar Singh's site [here](#), on how to make shapes. Tina Covelli recently reminded me of it when tracking her own HITT workouts.



Since 2020, I've simplified the blog – which you can now read below: By all means follow along the steps below to create your own data points from a vector image.

What you'll need:

1. **A dataset**, preferably between 500-4000 rows.
2. **An image vector** you want to put your points in the shape of. (such as the RF logo)

The original data I used for the visualisation can be [found here](#).

But the data can be found now in Github.

I stitched together the 2010-2020 datasets of where Roger Federer was named the winner or loser from the ATP records.

Step 0: Decide on the data you are using.

For my dataset, you will note there are three final sheets;

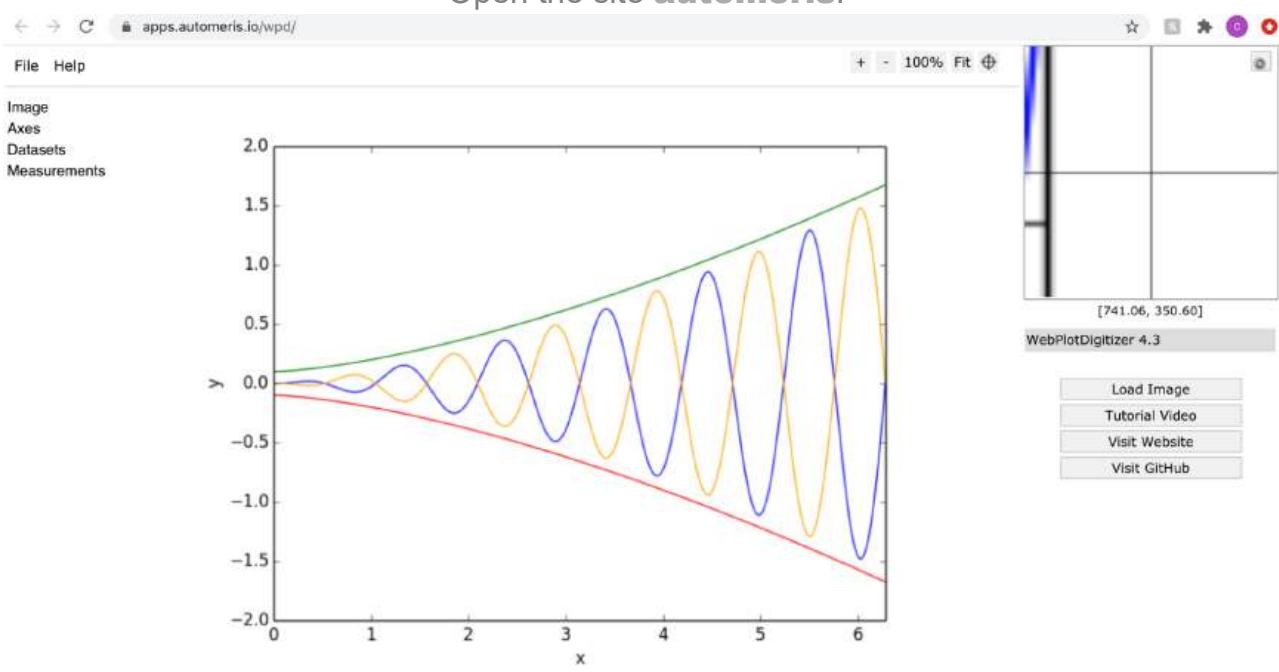
Original: Our original 659 games between 2010-2020.

RF: The logo co-ordinates of X and Y that we will create and then match to our dataset.
(659 points)

RF & Logo: Where I duplicate the dataset to have a view column that toggles between the logo and the RF logo. (659 points x 2 Views)

There are **659 games** he played during this time. This will become important later on.

Open the site **automeris**.



File, Load Image(s), Locate the Logo you want to use in this case I used a plain black vector of Roger Federer's brand. A simple Google of it will do, if not take the one from [here](#).

Load Image File(s)

Choose Files

Feel free to use your own image vector, but for simplicity sake, for now I'd advise using a one-block colour vector (If you start to use an image with various colours you will have to export multiple datasets as seen in step 5)



Step 4: Within the bottom left hand corner we see the Automatic Extraction Mask box. I've amended the (Foreground) colour to be the black as seen in the logo. This is where I want my dataset to create points for plotting.

The distance and ΔX and ΔY Px determine the spacing between the co-ordinates plotted. Have a play around with various combinations to see its functionality.

Automatic Extraction

Mask

Color **Foreground Color**

Distance

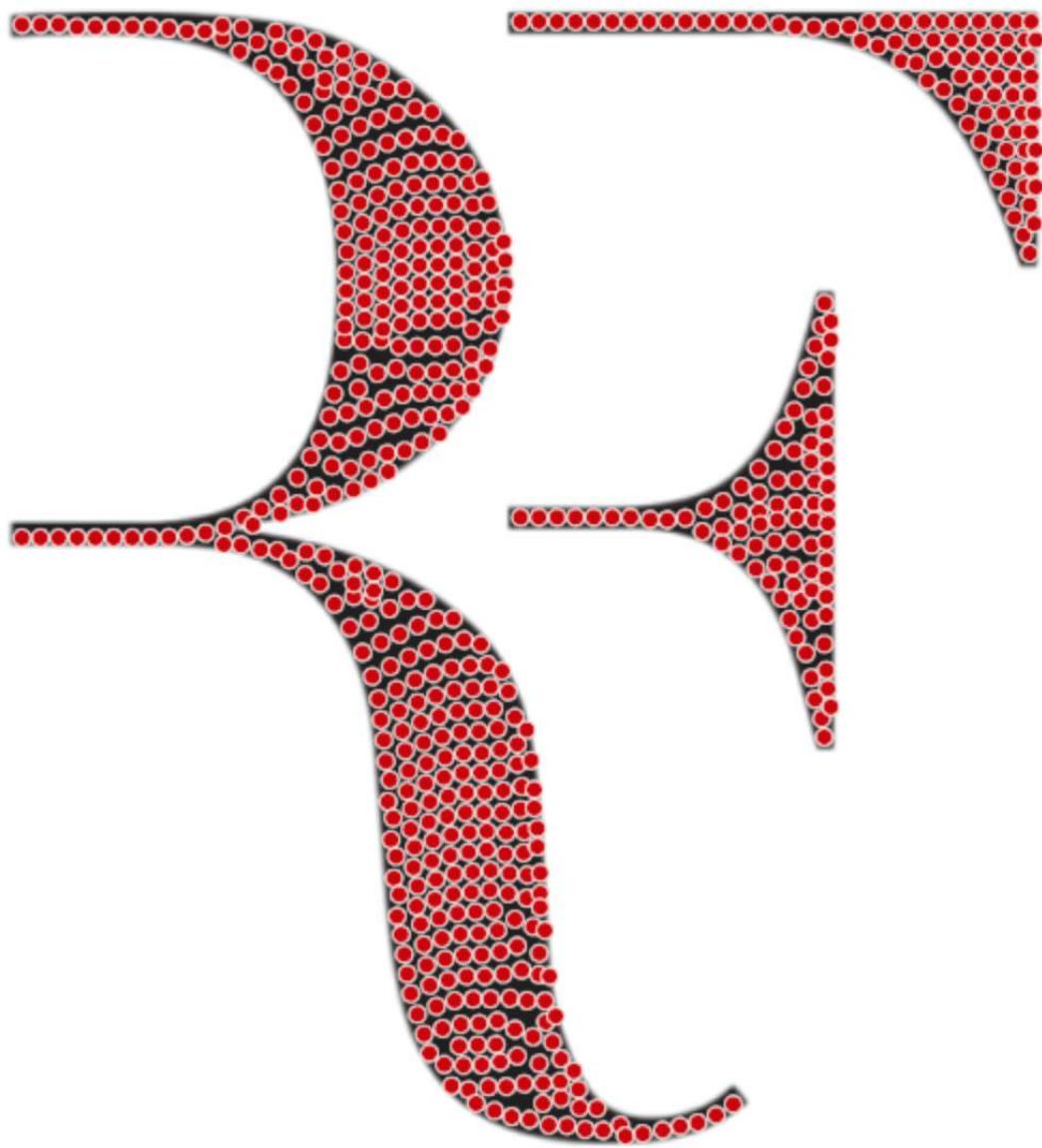
Algorithm **Averaging Window**

ΔX Px
 ΔY Px

Hit run!

You will see co-ordinates of X & Y appear across the chosen colour (Black) these co-ordinates we can later plot in Tableau.

Note: If you want to use various colours you will have to repeat this process over using the different foreground colours of your image and clicking run.



Time to readjust our X and Y values and distance until they look appropriately spaced and are as close to the number of data points in our data set. See within my print screen it shows I have only 609 points. Therefore I need to adjust my X and Y to be smaller to re-run with more datapoints until I am equal or slightly above 659. This is because when we come to plot the points we want to give each co-ordinate a specific game reference!

Note: It's super hard to get a perfect match. Better to get slightly too many points and delete a few using a Rand() function in excel.

[View Data!](#)

Once we view the data we will see our logo is a series of X-Y co-ordinates. It is personal preference how you download this. Personally I download to a .CSV

Acquired Data

Dataset: Default Dataset ▾

Variables: X, Y

44.5, 23.1
44.5, 189
49.5, 23
49.5, 189
54.5, 22.8
54.5, 189
59.5, 22.5
59.5, 189
64.5, 22.1
64.5, 189
69.5, 22
69.5, 189
74.5, 22
74.5, 189
79.875, 23.3125
79.5, 189
84.5, 23.75

Sort

Sort by: Raw ▾

Order: Ascending ▾

Format

Number Formatting:

Digits: Ignore ▾

Column Separator: Format

Copy to Clipboard **Download .CSV** **Graph in Plotly*** **Close**

*Plotly is a secure data analysis and graphing site with data sharing and access controls.
Visit <http://plot.ly> for details.

Assign X-Y Co-ordinates to our dataset.

Copy and paste into two new columns headered X & Y the new 659 points against our original data.

As you can see in the 'RF' sheet, we have now added the X and Y columns we have just downloaded into our dataset. This is essentially giving each of our games an X & Y co-ordinate.

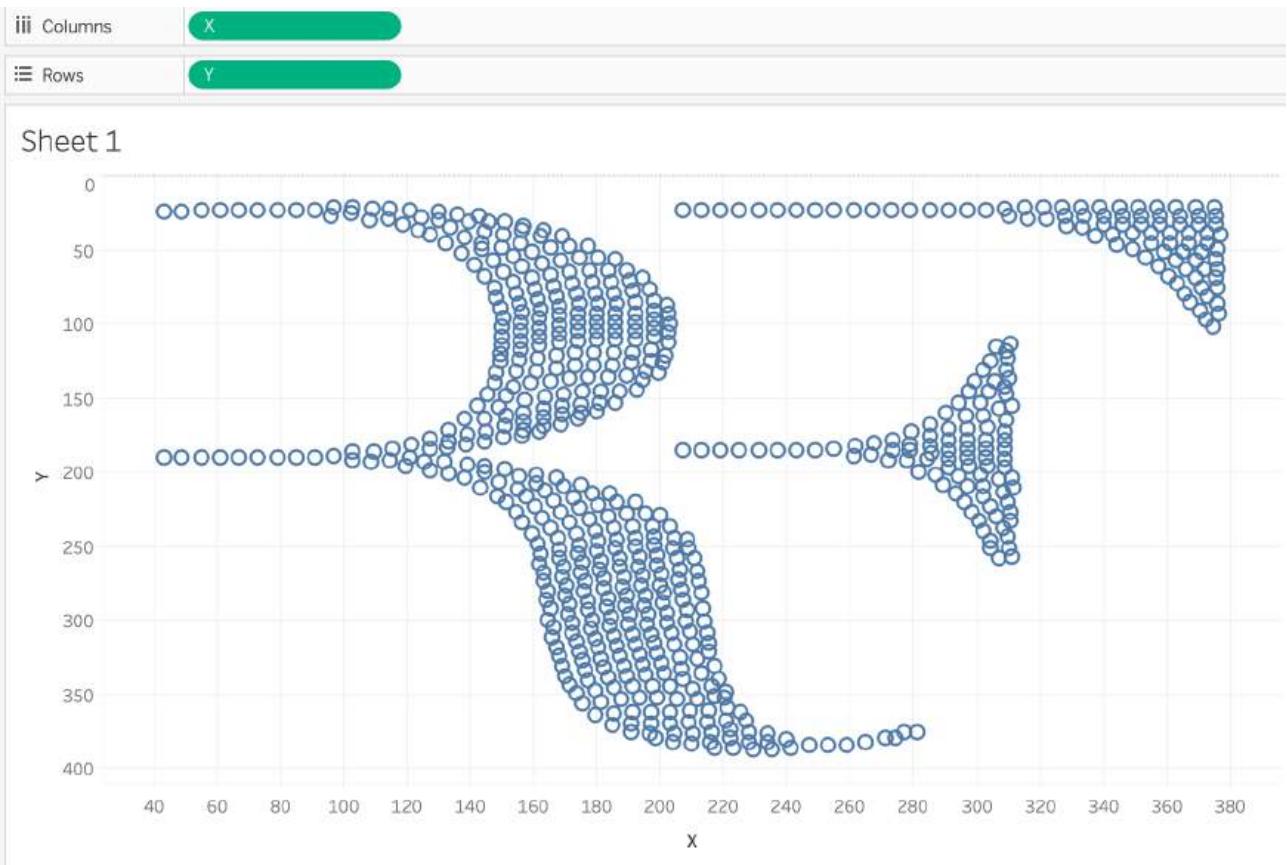
Open Tableau!

Connect to our new RF sheet that we just created.

Open a New Sheet

Drag X onto Columns, and Y onto Rows.

Make sure both are continuous dimensions. The marks type should automatically switch to being a circle, but you can also play around with other shapes.



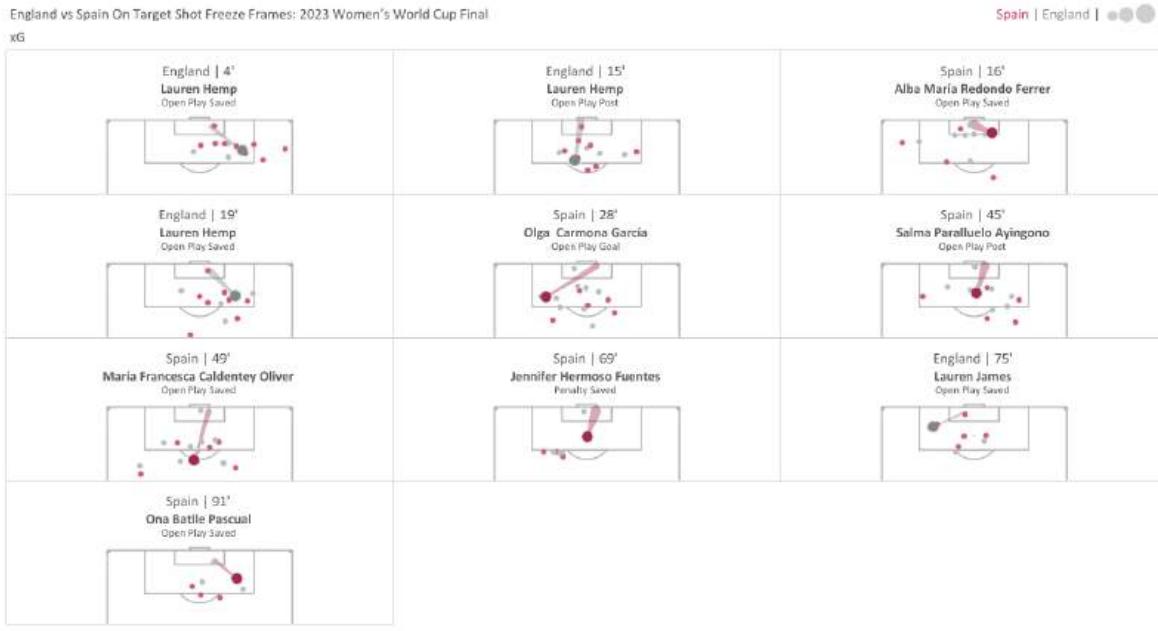
Because we have assigned each point an X and Y value, we can now add in the tool. I drag surface onto colour, add all the match details into the tooltip and resize to make the logo look slightly more uniformed. I put round onto size and sorted them by how far in the knockout stages Roger reached. Sizing can become a bit tricky dependent what you size on, so that your image doesn't lose shape.



Hope you enjoyed this blog, written 3 years ago, but still can be used today! **LOGGING OFF, CJ**

SHOT FREEZE FRAMES PT 2 (WITH VISIBILITY)

Hi all, A few weeks back I wrote a basic StatsBomb python script and spoke on how we can join some of the data ready, fit for use in Tableau to create freeze frame shot maps. I want to take that blog one step further and start to look at pitch visibility. This blog will feel like starting half way through a book – so please forgive me and read the previous StatsBomb Shot Freeze Frame post before moving onto this one.



Let's pick up from where we ended last time. We have our alteryx flow that has created our grid for our shot maps with freeze frames, we have identified our player teams and we have the shots that are on target. So how do we add in pitch visibility?



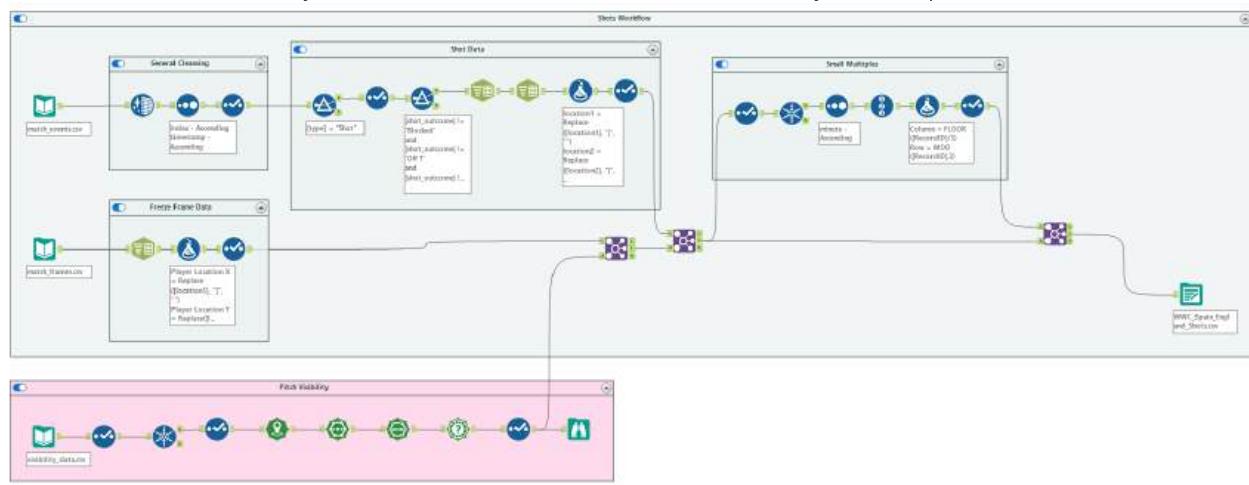
Well you may notice in the previous datasets we have visibility in the match frame data. These are a list of co-ordinates that make up the marks for what part of the pitch can be seen when the freeze frame was taken. (Hence why all our players remain within the polygon)

Results - Input Data (3) - Output

7 of 7 Fields | Cell Viewer | * 3,693 of 41,984 records displayed(partial results)

Record	id	visible_area	match_id	teammate	actor	keeper	location
1	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	True	False	False	[50.82277618592247, 37.80245852848263]
2	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	True	False	False	[52.19804610387511, 17.64345337797193]
3	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	True	False	False	[52.65557386491984, 47.29986238958084]
4	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	True	False	False	[59.73693096640113, 73.46007307611634]
5	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	True	False	False	[60.71575905832746, 58.110287183618546]
6	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	True	False	False	[60.86547524177565, 50.977082105653714]
7	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	True	True	False	[61.0, 40.099998474121094]
8	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	False	False	False	[61.07394249592433, 54.82980387636794]
9	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	False	False	False	[63.0003578454924, 30.048445579963023]
10	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	False	False	False	[63.00048696912367, 18.91242491323178]
11	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	False	False	False	[66.07222565151756, 48.839884751444292]
12	30728cd8-f0a6-4535-8f24-16a37a1d4f08	[374894703434483, 80.0, 51.3811438225292, 8.9...]	3906390	False	False	False	[66.1487272739169, 61.08248450962165]
13	b796b4f0-d521-4d6a-903e-b63c45120ec9	[34.7277452024827, 80.0, 50.3050199588541, 9.2...]	3906390	True	True	False	[51.099998474121094, 38.20000076293945]
14	b796b4f0-d521-4d6a-903e-b63c45120ec9	[34.7277452024827, 80.0, 50.3050199588541, 9.2...]	3906390	True	False	False	[51.653907329163346, 18.655811220735714]

You can see there are quite a few duplicates as each player for each ID will essentially replicate the same visibility fields. Lets revisit our Alteryx flow. (Of course, this is free to download from my GitHub repo under the title)



The amended workflow with "Pitch Visibility" now has a new dataset that we feed into it. To get that new dataset, I amended our previous python script to handle the X & Y co ordinates from the freeze frames dataset. Here is what that script looks like:

```
https://statsbomb.com/news/statsbomb-release-free-2023-womens-world-cup-data/
the competition id (72) and season id (107) to pull the data.
"""

import ast
import json

from statsbomypy import sb
import pandas as pd

# Get competitions
comp = sb.competitions()
#comp.to_csv('competitions.csv', index=False)

#Get Matches
df = sb.matches(competition_id=72, season_id=107)
#df.to_csv('matches.csv', index=False)

# Find a match_id required - final WWC
match = 3906390
match_events = sb.events(match_id=match)
#print(match_events)
#match_events.to_csv('match_events.csv', index=False)

# Find freeze frames
match_frames = sb.frames(match_id=match, fmt='dataframe')
#match_frames.to_csv('match_frames.csv', index=False)

# Create a DataFrame
df = pd.DataFrame(match_frames)

# Create a list to store the new rows
new_rows = []

# Iterate through the DataFrame
for index, row in df.iterrows():
    visible_area = row['visible_area']
    location = row['location']

Click to Download
visibility_data.csv
29.6 MB

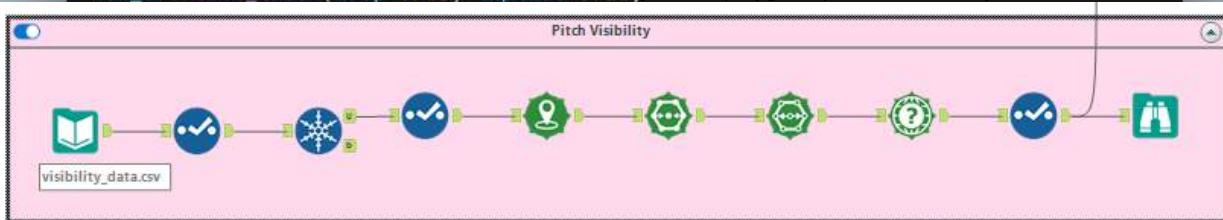
# Split the visible_area list into x and y coordinates
for i in range(0, len(visible_area), 2):
    x = visible_area[i]
    y = visible_area[i + 1]

    new_row = {
        'id': row['id'],
        'x': x,
        'y': y,
        'match_id': row['match_id'],
        'teammate': row['teammate'],
        'actor': row['actor'],
        'keeper': row['keeper'],
        'location': location
    }

    new_rows.append(new_row)

# Create a new DataFrame from the list of new rows
```

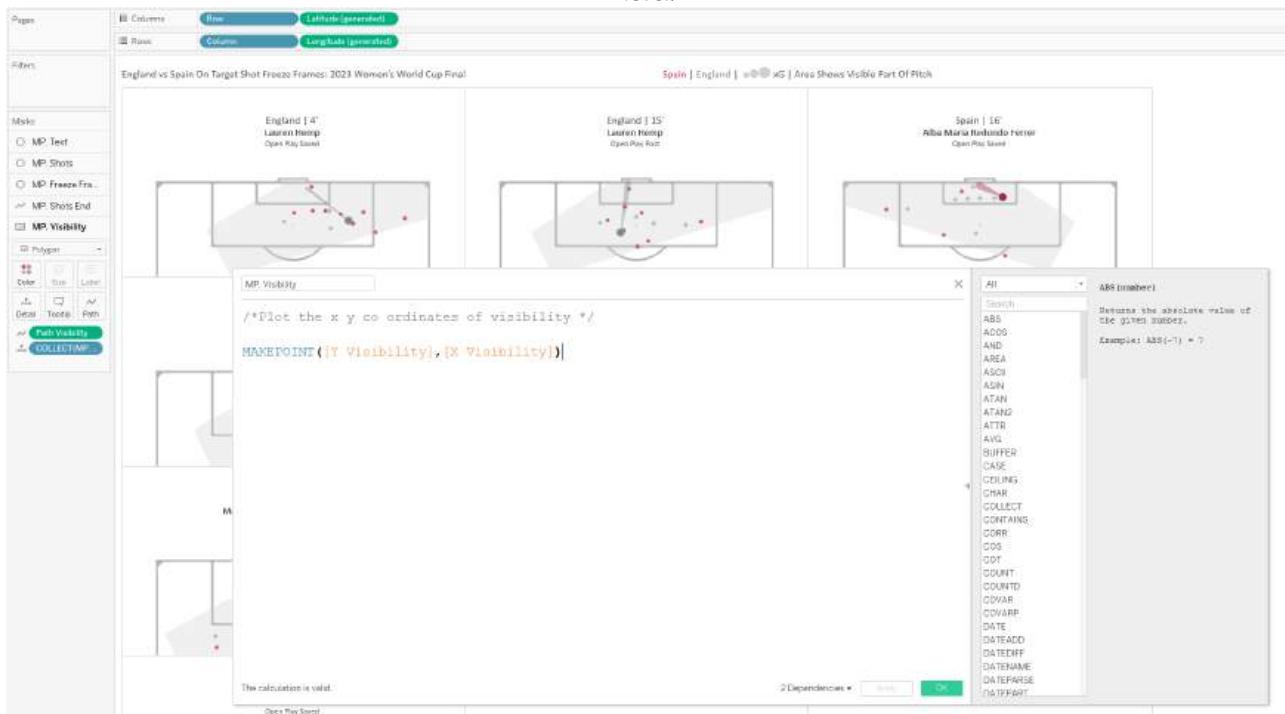
You will see `new_df = pd.DataFrame(new_rows)` the python code from last time, this is to split out the visible area to have new co-ordinates. For each co-ordinate we will have a new row. This code leaves us with one issue that we will tackle through Alteryx. That issue is keeping the order of the points joined together (our path) to create the equivalent of a convex hull. We can't `new_df.reset_index(drop=True, inplace=True)` join the polygon up the correct way.



The new steps in the workflow find the unique records from the visibility data. We then look to make a centroid mapping of the x and y point, reason being is we use Poly build to find the convex hull. Once we've found the sequence in which they are put together, given each ID, we then extract that information using Poly split and information tools.

5 of 5 Fields Cell Viewer * 11,260 of 17,888 records displayed(partial results) ↑ ↓					
Input Record	ID Visibility	SpatialObj	Path Visibility	X Visibility	Y Visibility
1	0018676d-3644-46c1-905e-d097ab96eec5	Point - View Browse Tool Map Tab	0	40.643348	80
2	0018676d-3644-46c1-905e-d097ab96eec5	Point - View Browse Tool Map Tab	1	77.352674	80
3	0018676d-3644-46c1-905e-d097ab96eec5	Point - View Browse Tool Map Tab	2	68.584915	21.85251
4	0018676d-3644-46c1-905e-d097ab96eec5	Point - View Browse Tool Map Tab	3	50.958368	21.702904
5	0018676d-3644-46c1-905e-d097ab96eec5	Point - View Browse Tool Map Tab	4	40.643348	80
6	001a9d70-b2cf-48ee-adc2-bc1961a71d71	Point - View Browse Tool Map Tab	0	5.143481	0
7	001a9d70-b2cf-48ee-adc2-bc1961a71d71	Point - View Browse Tool Map Tab	1	28.544499	53.52884
8	001a9d70-b2cf-48ee-adc2-bc1961a71d71	Point - View Browse Tool Map Tab	2	41.353927	50.370611
9	001a9d70-b2cf-48ee-adc2-bc1961a71d71	Point - View Browse Tool Map Tab	3	41.776561	0
10	001a9d70-b2cf-48ee-adc2-bc1961a71d71	Point - View Browse Tool Map Tab	4	5.143481	0

Above is an example of what that looks like – after that, it is a simple case of joining this data back into our original data on the common field of ID. You can find all the referring CSV's files in the repo to test for yourself. I am increasingly becoming a fan of the Poly Build and Split tool! Hopefully the python script and Alteryx flow really help you with taking your data prep to the next level.



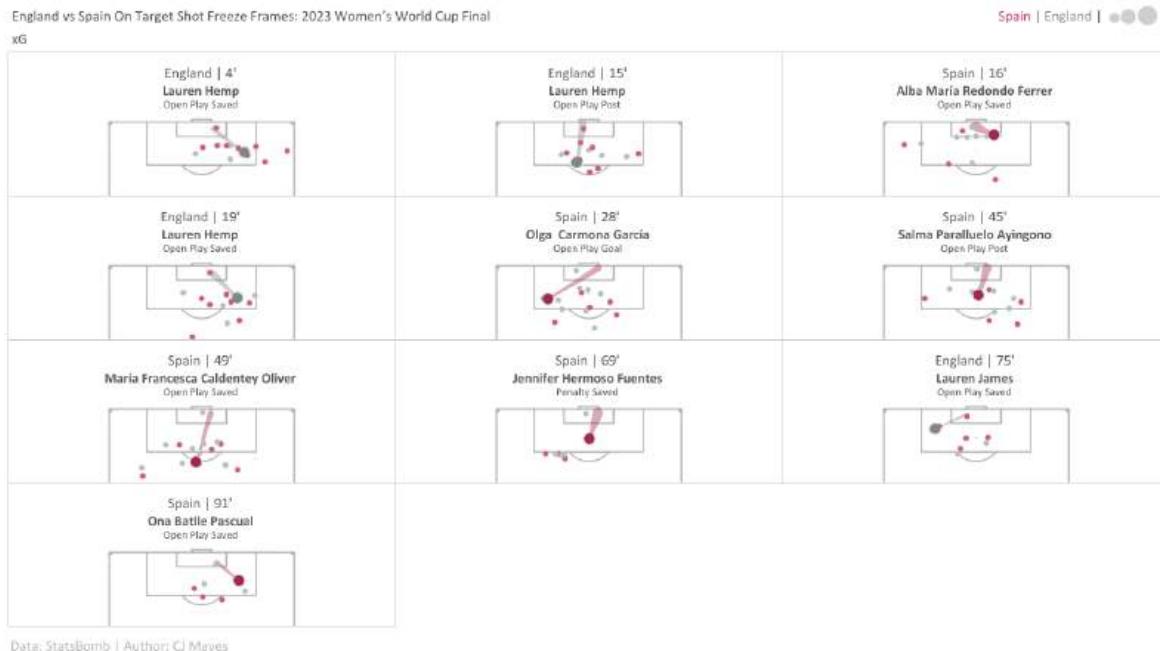
Once we have the data prepped for tableau, It's just a case of creating the new layer and making sure we get our correct field onto path. Let me know if you have any questions. Going further,

- Try write all the transformations in the data in Python
- Add in labels from the freeze frame data for each players name

LOGGING OFF, CJ

STATSBOMB SHOT FREEZE FRAMES (TABLEAU/ALTERYX)

Hi all back this week with some football content. Our SportsVsSunday team like to occasionally suggest different sports to be found [here](#). Read more about the initiative and the [weekly round ups here](#). Today we are going to look at some data processing to get our data ready for use within Tableau. I won't cover too much of the tableau side of things, because I've previously written large amounts on Tableau shot maps so the premise behind plotting the points with layers and creating layers would seemingly be repetitive. I am hoping this blog acts as a bridge between retrieving data and prepping it for visuals.



Here is what we will look to recreate. This is all the on-target shots as part of the Women's World Cup Final where Spain beat England 1-0. **Data Access** StatsBomb are kind enough to release data for free. The WWC was one example of this. To retrieve the data we will write a simple python script and export it to csv, but make no transformations to the data within the script.

```

from statsbombpy import sb

# Get competitions
comp = sb.competitions()
comp.to_csv('competitions.csv', index=False)

#Get Matches
df = sb.matches(competition_id=72, season_id=107)
df.to_csv('matches.csv', index=False)

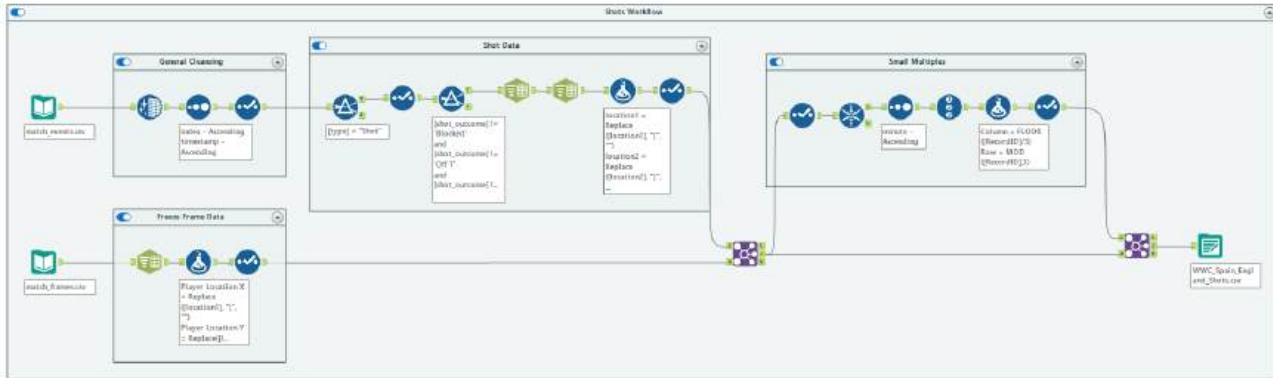
# Find a match_id required - final WWC
match = 3906390
match_events = sb.events(match_id=match)
match_events.to_csv('match_events.csv', index=False)

# Find freeze frames
match_frames = sb.frames(match_id=match, fmt='dataframe')
match_frames.to_csv('match_frames.csv', index=False)

```

The code can be downloaded from the repo, using the github icon under the title. Simply put, we look for the specific competition and season, and export all the matches. Then using this list, I narrow the search down just to the final and export all

the match events. The raw data for this is also within the repo. **Data Transformation** Next we look at transforming the data. Now you could do this in any way, but for today we will look to do it in Alteryx. Again, you can download the WWC.yxmd from the repo to follow along.



Above is a print screen of the workflow. Let's work through it bit by bit. The general cleansing part of the workflow looks to remove null columns, It then sorts the data based on the index file and timestamp, making all our match events in ascending order. After that I start to decide what columns I actually want, dropping a whole bunch of fields around clearances and fouls mainly because I know I'm just going to look at shots. The shots data section limits type to equal shot, meaning we are reducing our dataset, excluding things such as passes and carries. I also filter the shot outcome just to look at on target shots excluding things such as blocked, wayward and off target. If we look at the subset of data we need to clean up a location field as well as shot end field. Reason being is they come in pairs in the field. To do this I use the split to columns field as well as a replace formula to get rid of the brackets.

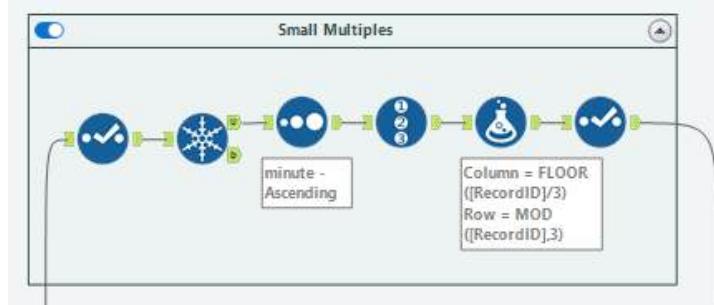
duration	id	location	minute	player	possession_team	shot_body_part
0.830634	794cf42d-828c-4b7a-8665-f682fe16572a	[107.9, 56.3]	4	Lauren Hemp	England Women's	Left Foot
0.910768	4d11233b-219f-4526-977f-2247792f6ff7	[104.0, 35.4]	15	Lauren Hemp	England Women's	Left Foot
0.432391	51072651-8ceb-4af3-b096-b3af9ba5b44d	[115.3, 46.7]	16	Alba MarÃ¡a Redondo Ferrer	Spain Women's	Right Foot
0.834736	29e307b6-8b66-41f1-ad9b-f28d7e055271	[107.3, 53.5]	19	Lauren Hemp	England Women's	Left Foot
0.912924	a7d3f03e-ae8e-4eb1-a70c-bbd9ce44d8e1	[107.1, 24.5]	28	Olga Carmona GarcÃ¡a	Spain Women's	Left Foot
0.925028	9b0cf7d1-86ce-427c-8cee-7934f7742ffe	[108.3, 40.7]	45	Salma Paralluelo Ayingono	Spain Women's	Right Foot
1.218355	522453c2-0ef4-4b97-94f3-d033200b51ae	[98.5, 37.9]	49	MarÃ¡a Francesca Caldentey Oliver	Spain Women's	Right Foot
0.515375	cbe15a9c-da5e-4005-a46f-4ded6839727b	[108.1, 40.1]	69	Jennifer Hermoso Fuentes	Spain Women's	Left Foot
0.611459	fe0c0079-ef00-4473-a384-3269a3627890	[112.6, 24.4]	75	Lauren James	England Women's	Left Foot
0.045529	42a6f141-e4a6-4425-b81a-d79729b5cf6b	[108.8, 54.2]	91	Ona Batlle Pascual	Spain Women's	Right Foot

Transforms to

shot_type	team	Location X	Location Y	Shot End Location X	Shot End Location Y
Open Play	England Women's	107.9	56.3	118.1	44
Open Play	England Women's	104	35.4	120	37.5
Open Play	Spain Women's	115.3	46.7	119.1	39.7
Open Play	England Women's	107.3	53.5	117.7	44
Open Play	Spain Women's	107.1	24.5	120	43.5
Open Play	Spain Women's	108.3	40.7	120	44
Open Play	Spain Women's	98.5	37.9	119.3	43.3
Penalty	Spain Women's	108.1	40.1	119.2	43.2
Open Play	England Women's	112.6	24.4	118.3	35.4
Open Play	Spain Women's	108.8	54.2	116.8	45.6

The Small

Multiples part of the workflow is personal preference. I like to assign the columns and rows of which each shot is before hand so it doesn't mean I need to calculate along anything in Tableau.



I filter to the unique ID's (Match shot events) assign them a record ID and then use a formula tool to allocate a column and row, before rejoining this into the data.

The screenshot shows an Alteryx Designer workspace. On the left, there's an 'Output Column' panel with three rows. The first row has a formula `FLOOR([RecordID]/3)` and a preview of 'Row' 0 with data type V_WString. The second row has a formula `MOD([RecordID],3)` and a preview of 'Row' 0 with data type V_WString. The third row is empty. To the right of the columns is a 'Data Preview' panel showing a table with 10 records. Below the preview is a 'Formula (4d) - Output' section with a table of 5 fields. The table has columns: Record, RecordID, Id, minute, Column, and Row. The data shows various player IDs and their locations over time.

You can see this uses the MOD and Floor calculation. **Freeze Frame Data** So interestingly enough the freeze frame data also appears in the match events dataset, but because I didn't want to do a whole bunch of json manipulation I thought it would be easier to extract the raw data and then show that is joined in on ID=ID. This means rather than have our data at one row per shot, we will have a row for each player captured within data. So a shot ID may have anywhere between 1 and ... well 22 (unlikely but possible) records. In a similar way to the shot location data, the only transformation here needed is a split to columns and formula tool to tidy up X and Y co-ordinates.

This screenshot shows another Alteryx Designer workspace. On the left, there's an 'Output Column' panel with three rows. The first row has a formula `Replace([location1], "[", "")` and a preview of 'Player Location X' with value 50.82277618592247. The second row has a formula `Replace([location2], "]", "")` and a preview of 'Player Location Y' with value 37.80245852848263. The third row is empty. To the right of the columns is a 'Data Preview' panel showing a table with 10 records. Below the preview is a 'Formula (4d) - Output' section with a table of 5 fields. The table has columns: Record, RecordID, Id, minute, Column, and Row. The data shows various player IDs and their locations over time.

Exported Data The final dataset is what we will take into Tableau – It can be found in the repo as WWC_Spain_England_Shots.csv. Thanks again to StatsBomb for making this data available. **Tableau** So a few notes on the Tableau side, but of course download the workbook for free from my profile. The first is if you'd like to create a shot direction the data will need to be with a union. Otherwise you can just connect once to the data.

Connections

Add

WWC_Spain_England_Shots

Text file

Files

Use Data Interpreter
Data Interpreter might be able to clean your Text file workbook.

working_notes.txt

WWC_Spain_E...d_Shots.csv

New Union

New Table Extension

WWC_Spain_England_Shots.csv is made of 2 tables. ⓘ

WWC_Spain_England_Shots.csv+

To achieve multiple layers within the visual, I end up using the Makepoint() function.



You'll see 4 layers, one is the shot line path, I put it at the back so you can see if there were players in the shot line. The freeze frame layer has all the locations of the surrounding players. The shot.... Well this is theoretically just the shot start mark, but I wanted to add in a circle to the visual. The Text field, which is an arbitrary point positioned strategically above the football pitch to add some labels and context to the shots. By all means, reach out if you have any questions on these elements, or refer back to some of the older work I've posted on soccer the past few years. Going further,

- Try write all the transformations in the data in Python
- Add in labels from the freeze frame data for each players name
- Add in some extra data marks
- Map pitch visibility from the dataset

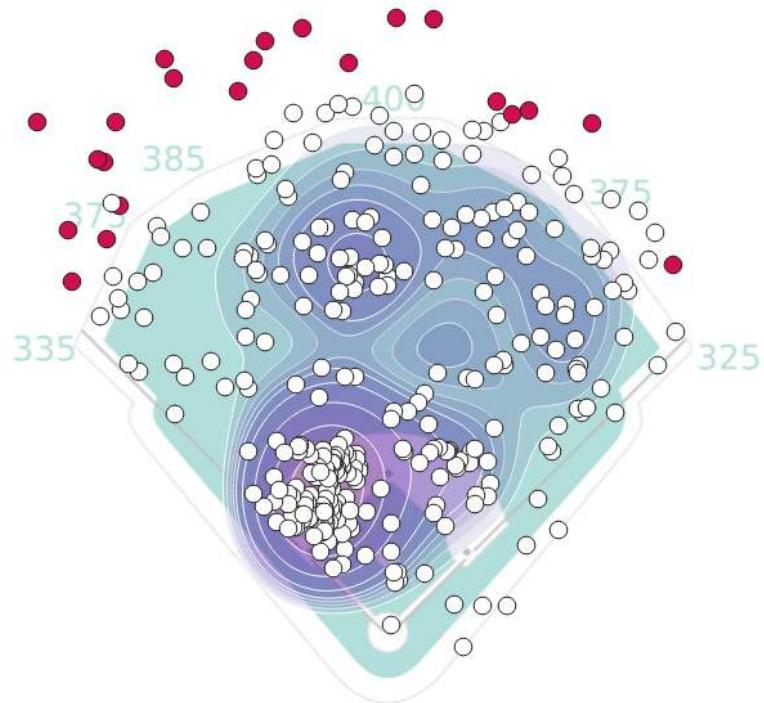
LOGGING OFF, CJ

BASEBALL SPRAY CHART HEAT MAP

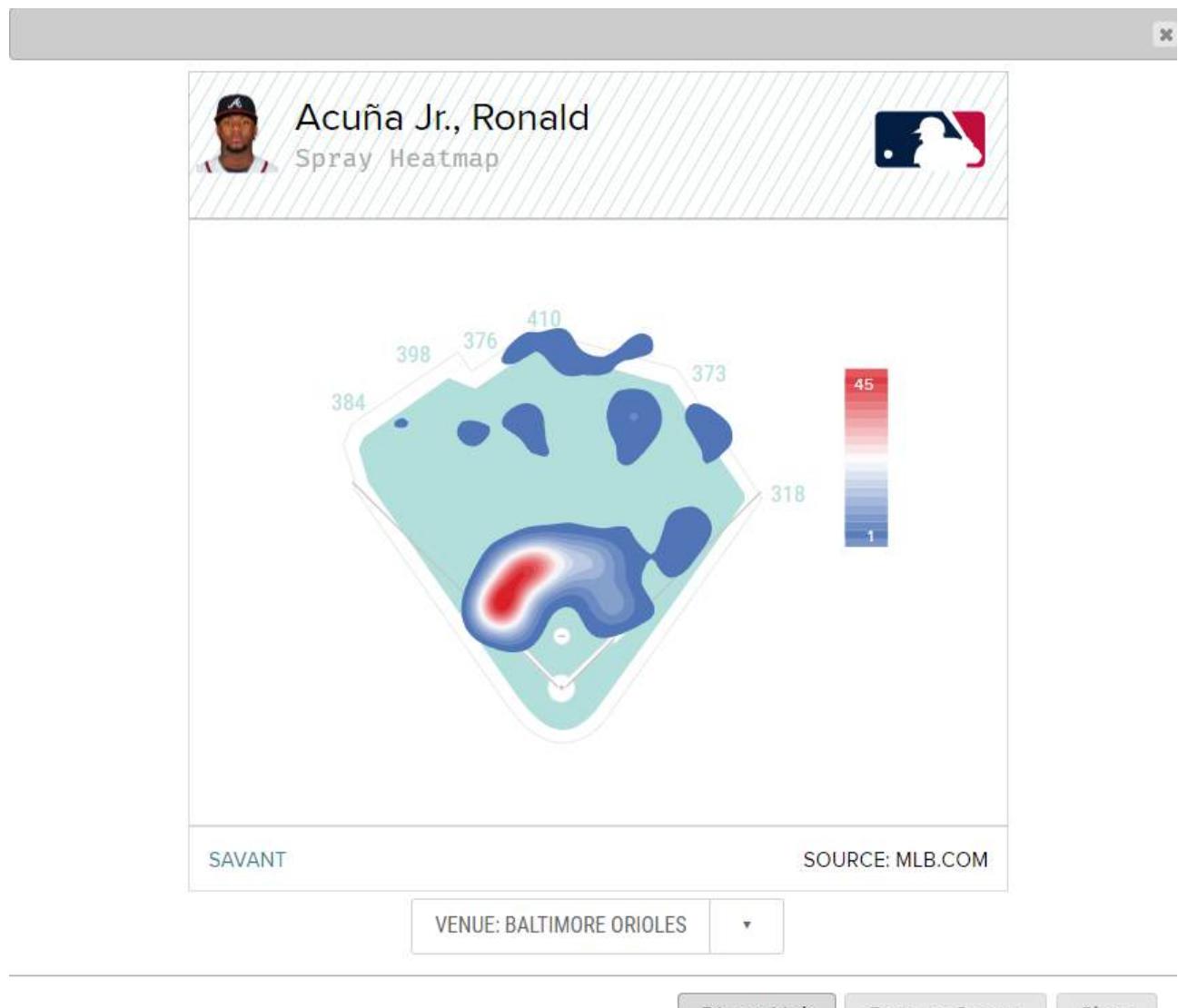
Hi all, It wasn't long ago that I wrote [my first tutorial on baseball](#). Baseball is still fairly new to me and the analytics behind it, but I've enjoyed the experimentation this far with charting some of it. Today we will revisit a previous blog and add one extra step in, and that is a heatmap to our spray chart.

Acuña Jr., Ronald

• Home Runs



For this tutorial we will use the same data as seen in our previous blog, and I recommend you [follow that blog](#) before giving this one a go. The data can be found in the Github repo at the top of the page. Heat maps aren't anything new, and baseball savant already showcase examples of spray heatmaps on their website. Our heatmap won't come out the exact same for a variety of reasons that we will cover off in the code.



To create our dataset we will need some code to create the contours. (Code credits: I've amended the below based on a snippet of code from Alexander Varlamov)

```

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.colors import rgb2hex
import sys
np.set_printoptions(threshold=sys.maxsize)

# Read data from file
data = pd.read_csv('data.csv')

# Create variables for X, Y coordinates
x = data['hc_x']
y = data['hc_y']

x = (x-125.42)/10
y = (198.27-y)/10

# Create a bivariate kernel density plot
cs = sns.kdeplot(x=x, y=y, legend=True, n_levels=10, thresh=.3, cmap='plasma')

#Set the limits to what is used in background mapping of tutorial
plt.xlim(-12.5, 12.5)
plt.ylim(-5, 19)
plt.axis('on')

plt.show()

# Create a list of all vertices, their colour and contours
lines = []

vertice_id=0
path_id=0
for collection in cs.collections:
    color = collection.get_edgecolor()
    for path in collection.get_paths():
        for vertice in path.to_polygons():
            path_id = path_id + 1
            for i in vertice:
                vertice_id = vertice_id + 1
                aa = rgb2hex(color[0]), path_id, vertice_id, i[0],i[1]
                lines.append(aa)
                print (rgb2hex(color[0]), path_id, vertice_id, i[0],i[1])

# Save values to file
np.savetxt("data_contour.csv", lines, header='color,contour_id,vertice_id,x,y',
fmt='%s', delimiter=",")

```

What the code does:

1. It reads data from a CSV file called “data.csv” (our data file from the previous blog, row level data of each hit) using pandas and assigns it to the variable data.
2. It creates variables x and y to store the X and Y coordinates respectively, extracted from the data DataFrame.
3. It performs some calculations on x and y to transform their values. ([These are transformational to follow suit of the previous blog where I made this in tableau](#))
4. It creates a bivariate kernel density plot using seaborn’s kdeplot function, passing x and y as inputs. It also specifies some additional parameters like legend, number of contour levels, threshold, and colormap.
5. It sets the limits for the x and y axis of the plot and turns on the axis.
6. It displays the plot using plt.show() This isn’t really needed, but always nice to see what your plot looks like before we go to export it.
7. It initializes an empty list called lines.

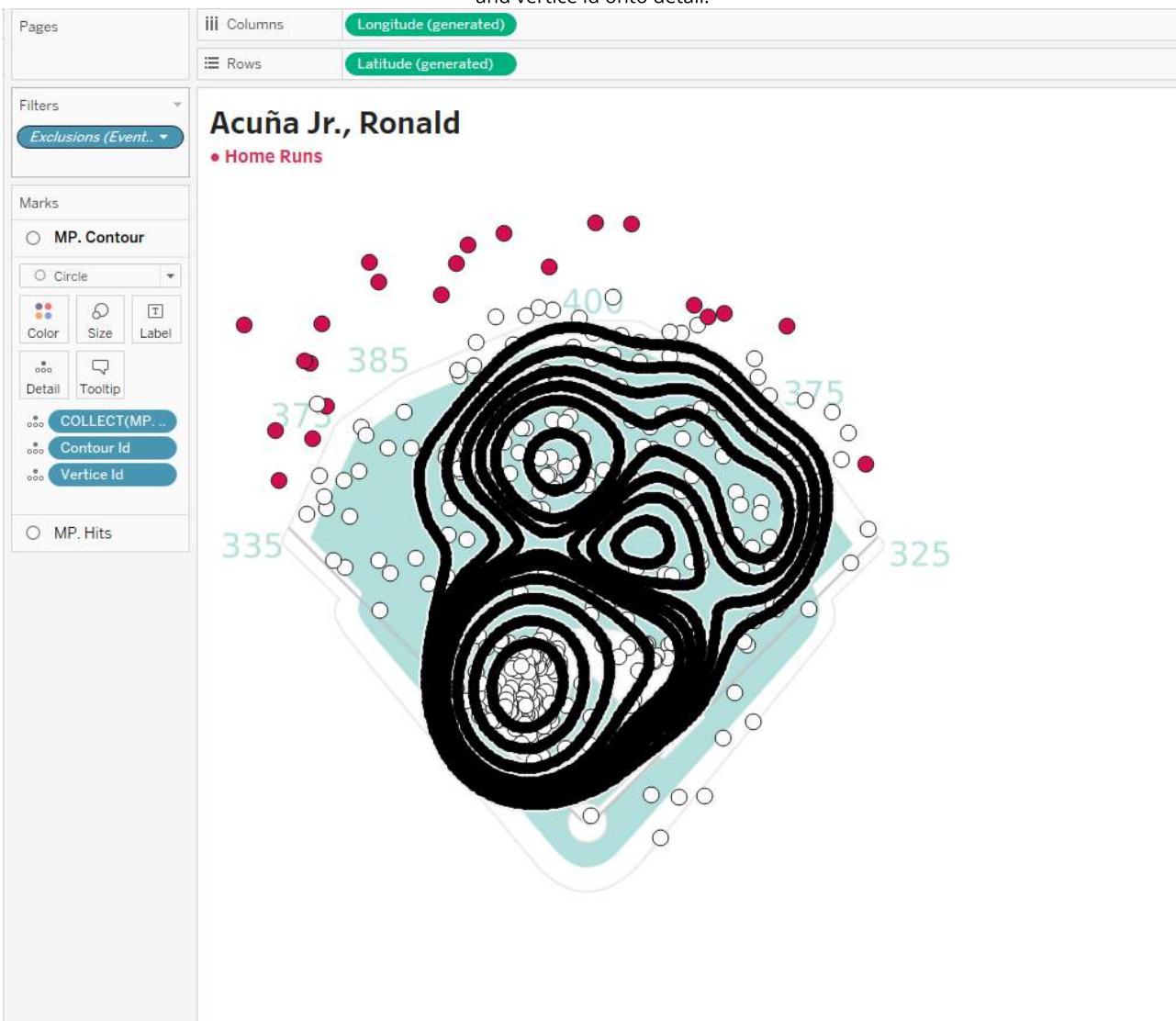
8. It iterates over the collections of the plot created in step 6, extract the edge color of each collection, and loop over the paths within each collection.
9. For each vertex in each path, it assigns a contour ID, a vertex ID, and stores the colour, contour ID, vertex ID, X-coordinate, and Y-coordinate in a tuple. This tuple is then appended to the lines list, and the information is printed to the console.
10. Finally, it saves the content of the lines list to a CSV file called "data_contour.csv" with a header row containing column names: color, contour_id, vertice_id, x, y. The values are saved in comma-separated format. (We will look to join this aggregate densified data into our original data.)

You will end up with two data files. data.csv (our original data) data_contour.csv (our heatmap data) Open the data and create a connection 1=2 with the new contour data.

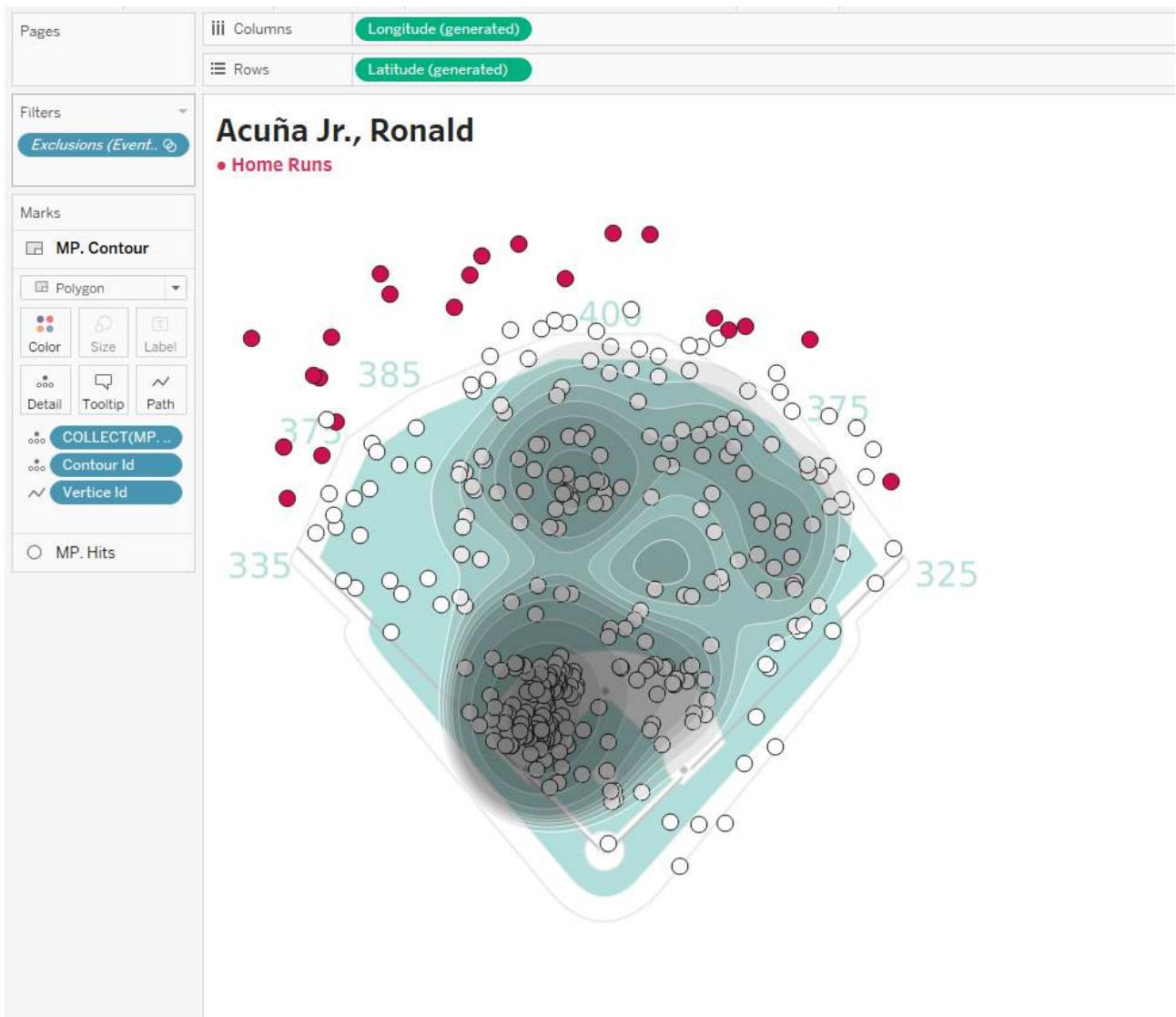
660670_data.csv is made of 2 tables. ⓘ

660670_data.csv → data_contour.csv

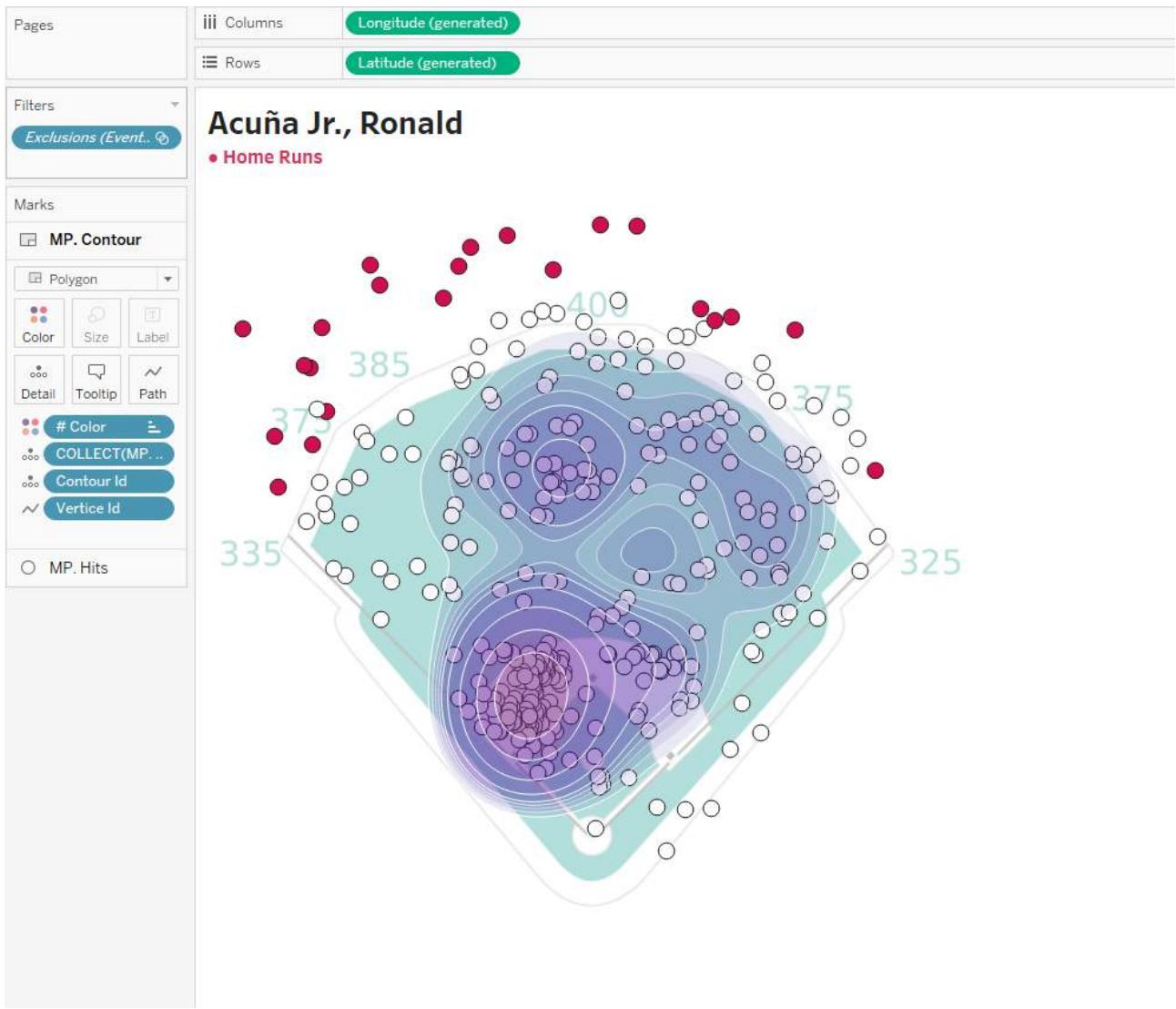
Create a new calculation MP. Contour Makepoint() If you've already followed me previous blog you can bring in this new layer onto your original chart. Here is where things get tricky. You must have used in your code the same x and y mappings as in your background image, and transformed the data in the same way. Otherwise your x and y points will not align well. Drag contour id and vertice id onto detail.



Let me know how you get on with this one – I can be found on the usual channels. Make it a line or polygon (depending on preference) and put vertice id onto path.



Add colour onto the color mark.



In your code, you will have specified a type of cmap, which is a good guide of color to use within your visual. You'll see I use a polygon, but reduce the opacity just to see some of the pitch. If you use polygon make sure to sort your colour field descending on the contour so that your 'Peaks' are at the top. & That's it. A really simple additional step to our previous mappings. Things to be cautious of:

- The x and y transformations in the code.
- Play around with the n_levels and thresh details as well as cmap to visualise different variations of the heatmap in the code.
- the xlim and ylim are based on my background imagery, build the original shot map first.

Going further,

- Try applying the code to a different sport
- Try create a small multiples of the best players this season.

LOGGING OFF, CJ

USING BOOLEANS DZV PT 2 (FOR BEGINNERS)

Hi all, A few weeks back we map an introductory blog for those that are new to dynamic zone visibility. The purpose of it was to introduce newbies to how a boolean parameter helps us leverage what is shown on the page. We will now take that to the next step which is how we can integrate it into a working dashboard. Download the [workbook](#) to follow along step by step, using the link at the top of the page.



In the [previous blog](#) we had our true false parameters showing on the page. Now we introduce sheet buttons to control what is on the page, making for a much more sleek interaction. Here is our basic example, here we have a map and when we click the purple box / sheet the map will hide,



We introduce the same logic for the bar chart, so when we click the red button, the bar chart will hide. Seems simple enough. Problem is, if we click both those sheets, then both our parameters end up as FALSE, so we have to introduce the reverse logic to set them as TRUE, for when the other one is FALSE? Kind of confusing right... It certainly isn't that intuitive to think about so lets take a look closer.



As a brief reminder of blog part one, we have a true/false parameter that drives the dynamic zone visibility. We have one for the map, and we can create a new one for our bar chart in the workbook.

Edit Parameter [P. DZV Bar]



Name

P. DZV Bar

Properties

Data type

Boolean

Display format

True

Current value

True False

Value when workbook opens

Current value

Aliases

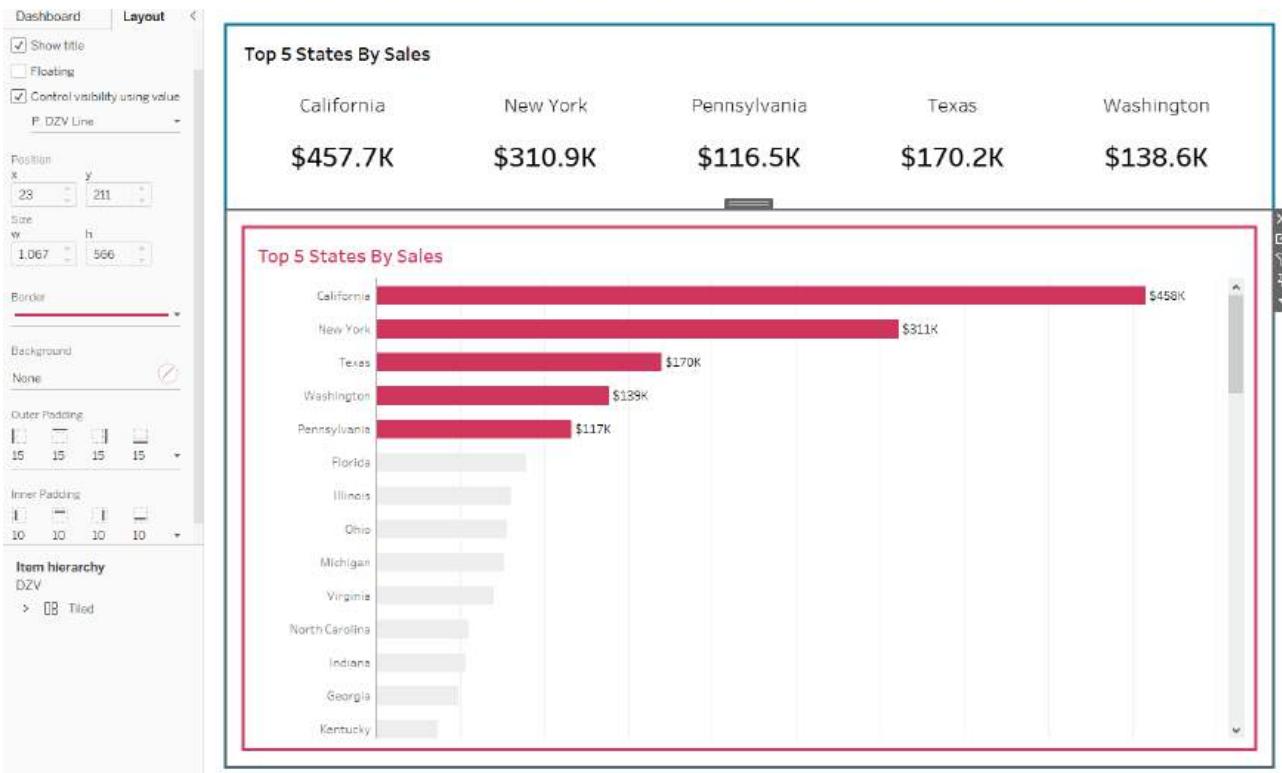
True True

False False

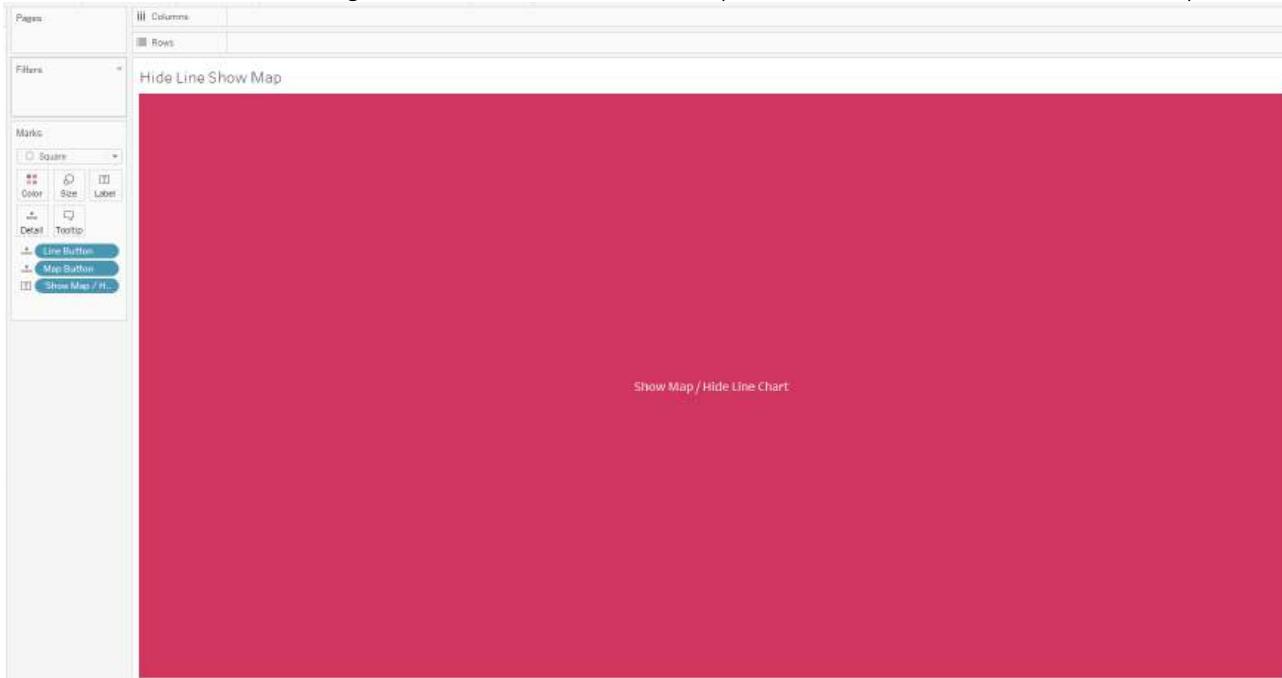
Cancel

OK

and we can of course set the dynamic zone visibility to use that parameter. Nothing new learnt so far, but now lets introduce the idea of our buttons.



Lets create a fake button, go to new sheet, make the marks a square, add the text of hide bar chart / show map.



Drag this new button onto the dashboard. Now we want to have the logic that our parameter (of true or false) is driven by our sheet. So you will see there are two calculations on detail. They are our button calculations, which just store the parameter.

Bar Button

[P. DZV Bar]

The calculation is valid.

4 Dependencies ▾

Apply OK

So now we need to set up the

Edit Parameter Action

Name: Bar To False

Source Sheets: DZV Single Button Drivers

Run action on: Select

Target Parameter: P. DZV Bar

Source Field: Bar Button ...

Aggregation: None

Clearing the selection will: Set value to False

Cancel OK

parameter action on the sheet.

This action is
the logical one – We are saying when we click our hide bar button, set the bar boolean parameter to false, source field, the field
that contains our bar You can test this by interacting with it on the dashboard. Following the same steps we can then complete
the same process for the map.



Now this works fine – but it is at a stage where 1 button is driving the map hiding and one button is driving the bar chart to hide. Problem is now – if we click both buttons, both our parameters are made false and everything disappears. So we actually now want to create the inverse logic, so that our map button doesn't just set our map to false, but also sets the bar to true.

Edit Parameter Action

Name: **Map To True** Insert ▾

Source Sheets: **DZV Single Button Drivers**

Run action on: Select

Target Parameter: **T|F P.DZV Map**

Source Field: **Map Butto...** Aggregation: **None** Aggregation: **True**

Buttons: **Cancel** **OK**

Here we take the map parameter and make it true, using our show map, using the map button. We can also create the same steps for our bar.

Name	Run On	Source	Fields	
Map To False	Select	DZV Single Button Dri...	P. DZV Map	
Map To True	Select	DZV Single Button Dri...	P. DZV Map	
Bar To False	Select	DZV Single Button Dri...	P. DZV Bar	
Bar To True	Select	DZV Single Button Dri...	P. DZV Bar	

Now with our four actions, we can then have the buttons flicking between the views. I have left this stage of the workbook as 'Single Button Drivers' so you can see the interaction's. The final thing to do now is to make the buttons 'switch out' so we can only have one button to click at one time. Go to the show map red button and control it using P. DZV Bar, and click the show bar purple button sheet and control it using P. DZV Map. Now we have a button that replaces itself in the view! Round-Up: For me dynamic zone visibility really opens up ways of interacting with dashboards. That being said there are a few 'gotchas' to consider.

1. The inverse logic when changing true and false parameters. This logic increases in complexity the more things you have on the page and what is showing on what screen. Sam Parsons work of items feeling like the 'minimise and maximise' onto the page is a great example of where you can take things.
2. Containers become so important – Your charts are render dependent on what container they are in. I've tried my best to colour code the different sheet and container borders for this tutorial but make sure your sheets are filling the container when they need to.
3. Padding!! If you are doing the 'sheet swap' or 'button swap' effect, make sure they line up with one another for a seamless transition!
4. Naming conventions: It can be quite confusing the idea of setting something true or false on multiple sheets. For example in this user case do I want my button to say 'Hide Map' or do I want it to say 'Show Bar Chart' – They are the same thing but when it comes to my calculations and dashboard this can muddle the brain.

If you have any questions on the workbook – do message me on socials. **LOGGING OFF, CJ**

DATA TOOLS FOR THE COMMUNITY WITH TRISTAN GUILLEVIN

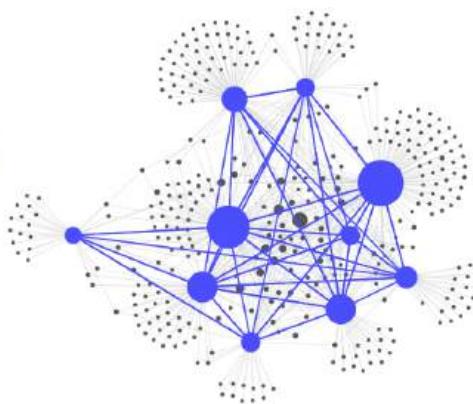
Hi all, The start to September brings with it another opportunity to share a guest blog. Someone that really needs no introduction in the data visualisation and technology space! I hope individuals enjoy reading a great conversation I had with [Tristan Guillevin](#) (@Ladataviz) Tristan is a Tableau ambassador, visionary and Iron Viz winner. He's since expanded his career to wider data design, and it was my absolute pleasure to get to chat all things data tools and see what exciting projects he's been working on.

DATA VISUALIZATION FREELANCE - TRISTAN GUILLEVIN

Helping you **level up**
your **data visualization**
projects and skills

Need an expert for a project?

[Contact me](#) [Newsletter](#)



CJ: Hey Tristan, great to have you be a part of the blog. The content you have been putting out is phenomenal and has really captivated the data visualization community as of late. I'd like to start a little bit with your background. You did various consultancy roles before diving into becoming a freelancer. What prompted the move? T: Hey CJ, thanks for giving me some space on your blog and for the kind words about my content! Before starting as an independent, I worked with two companies in two very different roles. For the first two years and a half, I was an analytics consultant for a Tableau partner. I worked with around thirty clients around the world. That taught me a lot about the job and the different tools, but ultimately the job was a bit repetitive, and I wanted to make a bigger difference. That's why I joined a start-up as a data visualization specialist. I worked on an embedded project as part of a team. I could take the time to focus on one big project, and I learned a lot of skills that were complementary to Tableau, like design in Figma or programming with Javascript. The switch to freelancer happened slowly. Everyone was stuck at home because of COVID, and my company accepted to change my position to part-time so I could work two days a week as a freelancer. I did that for six months, and then I became a full-time freelancer. It was a perfect transition for everyone. Nowadays, I consider myself less and less a freelancer since I created a company and will no longer work

alone from September! **CJ: I think many individuals in the community will have first come across your Tableau Public profile, especially after winning IronViz back in 2017. Since then it looks like your toolkit has expanded massively. Was that driven by 'right tool for the job', personal interests or client requirements?**

Vizzes 44 Favorites 19 Following 77 Followers 1,791

The Legend of Zelda - Network of heroes
Tristan Guillemin 4 163

Create a Beeswarm | Tutorial
Tristan Guillemin 19 586

Create a Bump Chart | Tutorial
Tristan Guillemin 42 1,350

Create a Sankey | Tutorial
Tristan Guillemin 28 594

Create a Network | Tutorial
Tristan Guillemin 100 1,941

Resume
Tristan Guillemin 237 19,294

The Marvel Cinematic Network
Tristan Guillemin 301 16,073

Bauhaus Sample Superstore
Tristan Guillemin 12 295

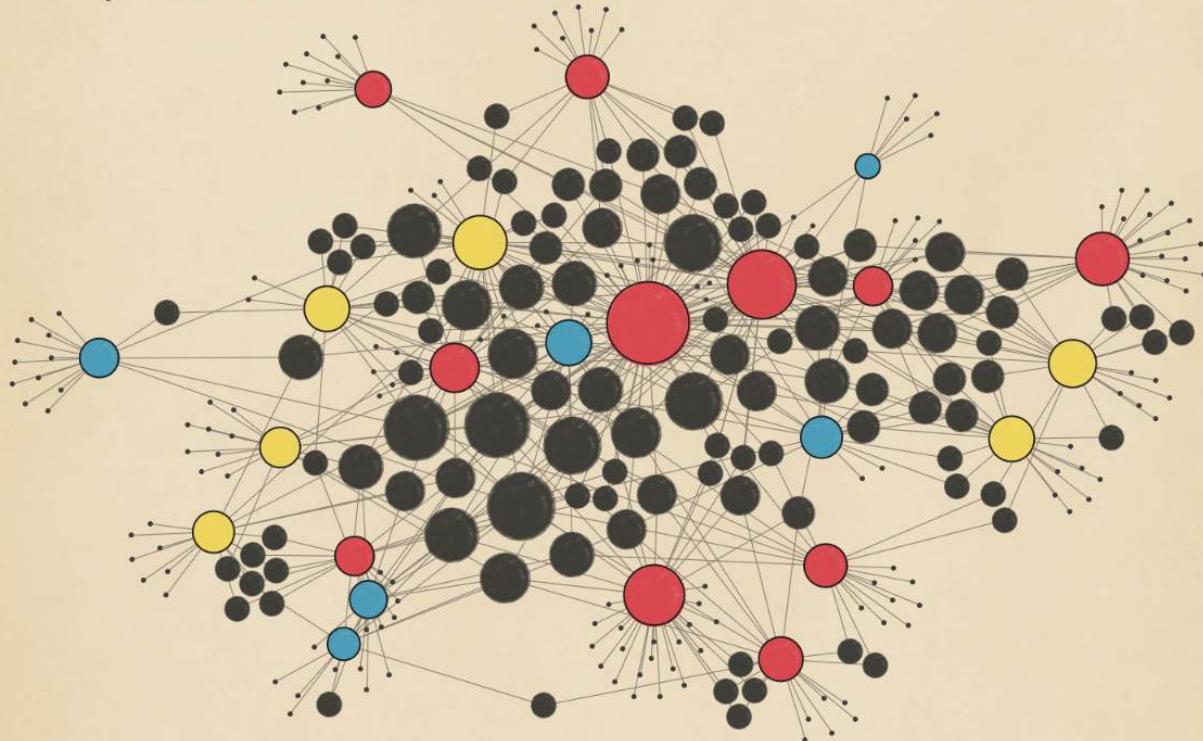
T: There were (*and still are*) multiple drivers. I think one of the most important ones is that I like to learn and challenge myself.

With every new project, even when it's just a Tableau Public visualization, I like to learn a new technique or way to create something. It makes it more fun and more interesting. Working for my clients, I like to have a horizontal approach to my toolkit.

Instead of learning multiple BI tools (*Tableau, PowerBI, Qlick, etc..*) I prefer to have a versatile toolkit. Depending on the task (*design, web development, dashboarding*), I can select the right tool for the job (*Figma, Svelte, Tableau*). So it's both driven by the fact that I like to discover new tools and a strategic approach to be a better consultant. **CJ: A considerable amount of your public facing work is advanced chart types. What is it about advanced technical charts that you enjoy?** T: 95% of the work for my clients is simple bar charts, line charts, and BANs. When I have the time to work on public personnel projects, I like to try something different and learn how to use and build a new chart type. I don't enjoy advanced charts more than simple ones! They are different, sometimes better, sometimes worst, but always different from my client-facing work. **CJ: I had the pleasure of watching you present at a Tableau London event a few months ago, showcasing your network diagram. You've made a complex chart available essentially at the click of the button. Why is making tools for the community important to you? Have there been any visuals from the community that have used your tools that you particularly like?**

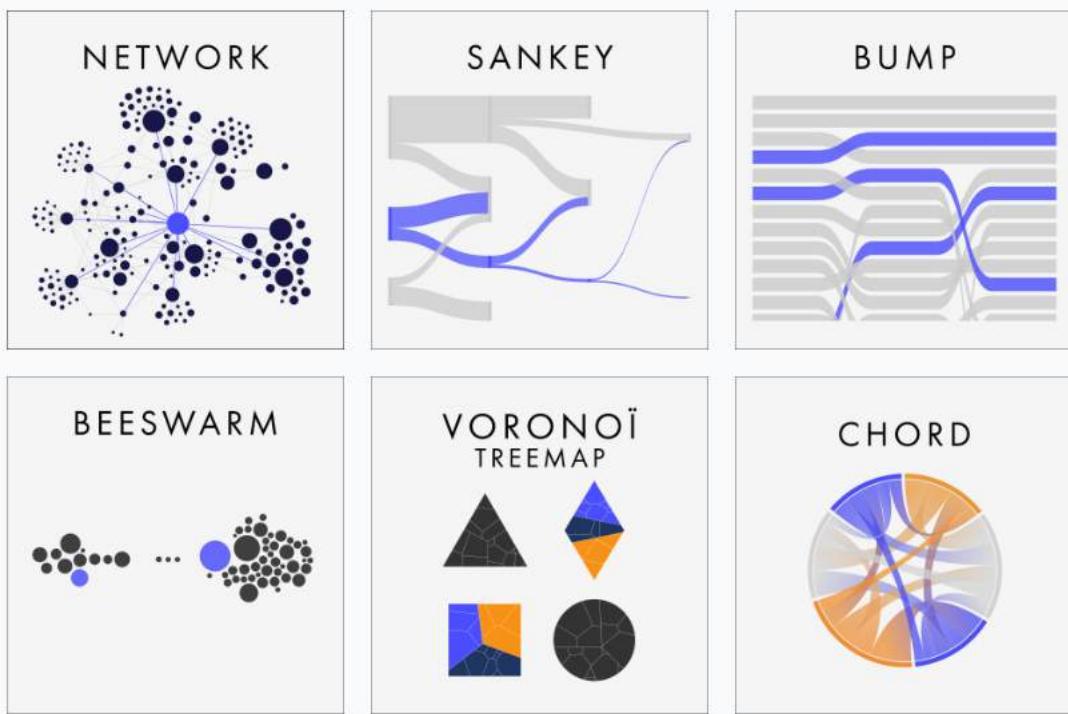
The MARVEL CINEMATIC NETWORK

Explore the Marvel Cinematic Universe like never before! Each main character and actor is connected to the movies they appear in, allowing you to see the connections between your favorite Marvel characters.



T: It is an intersection of multiple things. First, creating is the best way to learn. Second, the Tableau community has set an example when it comes to sharing content. It is why I'm still using the tool after eight years! People are helping each other all the time. I wanted to contribute too. Lastly, as an independent, I need to stay visible and relevant to find new clients and be able to help them the best I can. Building tools for the community is a perfect way to combine everything. I honestly cannot pick a single visualization that I liked more than the rest. A few of them have been awarded Viz of the Day. This truly makes me happy. It means people are seeking out new ways to represent data and being successful in doing so. Sometimes it does not follow the "best practices", but that's ok! What is important is to experiment and have fun!

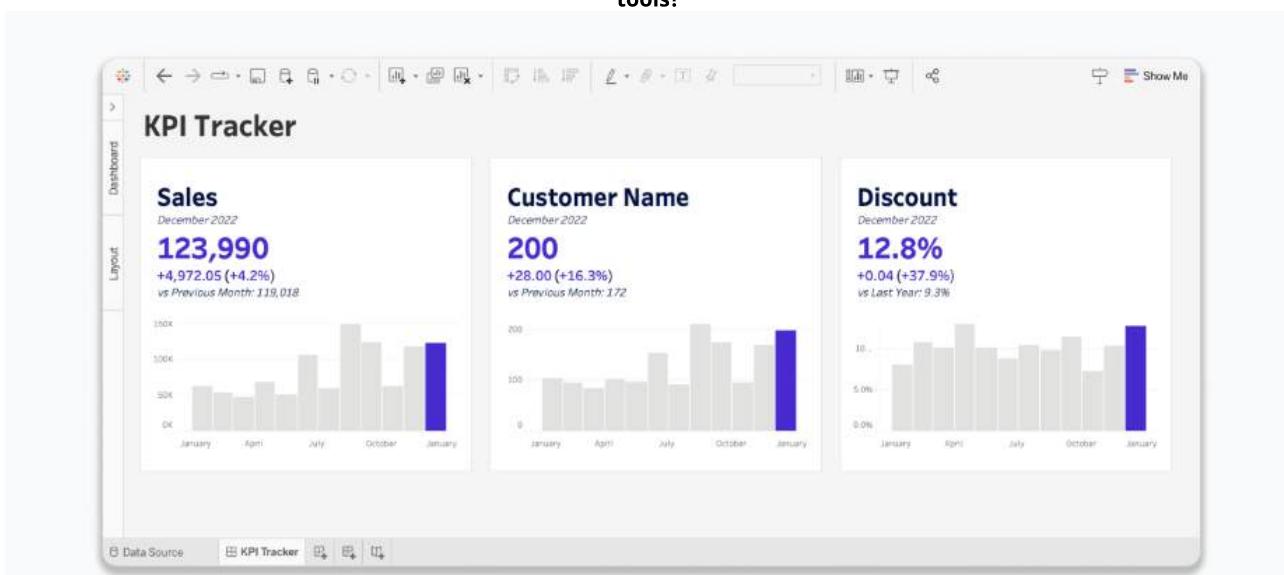
CJ: What resources would you steer community members towards that would like to learn more web based tools such as website designs, portfolio renderings and chart interactivity seen in your marvel visual? T: Svelte and Webflow are the tools that redefined what I could do on the web. Svelte is popular in the data visualization community, and it makes development fun again. I never enjoyed web development before using Svelte. For dataviz, it's much simpler than pure D3 and more intuitive than React. Webflow, on the other hand, is a web-based tool that I use mainly to design my websites. Writing CSS was always a bit of a nightmare for me, and with Webflow, I can modify CSS properties visually and then export the code. It makes the process much more enjoyable as I can see the impact of all changes, and I don't have to remember all the CSS properties. **CJ: The number of tools you have made available for free is incredible. When I first joined the community, I remember playing with the Voronoi treemap to start. Since then it's grown to 6 tools. What are some of the technical elements that go into creating these templates?**



My very first “tool” was the Voronoï Treemap, and it was hosted on Observable. There were a lot of drawbacks and manual work. The tool was generating the data, but you still had to build the viz in Tableau. Right now, everything is automated, and you get the visualization directly in a Tableau Workbook. The user experience greatly improved, but the core idea remains the same. All generators rely on the same principle. I’m using D3, a javascript library, to calculate positions (or paths) to build the charts on the web using SVGs. Once you have a result in SVG, it’s relatively easy to convert it into polygons with X and Y coordinates,

something that Tableau can understand and plot. That is roughly what the tool does in the background. Everything else is just about the UI/UX to make the experience smooth and secure for users. For example, when you download the result, the generated data is now automatically packaged in a template workbook. You directly get a visualization instead of just raw data.

It’s all about reducing friction and making the tools easy to use. **CJ: You recently released a new tool to generate KPIs in Tableau. Why did you decide to create a new tool instead of more advanced charts? Are you planning to build other tools?**



T: It all started because I was tired of building KPIs in Tableau manually. My day-to-day work with my clients consists mainly of building business dashboards. Tracking KPIs with Big Aggregated Numbers is often essential. Anyone who has ever built a KPI dashboard knows how annoying it is to do it in Tableau. You need a lot of calculated fields and, if you want it to look nice, a lot of tedious formatting. All those actions could be automated as they always follow the same pattern. This new tool, **BANG** (Big Aggregated Number generator), does just that. The tool generates all the calculations, the worksheets, and a nice dashboard directly in a Tableau workbook. It helps me save hours of work every time I need to build a KPI dashboard. I thought the tool could be useful to many people, so I worked hard on making it more robust and secure, and I released it as a product! It's a

great feeling when people trust you, purchase your tool, and find it useful. I really want to continue on that path. Before the end of the year, I'll release a new tool to simplify building any dashboard in Tableau! I can't wait to share more in the next months!

CJ: You've recently moved towards doing more Youtube video tutorials to help users better use your available tools.

What can we expect next for your channel?



T: I've always enjoyed video recording and editing. I was doing anime AMV in high school and later worked on different video productions at my university. I don't know why I waited so long to start publishing videos as I always liked this medium! Currently, I'm making tutorials on how to use my tools. Those are useful videos both to better explain how the tools work and also for me to be more confident and learn to make better videos. Eventually, I want to explore different formats but I don't yet have a clear plan for what's next. **CJ: Have you seen a change in the type of work that you are requested by a client? You are obviously someone who has been very committed to learning all types of new skills. What do you think will be the most important in years to come?** T: As I divide my time between working for my existing clients, building tools, and creating content, it's difficult to accept new work. I still provide mostly pure Tableau expertise to my recurring clients, but I do receive more and more inquiries related to developing solutions. For the years to come, I think the main game changer is AI. Building products is getting easier and easier, and pure technical skills will not be enough to be successful. I will personally focus on two things: learning how to use AI to build products more efficiently and providing high-quality content to build an audience.

Building well-made products for a reachable audience that needs them sounds like a good plan! **CJ: Thank you Tristan, I appreciate you taking the time to chat. Where can individuals reach out to you to continue the conversation?** In a lot of places! [Here's my link tree](#) where people can easily find me and decide how they want to reach out or keep up with my work. Thank you, CJ for giving me some space on your website! See you soon at a Tableau event! **CJ Round-Up:** The stand out things for me is that self belief to go from employee to freelancer to starting your own business. What an incredible feat. I want to wish Tristan the best of luck with his new company. Rest assured it is already on a path for success and I know how much the BANG & other generators Tristan has created have impacted our community. **LOGGING OFF, CJ**

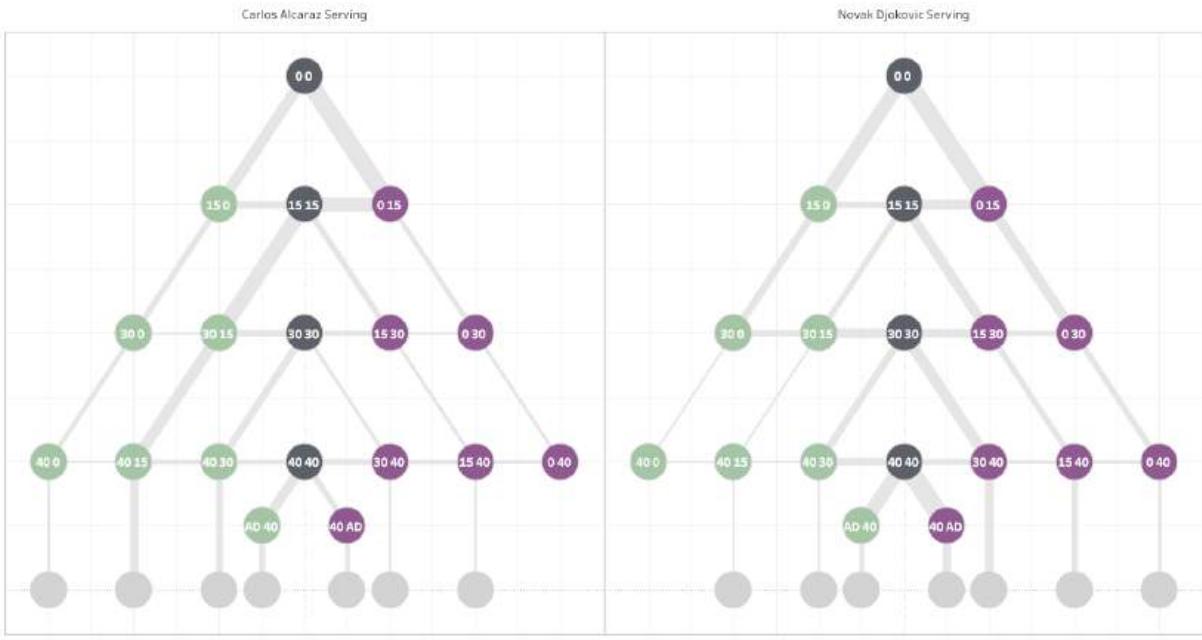
CREATE TENNIS GAME TREES IN TABLEAU

Hi all, Something new this week that I don't think individuals will have come across too frequently. It is the chart type called Tennis Game Tree's. Game Tree's aren't specific to Tennis, in fact you can read more about them [here](#).

Game Tree

Wimbledon 2023 Mens Final
Carlos Alcaraz vs Novak Djokovic

Game Tree is a depiction of point progression for a selection of games within a tennis match or across a series of tennis matches. Games start at '0-0' and 'progress' through the tree until a player wins a game. The thickness of the lines connecting any two scores indicates the number of points which 'progressed' between the two nodes.



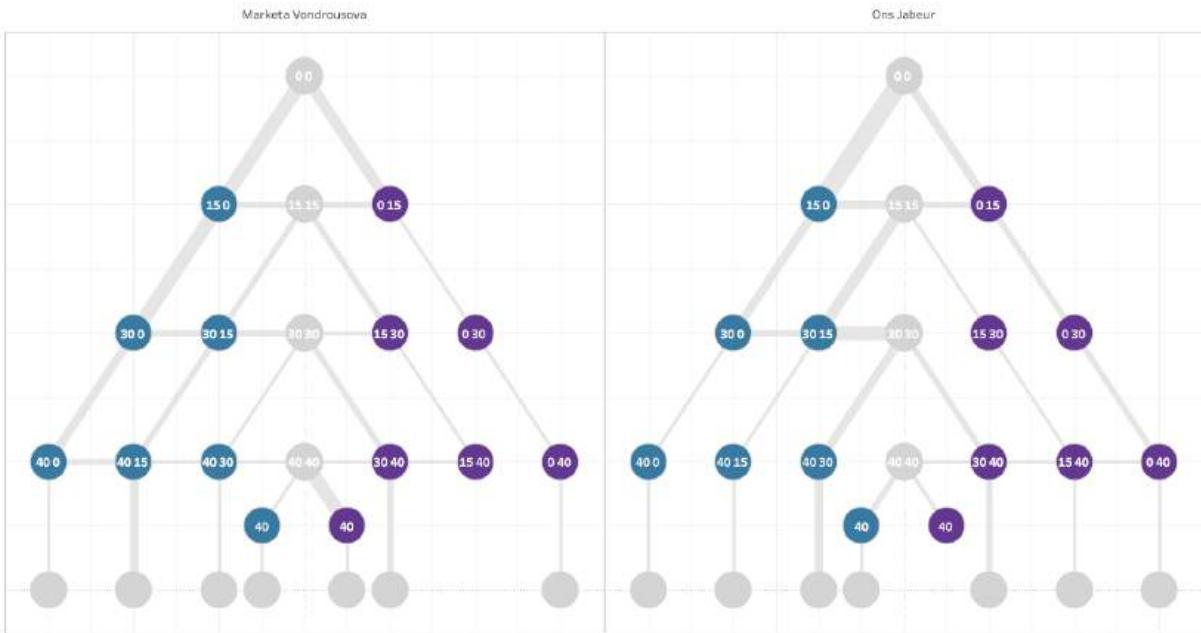
[Tutorial on CJ-Mayes.com](https://www.cj-mayes.com/tutorials/game-tree/)

What Is A Tennis Game Tree? "Game Tree" is a depiction of Point Progression for a selection of games within a tennis match or across a series of tennis matches; it is a [Sankey Diagram](#) and possesses the "[Markov property](#)", meaning that the set of future "states" that are possible are constrained by the current "state", the point score at any moment in a game. What that means in laymen terms is each point in tennis progresses you towards winning the game, i.e its 0-0 so the next point can only be 15-0 or 0-15... depending on that point winner determines what the next potential outcome of points can be, until eventually someone wins the game.

Game Tree

Wimbledon 2023 Womens Final
Marketa Vondrousova - Ons Jabeur

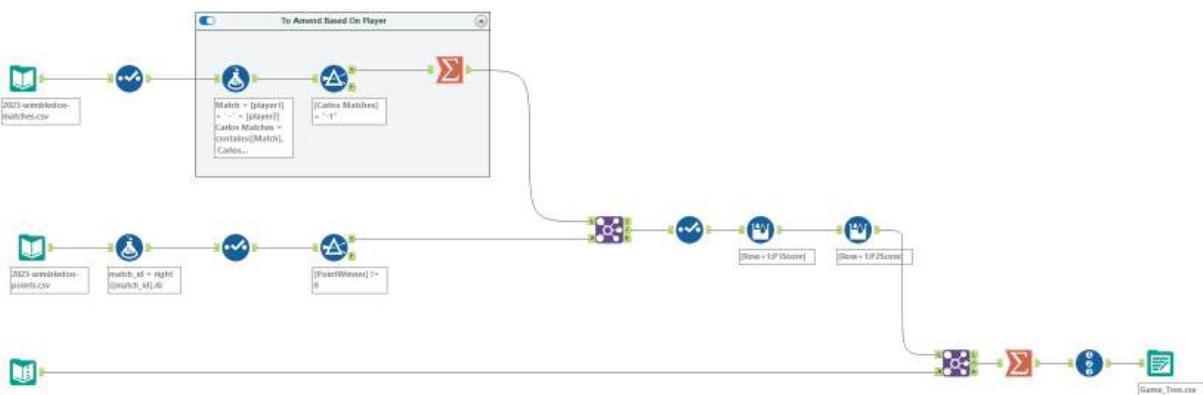
Game Tree is a depiction of point progression for a selection of games within a tennis match or across a series of tennis matches. Games start at '0-0' and 'progress' through the tree until a player wins a game. The thickness of the lines connecting any two scores indicates the number of points which 'progressed' between the two nodes.



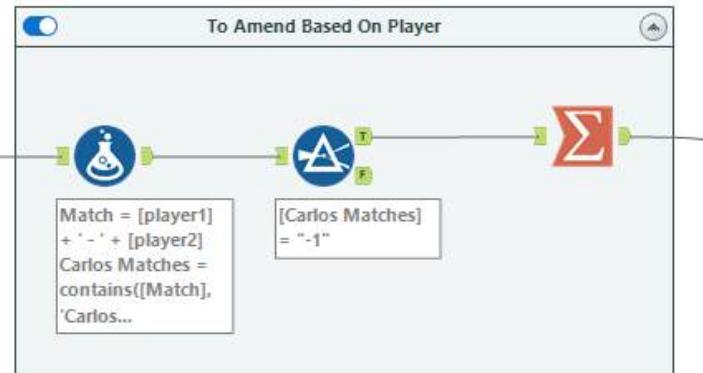
Tutorial on CJ-Mayes.com

Taking the above example of the Womens Wimbledon Final, we can highlight the progression of points for each player for when they are the server. It becomes increasingly interesting when you dive into the stories behind why the player may have been able to break serve / or had a good service match when looking at the spread of thicker lines. For example Carlos losing many of his first service points but yet didn't lose many service game points overall (when you reach the bottom of the chart) So how do we go about creating this chart? Well there is a workflow attached in the repository, but the good thing is you should be able to amend the flow for any game that has been charted within Jeff Sackman's repository! Let's talk through the data. The original tennis data can be found [here](#). You can find a copy of the files within the Github Repository at the top of the page, also found [here](#).

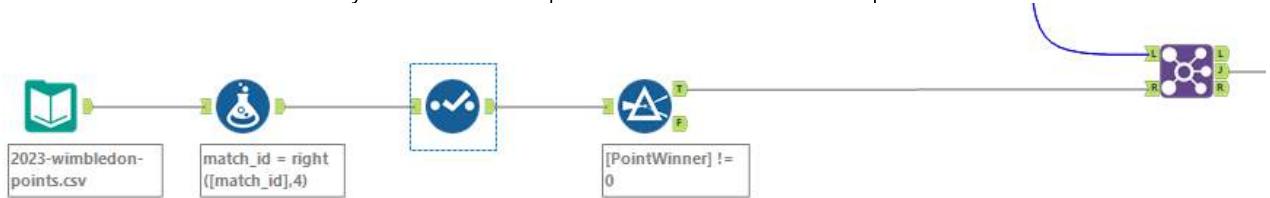
Alteryx Transformations I've added the Alteryx transformations but whether you want to replicate this in another tool lets take a look at the transformations required.



The match data – We connect to this and just select the non null fields, and amend the match number to be an integer.



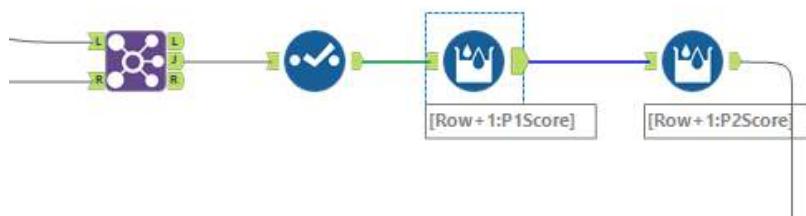
Next I create a match field, this is a simple concatenation, and then find matches for Carlos – This can be amended based on whichever player you are looking for. Finally I use the summarise tool to find the max match ID for Carlos (naturally, the ID's increase throughout the tournament so the maximum one would relate to the final!) None of this bit of the flow is that important as you could manually look up which match ID you need. The next part of the flow comes from the points data.



The matchID is the last four digits attached to the matchID column, we will need these to join together to reduce the file to just the final. The select tool just picks columns that are of use to me such as point, server, score etc. The filter looks to remove the first two records of data from the file where the match hasn't started. By joining these files together I have all the points of the final between Novak & Carlos.

Record	Max_Match	PointNumber	PointWinner	PointServer	P1Score	P2Score	P1PointsWon	P2PointsWon	P1Winner	P2Winner
1	Carlos Alcaraz - Novak Djokovic	1	2	2	0	15	0	1	0	0
2	Carlos Alcaraz - Novak Djokovic	2	1	2	15	15	1	1	0	0
3	Carlos Alcaraz - Novak Djokovic	3	1	2	30	15	2	1	1	0
4	Carlos Alcaraz - Novak Djokovic	4	2	2	30	30	2	2	0	0
5	Carlos Alcaraz - Novak Djokovic	5	1	2	40	30	3	2	0	0
6	Carlos Alcaraz - Novak Djokovic	6	2	2	40	40	3	3	0	0
7	Carlos Alcaraz - Novak Djokovic	7	2	2	40	AD	3	4	0	1
8	Carlos Alcaraz - Novak Djokovic	8	1	2	40	40	4	4	0	0
9	Carlos Alcaraz - Novak Djokovic	9	2	2	40	AD	4	5	0	0
10	Carlos Alcaraz - Novak Djokovic	10	2	2	0	0	4	6	0	0
11	Carlos Alcaraz - Novak Djokovic	11	2	1	0	15	4	7	0	0
12	Carlos Alcaraz - Novak Djokovic	12	2	1	0	30	4	8	0	1
13	Carlos Alcaraz - Novak Djokovic	13	2	1	0	40	4	9	0	0
14	Carlos Alcaraz - Novak Djokovic	14	1	1	15	40	5	9	0	0
15	Carlos Alcaraz - Novak Djokovic	15	1	1	30	40	6	9	1	0
16	Carlos Alcaraz - Novak Djokovic	16	2	1	0	0	6	10	0	0
17	Carlos Alcaraz - Novak Djokovic	17	2	2	0	15	6	11	0	0
18	Carlos Alcaraz - Novak Djokovic	18	1	2	15	15	7	11	0	0
19	Carlos Alcaraz - Novak Djokovic	19	1	2	30	15	8	11	0	0
20	Carlos Alcaraz - Novak Djokovic	20	2	2	30	30	8	12	0	0

Here is what that clean looks like so far. The part of the flow is by far one of the most important parts. We utilise the multi row



formula.

What the multi row formula allows us to do is find the players point for the next point. Notice in the data screen shot below we now have a field called P1 Next Score P2 Next Score.

Record	Max_Match	PointNumber	PointWinner	PointServer	P1Score	P2Score	P1PointsWon	P2PointsWon	P1Winner	P2Winner	P1 Next Score	P2 Next Score
1	Carlos Alcaraz - Novak Djokovic	1	2	2	0	15	0	1	0	0	15	15
2	Carlos Alcaraz - Novak Djokovic	2	1	2	15	15	1	1	0	0	30	15
3	Carlos Alcaraz - Novak Djokovic	3	1	2	20	15	2	1	1	0	30	20
4	Carlos Alcaraz - Novak Djokovic	4	2	2	30	30	2	2	0	0	40	30
5	Carlos Alcaraz - Novak Djokovic	5	1	2	40	30	3	2	0	0	40	40
6	Carlos Alcaraz - Novak Djokovic	6	2	2	40	40	3	3	0	0	40	AD
7	Carlos Alcaraz - Novak Djokovic	7	2	2	40	AD	3	4	0	1	40	40
8	Carlos Alcaraz - Novak Djokovic	8	1	2	40	40	4	4	0	0	40	AD
9	Carlos Alcaraz - Novak Djokovic	9	2	2	40	AD	4	5	0	0	0	0
10	Carlos Alcaraz - Novak Djokovic	10	2	2	0	0	4	6	0	0	0	15
11	Carlos Alcaraz - Novak Djokovic	11	2	1	0	15	4	7	0	0	0	30
12	Carlos Alcaraz - Novak Djokovic	12	2	1	0	30	4	8	0	1	0	40
13	Carlos Alcaraz - Novak Djokovic	13	2	1	0	40	4	9	0	0	15	40
14	Carlos Alcaraz - Novak Djokovic	14	1	1	15	40	5	9	0	0	30	40
15	Carlos Alcaraz - Novak Djokovic	15	1	1	30	40	6	9	1	0	0	0
16	Carlos Alcaraz - Novak Djokovic	16	2	1	0	0	6	10	0	0	0	15
17	Carlos Alcaraz - Novak Djokovic	17	2	2	0	15	6	11	0	0	15	15
18	Carlos Alcaraz - Novak Djokovic	18	1	2	15	15	7	11	0	0	30	15
19	Carlos Alcaraz - Novak Djokovic	19	1	2	30	15	8	11	0	0	30	30
20	Carlos Alcaraz - Novak Djokovic	20	2	2	30	30	8	12	0	0	30	40

The expression for this is:

0	15	0	1	0	0	15	15
15	15	1	1	0	0	30	15

Why do we need that? Well to create our decision tree / Game tree, we need to know from each point where it goes to out of the different possibilities. It is best to start treating our points as start and end nodes. Now I have that sorted, the final thing to do is

	Origin X	Origin Y	P1 Score	P2 Score	P1 Next Score	P2 Next Score	End X	End Y	*
1	0	4	0	0	0	15	1	3	
2	0	4	0	0	15	0	-1	3	
3	1	3	0	15	0	30	2	2	
4	1	3	0	15	15	15	0	3	
5	2	2	0	30	0	40	3	1	
6	2	2	0	30	15	30	1	2	
7	3	1	0	40	0	0	3	0	
8	3	1	0	40	15	40	2	1	
9	-1	3	15	0	15	15	0	3	
10	-1	3	15	0	30	0	-2	2	
11	0	3	15	15	15	30	1	2	
12	0	3	15	15	30	15	-1	2	
13	1	2	15	30	15	40	2	1	
14	1	2	15	30	30	30	0	2	
15	2	1	15	40	0	0	2	0	
16	2	1	15	40	30	40	1	1	
17	-2	2	30	0	30	15	-1	2	
18	-2	2	30	0	40	0	-3	1	
19	-1	2	30	15	30	30	0	2	
20	-1	2	30	15	40	15	-2	1	
21	0	2	30	30	30	40	1	1	
22	0	2	30	30	40	30	-1	1	
23	1	1	30	40	0	0	1	0	
24	1	1	30	40	40	40	0	1	
25	-3	1	40	0	0	0	-3	0	
26	-3	1	40	0	40	15	-2	1	
27	-2	1	40	15	0	0	-2	0	
28	-2	1	40	15	40	30	-1	1	
29	-1	1	40	30	0	0	-1	0	
30	-1	1	40	30	40	40	0	1	
31	0	1	40	40	40	AD	0.5	0.5	
32	0	1	40	40	AD	40	-0.5	0.5	
33	0.5	0.5	40	AD	0	0	0.5	0	
34	0.5	0.5	40	AD	40	40	0	1	
35	-0.5	0.5	AD	40	0	0	-0.5	0	
36	-0.5	0.5	AD	40	40	40	0	1	
	...								

to create the 'Tree Mapping' data. I have created every output possible from the scores and allocated them the required X & Y co-ordinates for our tree. Normally i'd be against this manual preparation but the tree never changes shape so our X & Y co-ordinates can be stationary. It is important to note that origin x and y values of 0 start at the top, but end values of 0 (i.e when the player has won the game) are seen at the bottom of the tree. We look to join this data in, and then we can do our final summary to be able to prep our data ready for

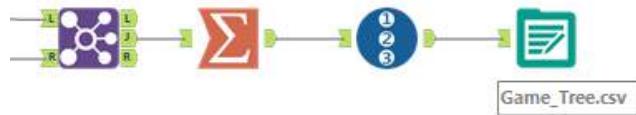


Tableau.

The most important part of this is the summarise

tool because we want the thickness of our lines to represent the number of times that movement from one point to another

Actions: Add ▾

	Field	Action	Output Field Name
▶	PointServer	Group By	PointServer
	PointNumber	Count Distinct	CountDistinct_PointNumber
	Origin X	Group By	Origin X
	Origin Y	Group By	Origin Y
	P1 Score	Group By	P1 Score
	P2 Score	Group By	P2 Score
	P1 Next Score	Group By	P1 Next Score
	P2 Next Score	Group By	P2 Next Score
	End X	Group By	End X
	End Y	Group By	End Y

Action Properties
This action has no properties

happened.

example data looks like, ready for exporting to CSV. You can find a copy of this final prepped dataset in the Github Repo.

Here is what that final

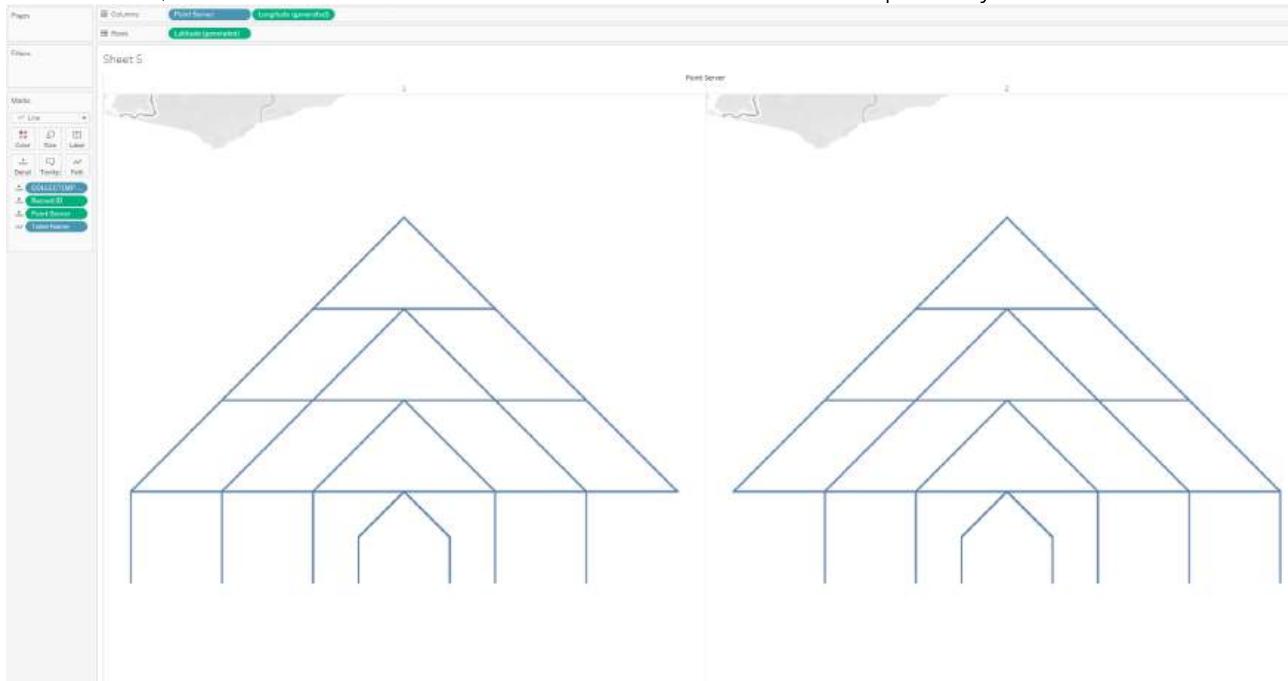
Record	PointServer	CountDistinct_PointNumber	Origin X	Origin Y	P1 Score	P2 Score	P1 Next Score	P2 Next Score	End X	End Y
1 1	2		-3	1	40	0	0	0	-3	0
2 1	1		-3	1	40	0	40	15	-2	1
3 1	6		-2	1	40	15	0	0	-2	0
4 1	3		-2	1	40	15	40	30	-1	1
5 1	1		-2	2	30	0	30	15	-1	2
6 1	3		-2	2	30	0	40	0	-3	1
7 1	5		-1	1	40	30	0	0	-1	0
8 1	3		-1	1	40	30	40	40	0	1
9 1	4		-1	2	30	15	30	30	0	2
10 1	8		-1	2	30	15	40	15	-2	1
11 1	5		-1	3	15	0	15	15	0	3
12 1	4		-1	3	15	0	30	0	-2	2
13 1	4		-0.5	0.5	AD	40	0	0	-0.5	0
14 1	3		-0.5	0.5	AD	40	40	40	0	1
15 1	4		0	1	40	40	40	AD	0.5	0.5
16 1	7		0	1	40	40	AD	40	-0.5	0.5

Tableau Build Once you've run your flow for your desired match, the only thing you will have to do is connect to your data and re-union it with itself, replacing the old data. However, below outlines how to build it from scratch just in case!

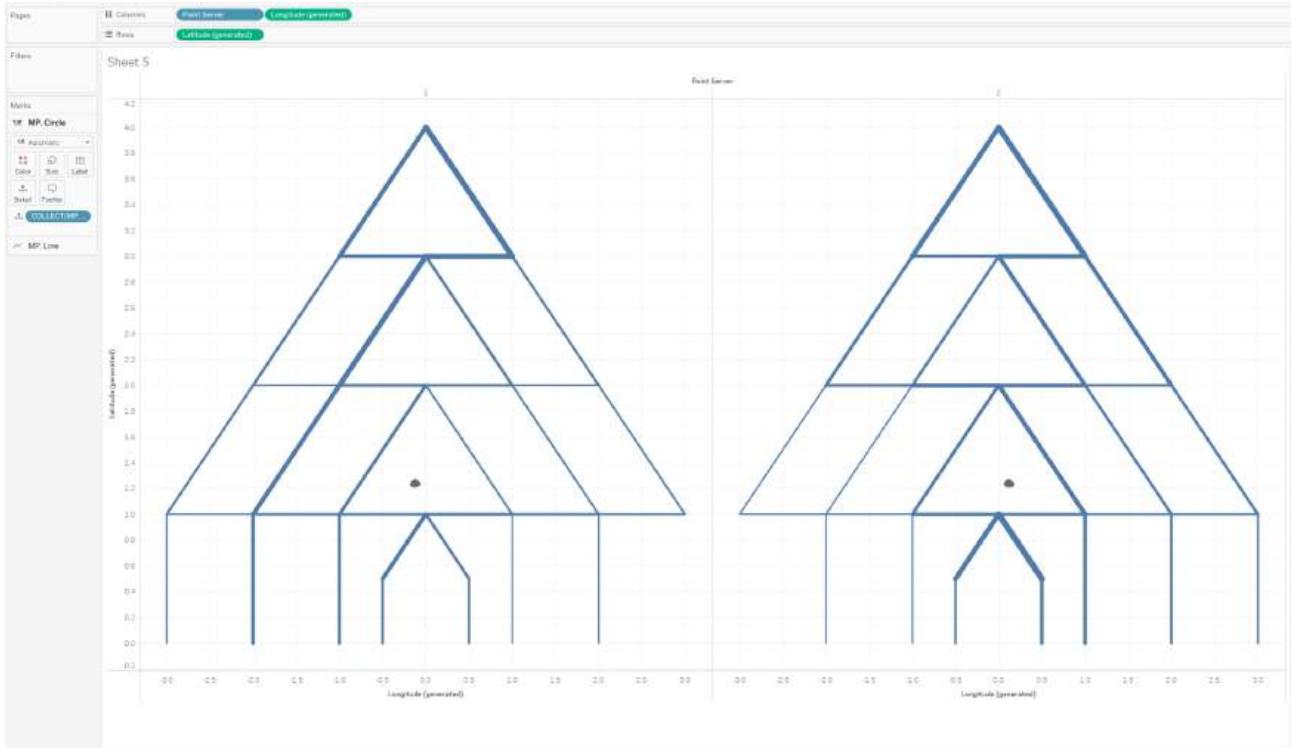
□ Game_Tree(2)

Game_Tree.csv is made of 2 tables. ①											
Game_TreeDetail											
Game_Tree.csv											
Game_TreeDetail											
Name	Game_TreeDetail										
Game_Tree.csv	Game_TreeDetail										
FIFO:	Game_TreeDetail										
Type	Field Name	Physka...	Re...
Record ID	Game_Tri...	Revi...
Point Server	Game_Tri...	Point...
Court/District Point Number	Game_Tri...	Court...
Orign X	Game_Tri...	Orign X
Orign Y	Game_Tri...	Orign Y
P1 Score	Game_Tri...	P1 Sc...
P2 Score	Game_Tri...	P2 Sc...
P1 Next Score	Game_Tri...	P1 Ne...
P2 Next Score	Game_Tri...	P2 Ne...
End X	Game_Tri...	End X
End Y	Game_Tri...	End Y
Table Name	Game_Tri...	Table...
Origin	End	Base	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9

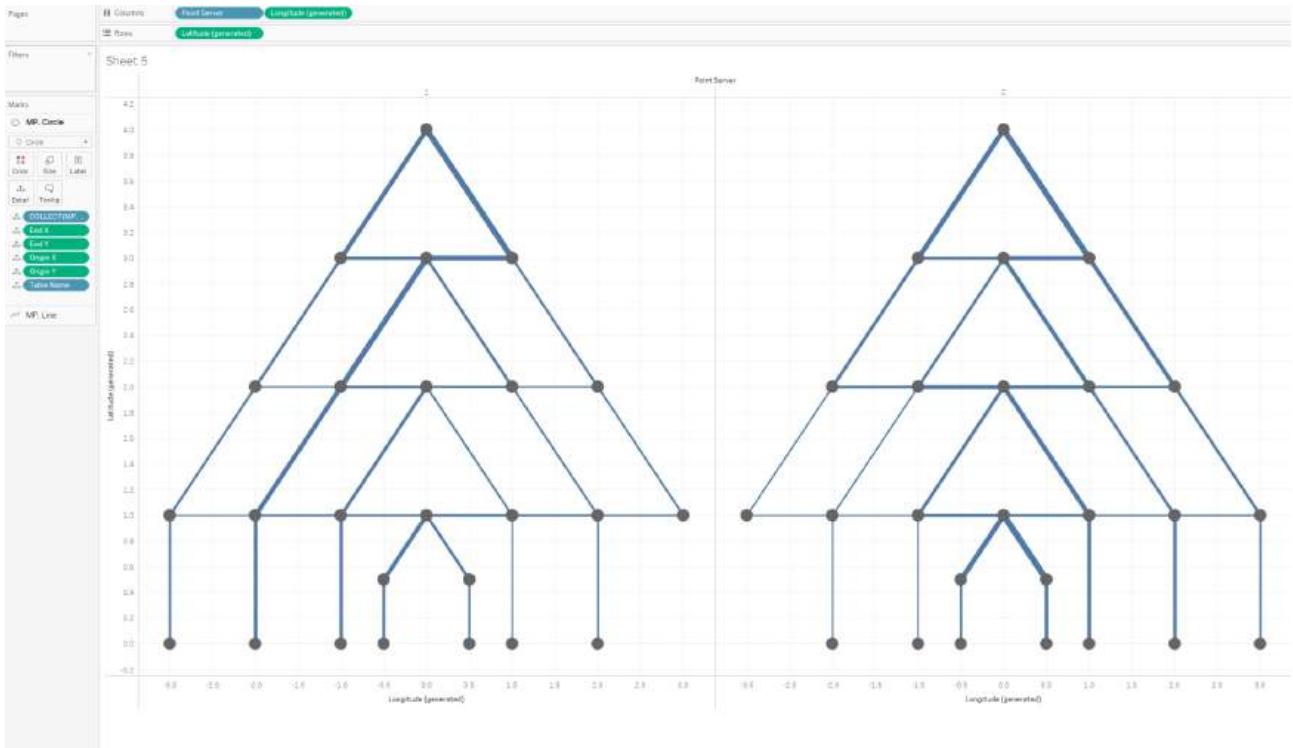
The first thing is to union your data file, this allows for us to have a start and end point. The first two calculations separate our X and Y coordinates. GT. X if = 'Game_Tree.csv' then ELSE END & GT. Y if = 'Game_Tree.csv' then ELSE END We want two layers to our chart, one is our lines, the other is our circles. Lets create them as makepoints. MP. Line MAKEPOINT(,) & MP. Circle MAKEPOINT(,) Double click MP Line, and then print server onto columns, as a discrete dimension. Change the mark to a line. Add Record ID, Point Server onto detail as continuous dimensions. Add Table Name to path and your visual should come alive.



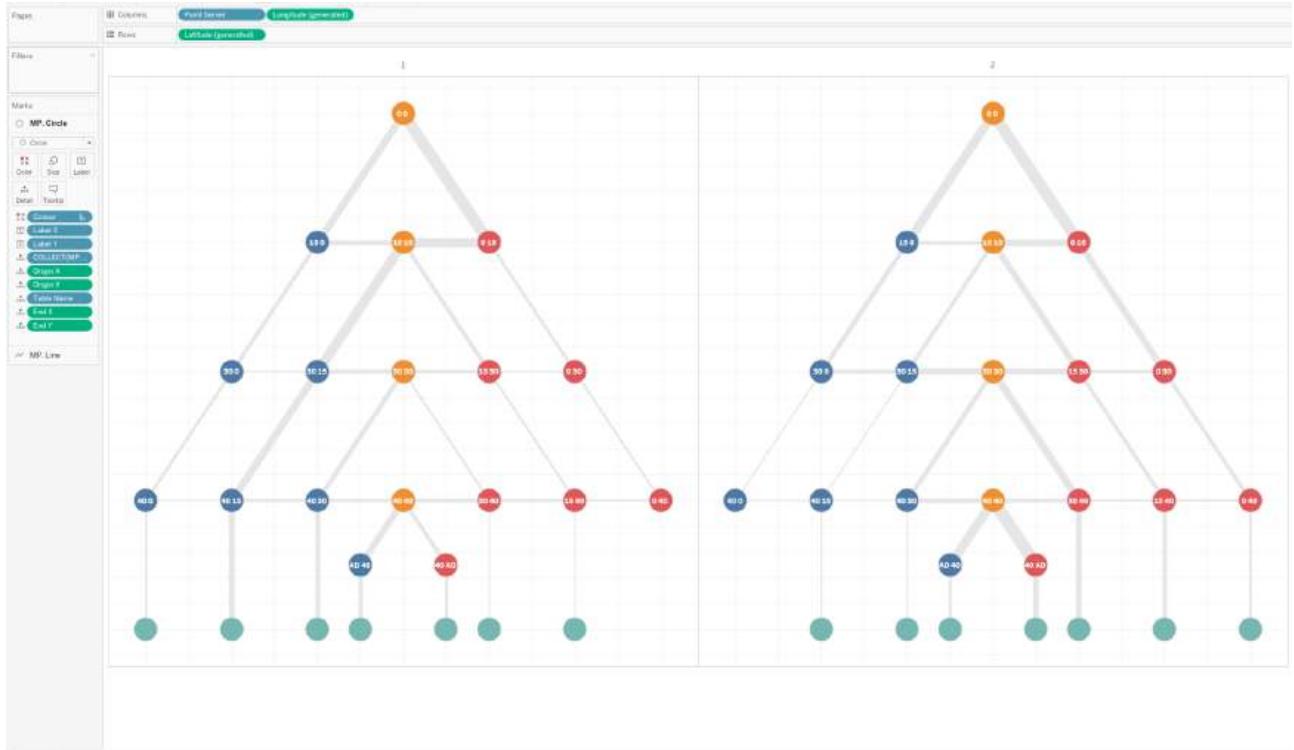
Add Count Distinct Point Number onto Size to showcase the frequency of that tree path. Next we want to add our MP Circle onto the pane. At this point we can then turn off our map background. Map -> Background -> None.



For this layer we want to make the mark a circle. Add Origin X, Origin Y, End X and End Y onto detail. Make them dimensions.
Add Table Name onto Detail.



This is the final chart in terms of how it is shaped. The final few touches help cosmetically to elevate the viz. Create the following Label calculations and bring them onto label. Label X if = 'Game_Tree.csv' then END Label Y if = 'Game_Tree.csv' then END Create a colour calculation to split the circles colours based on player. if = 'Game_Tree.csv' then sign() end

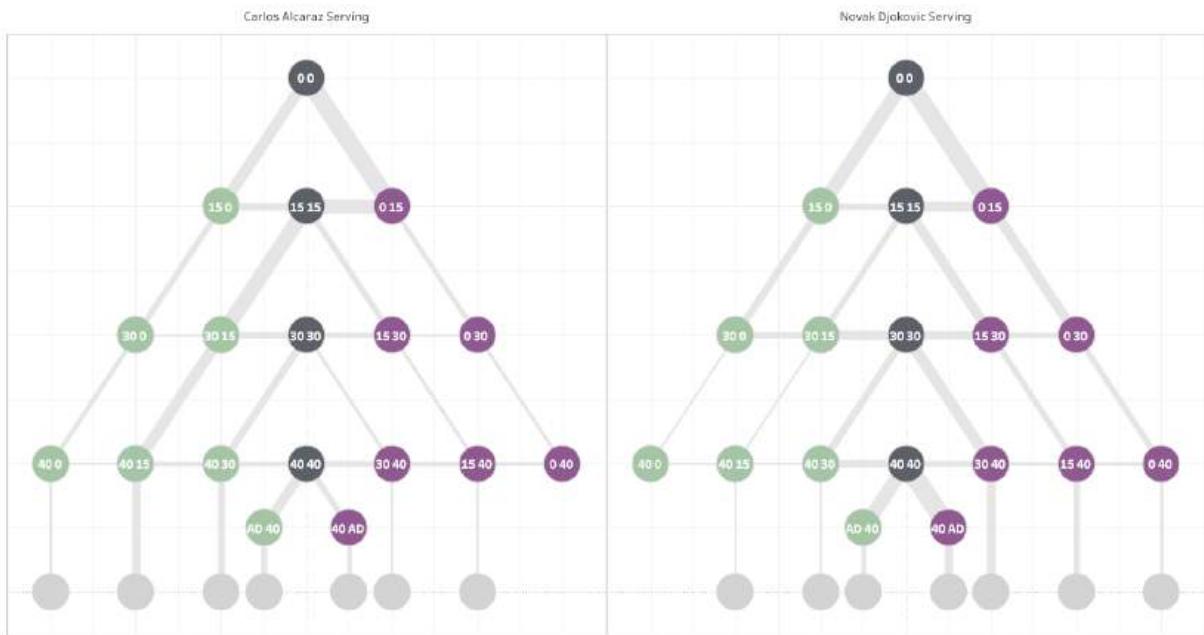


Amend as appropriate.

Game Tree

Wimbledon 2023 Mens Final
Carlos Alcaraz vs Novak Djokovic

Game Tree is a depiction of point progression for a selection of games within a tennis match or across a series of tennis matches. Games start at '0-0' and 'progress' through the tree until a player wins a game. The thickness of the lines connecting any two scores indicates the number of points which 'progressed' between the two nodes.



Tutorial on CJ-Mayes.com

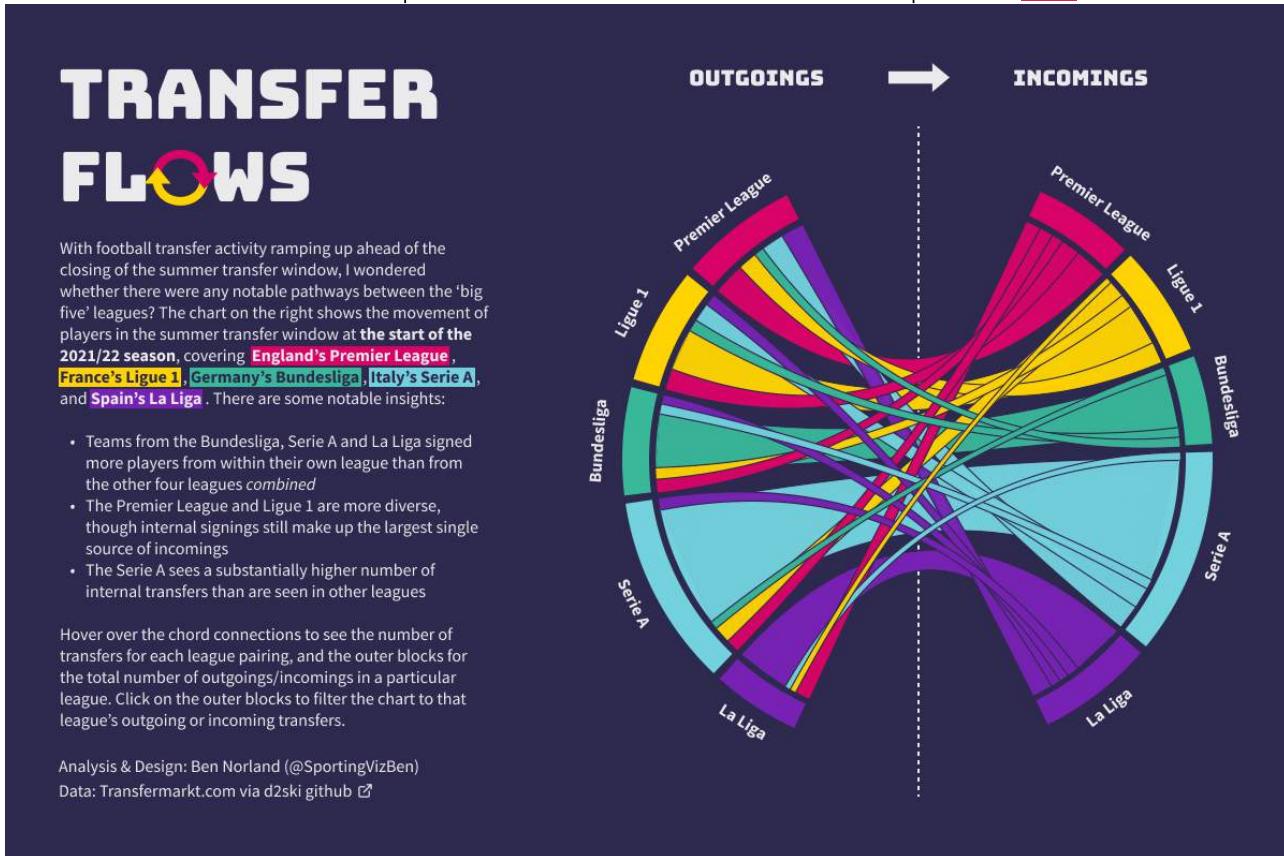
There we have it, our Tennis Game Tree. GOING FURTHER

- Why not run the workflow against a match of your choosing?
- Why not add some supplementary metrics to this chart and make it into a dashboard?
- Why not recreate the alteryx flow in a transformation tool of your choosing.

Any questions, just shoot me a message. **LOGGING OFF, CJ**

MAKING SPORTS DATA INTERACTIVE WITH BEN NORLAND

Hi all, This week I had the pleasure of sitting down with [Ben Norland](#), Ben is a stand out SportsVizSunday contributer as well as being recognised more recently as a Tableau featured author. Bens ability to show great stories, whilst incorporating technical charts and impressive design has really caught my eye. We sat down to chat a little more about where this passion comes from and some of the tips and tricks behind some of his visuals. Check his profile out [here](#).



CJ: Ben, 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 at Indeed or prior?



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

Ben Norland He/Him

Dublin, Ireland

Hi there! I'm a data analytics professional with a passion for visualization and communication of insight. I use Tableau to develop data tools that help business leaders make better-informed decisions. On my Tableau Public...

[Read more](#)

[Follow](#)

Vizzes 37

Favorites 96

Following 113

Followers 417

Modern Family Shared Storylines - #VotD
Ben Norland
139 ⚡ 8,375

Parameter Switch Button Example
Ben Norland
0 ⚡ 16

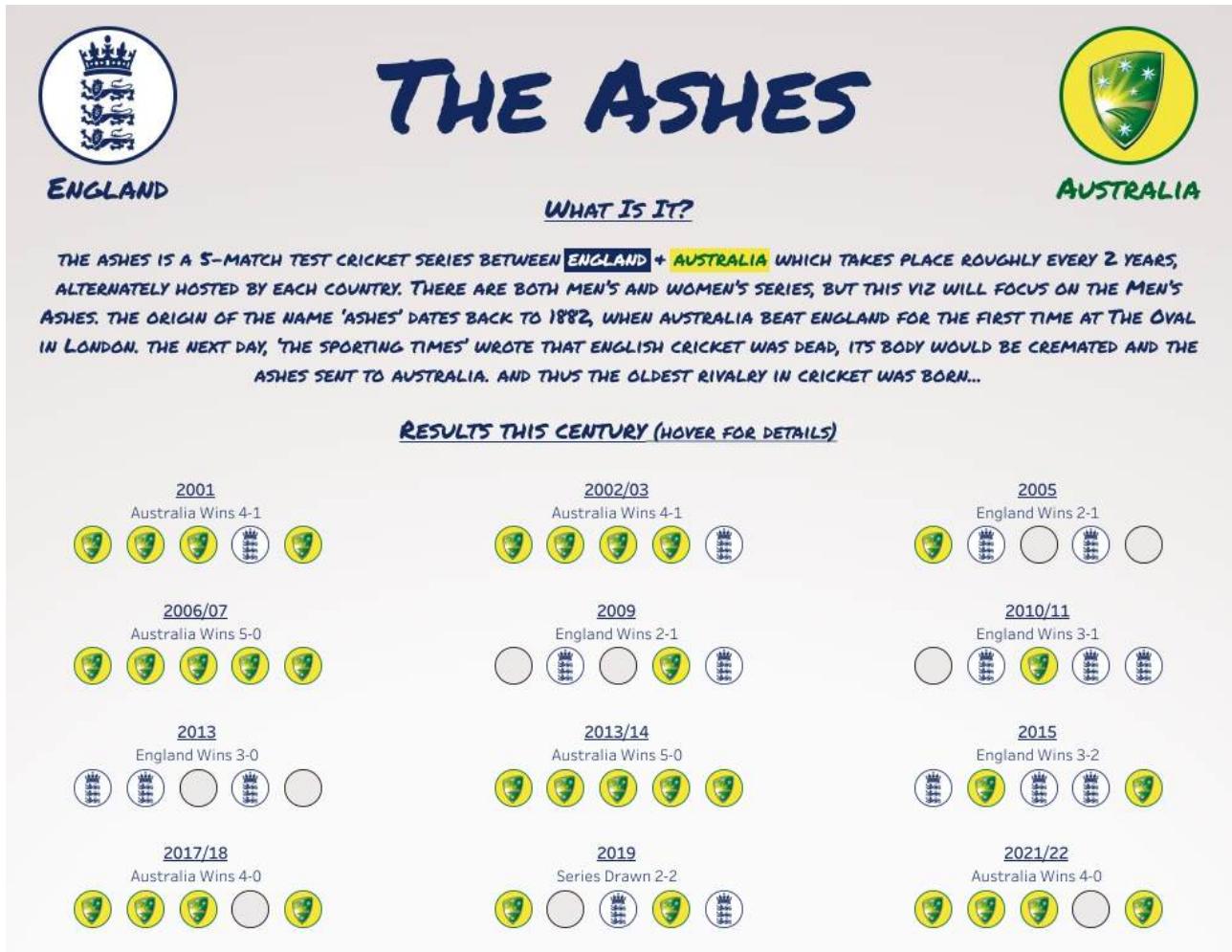
Quarterback Throw Zones
Ben Norland
15 ⚡ 689

The Destroyer of Worlds
Ben Norland
2 ⚡ 68

B: Thanks for having me, CJ! This blog is one of my regular go-to sites for inspiration, so to be asked to contribute is a real privilege! I've been working with Tableau for about 3 years, initially at Deutsche Bank and then at Indeed. But I've been in data-focused roles for about 8 years now. While my academic background wasn't data-related at all, I discovered pretty early on in my career that I really enjoyed building tools in Excel that surfaced data insights more easily. Around the same time, I read

Moneyball and became fascinated with the idea of using data to get a performance edge. That drove me to learn R, SQL, and eventually Tableau, and I now use all 3 on a daily basis to build data tools that help drive better-informed decision making. Away from the day job, I've developed a real love for Tableau Public and the DataFam community – I've drawn so much inspiration from it over the past few years, and I've taken a lot of joy from becoming a more active member of the community this year. CJ:

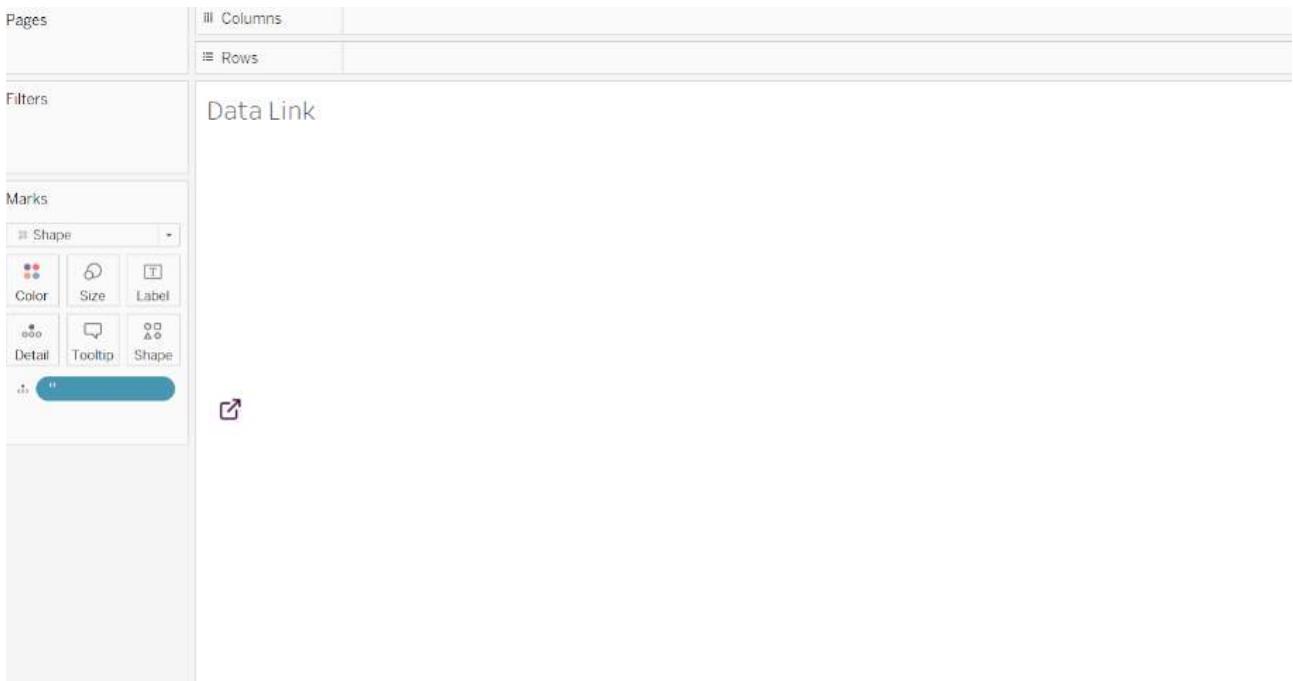
Are your favourite sports to visualise the same as the ones you follow? I see your profile considers 4 or 5 different sports. Do you have many go-to resources when it comes to sports analytics data? It's nice to see, you always add in an external link to the data!



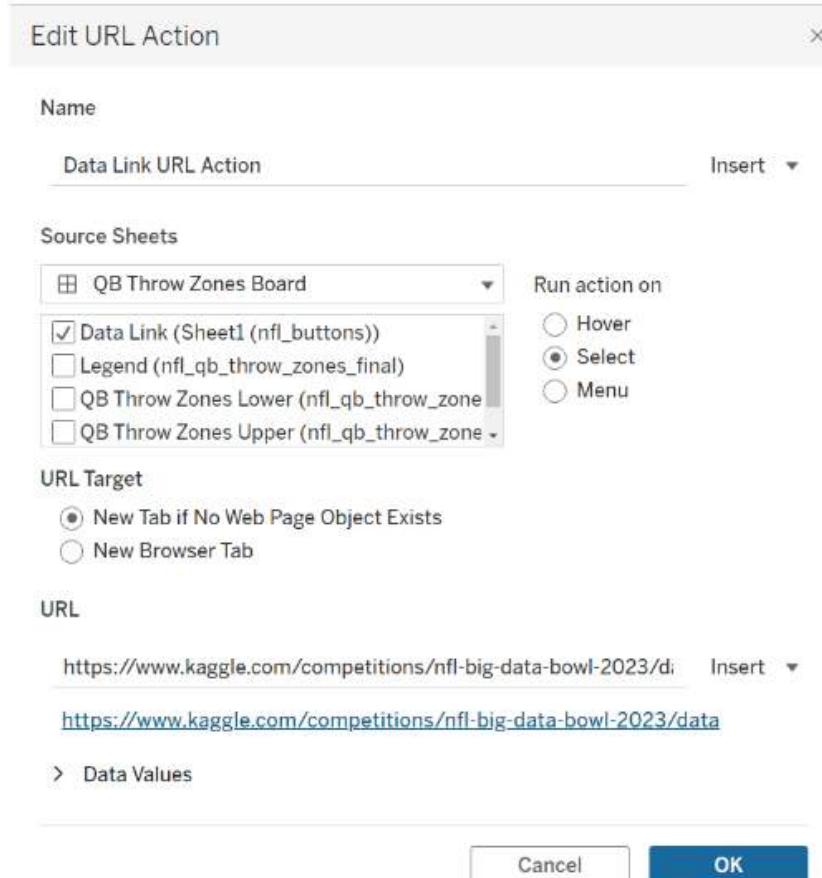
B: Yes, it would definitely seem that way, although it's not entirely intentional! I'm a huge football (soccer) fan, so I tend to cover that more than any other sport, but I also follow F1 and cricket closely, and have always had an interest in American sports too. I

think sports in general are such rich sources of data, and there's always an interesting angle to explore, be it simply learning more about a sport for yourself, or using data to challenge popular narratives. A recurring theme in the DataFam is 'viz what you love' – that is absolutely the case for me and why I repeatedly gravitate towards sports vizs! In terms of data sources, **Kaggle** is my go-to for finding interesting datasets (my **Premier League Ever-Presents**, **NFL**, **NBA** and **F1** vizs, amongst others, are all based on datasets found on Kaggle) – whatever topic you have in mind, you're pretty much guaranteed to find a relevant dataset on there. But I've also used data from the likes of **fivethirtyeight**, the **worldfootballR** R package, and **StatsBomb** in the past. I like to build self-serve tools that address questions I have, but that also might address different questions for others. With that in mind, I always like to add a link directly to the data – if I haven't answered their question, perhaps a viewer might want to go and dig in for themselves, or publish their own related viz. URL actions in Tableau are great for this purpose – my approach is as follows:

- Open a new worksheet and name it 'Data Link'
- Double click on the 'Marks' panel and type ":" in the pill that appears
- Change the mark type to 'Shape' in the dropdown, and select a relevant shape (I have a 'link' shape that I tend to use)



- Add this worksheet to your dashboard and position/resize accordingly
- In the 'Dashboard' menu, select Actions -> Add Action -> Go to URL
- Ensure your 'Data Link' worksheet is selected, and that 'Run Action On' is set to 'Select'. Then add the URL for your data source. Now your users will be sent to the datasource if they click on that area of the board!



There are almost certainly other ways to do this, but I like that this allows you to use a shape/image that can also be formatted to match your dashboard style, and offers a familiar, intuitive user experience. **CJ: When we caught up previously you mentioned how you set yourself a goal of building out your portfolio while on paternity. What new chart types or styles are you hoping to tick off in the remainder of the year? Can you explain a bit more about how you've gone about this?** B: Yes – my wife and I welcomed our second child earlier this year, and I've been very lucky to have quite a long period of paternity leave this time round. While chasing around after two young kids has occupied most of this time, I wanted to make sure I also used some of it to build out my

Tableau Public portfolio, partly so I didn't forget everything while away from work(!), and partly to push myself to learn new chart types and improve my design skills. Areas of focus have included Map Layers (I know, I was a little slow on the uptake...!), Chord Charts, Curved Bump Charts, and improving my understanding & usage of Polygons. I've also spent a good bit of time working on design elements in Figma – I tend to build backgrounds for my dashboards there, add them as images in Tableau, and then layer my charts over the top. Combining Figma with Tableau has a number of advantages, but one thing I particularly like is that it allows you to get around some of the challenges with publishing fonts to Tableau Public/Tableau Server – many fonts don't render well when published directly in Tableau, but you don't encounter the same issues when they're embedded in a background image, so you know viewers will see the final version in the way you intended. **CJ: You are an avid R coder, how has learning how to code helped you with transforming data for some of your sports visualizations? Do you tend to use R for retrieving, cleaning, or transforming data?**



B: I love how much control R gives you over transforming and analysing data. I use it for pretty much everything – from connecting to data (either via individual files, larger databases, or in some cases via specific R packages that contain data themselves), doing exploratory analysis, cleaning and transforming, joining or merging datasets, and prepping the output for Tableau. A good example is my [Player Comparison Dashboard](#). While the underlying data is available via worldfootballR, I needed to build an R script to ingest that data, join multiple files together, make adjustments so that all metrics are shown on a per-90-minute basis and can be re-scaled based on team possession, calculate each player's percentile rank for each metric, and pivot the data from wide to long to make it easier to work with in Tableau. In theory you could do a lot of that work in Tableau itself, but I find R gives me more control over these elements (it's easier to review the data after each step and isolate any bugs, for example), and then I can output a file that is much closer to the final version that Tableau needs, which is often helpful from

a performance perspective. I know that coding in R (or Python) can seem intimidating if you haven't done it before, but I would really recommend trying it. About 80% of what I do in R uses a few core functions and techniques that can be learned quite easily, and will give you outsized benefits in terms of productivity. I would recommend starting with the [Tidyverse](#) – this group of packages will give you a fantastic set of tools for connecting to, cleaning, transforming and analysing datasets. And if you have

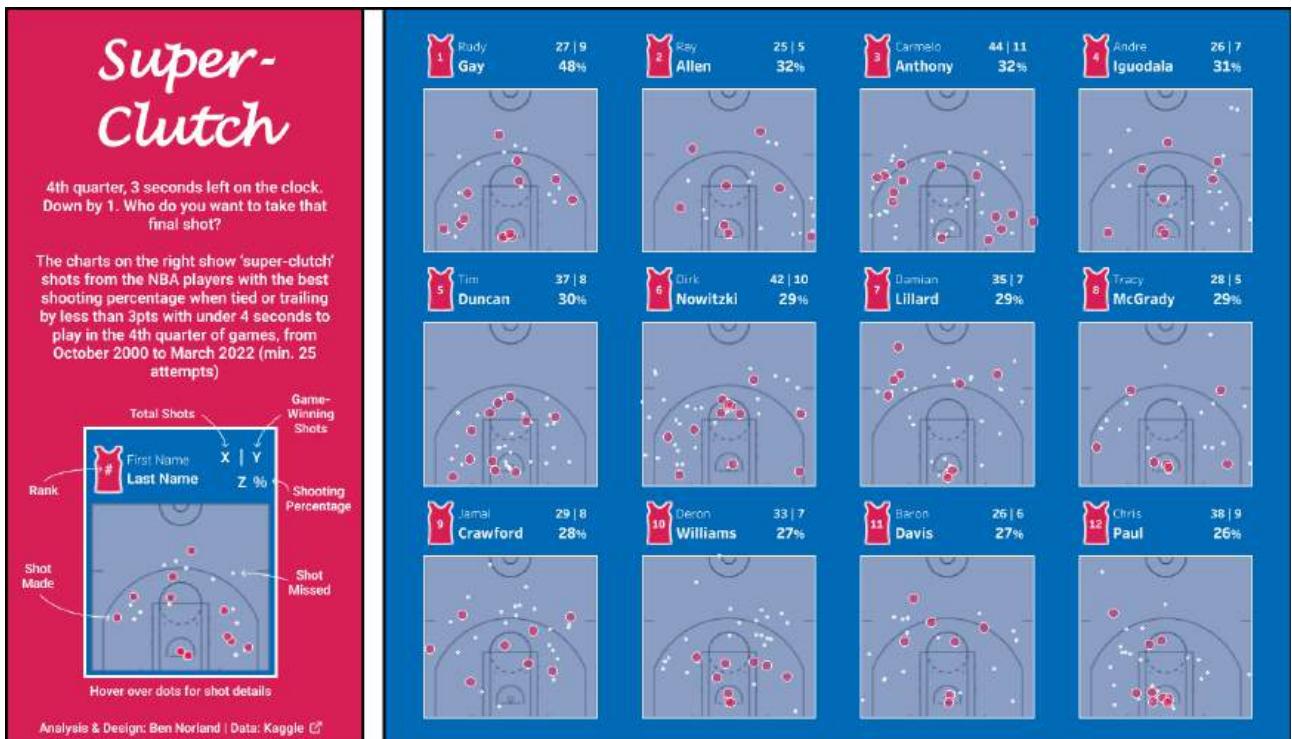
SQL experience, the syntax should seem pretty familiar. **CJ: Your Quarterback throw zone and your super clutch NBA visuals are a thing of beauty. Can you talk about the thought process behind your small multiple format to be able to add in extra details like jersey number or team name? What considerations go into the overall framing when you need a detailed key to go alongside the visual?**



B: I'm a big fan of small multiples because they allow you to show data through many different lenses within the same viz – be that change over time, entity by entity, or in the case of these two examples, performance differences between players. I must confess, though, that every time I build a small multiple, I find myself going back to the same guide to remind myself of the calcs

– they just don't stick in my head! I tend to use [Kevin Flerlage's calcs](#) for this, but [Andy Kriebel](#) also has a very similar approach. With the NFL and NBA vizzes, I've combined small multiples with map layers to add in those extra details. Before I worked with Tableau, I used to do a lot of viz work in R's ggplot2 package, which is fantastic for layering element upon element.

It actually frustrated me initially that this was quite difficult to do in Tableau, but map layers are an absolute game-changer in terms of enabling that. It can seem like a bit of a strange concept at first – literally transforming your data into lat-lon coordinates to enable this – but it's so effective. Luke Stanke has [a brilliant guide](#), and explains this far more elegantly than I could, but for anyone unfamiliar with the concept, in a nutshell it involves scaling all of your data on both your x and y axes (let's say we're scaling from 0 to 1 in both cases), and then turning each point into coordinates using Tableau's MAKEPOINT() function. Then, because you know your data won't go above 1 or below 0, you're free to use the space outside of this to place additional information. Take my [NFL Throw Zones](#) as an example: if I know that my X and Y axes both start at 0, I could use an additional map layer to place a team name label at MAKEPOINT(-0.25, 0.00). And you can use as many layers as you like for this purpose (though be wary of cluttering your viz!). For the keys, I like to use them to provide additional information or context in a way that minimizes clutter elsewhere in the viz. In the NFL viz, the key is used to highlight two important pieces of information: the boundaries of, and average throw rates for, each throw zone. Highlighting them here removes the need to include that information within the small multiples.



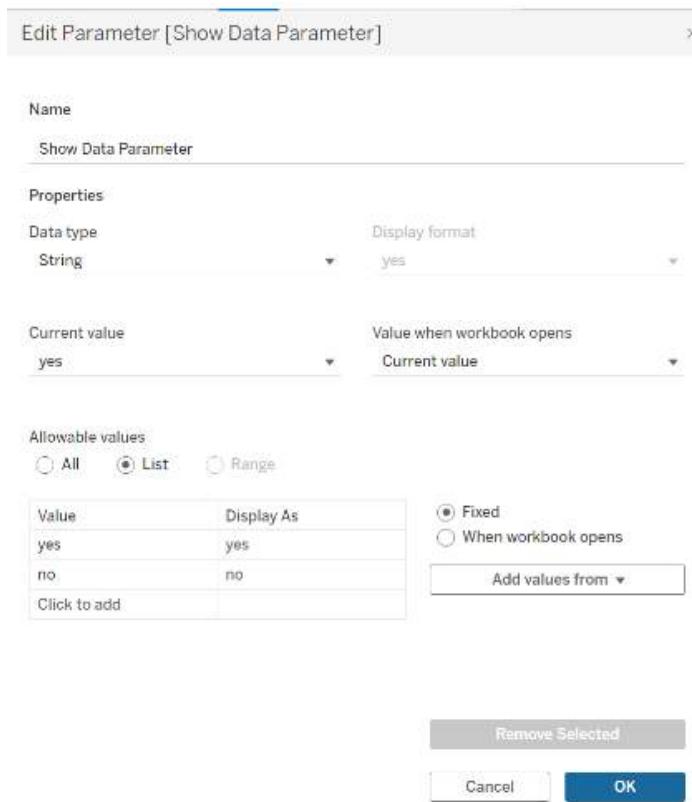
Similarly, in the NBA viz, the charts contain quite a lot of information, and adding labels within the small multiples would create a lot of clutter, but moving all of this information to the key means the user can still understand what's being displayed, while maintaining a clean look and feel overall. In a similar vein, I really like incorporating the colour key into other elements of the viz where possible. In the NFL viz, for example, I have incorporated it into the explanatory text at the top of the viz, which I think is both intuitive for the user and a great space-saver. **CJ: Your NFL viz and and your ashes viz share some similarities in terms of custom button and toggle design to change views on the page. Can you share some of the technical elements behind designing something like this using parameter actions?**

B: Parameter actions are probably my favourite feature in Tableau – they're the thing that turns a static viz into an interactive app. I use them in most of the vizzes I create, be it to highlight

particular data points, switch between metrics, or show/hide elements within the viz. However, from a UI perspective, parameters aren't the most aesthetically pleasing thing to add to a dashboard, so I tend to use a few tricks to make the design look a bit nicer. In my Ashes viz, I use parameter actions to switch between metrics in the 'Top Player Performances' section. As a default, Tableau will offer you a dropdown menu, but your options for formatting that menu would be pretty limited. So instead, I designed buttons for each metric in Figma, which allowed me to mirror the design style of the rest of the board. Then I open a new worksheet and assign these buttons as shapes to each metric. I add that worksheet to my dashboard, and set a parameter action to change my metric parameter each time a user clicks on one of the buttons. I've also created 'on-off' switch buttons in some of my more recent vizzes, using a very similar approach. In my NFL viz, I've created one of these to allow users to overlay each quarterback's individual throws over the zone map, should they wish to see them. They may look complicated at first glance, but they're actually quite simple: two circles that drive a parameter action, which both shows/hides the data, and changes the colour of the circles themselves to give the effect of them appearing/disappearing. I've created an example

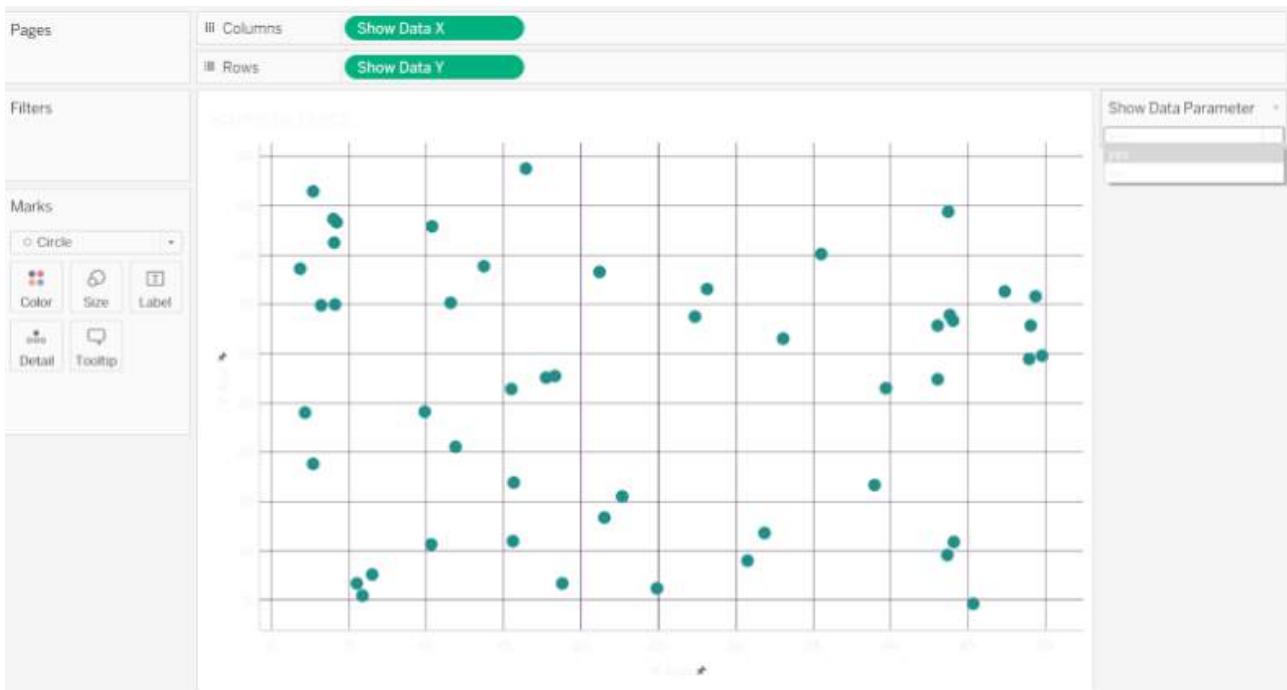
workbook for this [here](#), which uses the following steps:

1. I've created a dataset with some dummy and axis data, along with a column with 'Yes' and 'No' options. This column will be used to create my parameter and my control buttons
2. After bringing this data into Tableau, open a new worksheet (let's call it 'Data') and create a parameter (I've imaginatively named mine 'Show Data Parameter') from the dimension



3. Create two new calculations: , and . The idea here is that data should only appear if the is set to 'Yes', and the calculations to achieve this are as follows: Show Data X: if ='yes' then END Show Data Y: if ='yes' then END

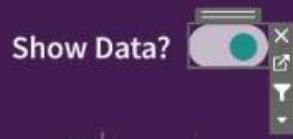
1. Drag the measure to the Columns shelf, and to the Rows shelf. You should now be able to toggle the to show and hide the data – we're halfway there!



2. Now to create our switch buttons: open a new worksheet (let's call it 'Parameter Switch'), drag the dimension to the Detail pill on the Marks shelf, and change the mark type to circle. You should see two circles appear (one for 'No', one for 'Yes'), which are going to be our buttons
3. Create a new calculation called , as follows: if = then else 'Other' END Drag to the Color pill and set your colours accordingly. Be sure to toggle the parameter to ensure you set colours for 'Yes', 'No' and 'Other'

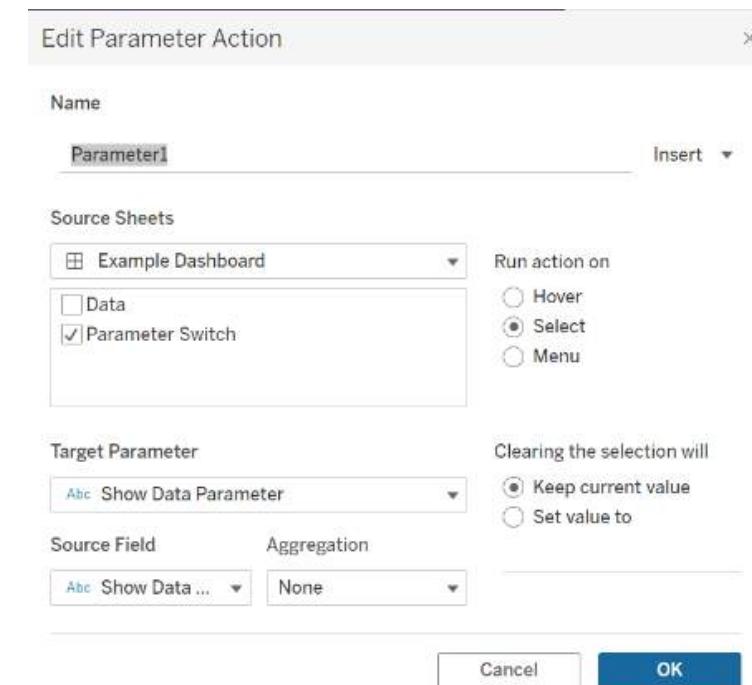
Open a new dashboard tab. For my board, I've created a simple background in Figma, which I bring into my board as a floating image, and position it to cover the full area of the board (I've used an 800x800 dashboard size in this case). I then drag and position my 'Data' worksheet to cover the bulk of the board. Finally, I drag my 'Parameter Switch' worksheet onto the board as a floating item. I remove the title, set the shading to 'None', and ensure the fit is set to 'Entire View'. You'll notice I have an oval shape next to the 'Show Data?' label on my background image – this is the background for the buttons, so I drag the 'Parameter Switch' sheet up, overlay it on that area, and resize accordingly. I ensure that within that worksheet, the colour for 'Other' matches that background oval so that the circle for the unselected parameter item 'disappears'.

Parameter Switch Button Example

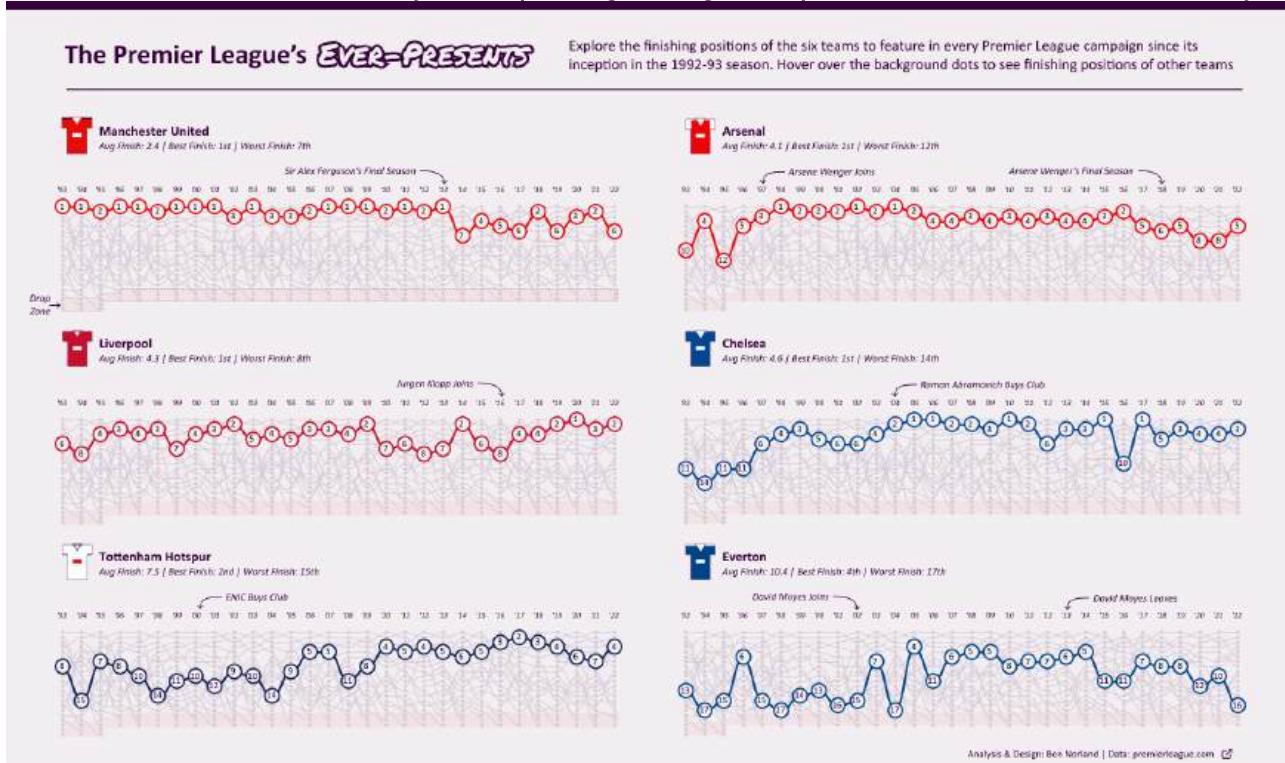


Go to the Dashboard drop-down menu and

select 'Actions'->'Add Action'->'Change Parameter'. Ensure the Source Sheet is set to your dashboard, that only 'Parameter Switch' is checked, and Run Action On is set to 'Select'. In the Target Parameter dropdown, select 'Show Data Parameter', and in the Source Field dropdown, select 'Show Data'. Click ok and go back to your dashboard. If you click on your 'Parameter Switch' buttons, you should now see the data appear/disappear accordingly



Finally, you'll notice I have and calculations dragged onto Detail on the 'Parameter Switch' sheet. This is to power another couple of worksheet/dashboard actions that ensure the buttons don't get highlighted when clicked on (which hampers the user experience). I learned this trick from a [Luke Stanke article](#) a few years ago, and have used it religiously ever since. With recent Tableau updates there may now be a quicker way to do this, but I've stuck with this extremely reliable method! **CJ: Your premier league visual 'ever presents' really caught my eye from its ability to showcase managerial and financial impact through labels without having to add huge amounts of written detail. How did you end up landing on using a table position rank chart to best show this story?**



B: In all honesty, this viz was probably more complex than it needed to be! I had seen Nir Smilga produce a great [background context viz](#), and wanted to do something similar. But I also decided to take the opportunity to try a curved bump chart. I had initially intended to show the season-by-season finishing positions of all teams to ever feature in the Premier League. But with the data densification required for both the context and the curved lines, it ended up being a huge dataset that wouldn't render when I published to Tableau Public! So I condensed it down to only the ever-present. It was only after the fact that I decided to add the manager/ownership labels to provide some additional context, but I think they worked out quite nicely. I like the way this chart shows progress (or, in some cases, decline) over time. From a personal perspective as a Spurs fan, it's interesting to highlight the progression the club has made during Daniel Levy's tenure, and the consistency we've achieved despite the criticism he gets. I was also pretty amazed to see that Manchester United never finished outside the top 3 under Fergie...! **CJ:**

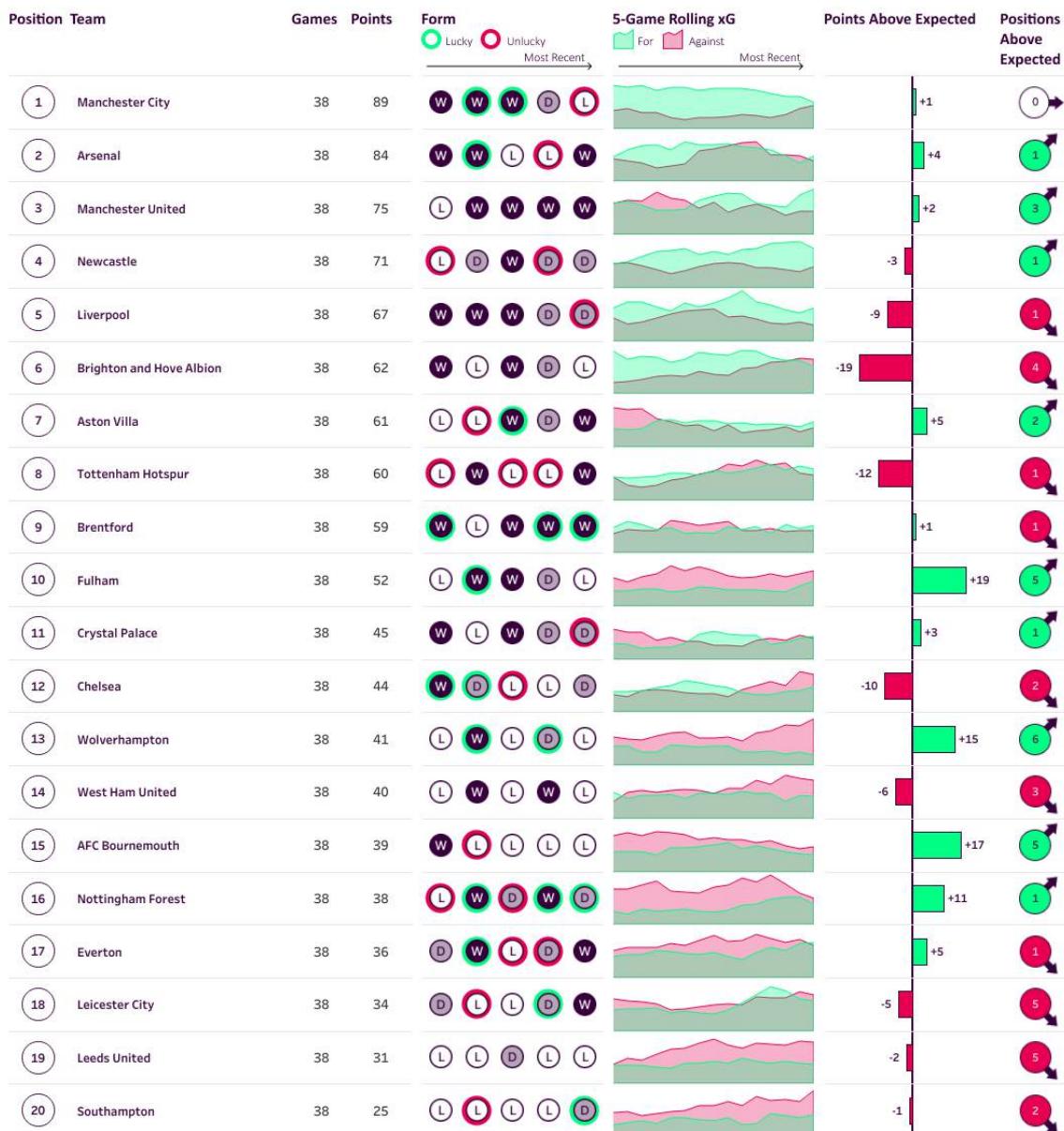
Sometimes I think people may skip over the technical expertise that goes into a visual creation when it is done so elegantly. I recognise some of the complexity within your player comparison dashboard with sheets such as the passing section having up to 9 layers! What thought goes into the preparation of a gauge bar with value overlays?



B: This viz was actually a re-make of one I did shortly after starting my Tableau journey, updated to make use of the worldfootballIR data source and also applying some of my more recent learnings. Once again, map layers are at the heart of this, and the gauge bar is actually a series of polygons. Just prior to this I had used Kevin Flerlage's excellent ["Yes Polygons!" tutorial](#) to build a viz looking at the most popular songs on Spotify from the past few decades, and so I had the idea to use that same approach to build the gauge. There are many ways to do most things in Tableau, but I found this approach solved some challenges around controlling the height/width of the gauge area (which would be harder to do with reference lines, for example), which in turn allowed me to more cleanly layer on other elements such as the metric name and absolute value, and the percentile dots. I think many experienced Tableau folks will know what I mean when I say it takes a little while to work out how Tableau "thinks". I feel like I'm getting better at that now, and can anticipate challenges that will be thrown up when trying to build particular chart types or work with particular types of data. That was definitely true with this viz – I was able to pull this together much more quickly than the original version, even though it contains more complex elements, because I could anticipate from the outset a lot of the challenges I would face, and how I would need to present the data to Tableau to make it all fit together properly. This is actually something I would really emphasise to anyone at an early stage in their Tableau journey: it can seem intimidating and frustrating when you see people producing really complex vizzes and you're struggling to even get it to display basic charts correctly (believe me, I've been there!), but those complex vizzes have been made possible by those authors going through the exact same journey you're on now. So stick with it, make use of the brilliant resources that are out there, and you'll be amazed at what you can achieve. And be sure to engage with the DataFam along the way – it is such a supportive and generous community. **CJ: Thank you so much for finding the time to dive into some of your SportsVizSunday vizzes with me Ben. Is there anything we should be keeping our eyes peeled for in the future?**

Premier League Table | Actual vs. Expected

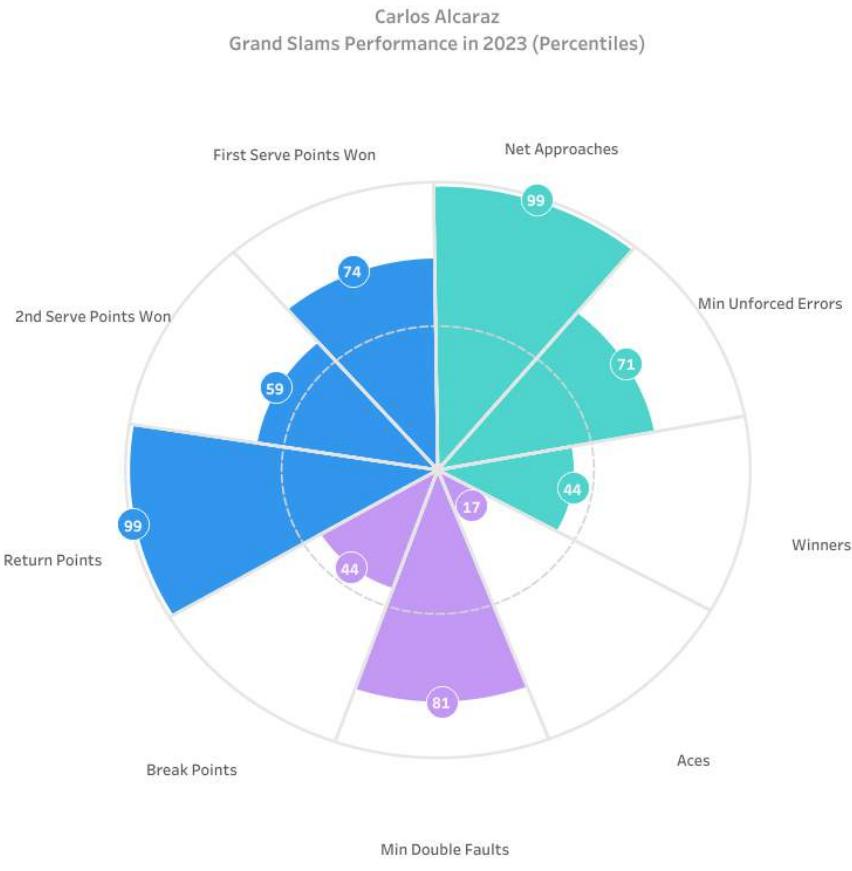
Analysis & Design: Ben Norland | Data: [Fivethirtyeight](#)



B: Not at all – thank you for inviting me! In terms of upcoming pieces, I'm back to work at the end of August, so I expect I'll be publishing on Tableau Public a bit less frequently (I think I've averaged a viz every 2 weeks over the past few months), but I will definitely still be publishing regularly. I don't tend to plan vizzes too far in advance – normally it's a case of being inspired by the work of others, or being struck by a particular event (usually sports-related) that prompts a new idea. But I am planning to build another version of my [PL Actual vs. Expected](#) table once we're a bit further into the new season. And I've just finished working on a piece at the moment looking at transfer flows between the big 5 leagues. [Check it out!](#) **LOGGING OFF, CJ**

RADIAL / POLAR CHARTS IN TABLEAU

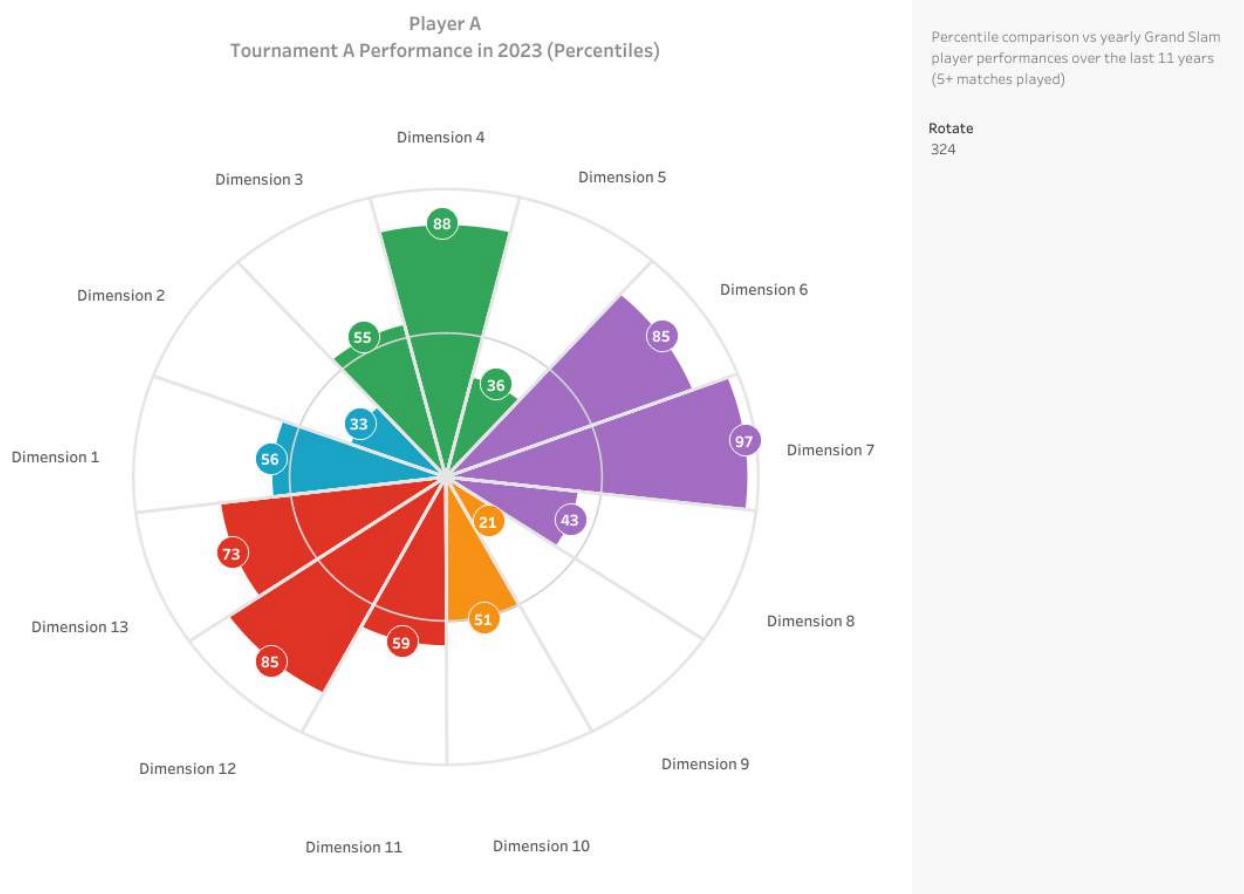
Hi all, This week we will look to create a polar chart in Tableau from scratch with all the calculations explained. This will be built in a way that it works as a template, so if you don't particularly care for how the template works under the hood but would like to just replace it with your own data, by all means just download the data template from github and download the workbook from Tableau Public.



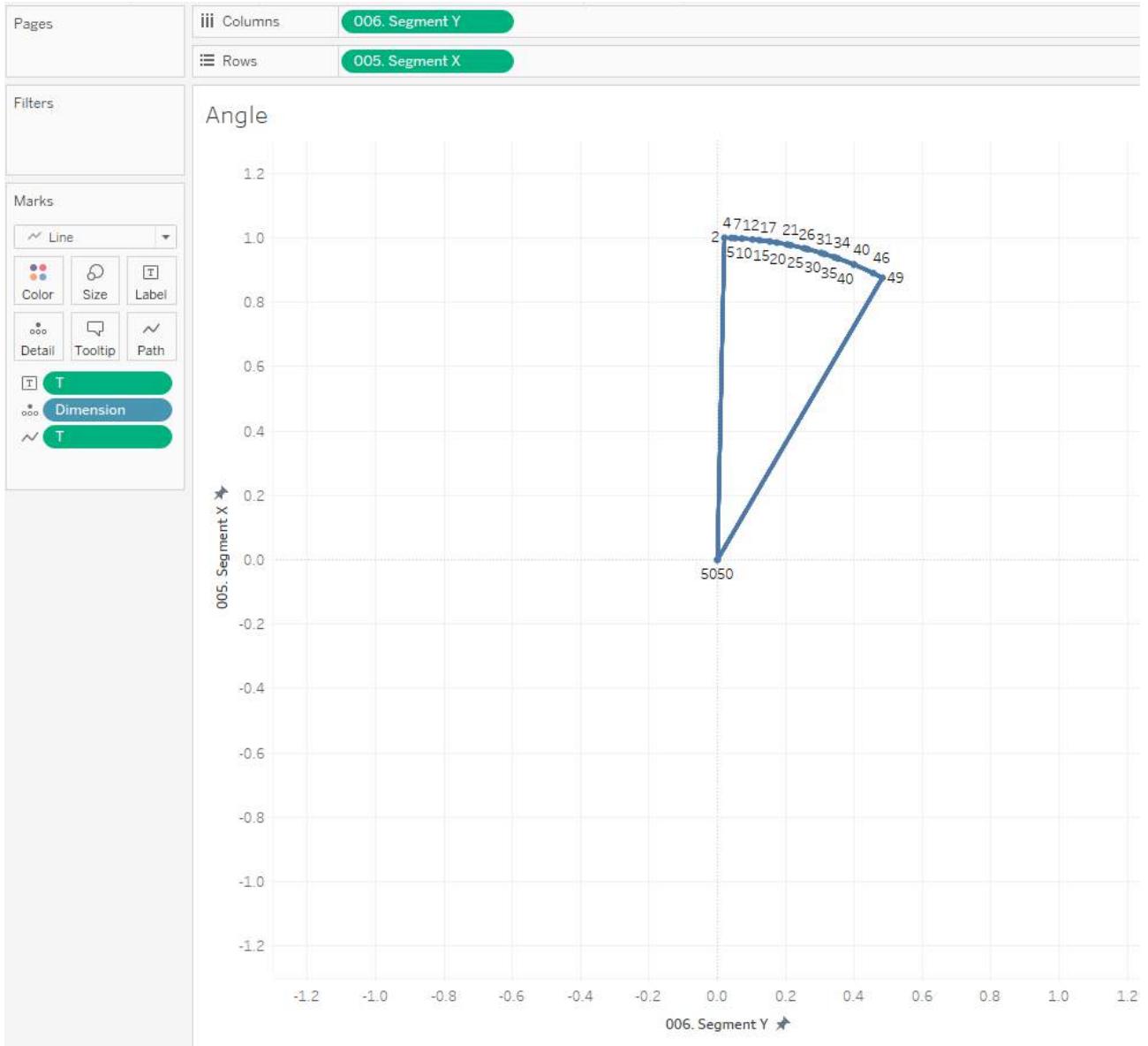
Percentile comparison vs yearly Grand Slam player performances over the last 11 years (5+ matches played)

Rotate
167.5

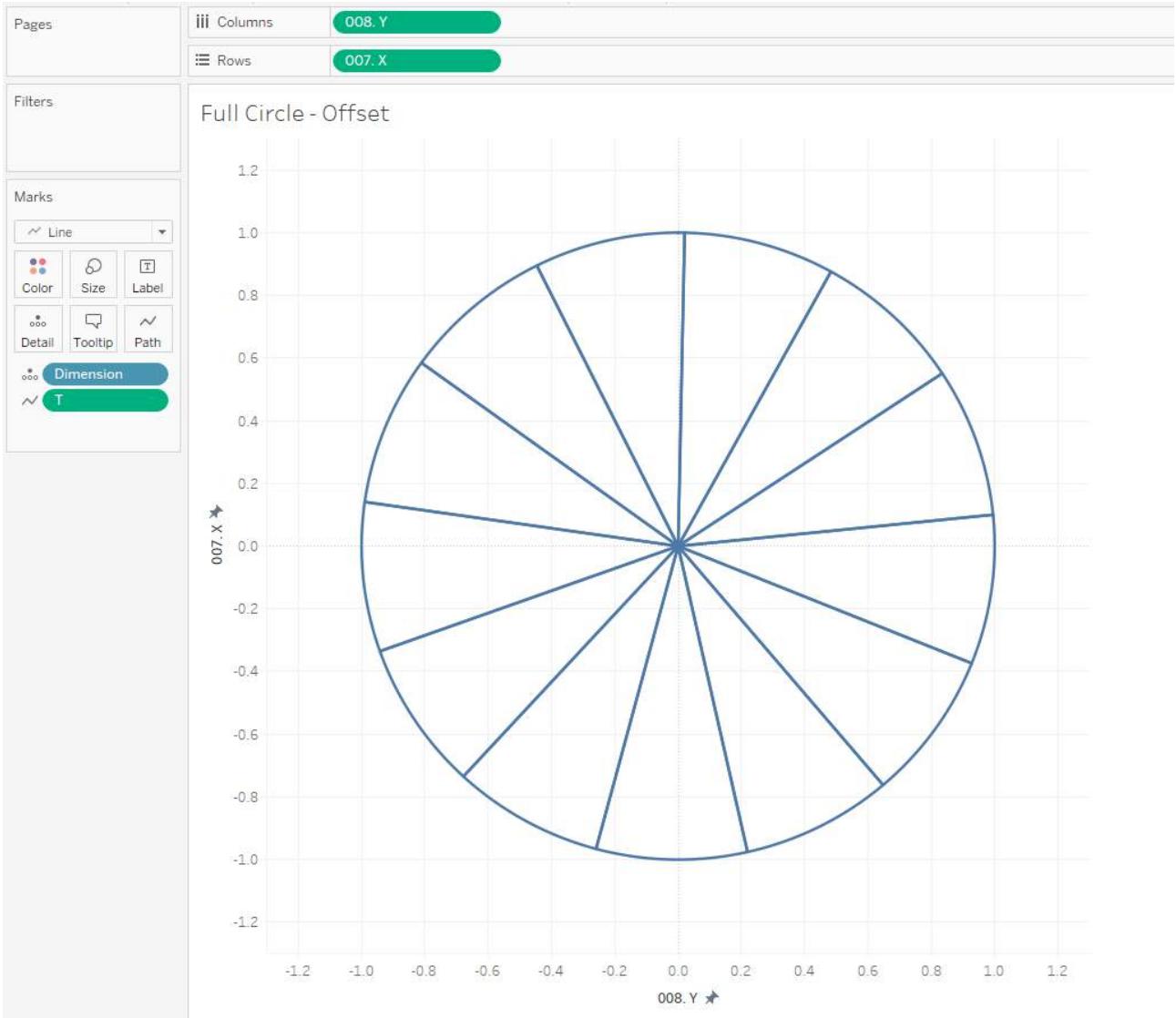
Here is the chart we are going to create. I've previously written about how to build charts very similar to this. You can read them [here](#), and [here](#). Both still work, but this goes to that next level with layers of adding labels and figures, and removes the need for a mapped background. I would say the other two templates I've created are intermediary and this one is advanced. If you want to use the template, simply download it from the top of the page, replace the data source with your own. Just remember to make your dimensions unique names and it should update. See below for where I've updated the template with 13 different metrics. **All metrics must be a value out of 100 for this to work.**



So.... now for those that want to understand the calculations. 001. Max Dimensions { Fixed : countD()} Find the maximum number of records, i.e dimensions we will want to plot. In the carlos viz this would mean it is 9, in the template this figure is 13. The number of slices. 001. Max T /*Find the maximum number of records, i.e dimensions we will want to plot. The example is 50 (you'll see the T values in the excel) - this doesn't need to be changed! We minus 2 because we use T to start and end at the origin... you'll see this in later calcs, and we take off the extra one so it aligns to the centre */ 002. Rank I just like to keep all my radial calculations the same terminology, you don't necessarily need this calculation. But we will rank all our T values, so we can start to move them to draw up that nice curved edge. 003. Angle 360/ /* The angle between each dot for data densification, 360/50 as the T values only go up to 50. */ /* Space the dots out. There should be equal space between each dot Divide by the number of segments we have. An alternative way to think about this is a full circle with 50 dots in it.... then saying woops we actually want 8 segments of that, so we squeeze all those 50 points in together and repeat it 8 times. Hopefully makes sense so far? 005. Segment X if = 1 then 0 elseif T = 50 then 0 else cos(radians()) END We are building each segment separately, so our start point is our origin, as is our end point. Anything in between use trigonometry to make it a circle. 006. Segment Y if = 1 then 0 elseif T = 50 then 0 else sin(radians()) end Do the same but use sin.... basic trigonometry. If maths isn't your jam just take it as gospel. If we then plot these values against one another.



So what this leaves us with is the centre point (where T is 1 and 50) and all the other values evenly spaced between T 2 up to T49. This however is the same segment 13 times over. What we now want is to rotate each slice around the point (0,0) dependent on whatever metric it is. 007. X The calculation for rotating around an arbitrary point is $x_1 = x_0\cos(\theta) - y_0\sin(\theta)$ So applied to our data that calculation is $* (\cos(\text{radians}(360/*))) - * \sin(\text{radians}(360/*))$ & for Y the calculation is: 008. Y $y_1 = x_0\sin(\theta) + y_0\cos(\theta)$ Which means applied to our Tableau calculation that is $* (\sin(\text{radians}(360/*))) + * \cos(\text{radians}(360/*))$



Plotting these new X and Y calculations against one another we get this wheel / Pizza type shape. See how it is the same segment just repeated 13 times. The next thing I noticed was that dependent on the number of slices my chart sometimes has a vertical line sometimes it doesn't so I wanted to add in the ability to rotate the whole wheel. I.e just spin it around up to 360 degrees around the centre point. What we need is a parameter that is a value between 0 and 360. We can call this P. Rotate.

Edit Parameter [P. Rotate]

X

Name

P. Rotate

Properties

Data type

Float

Display format

46.5

Current value

46.5

Value when workbook opens

Current value

Allowable values

All List Range

Range of values

Minimum

1

Fixed

When workbook opens

Maximum

360

Add values from ▾

Step size

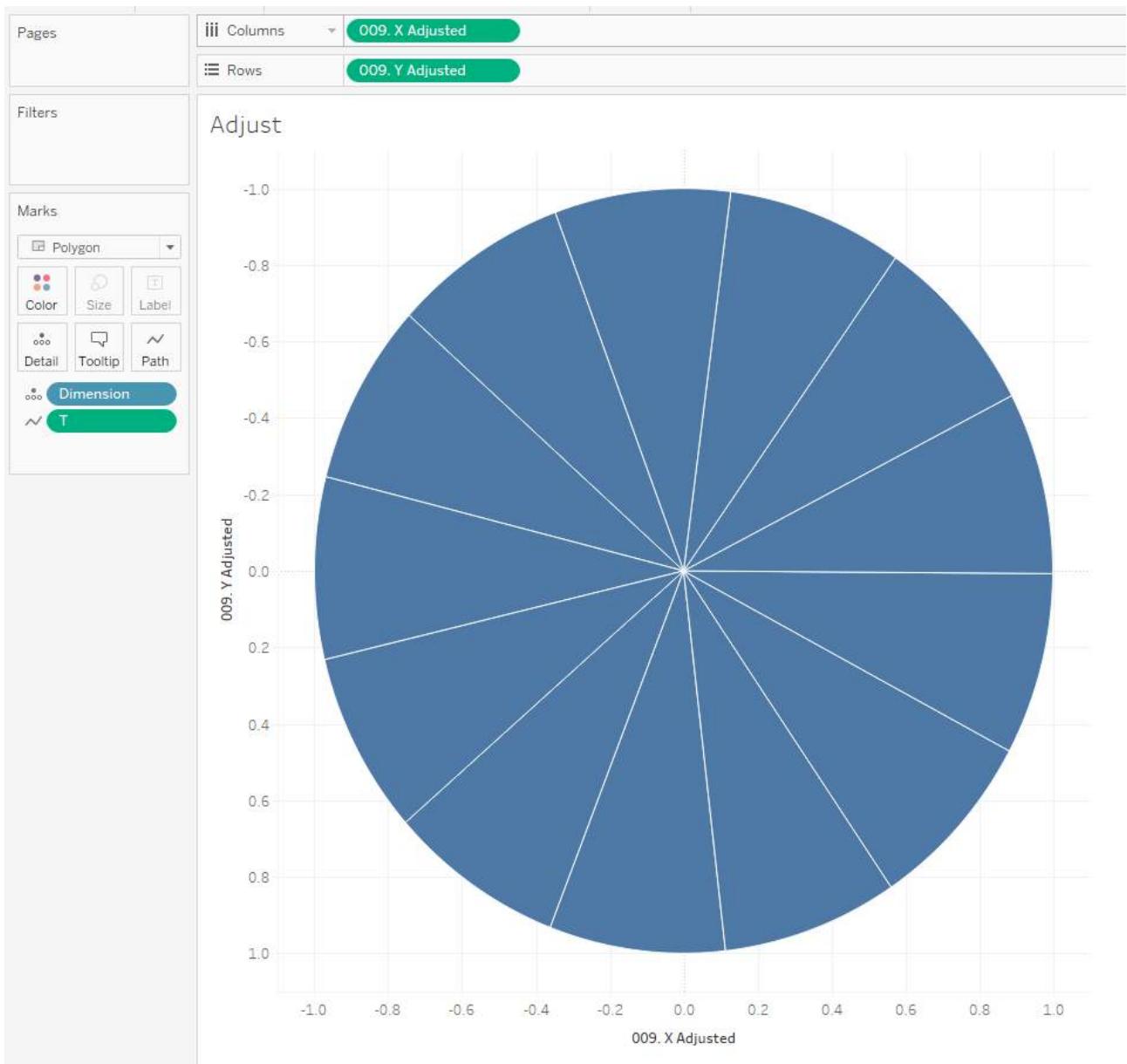
0.5

Cancel

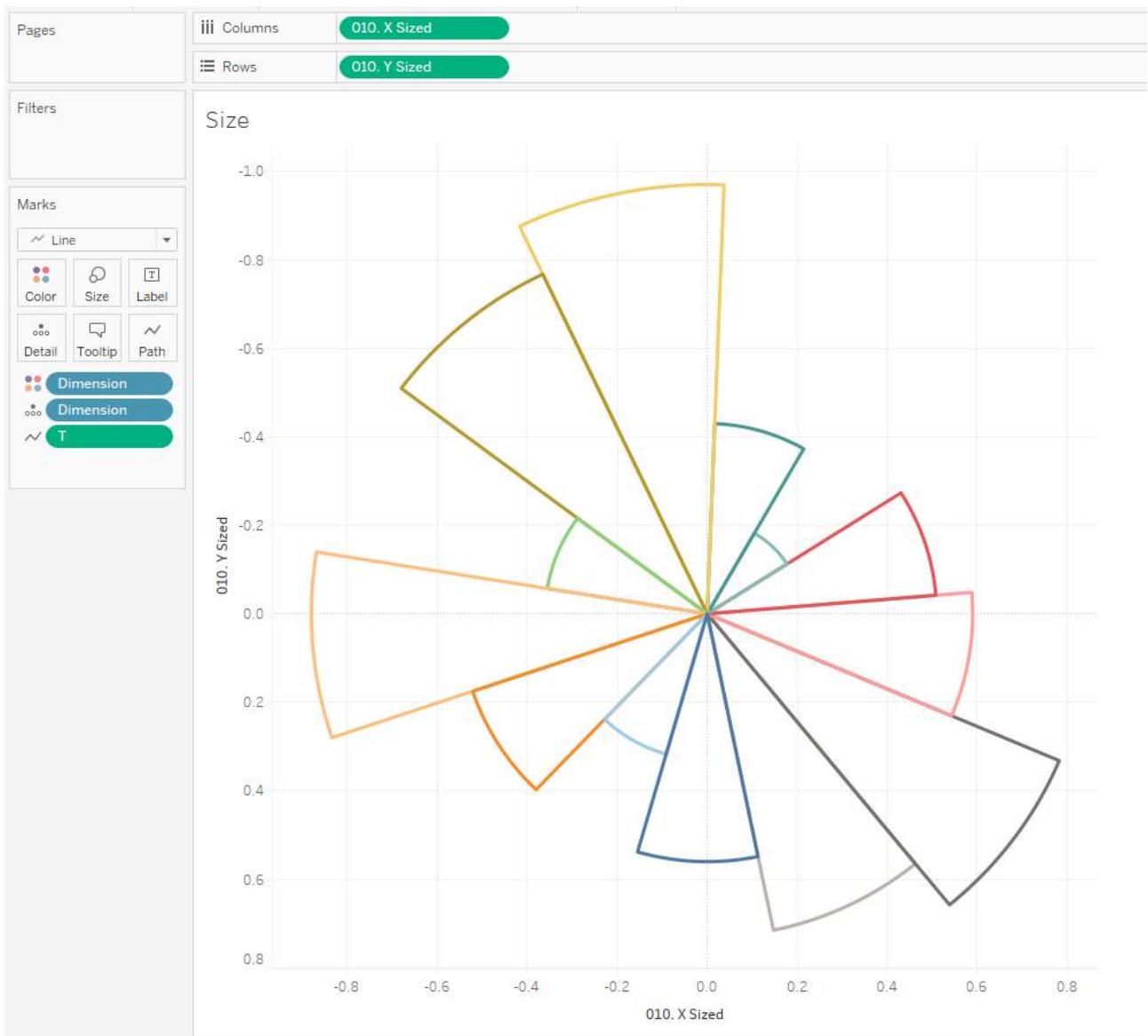
OK

To adjust our chart, we theoretically

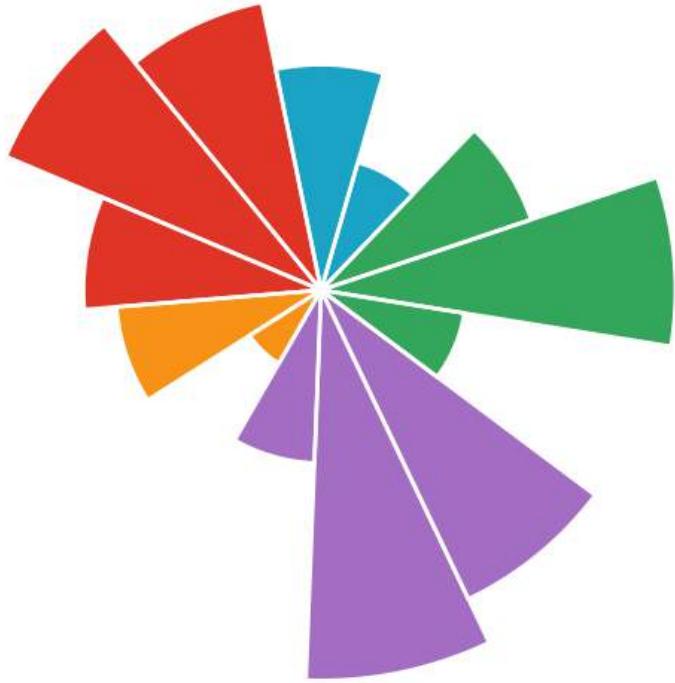
use the same x & Y calculation, but we build them off the previous so our new X is, 009. X Adjusted *cos(RADIANS()) – * sin(radians()) and our new Y calc is 009. Y Adjusted * (sin(radians())) + * cos(radians()) Now if you plot these two values against one another and show the parameter you will see how it spins when you change the degrees.



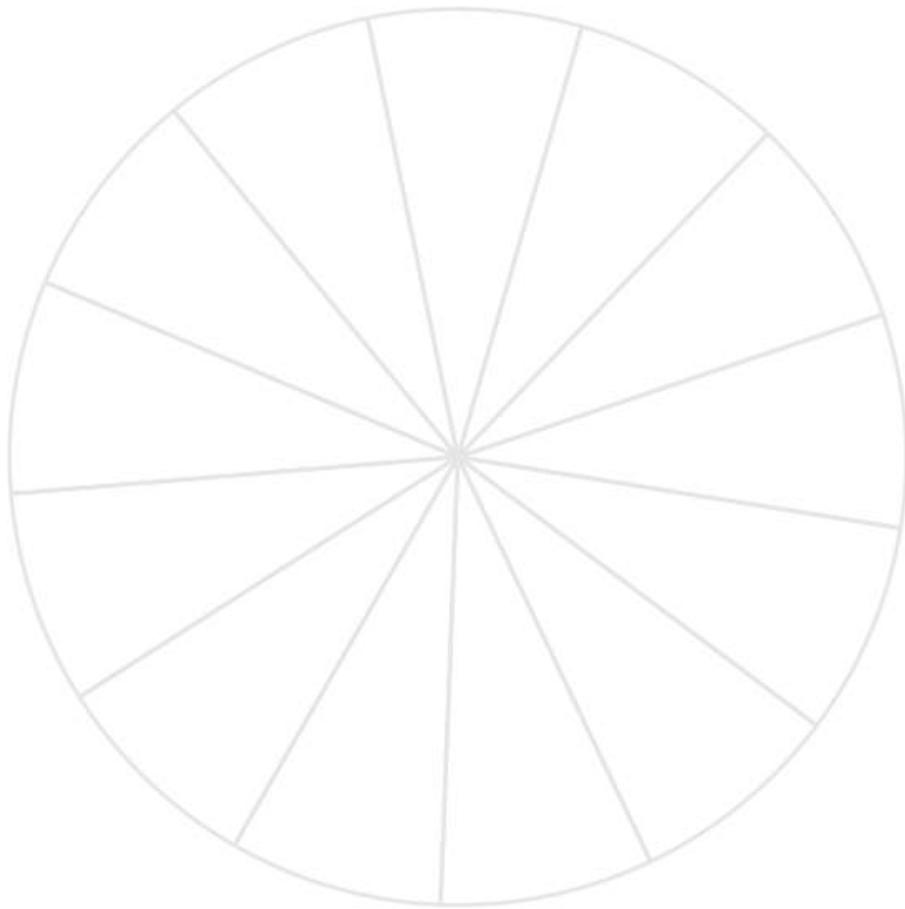
The next thing we need to consider is that when we use trigonometry to create these segment / circle that the radius equals 1. See how the circle goes between -1 and 1 in both x and y values. So now we need to scale these based on our different segments ratings out of 100! 010. X Sized * (/100) This therefore just makes it proportionally a value between 0 and 1. 010. Y Sized Same with Y * (/100)



Perfect, something that now is starting to represent our final product. So now you could leave it like this and flick the marks between either line or polygon but why not add some extra details? Let me talk through the different layers in the template. MP.
Poly MAKEPOINT(-,)

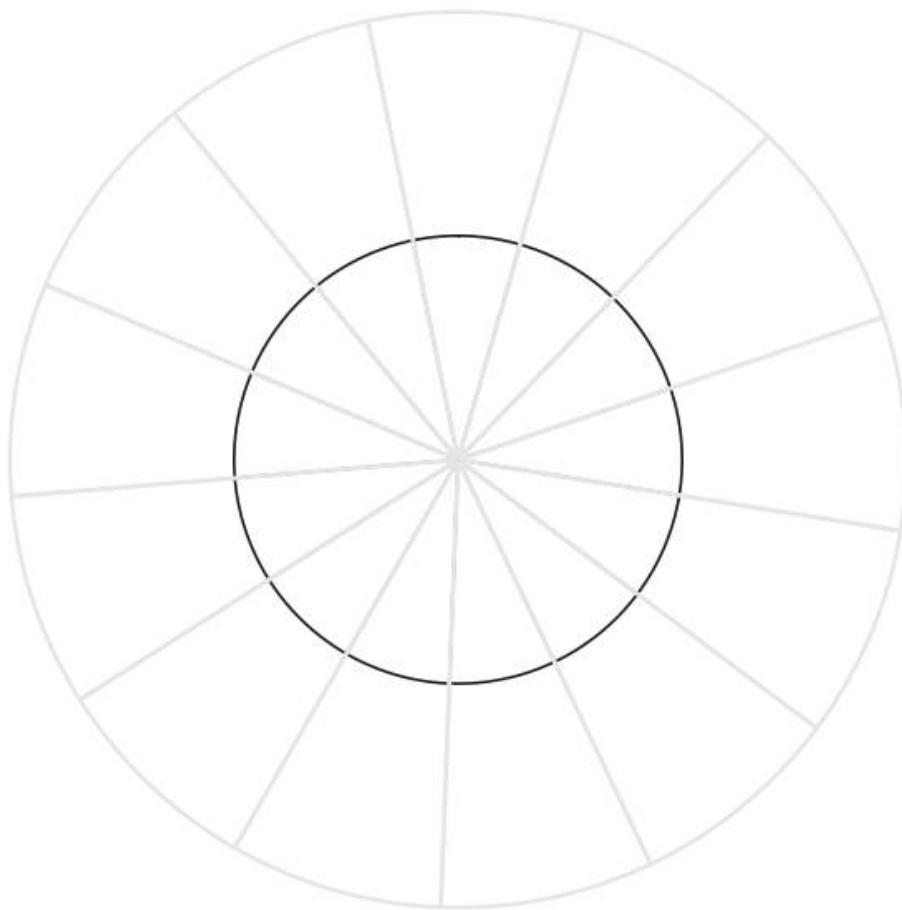


Plot our Y and X co ordinates, I put minus in front of Y so that we have the metrics clockwise when on the visual. Do remember if you do this for one, you may need to follow the same logic in your other calculations. MP Base MAKEPOINT(-) This is a line outline to help shape our visual by plotting where the maximum values could be i.e the full circle up to 100 ratings.

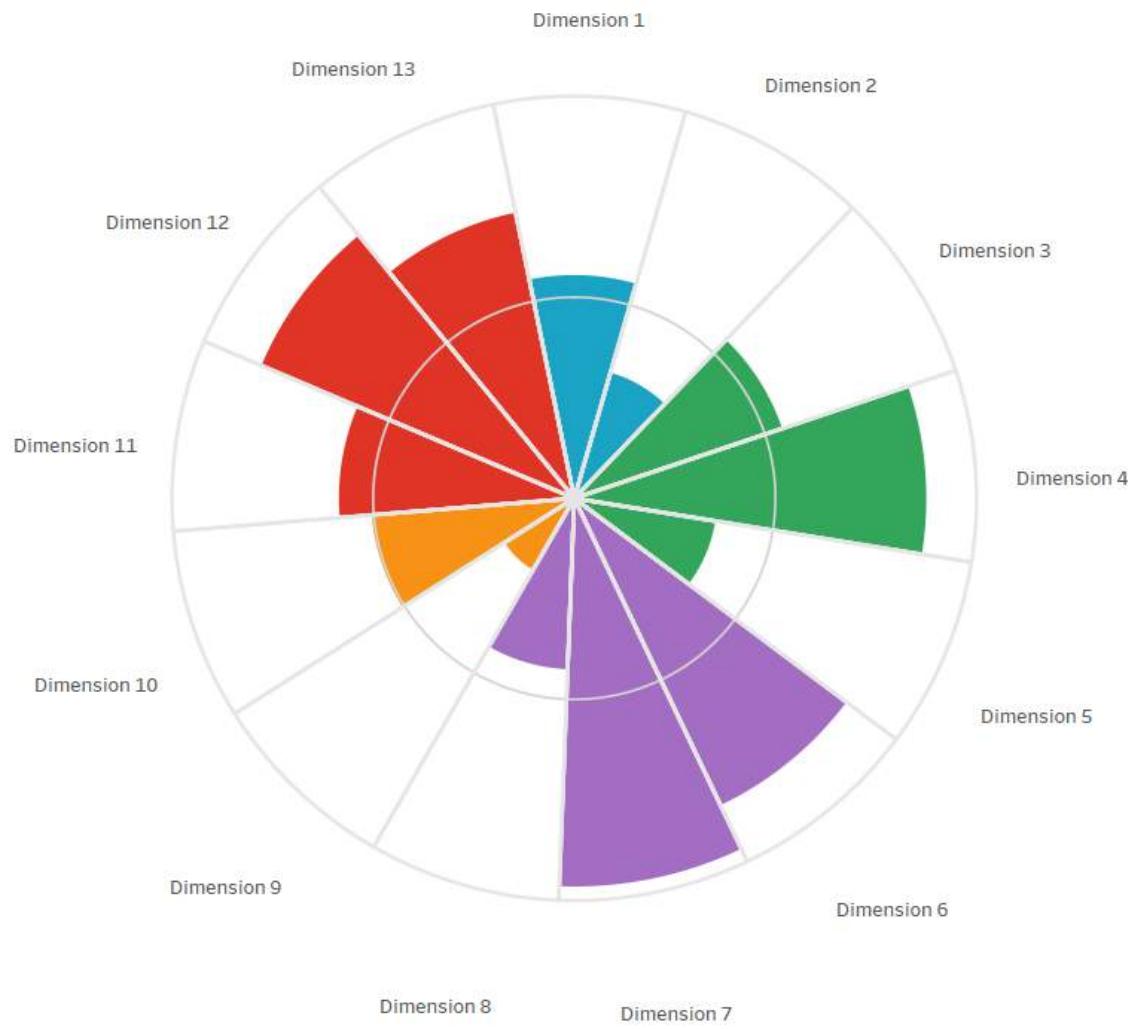


MP

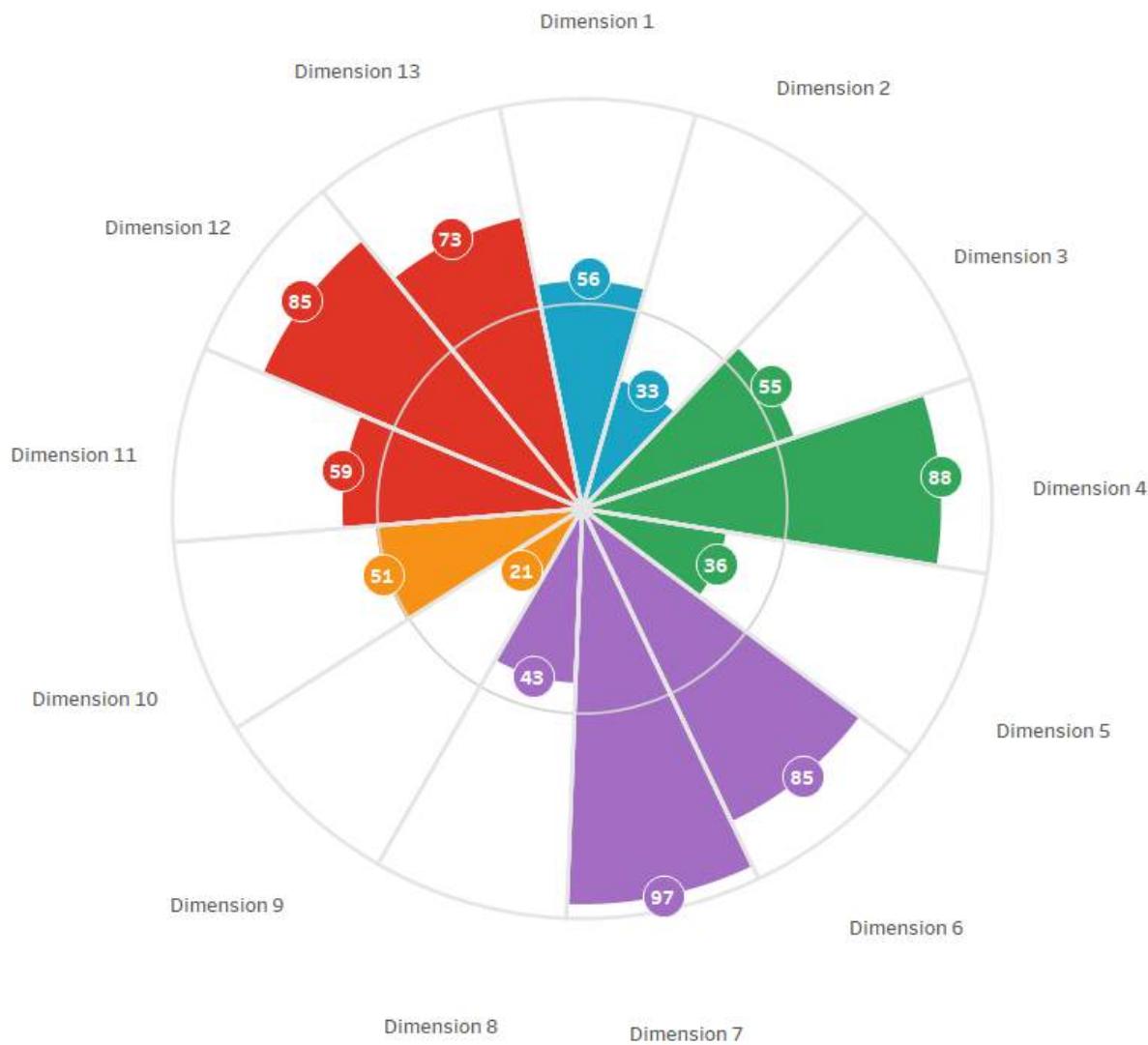
Base 50 MAKEPOINT(-/100*50,/100*50) Remember our full circle of slices? Here I am saying create one half the size. I bring this to the page to show the 50 rating circle and then make this a dashed line. In the screenshot below it is the black circle.



MP. 0 makepoint(0,0) This is just a centre point, It is not required, but maybe if you are looking at a specific player you may want to add a shape layer with the persons face in? MP. Label if = 25 then MAKEPOINT(-*1.25,*1.25) END Where T is 25, that is roughly the mid point of each segment. the *1.25 moves the label outside of the circle.



Finally lets use the same logic but add in those values onto the chart MP. Value if = 25 then MAKEPOINT(-,) END



There we have it. Lots more customisation you could bring into this but I hope it gives you more flexibility when it comes to radial charts. Do ping me if you get stuck, I'm always happy to help. **LOGGING OFF, CJ**

BASEBALL SPRAY CHARTS IN TABLEAU (STATCAST)

Hi all, How are you? These past few months have been really good for the site content wise. To overshare a little we passed our 2020 views count for the year back in June. In my head the best part of this accomplishment is the idea my life feels more in equilibrium. It's been a good few months, moving house, celebrating best friends birthdays, milestones etc. Anyway, todays blog a little different in the sense that I've never done any content on baseball before. Mainly because it isn't as big here as in the US, but nonetheless I have always admired the idea of mapping where shots / runs are created. In fact I first came across this idea years ago, by Jacob Olsufka who produced this awesome hex map of hits.



PICK A BATT.. Robin..

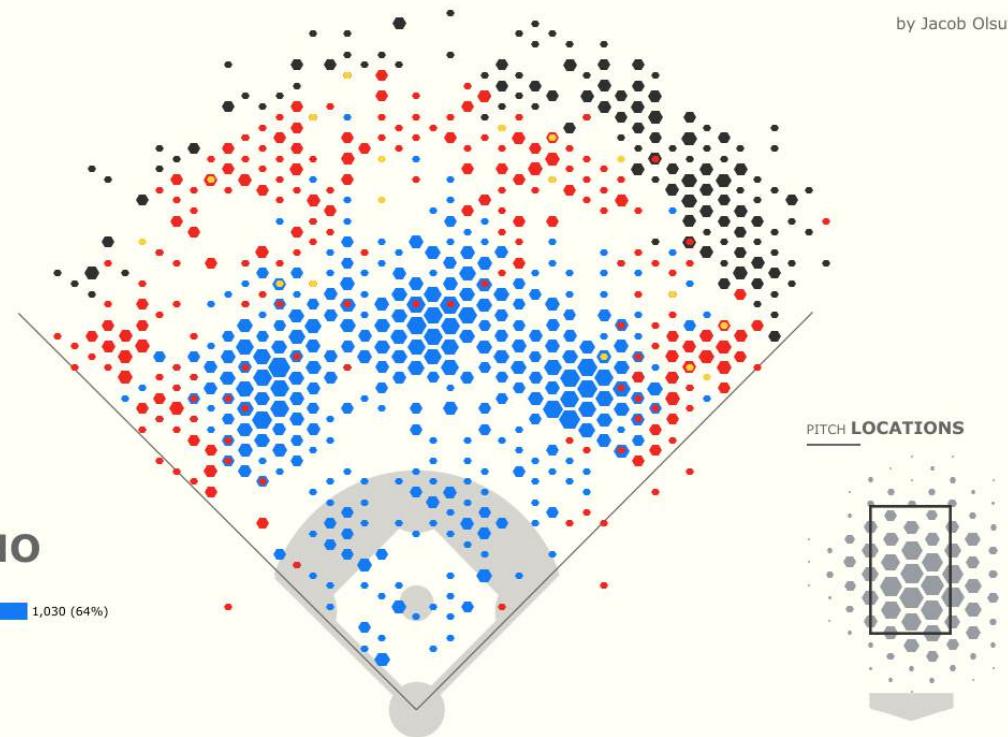
CLICK TO HIGHLIGHT
Yellow Triple
Red Double
Blue Single
Black Home Run

NUMBER OF HITS
5
10
15
1
20



ROBINSON CANO

Single	1,030 (64%)
Double	340 (21%)
Home Run	231 (14%)
Triple	17 (1%)

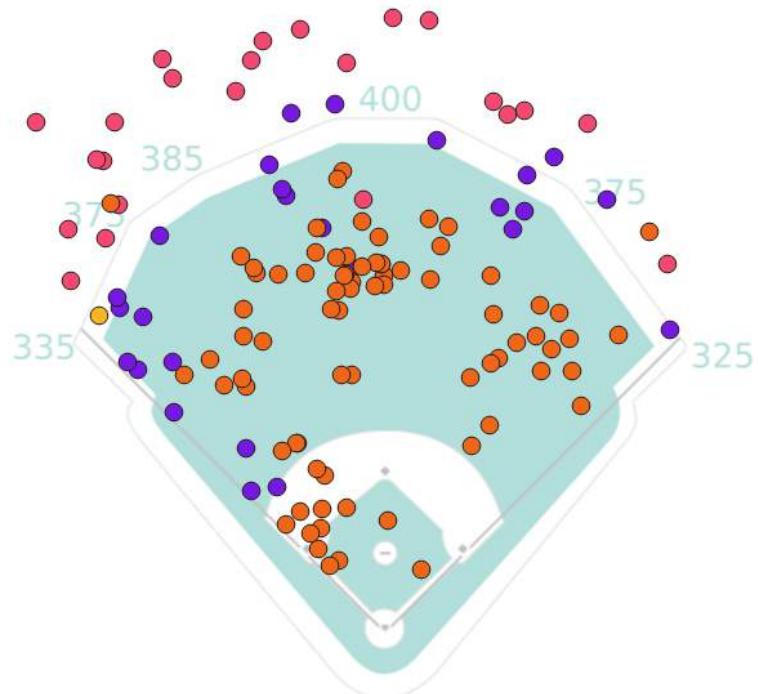


Data from the inning_hit.xml and inning_all.xml files of MLB Gameday data. Images from ESPN

We will look to strip it back a little and rather than aggregate shots into hex's just plot the original shots for a player for a season.

Acuña Jr., Ronald

Spray Chart

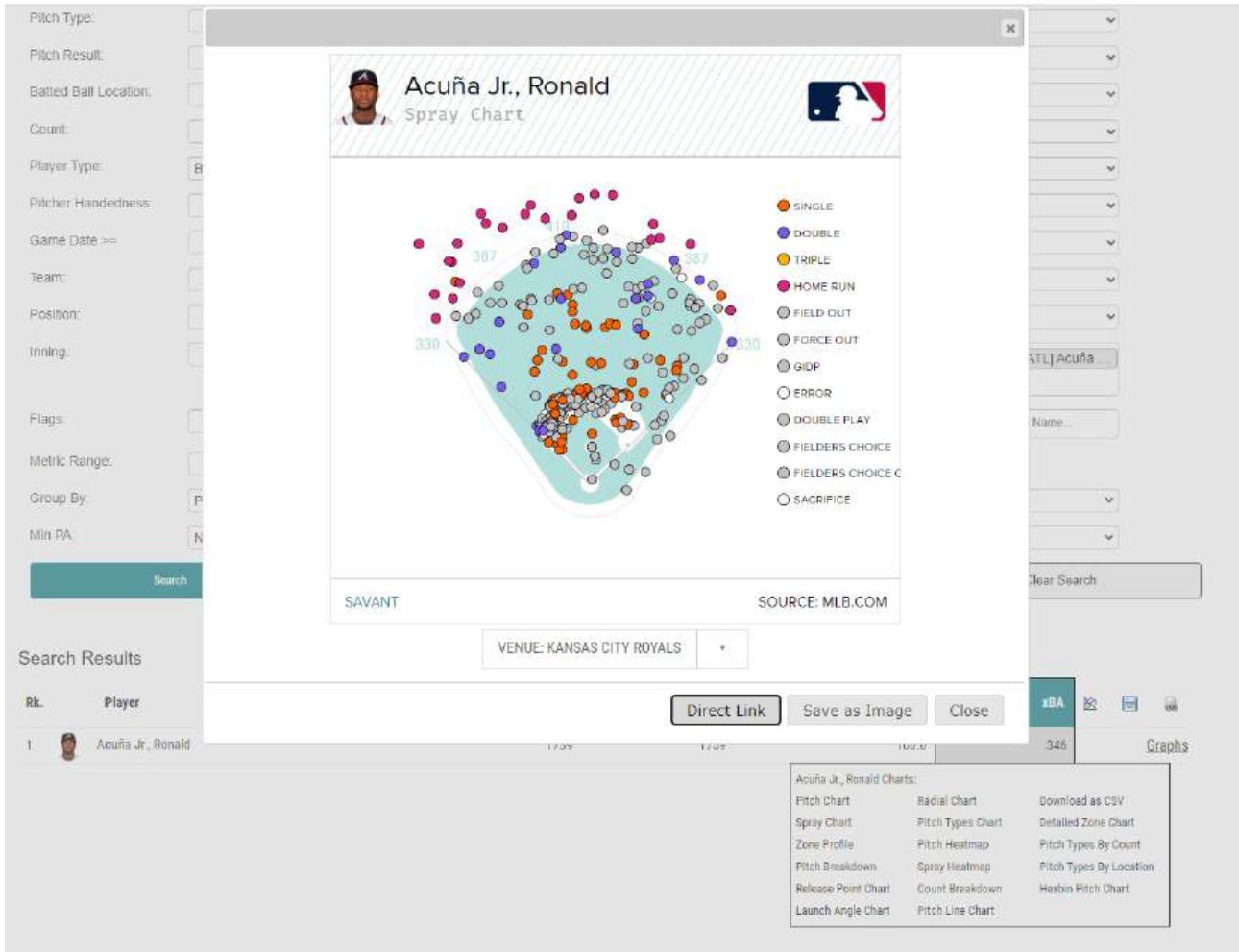


You will see the chart we will look to recreate is very similar to that seen on [Statcast](#)

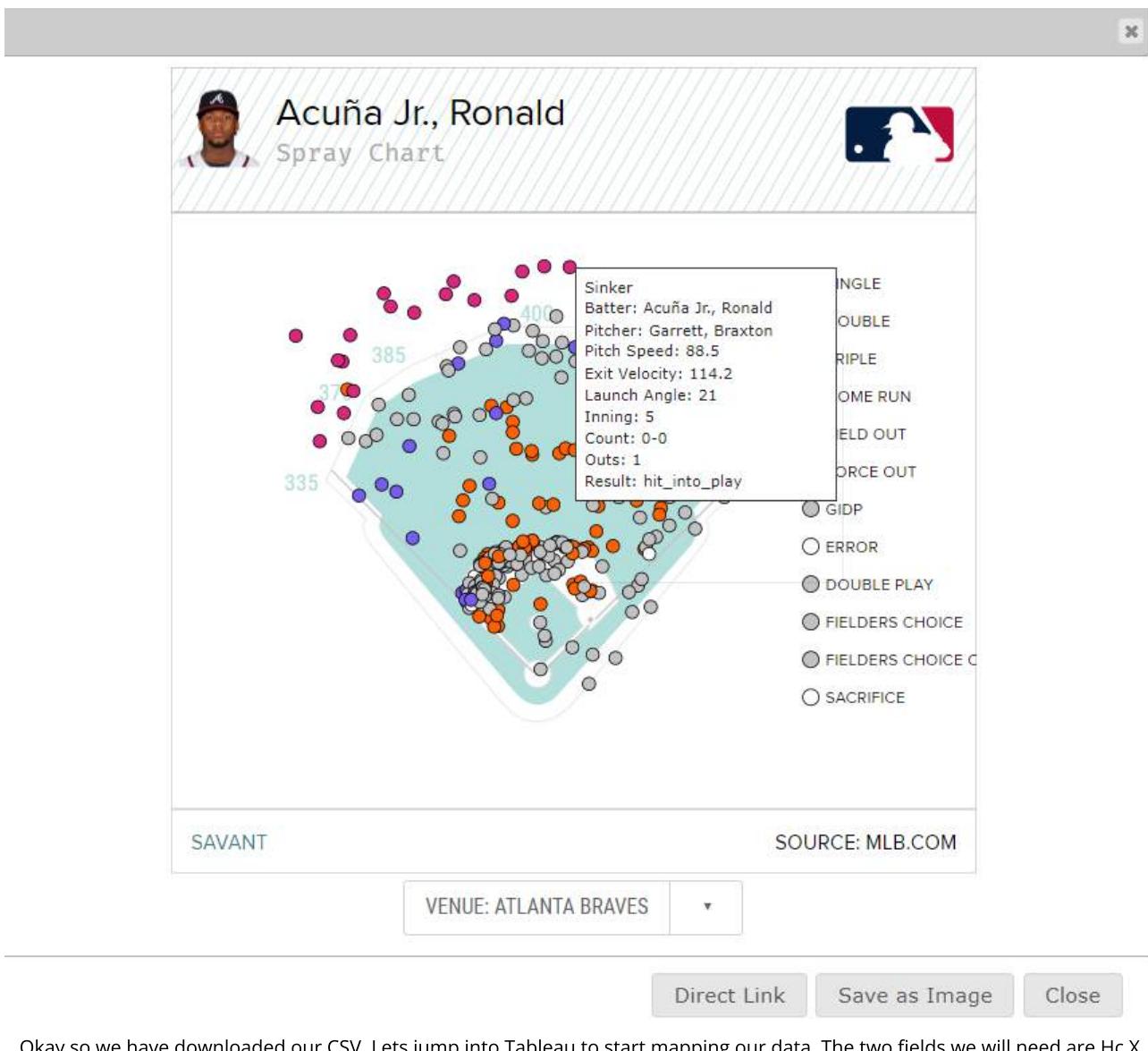
Search Results

Rk.	Player	Pitches	Total	Pitch %	xBA	Graphs
1	Acuña Jr., Ronald	1759	1759	100.0	.346	
Acuña Jr., Ronald Charts:						
	Pitch Chart	Radial Chart	Download as CSV			
	Spray Chart	Pitch Types Chart	Detailed Zone Chart			
	Zone Profile	Pitch Heatmap	Pitch Types By Count			
	Pitch Breakdown	Spray Heatmap	Pitch Types By Location			
	Release Point Chart	Count Breakdown	Hexbin Pitch Chart			
	Launch Angle Chart	Pitch Line Chart				

The baseball savant website allows you not only to visually see the spray chart for the selected filters you've applied but also download the csv file. This is what we will do to retrieve our data for this tutorial. I've posted the example data on the Github repo. Here is how the chart renders on the site.



It's worth noting that this chart is interactive in the sense that it offers tooltips as well as the option to change the venue background chart, see how we can switch it to the Atlanta Braves venue. We will revisit this when we build our chart.



Okay so we have downloaded our CSV. Lets jump into Tableau to start mapping our data. The two fields we will need are Hc X

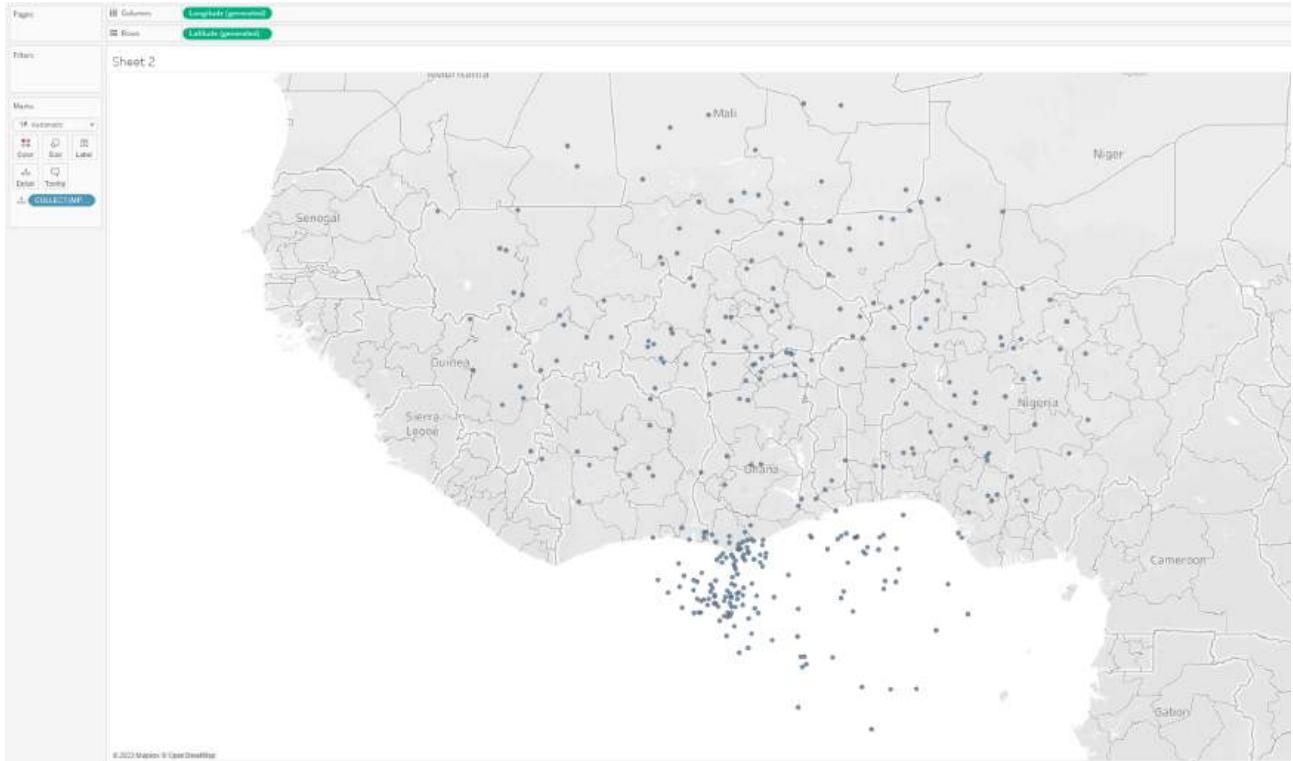
Search ▼

Folders

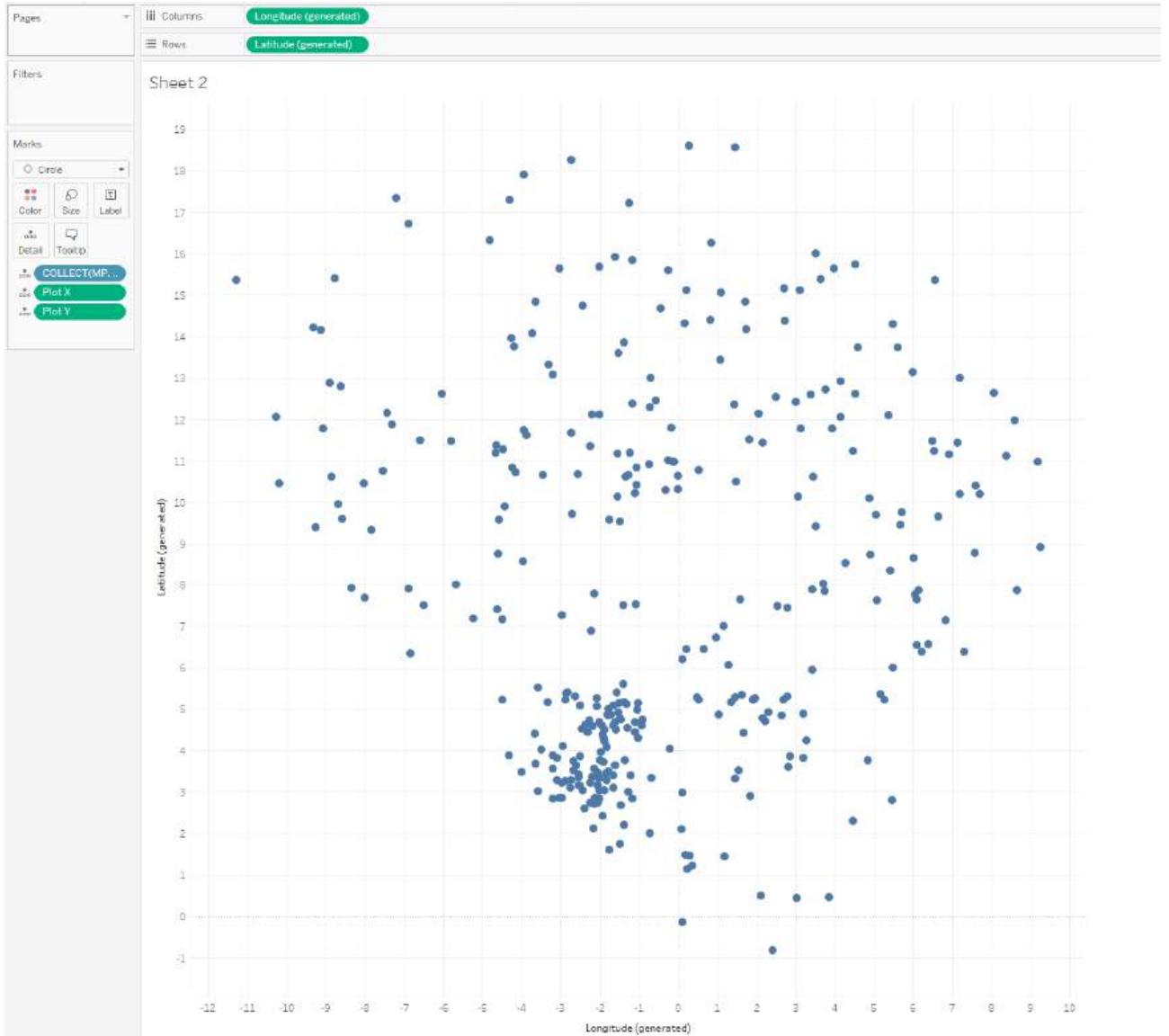
- > Data
- Events (group)
- Measure Names
- # **Hc X**
- # **Hc Y**
- =# MP. Hits
- =# Plot X
- =# Plot Y
- # 660670_data.csv (Count)
- # Latitude (generated)
- # Longitude (generated)
- # Measure Values

and Hc Y these are the positional points of where the shot has been hit.

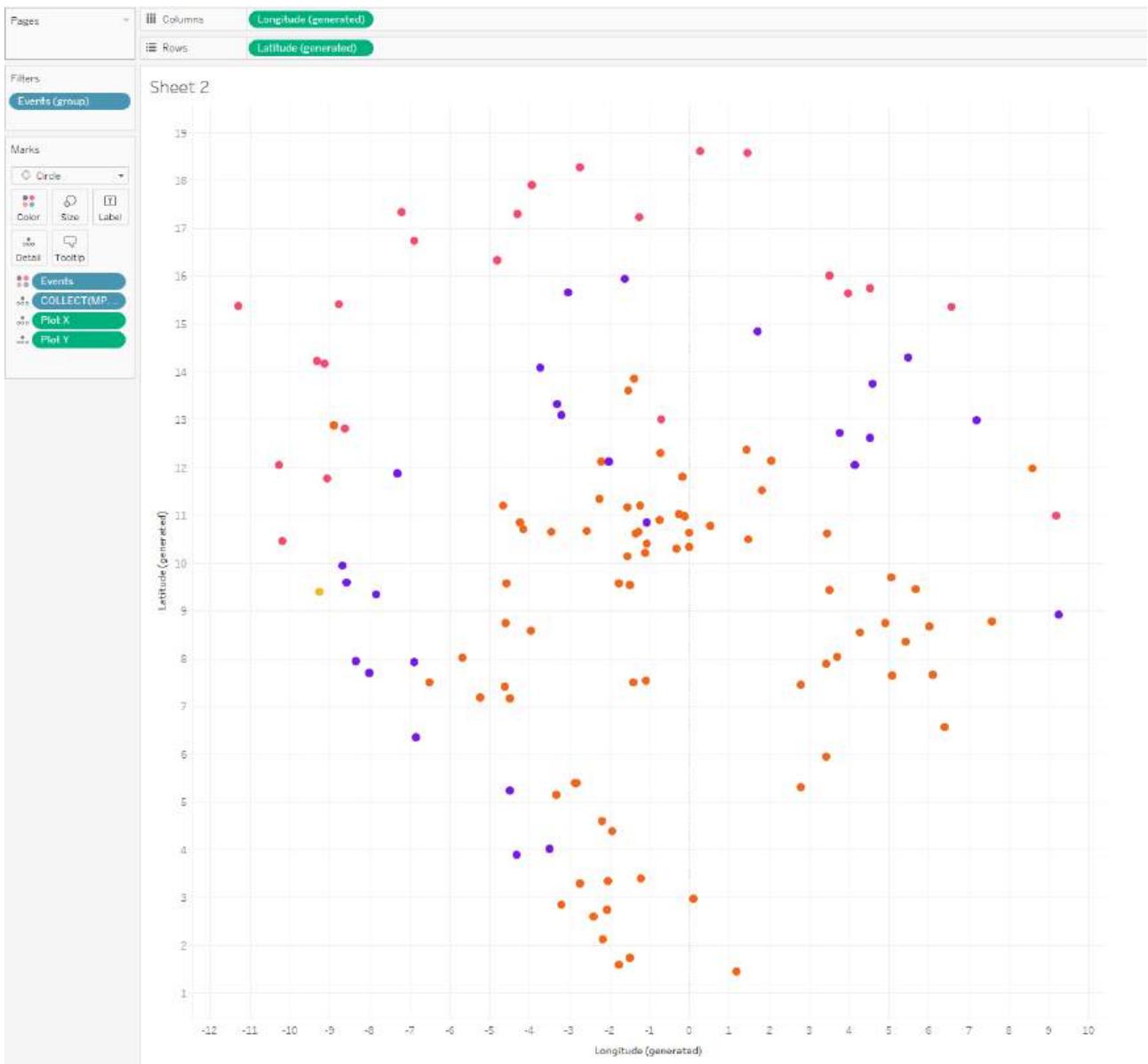
Lets make two transformations. Plot X ($- 125.42 / 10$) and, Plot Y ($198.27 - / 10$) These are calculations to make sure our data is in the right shape, dividing by 10 isn't necessary per se, but because I wanted to use the makepoint function incase I required more layers at some point its better to have these scaled down between geographic points. Off the back of those calculations we can create MP. Hits MAKEPOINT(,) Double click the makepoint calculation.



Turn the map off, Map > Background Maps > off Drag Plot X and Plot Y onto detail (make them dimensions) and change the marks to a circle.



This is the vague structure of the baseball map that we need. We could drag events onto colour to bring the visual to life more and cross reference some data points with whats seen on the site. What I will do is filter out anything that isn't a single, double, triple or home run.



The final thing we will want to do is map onto the background an image of the pitch. Now this is the tricky bit, mainly because.... all stadium layouts are slightly different. So it is a case of what to choose. Now by all means you could go find a standardised one similar to Jacobs at the start of the blog, but for today lets find a method to take it from the site directly.

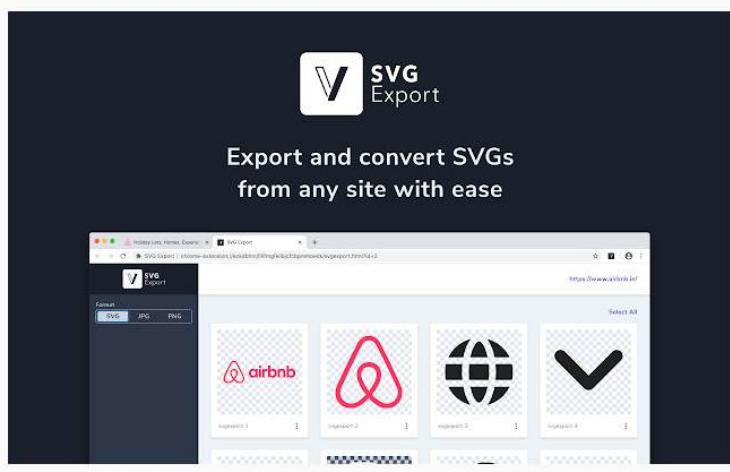
SVG Export

svgexport.io  Featured

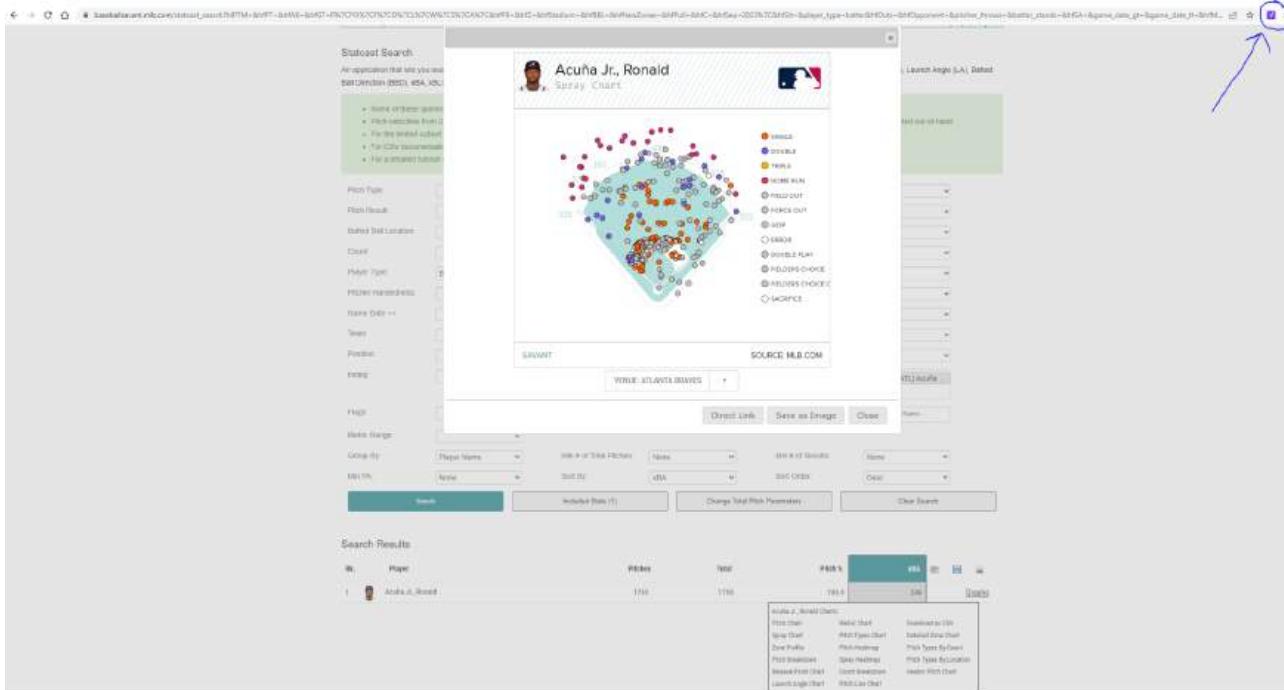
★★★★★ 135  | Productivity | 100,000+ users

[Remove from Chrome](#)

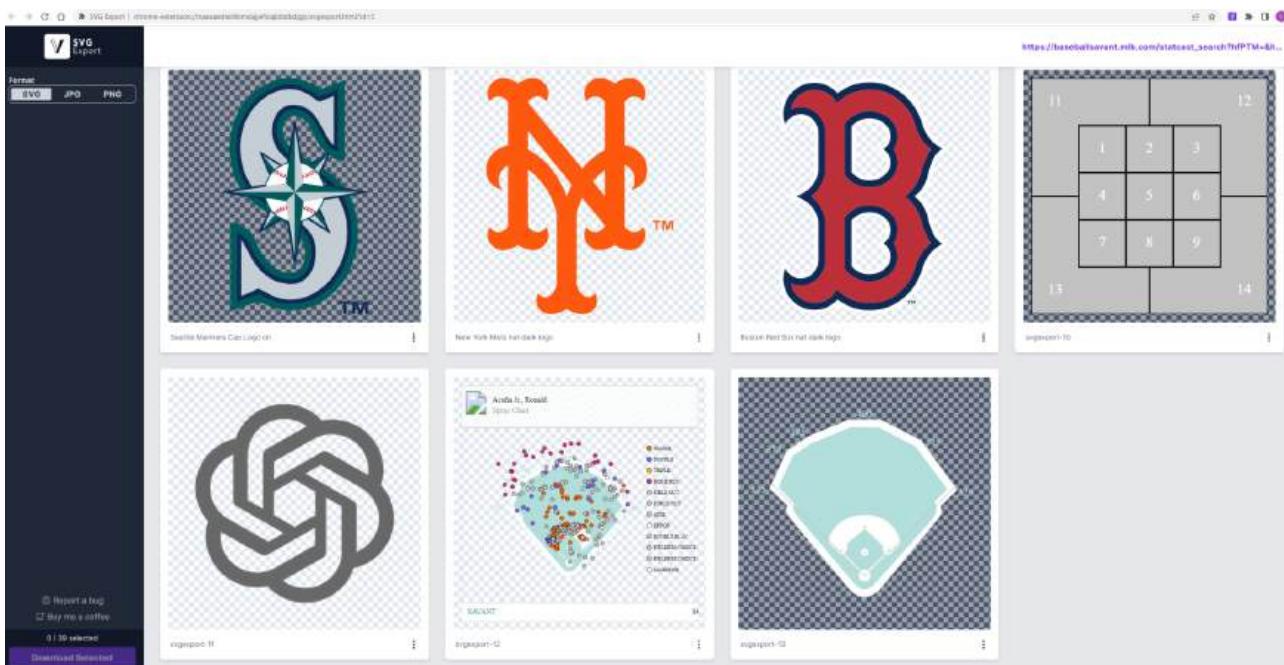
Overview Privacy practices Reviews Support Related



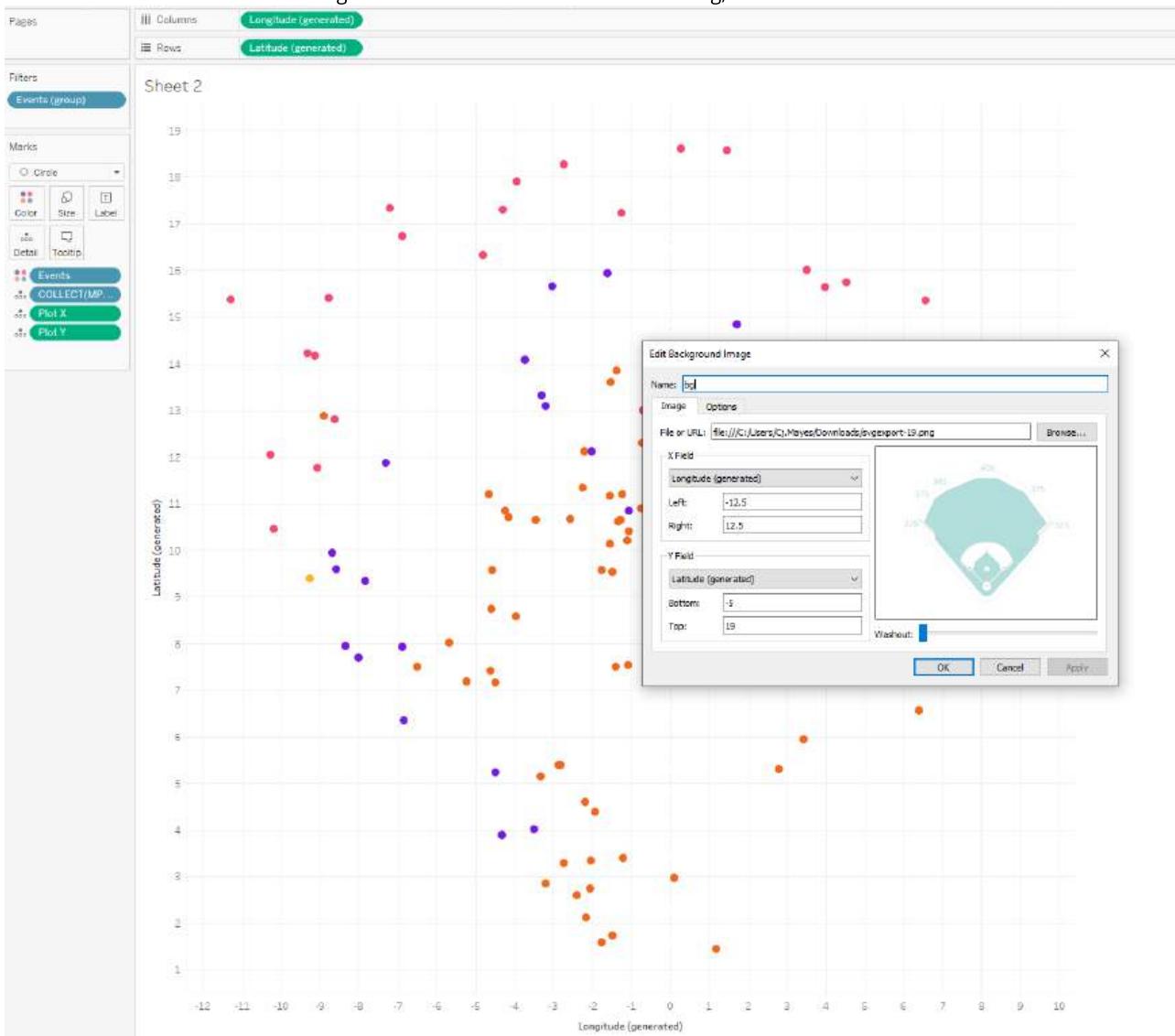
To do this I am going to use the SVG Export extension (It is one of my favourite extensions to help mirror branding!) All you have to do is add your extension to the page you are looking at,



When you run this, it will give you all the SVG icons from the desired webpage. It's worth noting at this stage, you will want to change your selection of what pitch background you want to map.

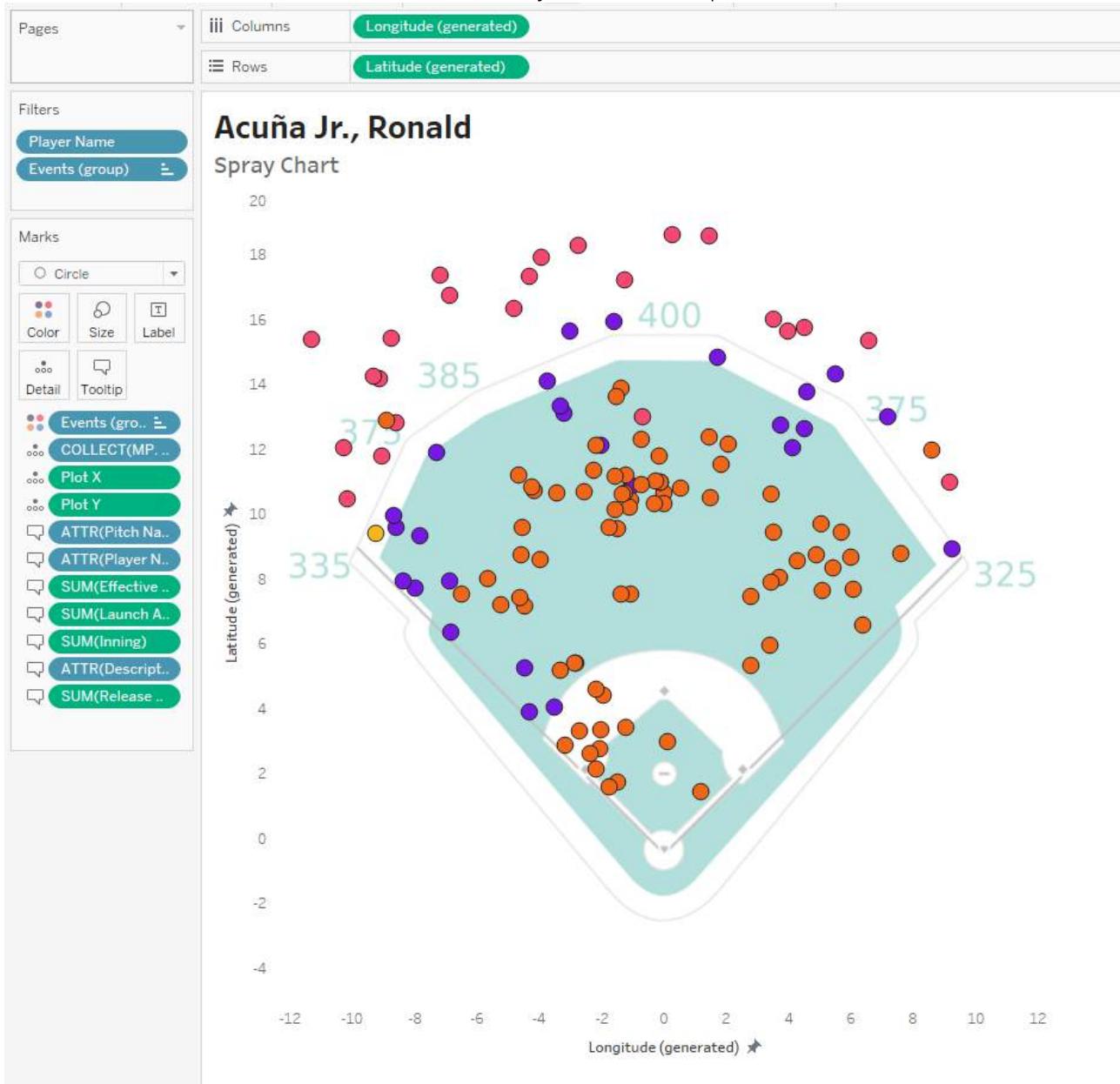


One of these images will be the pitch that we will use to map. Click the ellipse and download the file with a height and width of 500. Make sure to change the format to PNG before downloading, so that we can use it within Tableau.



I've used co-ordinates that as best accurately map the points onto the pitch as possible. Of course these may change dependent on the scale of your own pitch and co-ordinates, but I hope they provide good guidance. The main final thing is to fix your axis as

well as add in any final desired tooltips.

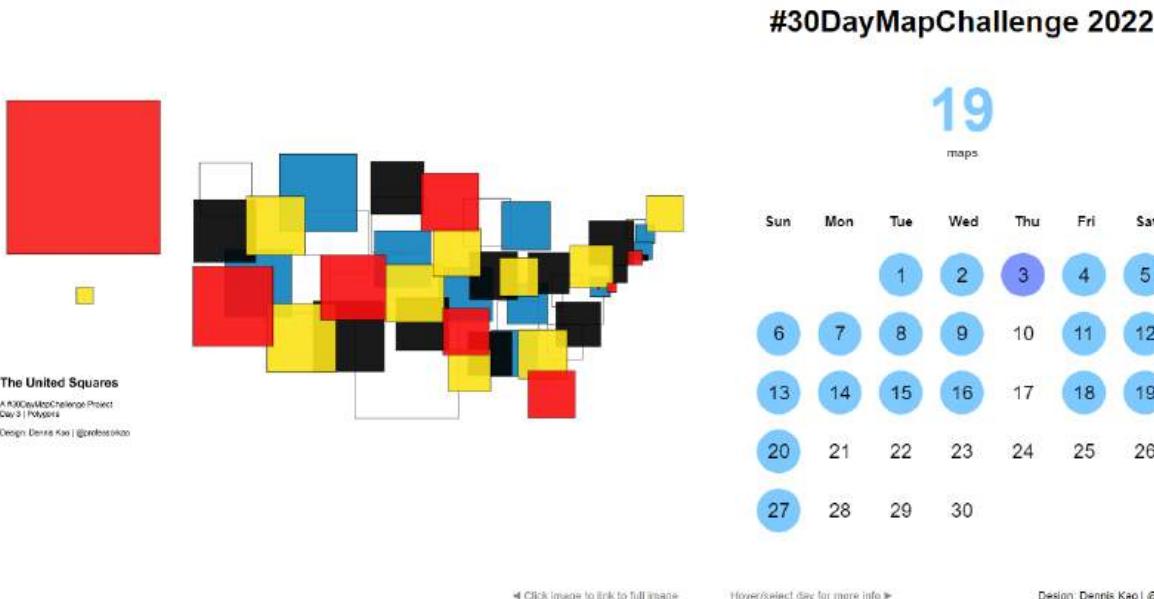


and there we have it, our first baseball chart in Tableau. Perhaps in future weeks we look at retrieving the data in a more automated way or having the option to map multiple different pitches. Let me know how you get on with this one – I can be found on the usual channels. **LOGGING OFF, CJ**

DESIGNING BETTER MAPS WITH DENNIS KAO

Hi all, Back to back guest blogs this month. Thank you to the 12 individuals that have got involved with sharing their expertise this year – I love that I have both the opportunity and privilege to host them, and no doubt as a community we appreciate the efforts that you go to create them. This week, we change our attention to all things maps, and that means I had to turn to Dennis Kao. CJ: Dennis, what a delight to have you join the site to be able to talk through cartography, map making and all things Tableau. Many of you will know Dennis from the community for his impact as a Tableau Public Ambassador, multiple vizie award winner as well as co-leading the HeathTUG and EduVizzers. More recently I've been really impressed with Dennis' commitment to the array of different maps he has been creating in Tableau. **C: Before we dive in, where does the passion for maps come from? Of your 150+ Public vizzes it looks like they've grown in frequency especially the past year? Was this from doing challenges like the 30DaysMapChallenge?** D: Not entirely sure... Maps have always had a significant place in my life. I just think maps are so cool... As a kid, I remember spending hours combing through my dad's National Geographic collection. I also grew up and started driving during a time when you had to read a paper map to get anywhere. Every year, I looked forward to getting the new Thomas Bros guide – which was a widely-used atlas in Los Angeles at the time. With a paper map, figuring out how to get somewhere always felt like a challenge... a puzzle. In my work, I've seen firsthand the power of maps and how they're designed (and what information they include or leave out) have a significant impact on people and communities. As a social worker, we're trained to think about the health and well-being of a person in the context of their environment. For example, depending on where a person lives, there may be more or less available healthcare services within a

reasonable walking or driving distance. In this case, I think maps could serve as a useful tool to examine and visualize the geographic gaps in the health system. It's only been in the past few years, since discovering Tableau, when I've focused more on the artistic side of mapmaking. And I'm absolutely loving the ride! This journey has really brought out a creative side that I never thought I had. This past year, I've been really trying to explore mapmaking design in Tableau Public... and yes, creating 20-30 maps during the #30DayMapChallenge each November. **CJ: You can view [Dennis round-up here](#). Here are a few of my favourites,** Day 3: The United Squares



Day 7: Redlining in Los Angeles



Day 13: Jeju Island

#30DayMapChallenge 2022



19

maps

Sun	Mon	Tue	Wed	Thu	Fri	Sat
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

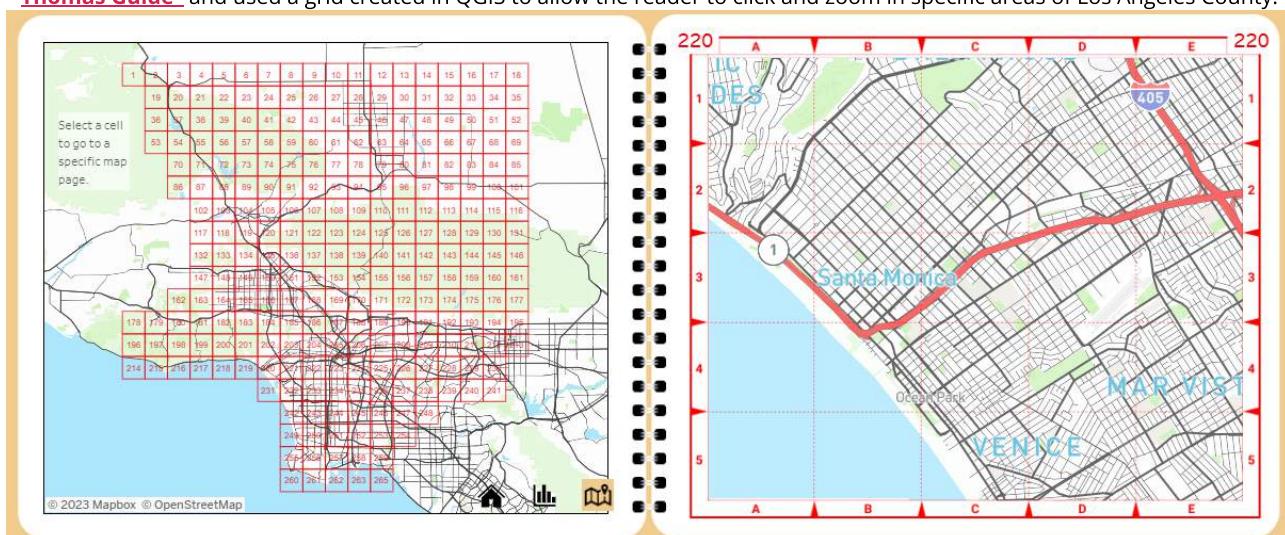
◀ Click Image to link to full image

Hover/select day for more info ▶

Design: Dennis Kao | @professorakao

C: What advantages come with learning multiple tools, like QGIS and other tools that can become compatible with Tableau?

D: Prior to discovering Tableau, I spent over 10 years mapping and conducting spatial analysis. I was initially trained using ArcGIS, but now mostly use QGIS (because it's free and available for the Mac). And while Tableau continues to up its spatial & mapping game, I do think there are still advantages in having GIS in your toolkit. First and foremost, GIS software, such as QGIS or ArcGIS, are designed specifically to deal with the unique properties of spatial data. So, from adding layers to exploring the data to editing and creating new features, I believe the process is just easier and faster. They're both designed as point-and-click applications and thus, have a lower learning curve. But you also have the option to use coding, e.g. Python, if preferred. GIS software provides a bunch of different tools that allow you to clean or edit the data and create new spatial features. Now, I almost exclusively use Tableau Public to visualize my final maps, but typically use QGIS to clean and prepare the data. At the beginning of this year, I started #RandomQGISToTableauStuff to explore and share some of these features, e.g. grids or water lines. The key advantage – in my mind – for creating these features in GIS is that they're hard-coded to the actual geography. So, when you bring them into Tableau, they function just like any other other spatial data file (e.g. shapefiles). For example, a grid is essentially a series of polygons that you can layer them, filter, zoom in and out, select, symbolize with color, label, and visualize in the same way you might do so with a map of the United States or Canada. In one of my visualizations, I paid homage to ["The Thomas Guide"](#) and used a grid created in QGIS to allow the reader to click and zoom in specific areas of Los Angeles County.

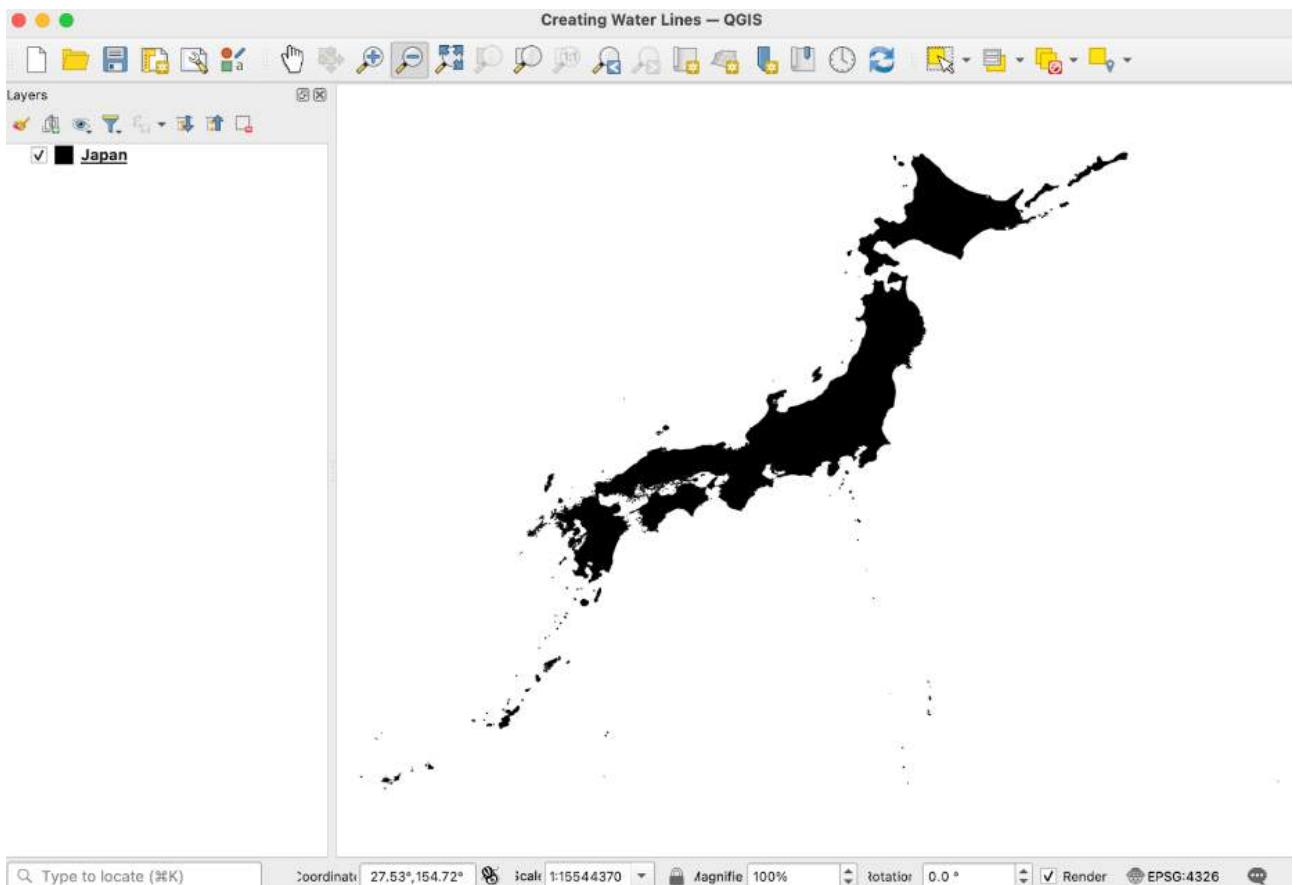


C: For those looking for map viz inspo, what are some of your favourites be it from Tableau or further afield? I'm continually inspired by the #DataFam, but a few that immediately come to mind are: [Wendy Shijia](#) (@ShijiaWendy), [Naresh Suglani](#) (@SuglaniNaresh), and [Damola Ladipo](#) (@AdedamolaLadipo). I also love the earlier stuff by [Ken Flerlage](#) (@flerlagekr) and [Adam McCann](#) (@adademccann). I would also encourage you to check out the 39 maps that folks did for the #IronQuest challenge back in March 2020, which focused on maps, [here](#). When I was first making the transition from GIS to Tableau, I found [Marc Reid](#) (@marcreid) particularly helpful. Finally, my approach has also been informed by the work of [Sarah Battersby](#), who wrote extensively about [QGIS and Tableau](#). And be sure to follow the #30DayMapChallenge that occurs in November...and be prepared to be absolutely blown away by some beautiful mapmaking. CJ: Now you know I'm a huge fan of your typefully and the post you put up around 10 things you can do in QGIS to enhance your maps in Tableau. Could you give us a

beginner's guide to how you created your water lines visual? D: Back in January of this year, I started a series called #RandomQGISToTableau, which I describe – over a thread of 4-5 short tweets – how to create spatial features in QGIS that can be used to enhance your maps in Tableau Public. One of my earliest posts focused on creating "[water lines](#)". And thanks to CJ, I have more space in which I can actually write complete sentences, so I'm going to describe the process in a bit more detail.

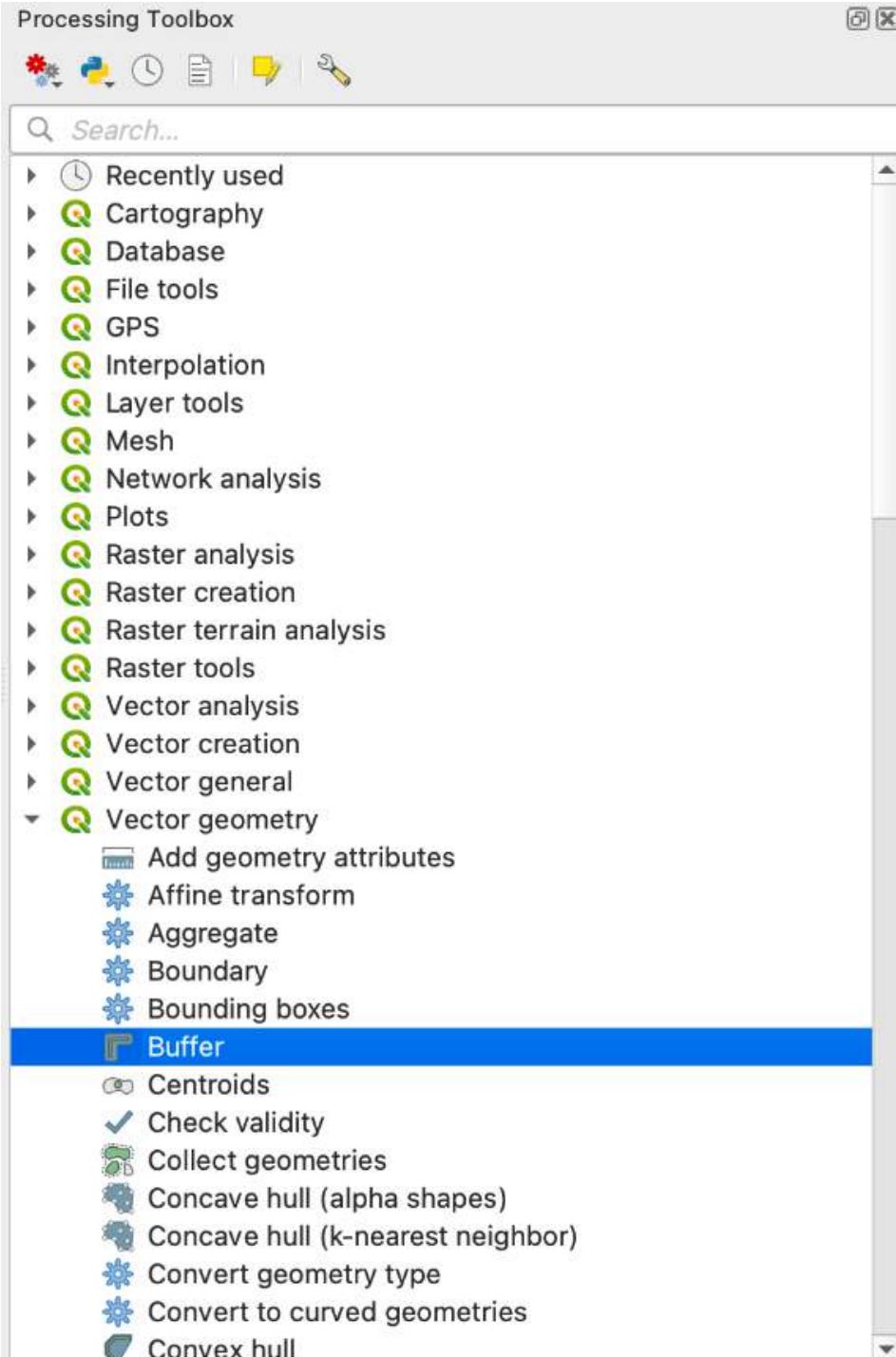


Have I mentioned that I love maps? The inspiration for this came from looking at old maps, such as [this one](#). I later came across [Olivia Vane's writings on water lines](#). So, the general approach is use QGIS to create a shapefile that has the desired geographic feature. Then we simply have to bring in and add it as a map layer in Tableau. In the case of water lines, we can use the BUFFER tool to create a series of polygons that surround the target geography at different distances. Tableau also has a buffer function, BUFFER(), but as far as I know, you can only create buffers around points right now (hint, hint...). The first step is to add your target geography. For this example, I will be starting with a shapefile of Japan (Source: Humanitarian Data Exchange).



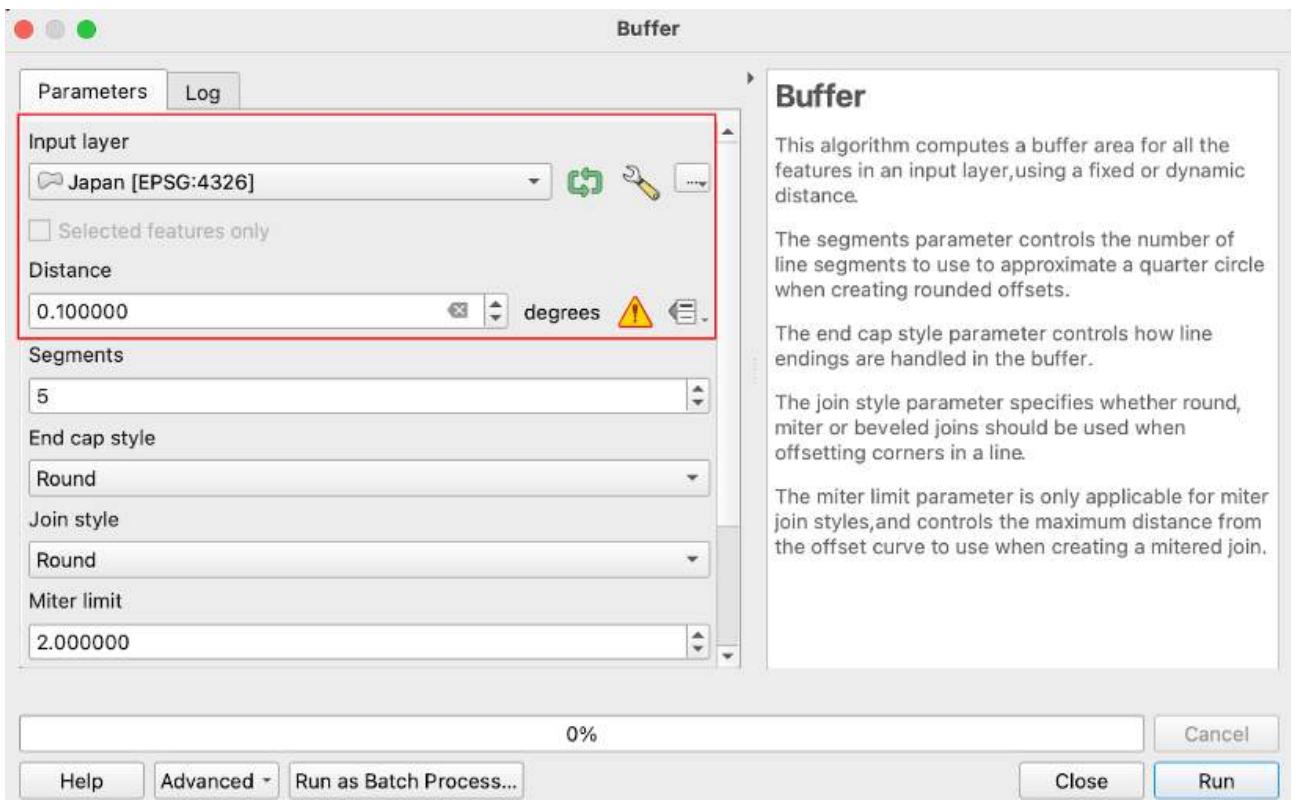
To create our first buffer, select the BUFFER tool either through the menus (Vector → Geoprocessing Tools) or the Processing Toolbox (under Vector Geometry).

A screenshot of the QGIS Processing Toolbox. The main menu bar shows "Vector", "Raster", "Database", and "Web". Below the toolbox, there is a list of tool categories: "Shape Tools", "Geoprocessing Tools", "Geometry Tools", "Analysis Tools", "Research Tools", and "Data Management Tools". A sub-menu for "Geoprocessing Tools" is open, displaying several tools with icons: Buffer..., Clip..., Convex Hull..., Difference..., Dissolve..., Intersection..., Symmetrical Difference..., Union..., and Eliminate Selected Polygons... . The "Buffer..." tool is highlighted with a blue background and a yellow icon.

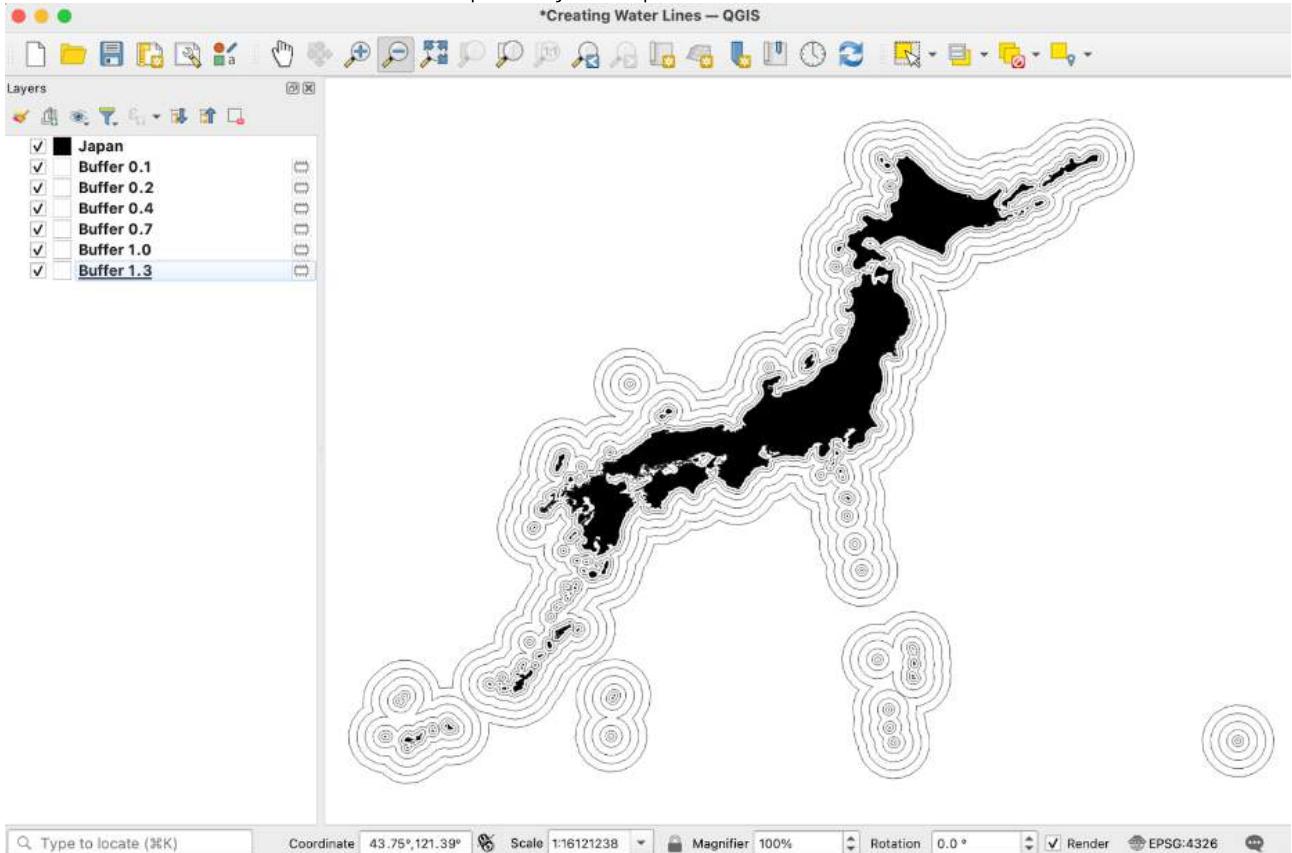


In the Buffer dialog window, select your target geography in the "Input Layer" field and input your desired distance in the "Distance" field. You may need to play around with the values. In this case, I wanted to mimic the ripple effect, so I increased the distance slightly with each buffer.

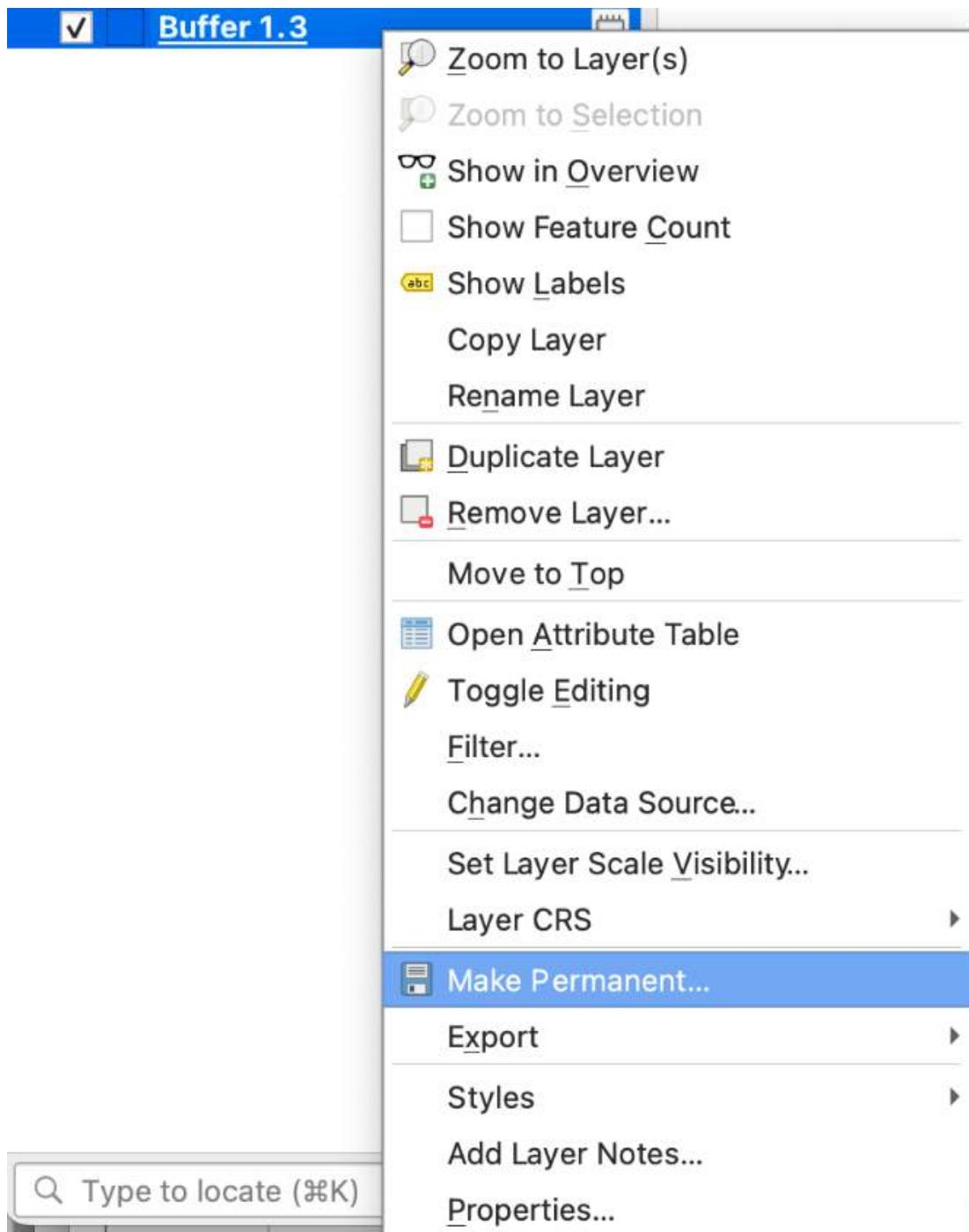
Repeat this step until you get the desired number of buffers. For example, I decided to go with six buffers at the following distances: 0.1, 0.2, 0.4, 0.7, 1.0, and 1.3 degrees.



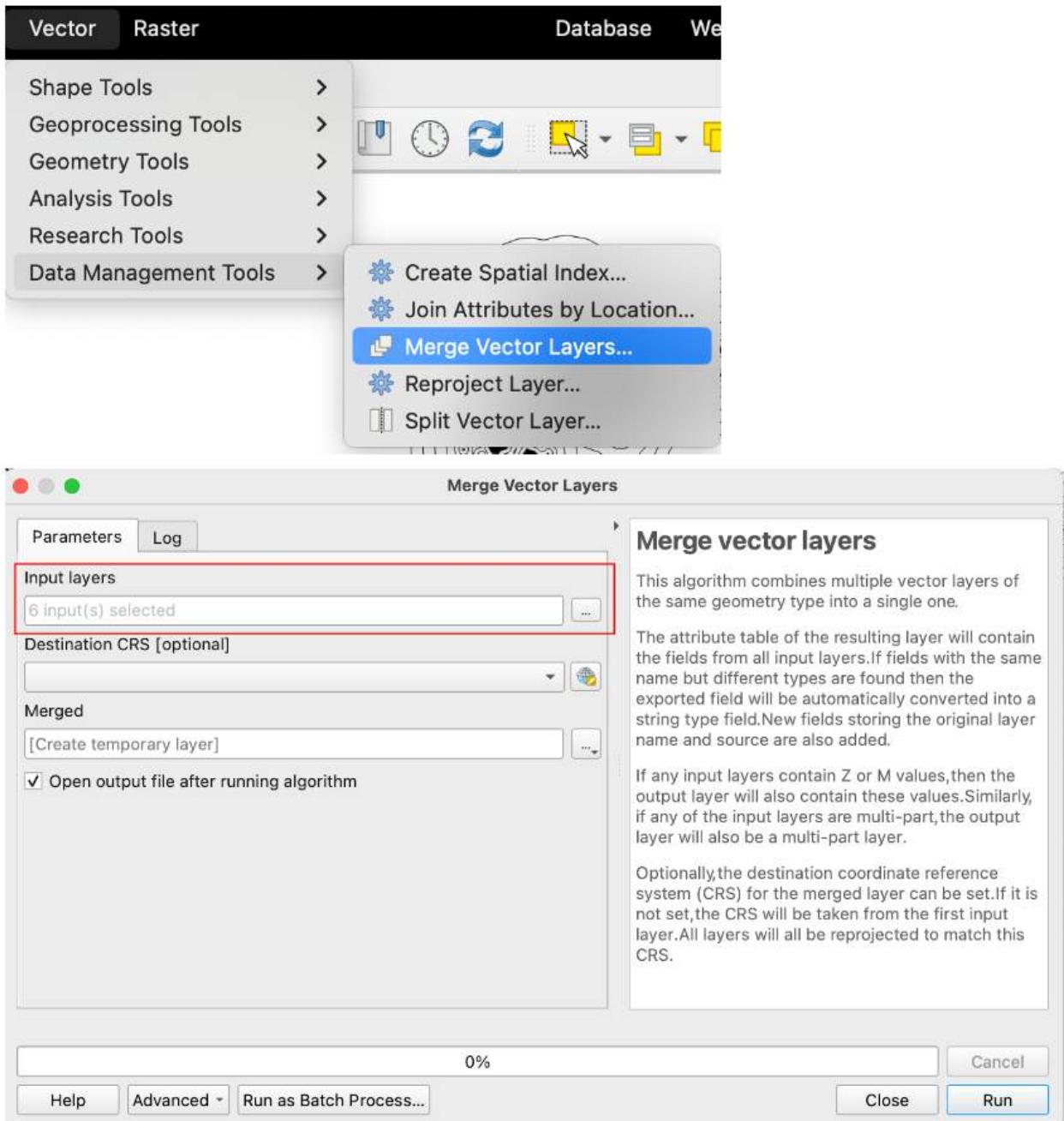
The result is 6 polygon feature layers that surround the geography of Japan at varying distances. A few important things to note here. First, buffers are essentially polygons so you can play with the fill and line (border) colors to get the desired effect. Second, if you have a water body as your starting geography, you will want to use negative values for your distances to obtain internal buffers. For example, in my earlier post, I started with the Great Lakes.

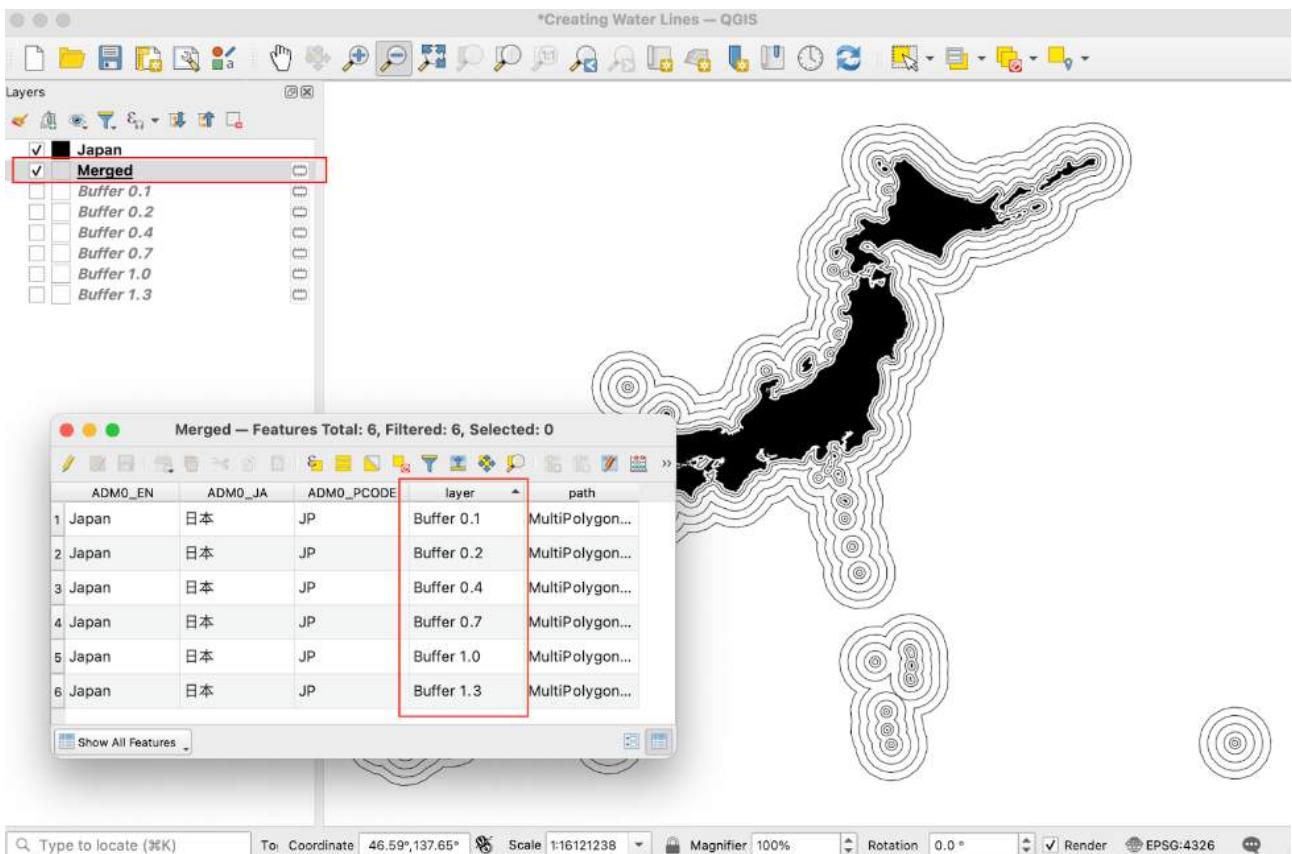


At this point, you have two options. On one hand, you can save each of these layers as a shapefile to be used in Tableau by right-clicking on each layer and selecting "Make Permanent". While it's a bit more trouble to bring in the various files into Tableau, keeping them as separate files will allow you to manipulate each buffer (formatting, color, opacity, etc.) as separate map layers.

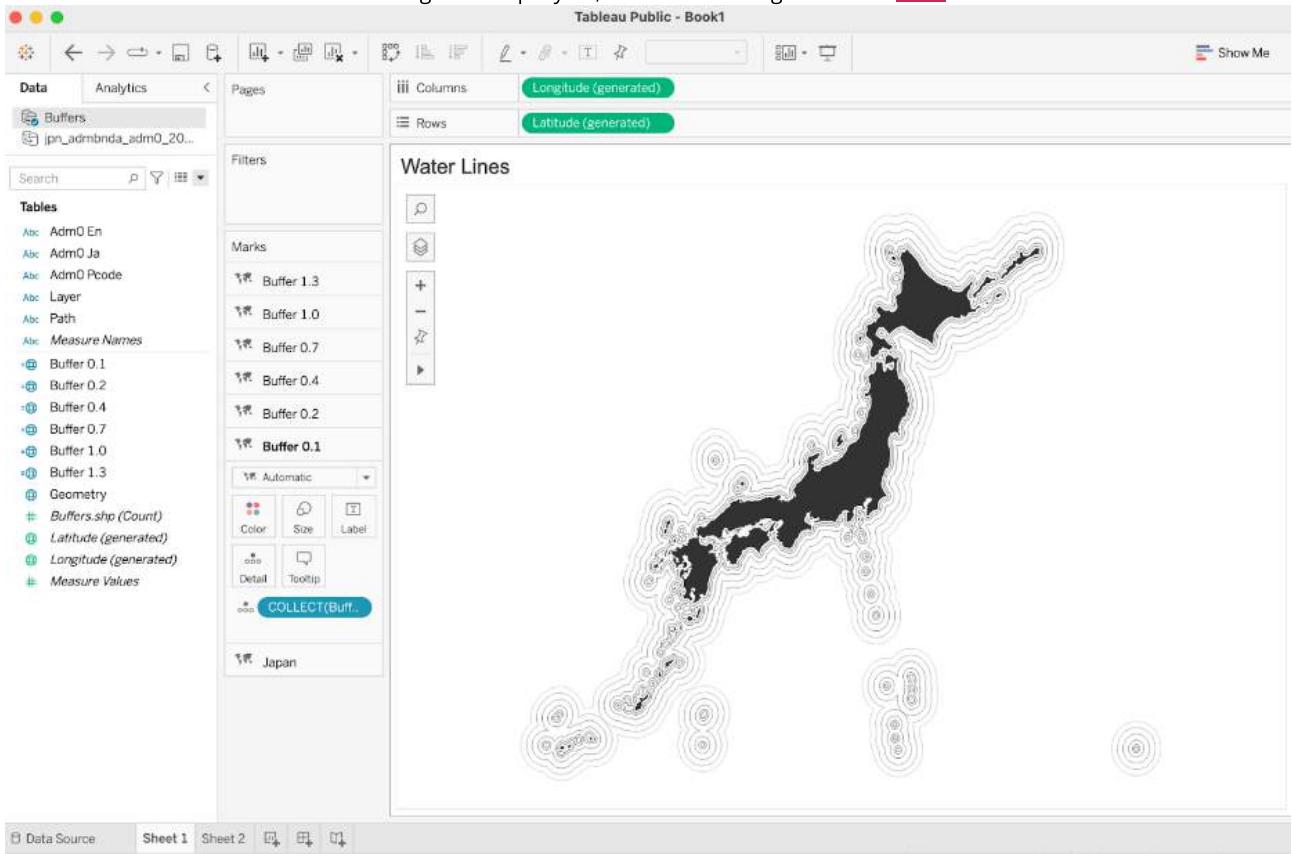


On the other hand, you may want to first use the Merge Vector Layers tool to merge the layers into a single layer that you can then save to be used in Tableau. And if we look at the attribute table, you will notice a "Layer" field that has the name of each layer. This will streamline the process of bringing in the data, but you may still want to separate each buffer as separate map layers for formatting purposes. You would then have to create a series of calculations, such as the following: IF = "Buffer 0.1" THEN END





Now you're ready to bring-in the data and visualize the water lines in Tableau Public. If you need more information about working with map layers, Marc Reid has a great article [here](#)

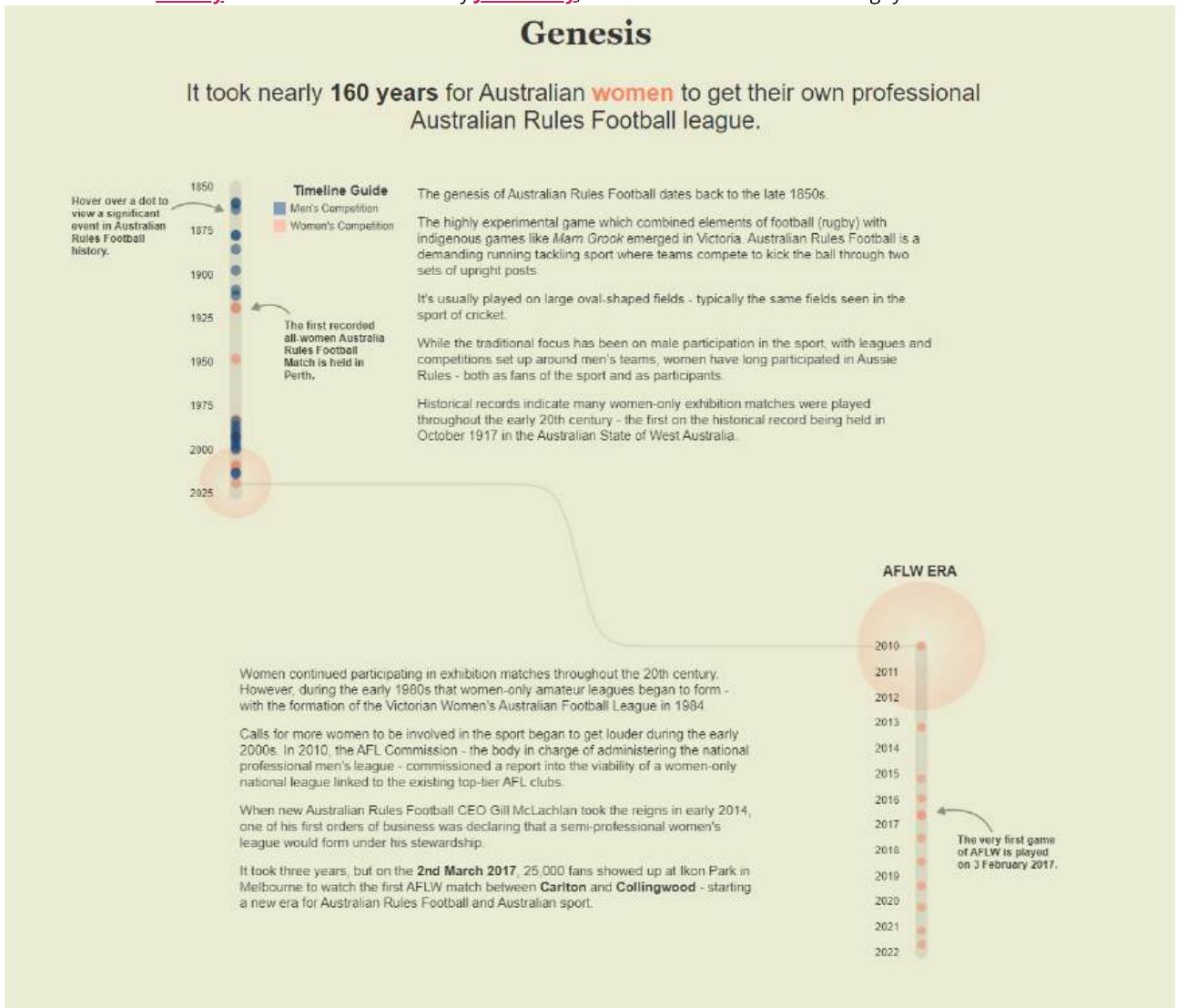


Above is the different buffer layers I've used within my visualisation! I hope you found this helpful! Thanks CJ for the opportunity to share some thoughts. **CJ Round-Up:** I always love to see how individuals start to mix different tools to produce something quite spectacular. Dennis does just this with his use of QGIS & Tableau. Be sure to connect with Dennis on Twitter, [here](#).

LOGGING OFF, CJ

AFL STANDINGS WITH DARRAGH MURRAY (R STUDIO)

Hi all, Its a great pleasure to invite good friend and previous colleague Darragh Murray onto the blog today. Darragh has really impressed me over the years from his T shape skills – the AFL blog he shares with us today reflects that. If you'd like to connect with Darragh he can be found on [Twitter](#) and his own site [here](#). **Pursuit of the Pennant | Using R and fitzRoy to visualise historical AFL ladder standings**. Last year, I was fortunate to be selected as one of the top 15 entrants in the 2023 edition of Tableau's Iron Viz competition. My visualisation of the [evolution of women's participation in Australian Rules Football](#) relied on historical Australian Rules football statistics obtained via a handy API made available through an R package called [fitzRoy](#) written and maintained by [James Day](#), and co-authored with Robert Nguyen and Oscar Lane.



In this tutorial, I want to demonstrate how one can access this API via R, obtain data to gain a historical understanding of the finishing positions of various Australia Rules football teams over time and visualise this information in a platform such as Tableau Public. **So remind me, what is Australian Rules Football?** Australian Rules Football (often called Aussie Rules or 'AFL' or simply 'footy') is a full contact ball sport played on a large oval field between two contending teams of 18 players. Footy is a very popular sport in Australia, with some regular season matches hosting up to 100,000 spectators depending on the teams playing. [Here's a short video that explains the game concept, complete with flashy highlights](#). Each year 18 AFL teams compete for the AFL 'flag' or premiership. This involves playing up to 23 rounds of regular season matches, followed by a knockout final series culminating in the AFL Grand Final – usually held in late September at the Melbourne Cricket Ground. 'Footy' has a long history in Australia and is one of the world's oldest organised team sports, tracing its origins back to the mid-19th century. Fortunately for us data nerds, the footy fans of the time kept highly detailed records, and this tradition has continued well into the 21st century. And thanks to the fitzRoy API, getting comprehensive game and player data is highly straightforward. **What exactly is in fitzRoy?** "The goal of fitzRoy is to make it easy to access data from the AFLM and AFLW competitions. It provides a simple and consistent API to access data such as match results, fixtures and player statistics from multiple data sources." The fitzRoy API provides access to several data sources that compile statistics on both the men's and women's Australian rules competition. Data available include:

- Fixturing and match results.
- Player lineups and statistics.
- Team ladder positions by round.
- ..and a whole lot more.



Overview

fitzRoy aims to provide a consistent and reliable API to various data sources of both the Mens and Womens competitions of the AFL. These functions provide easy and tidy access to data such as fixtures, results and statistics from various data sources.

Installation

Install the released version of fitzRoy from CRAN:

```
install.packages("fitzRoy")
```

Or install the development version from GitHub with:

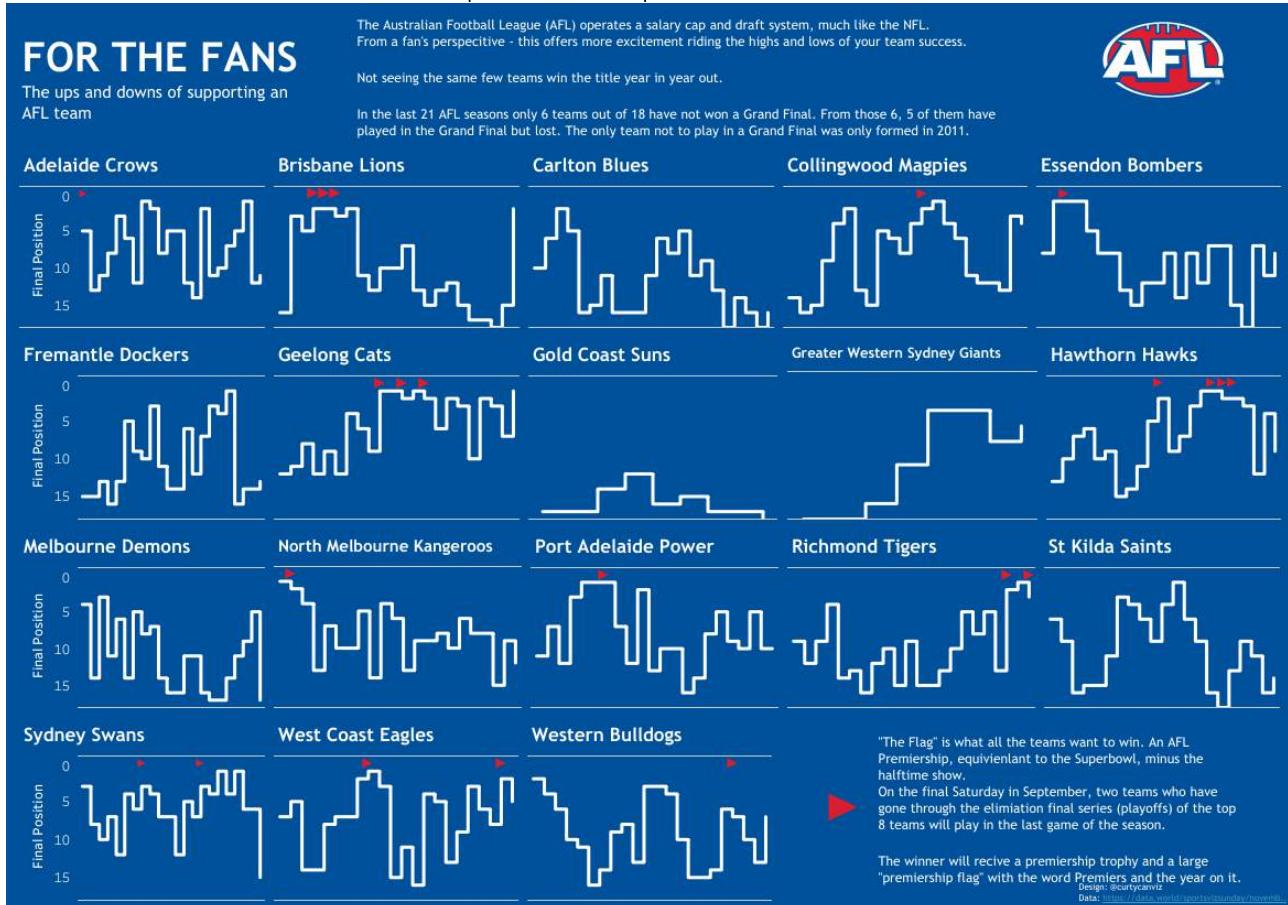
```
# install.packages("devtools")
devtools::install_github("jimmyday12/fitzRoy")
```

Essentially fitzRoy provides people access to:

- Official data from the official AFL website.
- AFL Tables, which contains match results dating back to 1897.
- Footywire, which has detailed player statistics from 2012 onwards.
- Squiggle, home to the best AFL predictive models.
- Fryzigg, a further source of player statistics.

I feel like I'm regurgitating much of the useful information in the '[Introduction to fitzRoy](#)' vignette, but my point here is that fitzRoy provides users access to various resources with varying levels of detail. Older data will have fewer details, whereas more recent seasons and matches will have more detailed data for users to manipulate and shape.

Visualising Ladder Positions: Showing AFL Team Performance Over Time. Inspired by a [Tableau Public visualisation](#) completed by my friend [Kris Curtis](#), I wanted to quickly and easily build a dataset that allowed me to visualise the historical finishing position of each Australian Rules football team since the formation of the national competition in 1990 up until the end of 2022.



Manual data collection would be relatively straightforward, and I imagine someone could quickly assemble a dataset within an hour or two. Still, I wanted to show how easy it is to complete with a few lines of R-code using fitzRoy. I'll use R and run code for this walkthrough using the defacto standard R IDE, [RStudio](#). I won't cover the step-by-step in installing these platforms, **but if**

[you're stuck, here is an excellent rundown on how to do so.](#) The next part of this blog will run you through the logic of getting all the footy data required to build a visualisation similar to the above using the fitzRoy API in a step-by-step fashion.

However, if you'd prefer just to see the commented code, you can do that by looking at it on [github](#). **Setting Up Your Environment: R Packages**

The first task one needs to do after installing R (and likely an IDE like the aforementioned RDE Studio) is to make sure you have the appropriate packages installed in your R environment. Fortunately, this is dead easy to do:

```
install.packages("fitzRoy") install.packages("dplyr")
```

If you don't have the above packages installed, this code is what you need to get sorted. Calling these in the script will install them. Once you've run them once and installed them, you can comment these lines out using a hashtag (#). We then load a package into our R session by calling the library function.

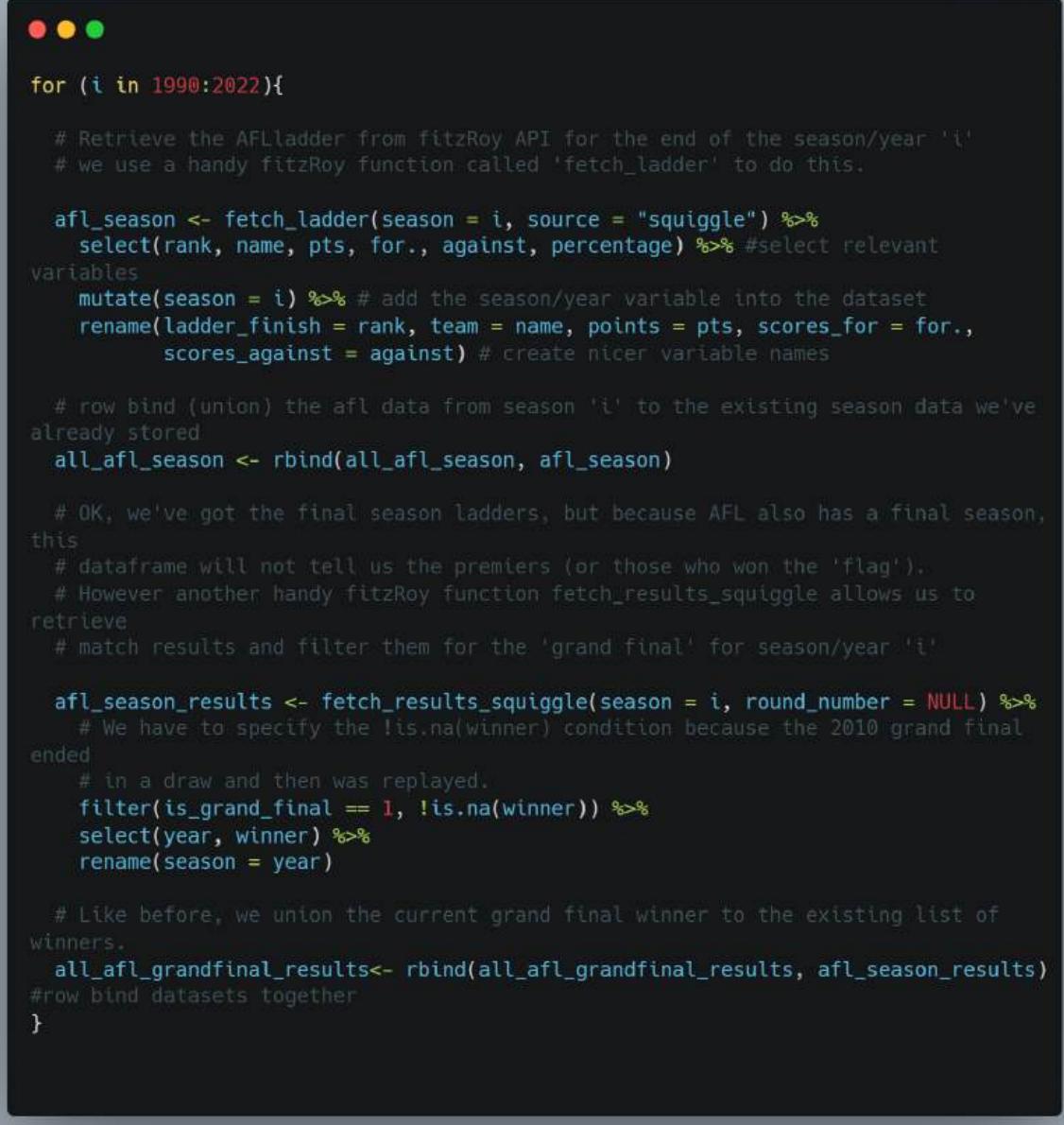
```
library("fitzRoy") library("dplyr")
```

Our basic environment is now set, so we can get cracking. The first step is creating some empty data frames to store our afl season data.

```
all_afl_season = data.frame() # a dataframe to store regular season data
```

```
all_afl_grandfinal_results = data.frame() # a dataframe to store the results of the AFL grandfinal
```

We're using empty data frames because we will create a loop and append new data to these data frames each iteration. We're starting with the 1990 season and continuing year by year, adding new data until we hit 2022. The data we get for each AFL team in each season is their final ladder position. However, unlike sports like premier league football, AFL teams play in a final series with a top-eight ladder position stamping your ticket. Theoretically, anyone in the final eight can win the premiership, though the higher your position gives you certain advantages (home ground finals and so forth). However, the team that finishes top of the ladder often fails to win the premiership. Here's the entire loop code with comments to guide you on what we're doing. You can download it from the GitHub Link at the top of the page.



```
for (i in 1990:2022){

  # Retrieve the AFL ladder from fitzRoy API for the end of the season/year 'i'
  # we use a handy fitzRoy function called 'fetch_ladder' to do this.

  afl_season <- fetch_ladder(season = i, source = "squiggle") %>%
    select(rank, name, pts, for., against, percentage) %>% #select relevant variables
    mutate(season = i) %>% # add the season/year variable into the dataset
    rename(ladder_finish = rank, team = name, points = pts, scores_for = for.,
           scores_against = against) # create nicer variable names

  # row bind (union) the afl data from season 'i' to the existing season data we've already stored
  all_afl_season <- rbind(all_afl_season, afl_season)

  # OK, we've got the final season ladders, but because AFL also has a final season, this
  # dataframe will not tell us the premiers (or those who won the 'flag').
  # However another handy fitzRoy function fetch_results_squiggle allows us to retrieve
  # match results and filter them for the 'grand final' for season/year 'i'

  afl_season_results <- fetch_results_squiggle(season = i, round_number = NULL) %>%
    # We have to specify the !is.na(winner) condition because the 2010 grand final ended
    # in a draw and then was replayed.
    filter(is_grand_final == 1, !is.na(winner)) %>%
    select(year, winner) %>%
    rename(season = year)

  # Like before, we union the current grand final winner to the existing list of winners.
  all_afl_grandfinal_results<- rbind(all_afl_grandfinal_results, afl_season_results)
  #row bind datasets together
}
```

Our previously two empty data frames are now filled with the good stuff! Each team's finishes positions plus the data on who won the flag that year! Here's what some sample data looks like for each data frame.

	ladder_finish	team	points	scores_for	scores_against	percentage	season
1	1	Essendon	68	2526	1815	139.17355	1990
2	2	Collingwood	64	2376	1825	130.19178	1990
3	3	West Coast	64	2274	1920	118.43750	1990
4	4	Melbourne	64	2339	2066	113.21394	1990
5	5	Hawthorn	56	2414	2002	120.57942	1990
6	6	North Melbourne	48	2519	2210	113.98190	1990
7	7	Western Bulldogs	48	2016	2031	99.26145	1990
8	8	Carlton	44	2277	2187	104.11523	1990
9	9	St Kilda	36	2328	2313	100.64851	1990
10	10	Geelong	32	2248	2398	93.74479	1990
11	11	Richmond	28	1988	2530	78.57708	1990
12	12	Fitzroy	28	1874	2389	78.44286	1990
13	13	Sydney	20	1904	2704	70.41420	1990
14	14	Brisbane Lions	16	1733	2426	71.43446	1990

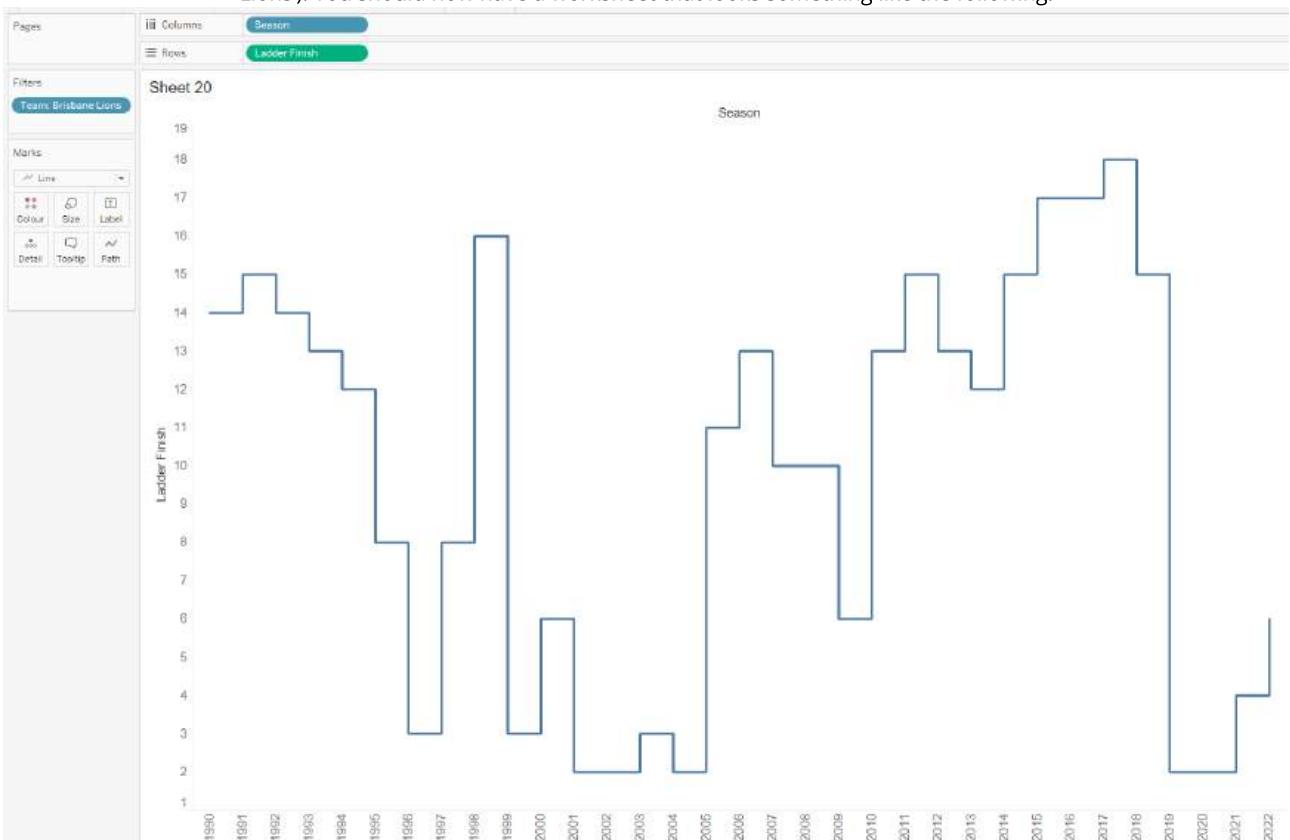
	season	winner
1	1990	Collingwood
2	1991	Hawthorn
3	1992	West Coast
4	1993	Essendon
5	1994	West Coast
6	1995	Carlton
7	1996	North Melbourne
8	1997	Adelaide
9	1998	Adelaide
10	1999	North Melbourne
11	2000	Essendon
12	2001	Brisbane Lions
13	2002	Brisbane Lions
14	2003	Brisbane Lions

The final preparation step is merging these two data frames into a singular structure that contains both pieces of information in one easy-to-visualise dataset. This is quite straightforward using the `left_join` function.

```
afl_ladder_with_flags <- left_join(all_afl_season, all_afl_grandfinal_results, by='season') %>%
  mutate(premiership_flag = ifelse(team == winner, TRUE, FALSE)) %>% # use true/false for winner variable
  rather than team name
  select(-winner) # do not need this field any more
And you can then see the final ladder positions with an indicator to show which team won the premiership (or 'flag') in each season.
```

	ladder_finish	team	points	scores_for	scores_against	percentage	season	premiership_flag
1	1	Essendon	68	2526	1815	139.17355	1990	FALSE
2	2	Collingwood	64	2376	1825	130.19178	1990	TRUE
3	3	West Coast	64	2274	1920	118.43750	1990	FALSE
4	4	Melbourne	64	2339	2066	113.21394	1990	FALSE
5	5	Hawthorn	56	2414	2002	120.57942	1990	FALSE
6	6	North Melbourne	48	2519	2210	113.98190	1990	FALSE
7	7	Western Bulldogs	48	2016	2031	99.26145	1990	FALSE
8	8	Carlton	44	2277	2187	104.11523	1990	FALSE
9	9	St Kilda	36	2328	2313	100.64851	1990	FALSE
10	10	Geelong	32	2248	2398	93.74479	1990	FALSE
11	11	Richmond	28	1988	2530	78.57708	1990	FALSE
12	12	Fitzroy	28	1874	2389	78.44286	1990	FALSE
13	13	Sydney	20	1904	2704	70.41420	1990	FALSE
14	14	Brisbane Lions	16	1733	2426	71.43446	1990	FALSE

We can now export a .csv file, and that's the hard work all done! `write.csv(afl_ladder_with_flags, "output/afl_ladder_with_flags.csv", row.names=FALSE)` **Visualising AFL Ladders & Flags in Tableau** Now that you have nicely prepared dataset, you can visualise this information in a number of different platforms. I'm going to throw together some basic instructions in Tableau, but feel free to use other platforms should you choose. For this part of the tutorial, I'm going to assume some basic familiarity with Tableau. I'm also going to use the clever *transparent shapes* trick to visualise premiership flags, but I'm not going to cover how to use them in any depth (for that, I recommend reading Kevin Flerlage's blog "[14 Use Cases for Transparent Shapes and Images](#)"). Open up your install of Tableau, either Desktop or Public and then use the 'connect to data' option to locate the .csv file you outputted from RStudio. We will build a series of simple line charts – one for each team, showing the ladder position for each season between 1990 and 2022. If a team wins the premiership, that season will have a right-facing triangle to indicate that in the space above the relevant season. As previously mentioned, the 'flag' is what footy fans colloquially call the premiership, and coincidentally the right-facing triangle does look like a flag). Open a Tableau Worksheet and drag 'season' to columns and ladder finish to rows. Right-click 'ladder finish' and make it *continuous*. Change the Marks type to 'line'. Then set the path marks 'line type' to 'step'. Drag 'team' to the filters and select a team (for example 'Brisbane Lions'). You should now have a worksheet that looks something like the following:



Hold up! Things look a bit odd. Ladder position is in the opposite direction to where it suppose to be, so we need to *reverse* the Y axis so that teams that finish in top spot are at the top of the graphic. This is easily done by editing the axis parameters of the 'ladder finish' pill. Once you've reverse the axis, the next step is to add in the *flag* symbols above the relevant season. We're going to need a Tableau calculation here called '**Flag Position**'. Here's the calculation: **Flag Position** IF = TRUE THEN 0.3 ELSE 19 END

This calculation is just telling Tableau where to place a 'flag' on the Y axis depending on the value of the '*Premiership Flag*' variable. Drag the 'Flag Position' calculation to rows and make it a dual axis. Synchronise the axis. You'll now note that your view will look something like the following:

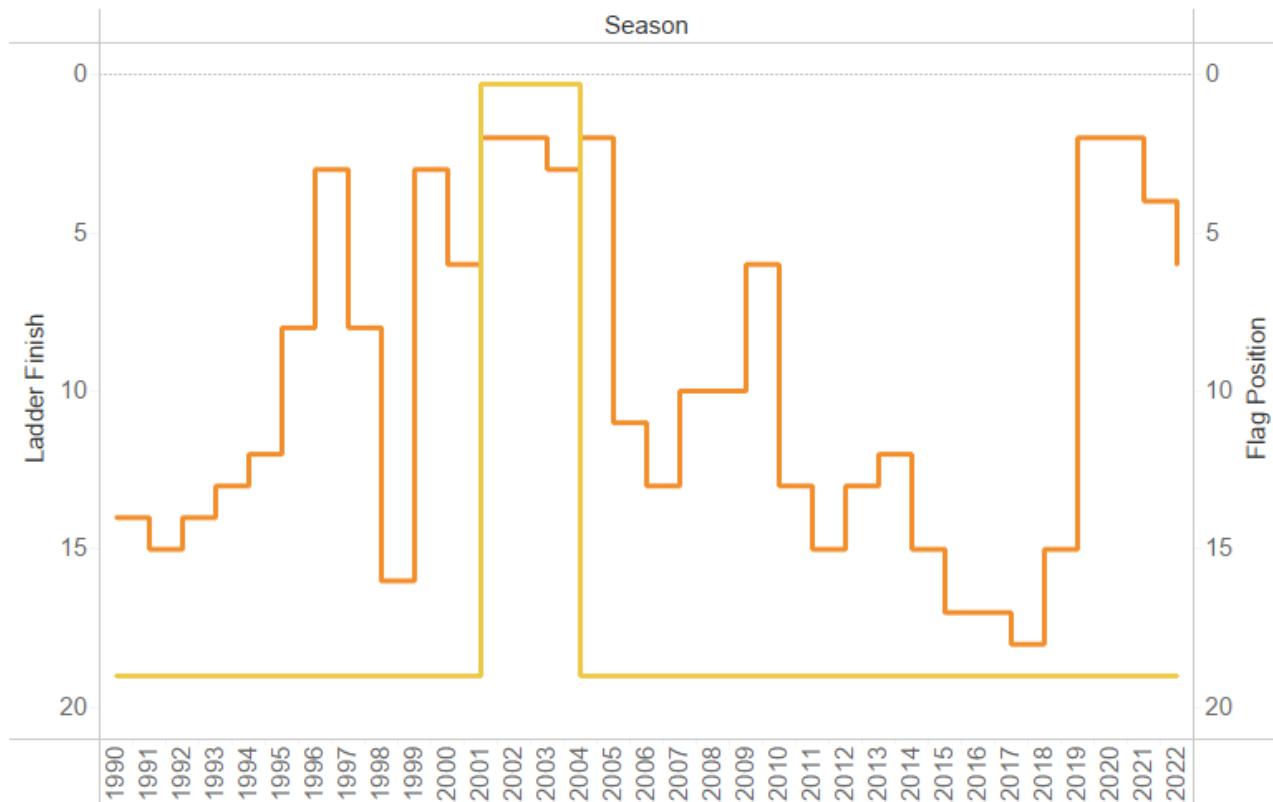
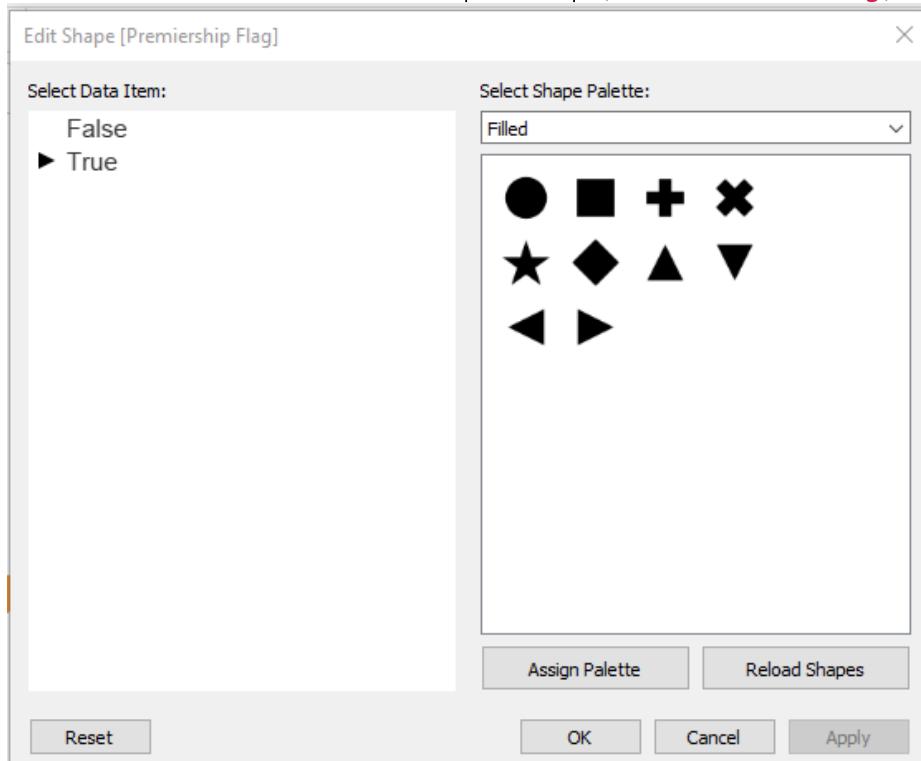


Tableau has by default set the '*Flag Position*' measure to line. You'll note that where Brisbane won the premiership between 2001 and 2003, it has set the '*Flag Position*' value at 0.3. We're going to change this to our nice little premiership flag icons! Click on the Flag Position mark, and change the mark type to shape. Then drag the '*Premiership Flag*' dimension to the shape mark, then click on the Shape mark to edit the shape. We're going to set the 'true' value to be a filled right facing arrow, and the 'false' value to be a transparent shape (remember: [read that blog!](#)).



You can also set the colours of the flag by using a similar approach except dragging the '*Premiership Flag*' dimension to colour rather than shape. A tiny bit more formatting (removing headers and duplicate axis etc.) and you can have something like this:



Of course, you can iterate this approach for each team if you wish and create a fully formatting dashboard that shows each team individually, which is what Kris did in his original dashboard. Here's what I ended up producing:

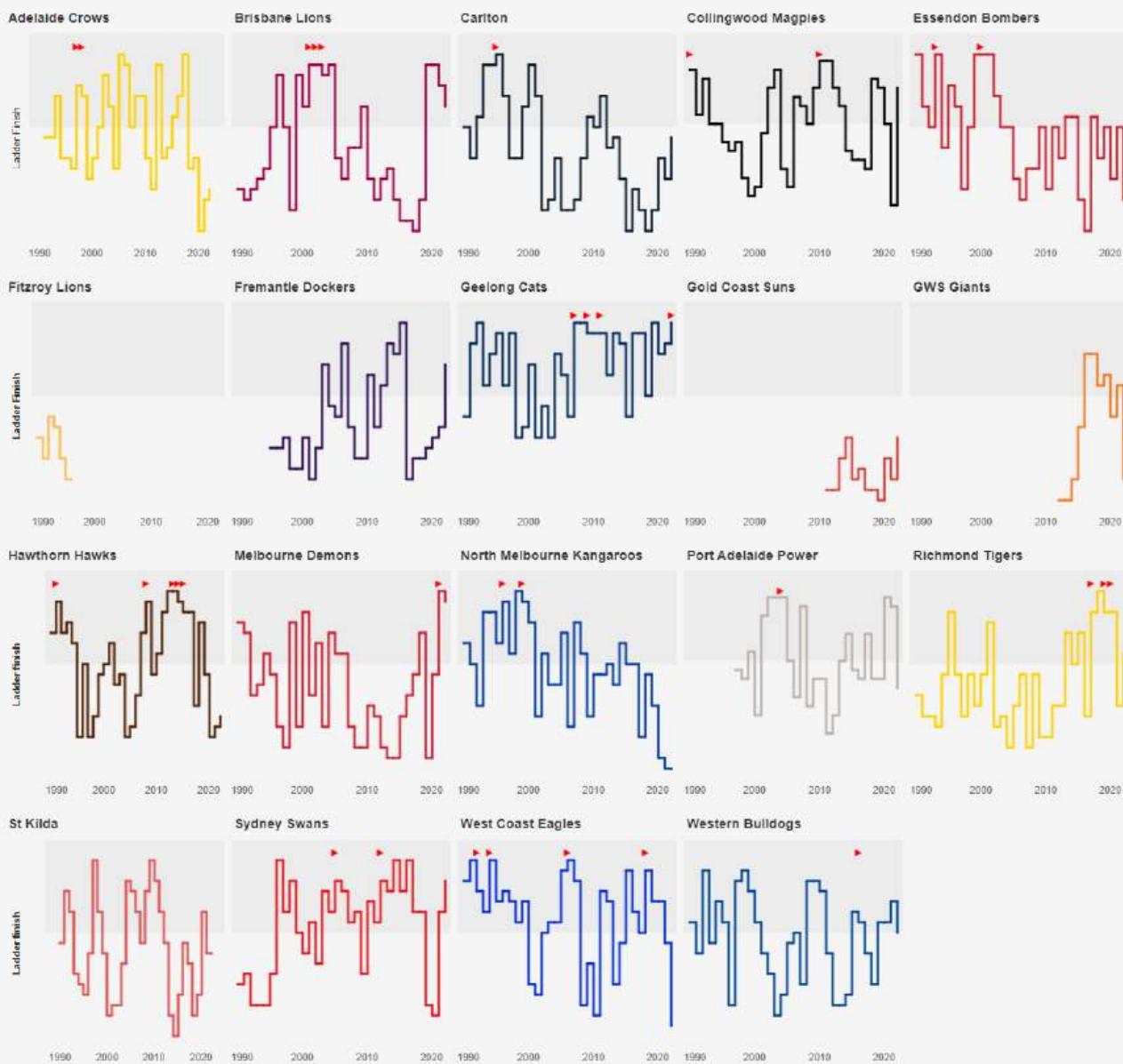
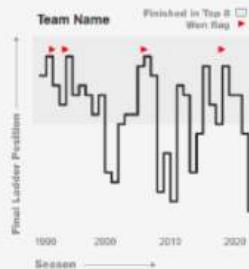
PURSUIT OF THE PENNANT ►

AFL team ladder positions & flag success | 1990 - 2022

Winning the AFL premiership flag (or pennant) is the ultimate prize in Australian Rules Football. Since the beginning of the modern Australian Football League (AFL) in 1990, 19 teams across 32 seasons have competed for the flag (with one team, Fitzroy, eventually merging with Brisbane to form the Brisbane Lions in 1996).

The charts below outline the success of each of these teams, showing ladder position, whether they finished in the top 8 and, of course, whether they managed to nab the flag!

Consult the key to the right to understand how to interpret each line chart.



Created by Darragh Murray (Twitter: @dofmurray; Threads: @thedataviz). Viz concept borrowed from Kris Curtis (@CurlyCanViz).

Data sourced from fitzRoy API
Data prepared using #rststats

For this layout, I essentially built the visualisation for one team and then duplicated that visual for all teams, changing the team filter as I went. I then plonked that into a 4x5 grid of containers on a dashboard. You're more than welcome to [download the dashboard](#) and check out the exact mechanics of how it works. **BONUS: Visualising AFL Ladders & Flags in R** I know I showed you how to show this in Tableau, but R has a handy visualisation engine called ggplot2(). I'm not making the same dashboard as I did in Tableau, but I'll throw in more code you can add to your script, achieving a vaguely similar outcome to what we did in Tableau but via a few lines of R code! Return back to the R environment we had set up earlier. We're going to load a few more packages (install them if you haven't got them already installed!).

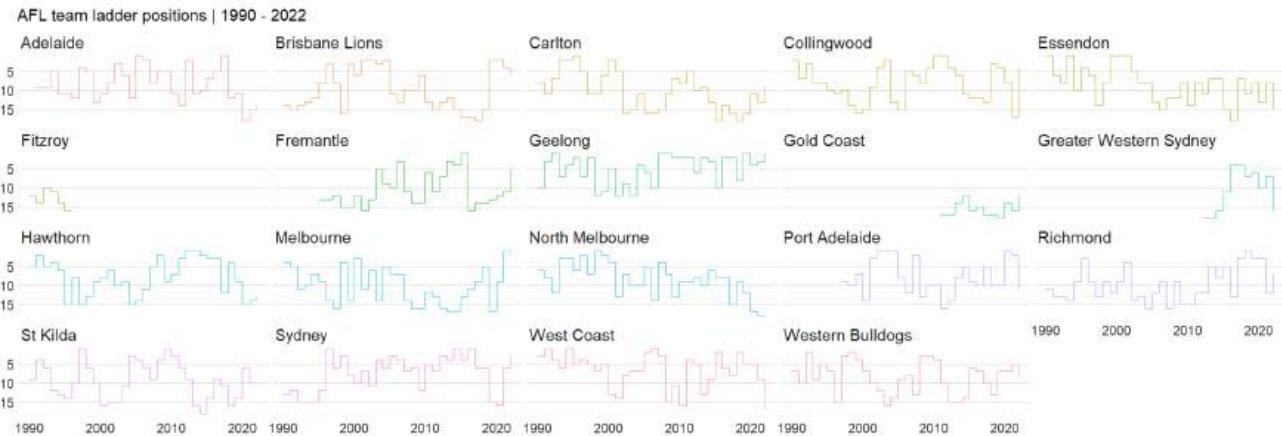
```
library("ggplot2") # ggplot2 allows us to chart data  
library("bbplot") # bbplot allows us a bit more functionality to pretty up our graphs BBC style  
And it turns out, we can do a lot of the hard graphing work in about 6 lines of R code # now we can chart using ggplot and facet_wrap
```

```

by team (small multiples) chart_seasons <- ggplot(all_afl_season, aes(x=season, y=ladder_finish,
colour=team)) + labs(title="Pursuit of the Pennant", subtitle="AFL team ladder positions | 1990 - 2022") +
  geom_step() + facet_wrap(~team) + scale_y_reverse() + bbc_style() + theme(legend.position = "none")
chart_seasons #output the seasons chart The last command then outputs to the plots tab within RStudio, and should look
something like this:

```

Pursuit of the Pennant

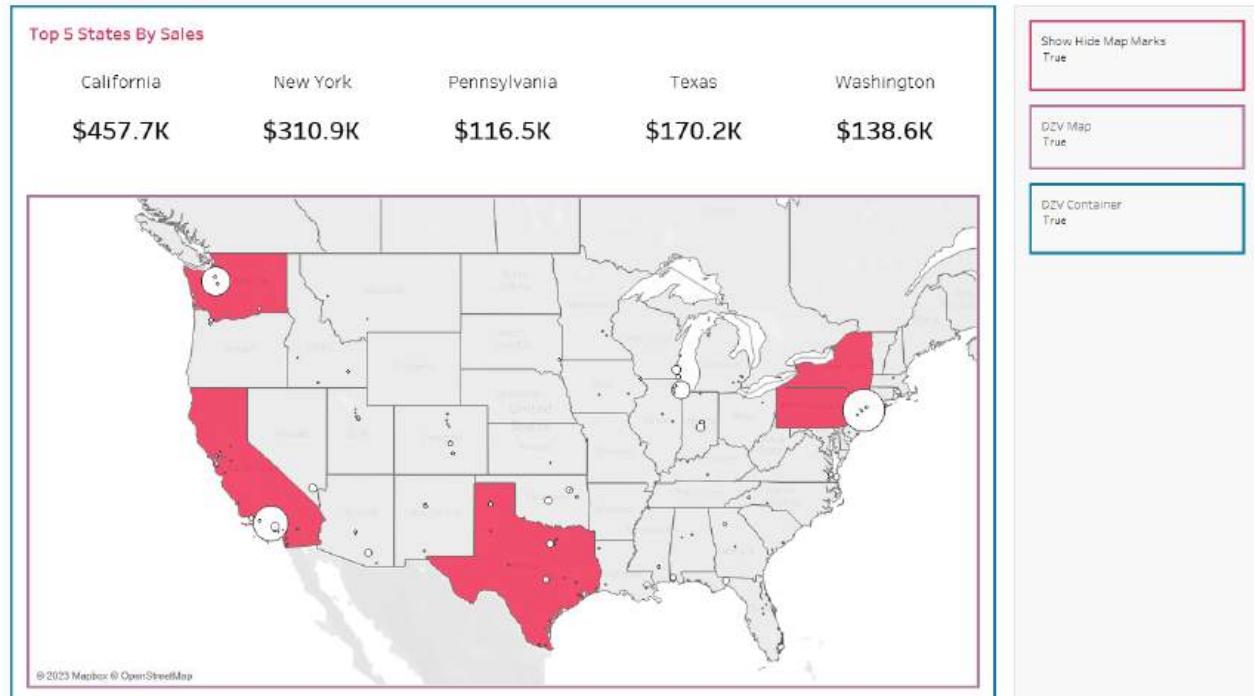


While it's not totally the same as the Tableau dashboard above, it shares similar characteristics. With a bit more code, one could replicate that Tableau graphic pretty closely, but I'm going to leave it at that for today. R's visualisation engine is potent, and there are many great guides on how to build pretty graphs using ggplot2() and Tidyverse() – and my favourite is this [guide to beautiful plotting by Cedric Scherer](#), so if you're interested in ggplot2(), you should definitely check that link out. Anyway, thanks for reading all the way through. I hope it's been useful and that you've learned something about the power of both fitzRoy and #rstats! **CJ Round-Up:** Thank you Darragh for sharing both the visual in Tableau and the bonus version in R Studio. What a fantastic way of retrieving, cleaning and visualising AFL data. As always the code has been stored on Github, you can find it at the top link under the title, as well as a link to Darraghs visual on his page. **LOGGING OFF CJ**

USING BOOLEANS (DYNAMIC ZONE VISIBILITY FOR BEGINNERS)

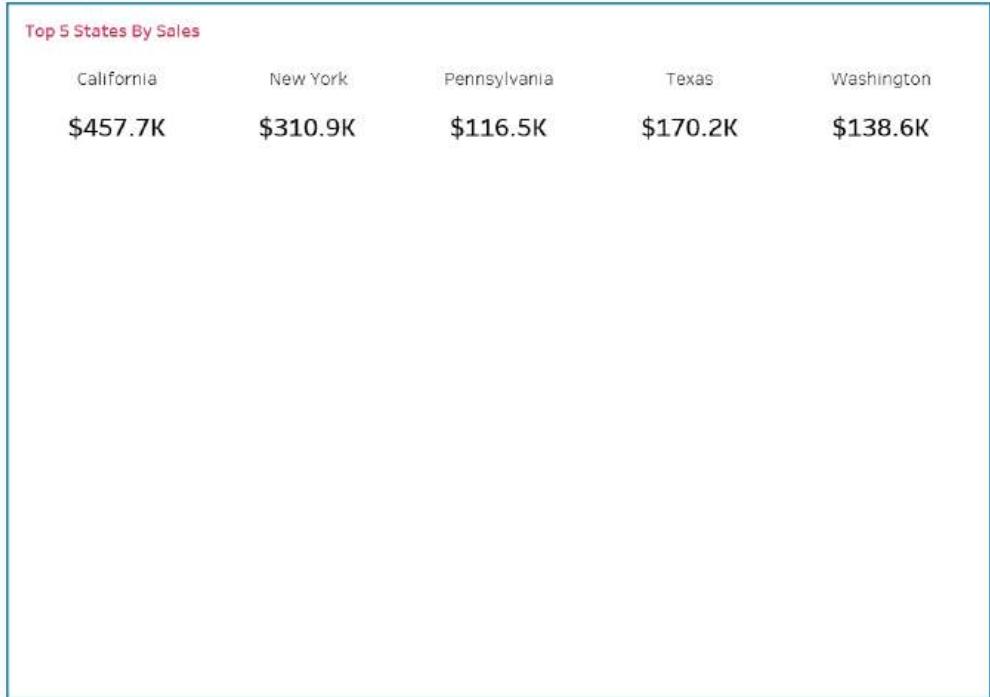
Hi all, I hope you are well? I must say, things have been really positive at the moment. I've been in a really good frame of mind recently, and found my 'burners' between work, personal commitments, sport and a new found interest of doing up my new home has brought me a lot of joy. Anyway, enough of my internal mental reflections. What are we here today for? We are going to be taking a look at the use of booleans in Tableau to show and hide aspects of a visualisation. This will be a beginners guide!

Download the workbook to follow along step by step.

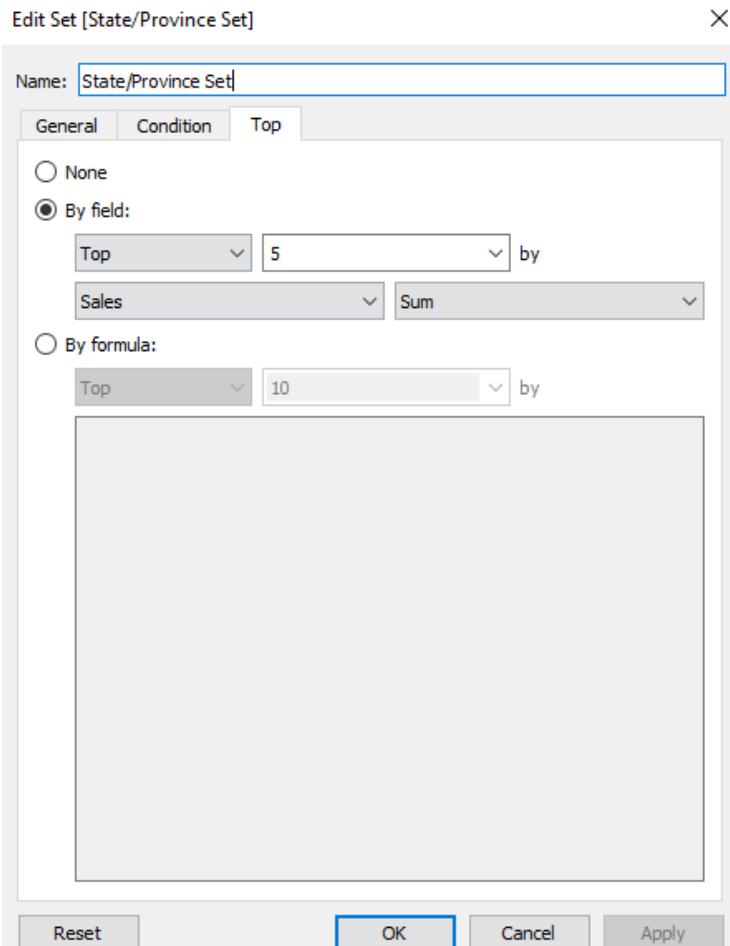


Mainly we will look at two examples. **User case 1** that looks to impact the chart itself. We will look to use a boolean to hide layers on a map! **User case 2** that uses dynamic zone visibility that we will use to hide the map, as well as hide containers. Hopefully a good intro for those just finding their feet with DZV. It goes without saying, if you want alternative

learning tutorials that vary in complexity, the likes of Marc Reid, Tim Ngwena and Sam Parsons have all released exciting material on leveraging dynamic zone visibility. A few of my favourites can be found here: Marc Reid: [Dynamic Zone Visibility – Full Dashboard Tutorial with instructions](#) Tim Ngwena [Swap sheets, containers and more! Dynamic Zone Visibility : New in Tableau 2022.3](#) Sam Parsons [Tableau Best New Feature Dynamic Zone Visibility](#) Okay lets dive in, the Tableau workbook can be downloaded for free at the top of the page. Our starting basis will be the build along dashboard – this is so that we don't need to faff with containers.



First we will look at user case 1, and create a map. Double click on the Latitude and Longitude calculations to bring them onto columns and rows. Drag country/Region onto detail as well as State/Province. By changing your mark to map you should get a filled in map, in Tableaus automatic blue colour. Off-piste, but now lets create a quick set from the state/province field called State/Province Set and select the top 5 sales field, using the sum aggregation. We will use this to colour the states that are

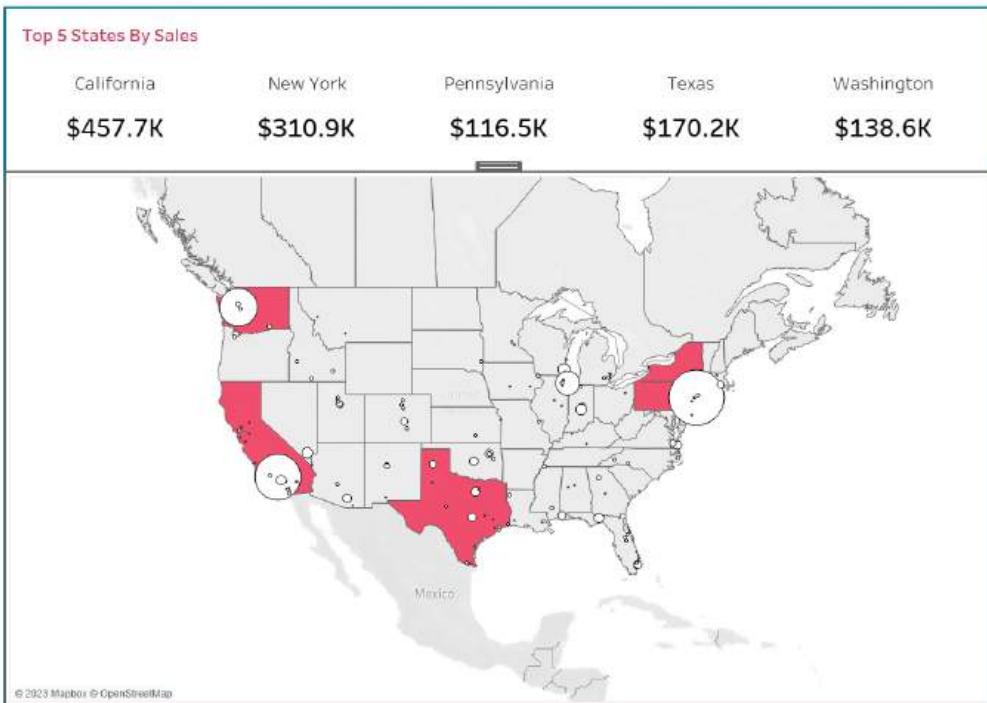


shown in our BANS. Drag that field onto colour. Next thing we want to do is add a layer that we can turn on and off at the click of a filter. To do that we need a parameter. Create a parameter **P. Show Hide Map Marks**. It should be a boolean.

Edit Parameter [P. Show Hide Map Marks]

Name	P. Show Hide Map Marks	
Properties		
Data type	Display format	
Boolean	True	
Current value	Value when workbook opens	
<input checked="" type="radio"/> True	<input type="radio"/> False	Current value
Aliases		
True	True	
False	False	
Cancel		OK

Next we want to create a layer that is driven by this parameter. Create a calculation MP. City if then END. This means when the parameter is true, use the city field (a geographic location). Add Sales to size and recolour in your choosing. Re-adjust the map to fit the center of the screen, and then this is ready to add to our dashboard.



Click on the sheet, go to parameters and click our new parameter of Show Hide Map Marks. Now as you toggle between True (showing) and False (Hidden) The map marks from our layer will change between showing and hiding. How simple was that! Our first user case of using a boolean to show and hide content. Lets look at the next user case.



Can we use a boolean to show and hide the map at a sheet level? Lets create a new parameter. Again, a boolean field. Call this

Edit Parameter [P. DZV Map]

X

Name

P. DZV Map

Properties

Data type

Boolean

Display format

True

Current value

True False

Value when workbook opens

Current value

Aliases

True True

False False

Cancel

OK

Add this

parameter DZV Map.

parameter to our dashboard by clicking on our map sheet, going to parameter and clicking DZV Map.

Top 5 States By Sales

State	Sales
California	\$457.7K
New York	\$310.9K
Pennsylvania	\$116.5K
Texas	\$170.2K
Washington	\$138.6K

© 2023 Mapbox © OpenStreetMap

Show Hide Map Marks
True

DZV Map
True

- Go to Sheet
- Duplicate Sheet
- Title
- Caption
- Legends
- Filters
- Sets
- Highlighters
- Parameters**
 - Show Hide Map Marks
 - P. DZV Map**
 - P. Show Hide Map Marks**
- Use as Filter
- Ignore Actions
- Add Show/Hide Button
- Floating
- Fix Height
- Edit Height...
- Select Contain in Vertical
- Direct
- Remove from Dashboard
- Rename Dashboard Item...

But at this point our map has nothing that links it to the parameter. This is where dynamic zone visibility comes into play.



If we highlight the sheet and then go to layout, we can tick the control visibility using value and use our newly created DZV Map boolean. When you change the parameter to false, the map will hide from the screen. Lets look at one final user case, what if we wanted to completely hide the whole container in blue? We should follow the same logic. Create a final new parameter this time

Edit Parameter [P. DZV Container]

X

Name

P. DZV Container

Properties

Data type

Boolean

Display format

True

Current value

True False

Value when workbook opens

Current value

Aliases

True: True

False: False

Cancel

OK

called **DZV Container**

same steps from previous to add the parameter to our dashboard

Follow the



This time we will want to select our container (blue) when you're on the container level. Go to Control Visibility Using Value and this time click on our new parameter DZV Container.



Flicking the boolean to false hides the whole container.



That's it for this week. Hopefully an easy one for beginners to follow along to and try different ways of leveraging booleans to show and hide both layers, sheets, and containers. I would really recommend checking out the videos referenced at the start of the blog to take this to the next step. GOING FURTHER

- Why not try and use parameter and DZV to show and hide multiple sheets using icons.

Speak soon, **LOGGING OFF, CJ**

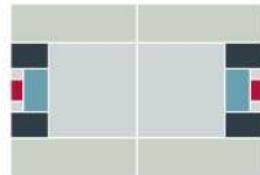
SOCKER SHOT ZONES IN TABLEAU

Hi all, Been on a bit of a soccer streak recently blog wise. Hoping to continue that idea with this blog. Today's blog will look to create 'Shot Zones' using an Alteryx workflow and the output visualised in Tableau. You can see what we will end up with below.

How To Build A Shot Zone Pitch

Tutorial on CJ-Mayes.com

Circles Sized on xG | ● Goal | ○ Shot Missed
----- Pitch Split Into 6 Zones -----



1 [] 10 Data: Understat - Manchester United (4.92 xG) 4-1 win against Chelsea (1.68 xG) on May 25th 2023

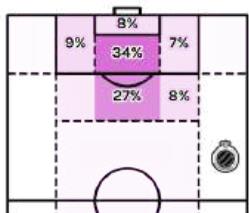
Where did this idea originate? Well I came across [this paper](#) looking at flanks and out of the box shots and started digging a bit further. This lead me to this [shot matrix created](#) using Opta data. This concept isn't that new, in fact I came across this beautiful visual by sonofacorner below.

Shooting Areas in the Jupiler Pro League

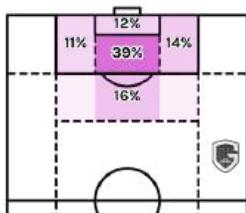
Percentage of non-penalty shots taken in each area. Viz by @sonofacorner.

Only areas with more than 5% of shots are labeled.

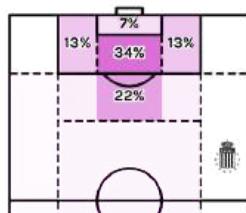
Club Brugge | shots 105



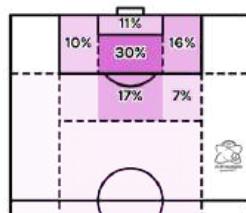
Genk | shots 93



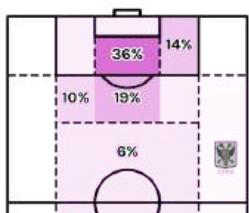
Sporting Charleroi | shots 83



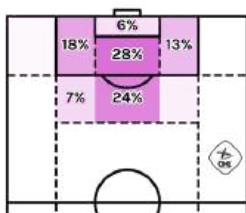
Zulte Waregem | shots 70



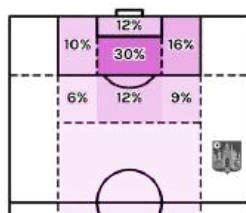
St.Truiden | shots 69



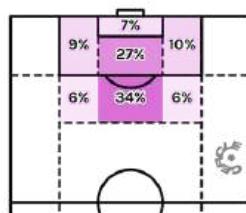
OH Leuven | shots 68



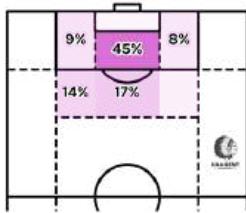
Westerlo | shots 67



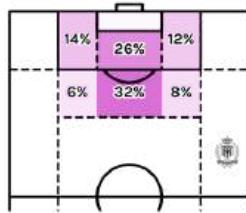
Cercle Brugge | shots 67



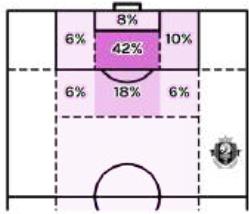
Gent | shots 66



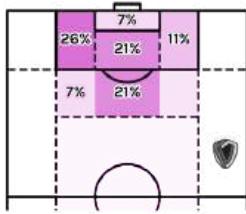
KV Mechelen | shots 65



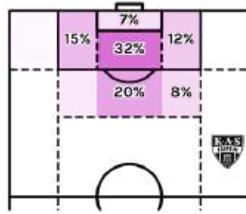
RFC Seraing | shots 62



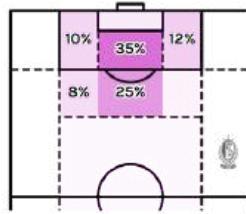
Oostende | shots 61



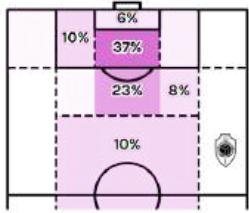
Eupen | shots 60



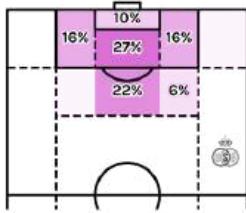
Standard Liege | shots 60



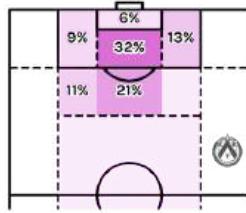
Royal Antwerp | shots 52



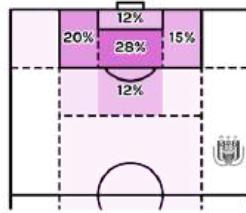
Union St.Gilloise | shots 51



Kortrijk | shots 47



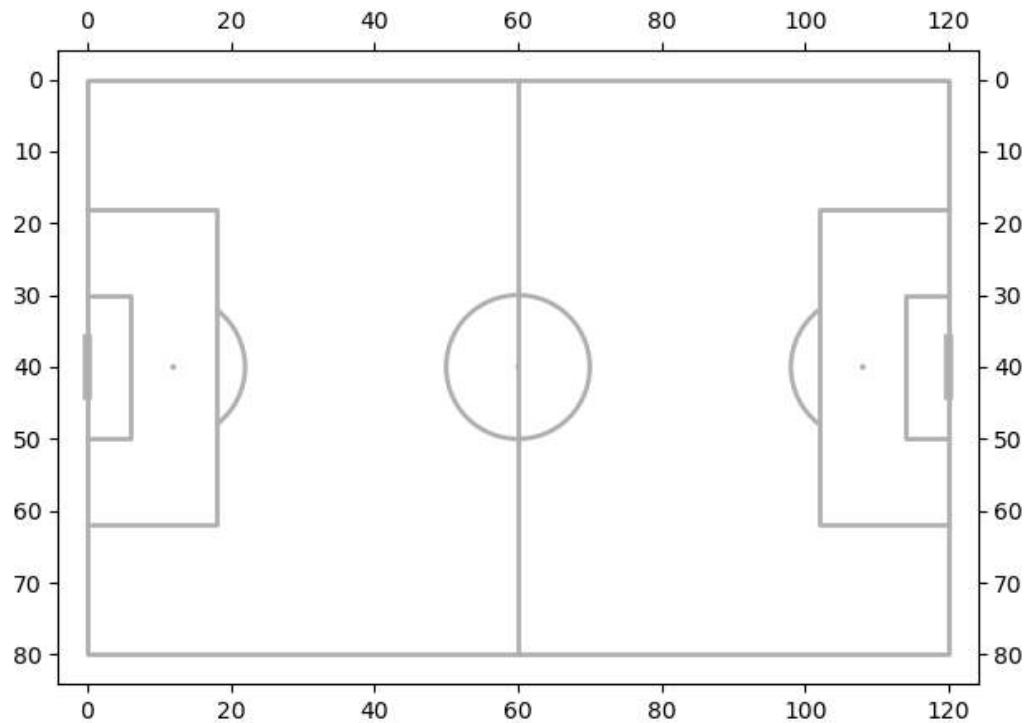
Anderlecht | shots 40



Which brings us onto how to create it, Realistically to get to our finished point we need to follow the steps below:

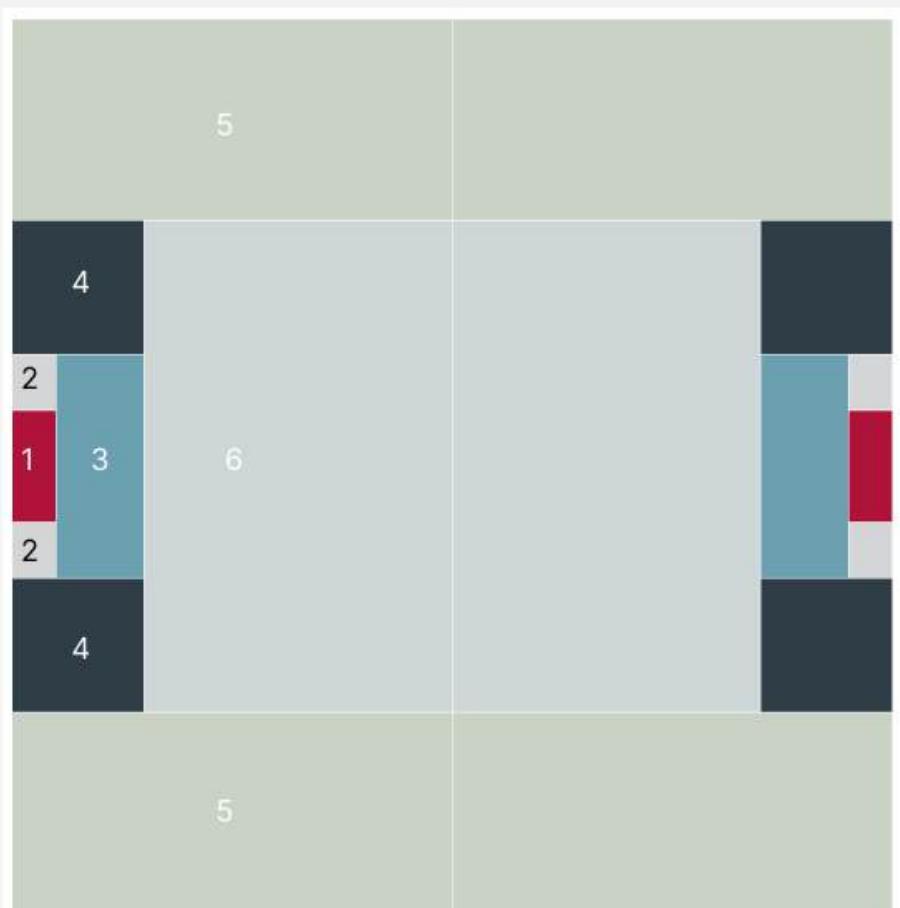
- Figure out the X & Y locations of the pitch and the shots data
- Draw a path that creates polygons for the different shot boxes
- Find out what zone each shot is in
- Visualise that data on a pitch

Seems simple enough? First thing we do is figure out the shape of our zones. I got the dimensions of a pitch from the [mplsoccer](#) documentation.



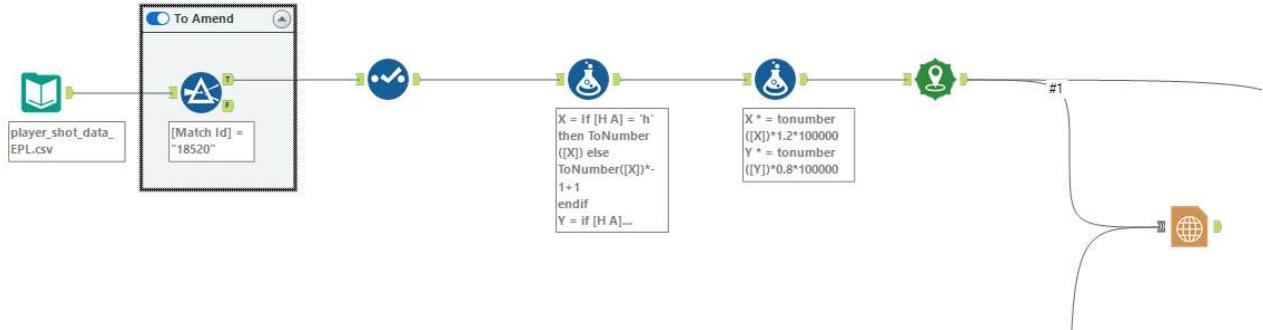
[documentation](#).

Then I decide what the zones would be split into. You will see that it contains 6 zones.



The beauty about the workflow is you will only need to amend the input data for the shot locations at the start to be able to change how the output looks. In this case i've broken the pitch into 6 zones. Zone 1 is in front of the goal. Zone 2 is in the 6 yard box but not in line with the goal, Zone three backfills to the top of the box. Zone 4 either side fills the remainder of the box. Zone 5 is within the flanks, and Zone

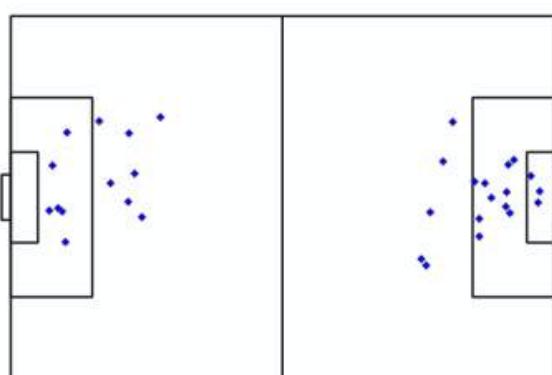
6 is anything behind the box. Perhaps if you were to take this further you may split zone 6 into a 7th zone if you want to emphasize shots taken just outside of the box. Given the above diagram, this can easily be done! Let's look at the Alteryx Workflow.



We connect to our EPL shot location data and choose a specific match, in this case I've chosen the match between Manchester United and Chelsea [here](#).



The top part of the flow is used to split the home and away shots in two to mirror the above. The second formula tool just scales up our values to be able to fit on the pitch. The really important part here is the spatial tool of create points where we can start



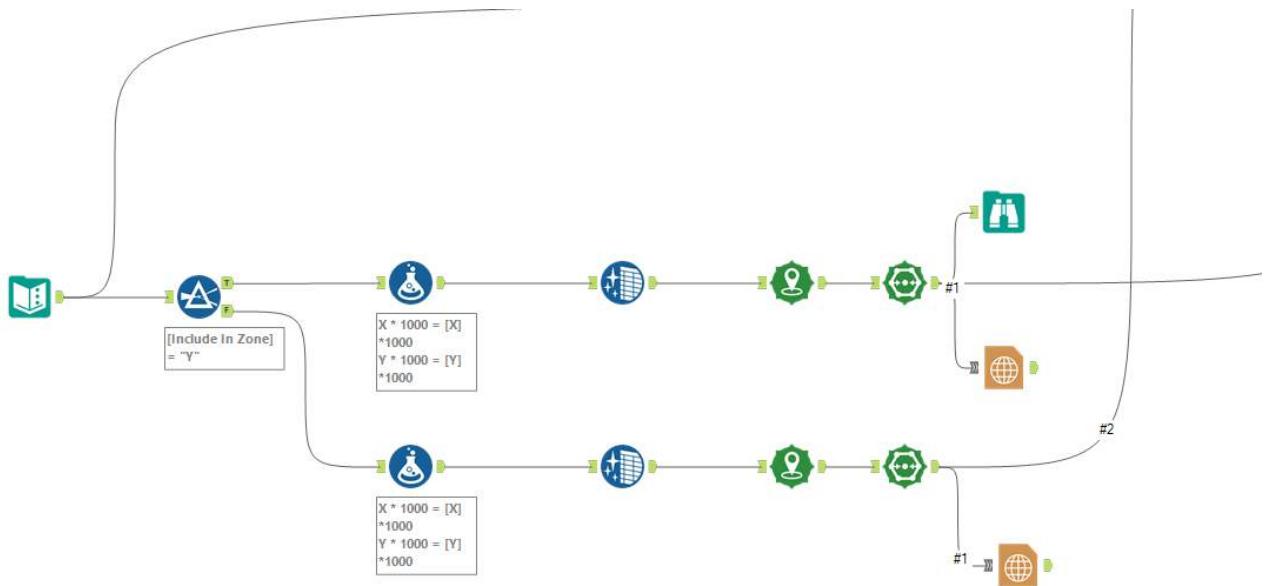
to see our points mapped out.

The bottom part of the flow

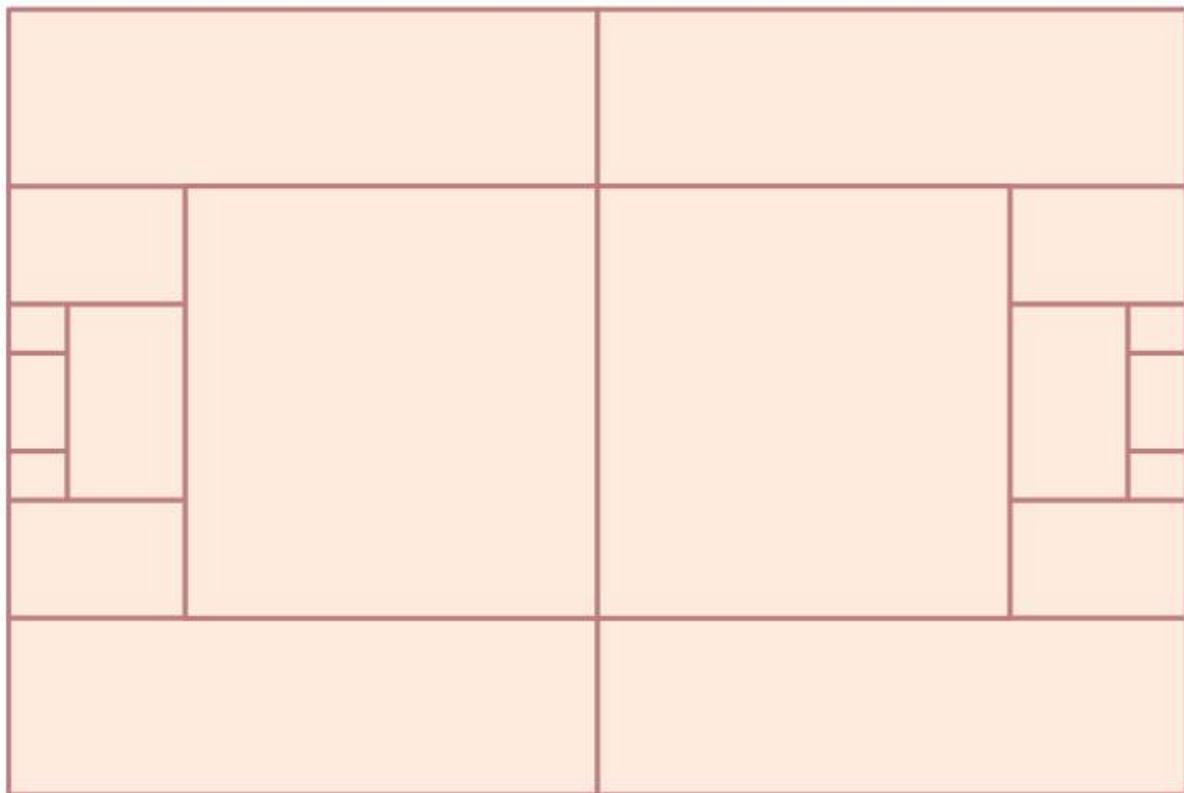
	X	Y	Gr	SN	Pitch Part	Include In Zone	Zone Label	*
1	0	0	1	99	Left Half Pitch	N		
2	0	80	2	99	Left Half Pitch	N		
3	60	80	3	99	Left Half Pitch	N		
4	60	0	4	99	Left Half Pitch	N		
5	0	0	5	99	Left Half Pitch	N		
6	60	0	1	98	Right Half Pitch	N		
7	60	80	2	98	Right Half Pitch	N		
8	120	80	3	98	Right Half Pitch	N		
9	120	0	4	98	Right Half Pitch	N		
10	60	0	5	98	Right Half Pitch	N		
11	0	30	1	97	Left 6 Yard Box	N		
12	0	50	2	97	Left 6 Yard Box	N		
13	6	50	3	97	Left 6 Yard Box	N		
14	6	30	4	97	Left 6 Yard Box	N		
15	0	30	5	97	Left 6 Yard Box	N		
16	0	35	1	96	Left Goal	N		
17	0	45	2	96	Left Goal	N		
18	-2	45	3	96	Left Goal	N		
19	-2	35	4	96	Left Goal	N		
20	0	35	5	96	Left Goal	N		
21	120	35	1	95	Right Goal	N		
22	120	45	2	95	Right Goal	N		
23	122	45	3	95	Right Goal	N		
24	122	35	4	95	Right Goal	N		
25	120	35	5	95	Right Goal	N		
26	120	30	1	94	Right 6 Yard Box	N		
27	120	50	2	94	Right 6 Yard Box	N		
28	114	50	3	94	Right 6 Yard Box	N		
29	114	30	4	94	Right 6 Yard Box	N		
30	120	30	5	94	Right 6 Yard Box	N		
31	0	18	1	93	Left Box	N		
32	0	62	2	93	Left Box	N		
33	18	62	3	93	Left Box	N		
34	18	18	4	93	Left Box	N		
35	0	18	5	93	Left Box	N		
36	120	18	1	92	Right Box	N		
37	120	62	2	92	Right Box	N		
38	102	62	3	92	Right Box	N		
39	102	18	4	92	Right Box	N		
40	120	18	5	92	Right Box	N		
41	0	35	1	91	Left Zone 1	Y	1	
42	0	45	2	91	Left Zone 1	Y	1	
43	6	45	3	91	Left Zone 1	Y	1	

So

the input data here is all the pitch points (Where Include in Zone = N) as well as our mapped out zones (Where Include in Zone = Y) We then split these apart (as we will later want to say which of the zones the shot is taken in) The formula tool scales our pitch to the same amount as our shot data.



The flow mirrors each other, the top section building the 'zones' the bottom part drawing our pitch outlines. Create points and build spatial tools help us build our pitch and zones. Here is what our zones output looks like:



and

here is our basic pitch drawing:

Report Map (148) - Configuration

Preview (Click to enlarge)

The window displays a map preview with two rectangular zones. A scale bar at the bottom indicates distances from 0 to 1.8 miles. Below the map is a 'Refresh' button. The settings panel contains the following configuration:

Settings	Data	Layers	Legend
Map Size (W x H):	8	by	6 Inches
Resolution:	1x (96 dpi)		
Scale:	Miles		
Reference Base Map:	[None]		
Background Color:	R=253, G=254, B=255		
Map Drop Shadow:	No		
Expand Extent:	10	% w/ a Minimum Width of	0.5 Miles

Both being visualised using the report map tool. The next step is to find which zone the shot intersects so we can use the spatial match tool.

Targets (T Input)

Spatial Object Field: **Centroid**

Universe

Use Records from U Input

Use Records from File or Database:

Spatial Object Field: **SpatialObj_Built**

Where Target Intersects Universe

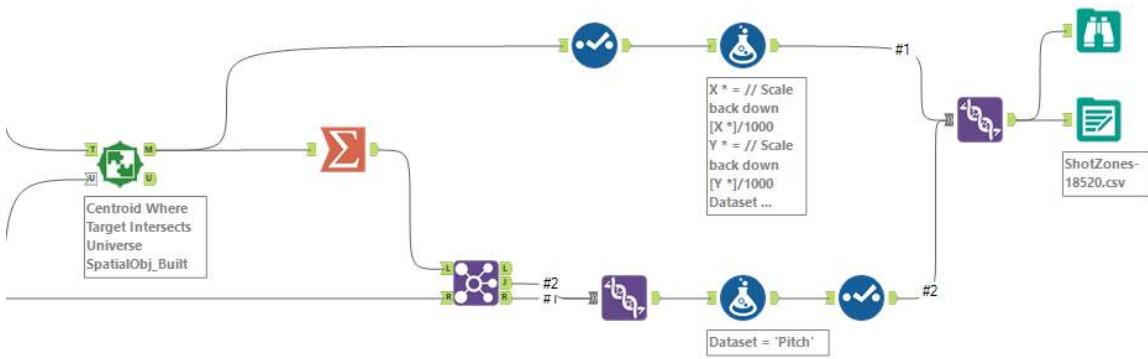
Output Intersection Object (Intersects Only)

Options ▾ |

TIP: To reorder multiple rows: select, right-click and drag.

	Input	Field	Type	Size	Rename	Description
<input checked="" type="checkbox"/>	Target	A Team	V_WString	254		
<input checked="" type="checkbox"/>	Target	Date	V_WString	254		
<input checked="" type="checkbox"/>	Target	H A	V_WString	254		
<input checked="" type="checkbox"/>	Target	H Team	V_WString	254		
<input checked="" type="checkbox"/>	Target	Id	V_WString	254		
<input checked="" type="checkbox"/>	Target	Match Id	V_WString	254		
<input checked="" type="checkbox"/>	Target	Player	V_WString	254		
<input checked="" type="checkbox"/>	Target	Player Assisted	V_WString	254		
<input checked="" type="checkbox"/>	Target	Player Id	V_WString	254		
<input checked="" type="checkbox"/>	Target	Result	V_WString	254		
<input checked="" type="checkbox"/>	Target	Season	V_WString	254		
<input checked="" type="checkbox"/>	Target	Shot Type	V_WString	254		
<input checked="" type="checkbox"/>	Target	Situation	V_WString	254		
<input checked="" type="checkbox"/>	Target	A Goals	V_WString	254		
<input checked="" type="checkbox"/>	Target	H Goals	V_WString	254		
<input checked="" type="checkbox"/>	Target	Minute	V_WString	254		
<input checked="" type="checkbox"/>	Target	X	V_WString	254		
<input checked="" type="checkbox"/>	Target	X G	V_WString	254		
<input checked="" type="checkbox"/>	Target	Y	V_WString	254		
<input checked="" type="checkbox"/>	Target	X *	Double	8		
<input checked="" type="checkbox"/>	Target	Y *	Double	8		
<input checked="" type="checkbox"/>	Target	Centroid	SpatialObj	536...		
<input checked="" type="checkbox"/>	Universe	Pitch Part	String	16		
<input checked="" type="checkbox"/>	Universe	SpatialObj_Built	SpatialObj	536...		
<input checked="" type="checkbox"/>		*Unknown	Unknown	0		Dynamic or Unknown Fields

Side note, This may be my new favourite Alteryx tool!!



The final bit of the flow looks to summarise what zone our points in. This is at an aggregated level. I group this on team, so we can split our zones out for Manchester United and Chelsea

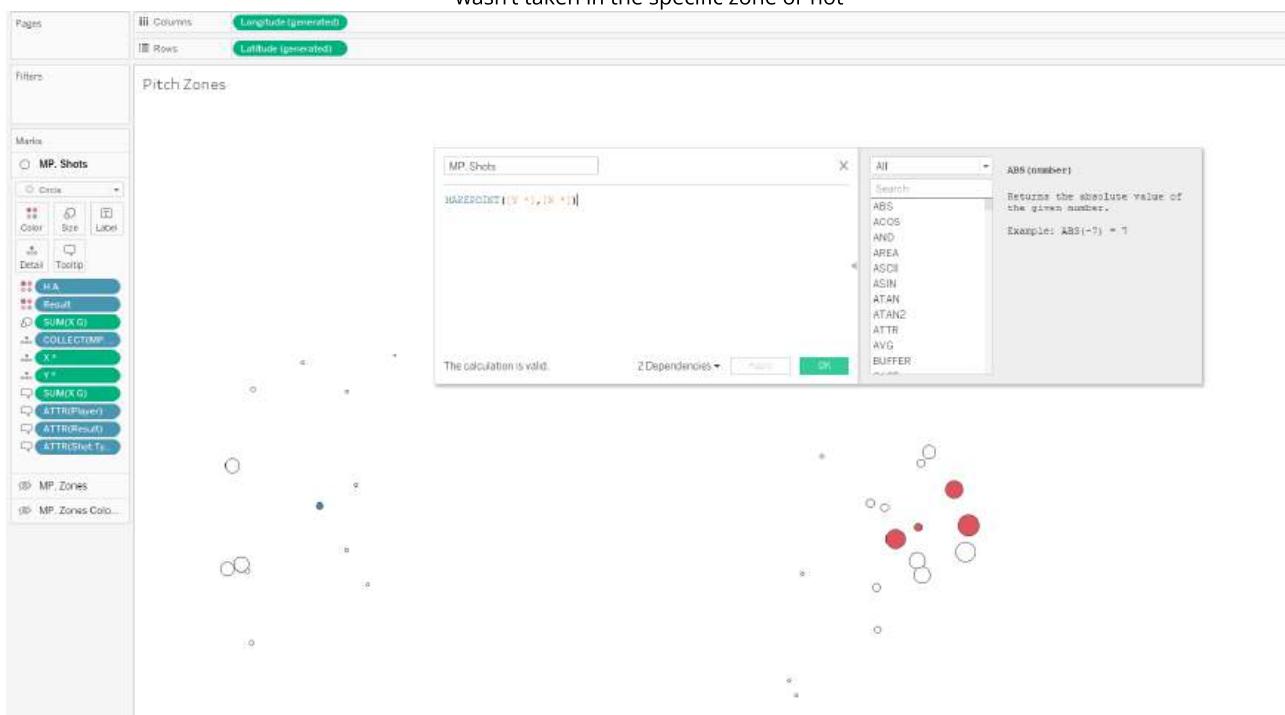
3 of 3 Fields | Cell Viewer | 6 records displayed | ↑ ↓ |

Record	H A	Pitch Part	Count
1	a	Left Zone 3	5
2	a	Left Zone 4A	1
3	a	Left Zone 6	7
4	h	Right Zone 1	3
5	h	Right Zone 3	10
6	h	Right Zone 6	5

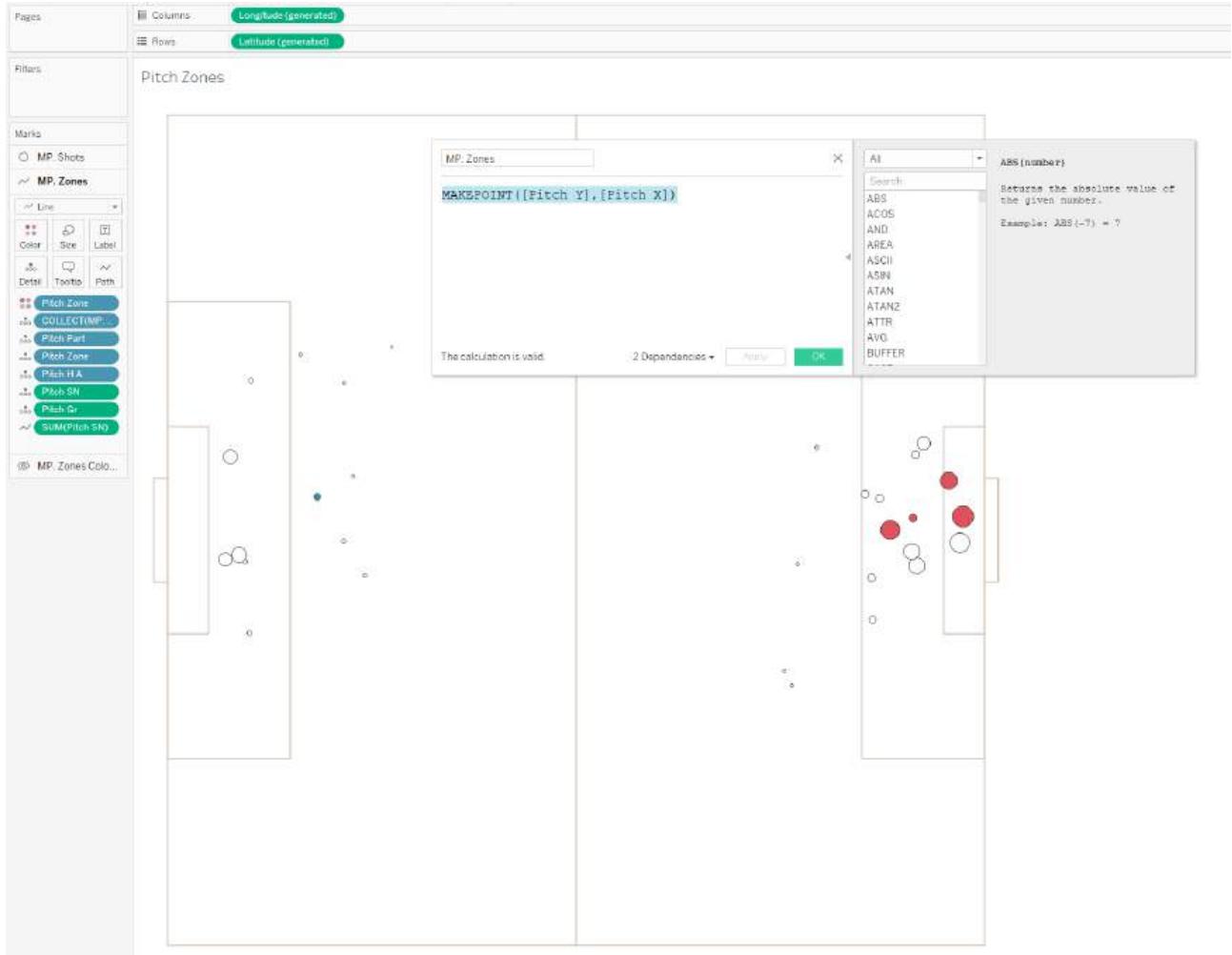
To make life easier in Tableau later i do a bit of a funky full outer join to be able to have row level shot data but also then our aggregated pitch polygon data.

H Goals	Minute	X	X G	Y	Z	X*	Y*	Dataset	Pitch X	Pitch Y	Pitch Gr	Pitch SN	Pitch Part	Pitch Zone
26	4	51	0.915999985	0.124078430	0.590999985	109.919998	47.279999	Raw	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
27	4	83	0.954000015	0.120901421	0.542999992	102.480002	43.439999	Raw	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
28	4	51	0.771999969	0.019588637	0.459000015	92.659995	35.720001	Raw	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
29	4	92	0.918000031	0.0557787061	0.457000008	110.160000	36.560001	Raw	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
30	4	46	0.961999969	0.123000711	0.442000008	103.419995	35.360001	Raw	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
31	4	72	0.985	0.161168838	0.5	106.2	40	Raw	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]
32	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	Pitch	0	30	1	97	Left 6 Yard Box	N
33	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	[Null]	Pitch	0	50	2	07	Left 6 Yard Box	N

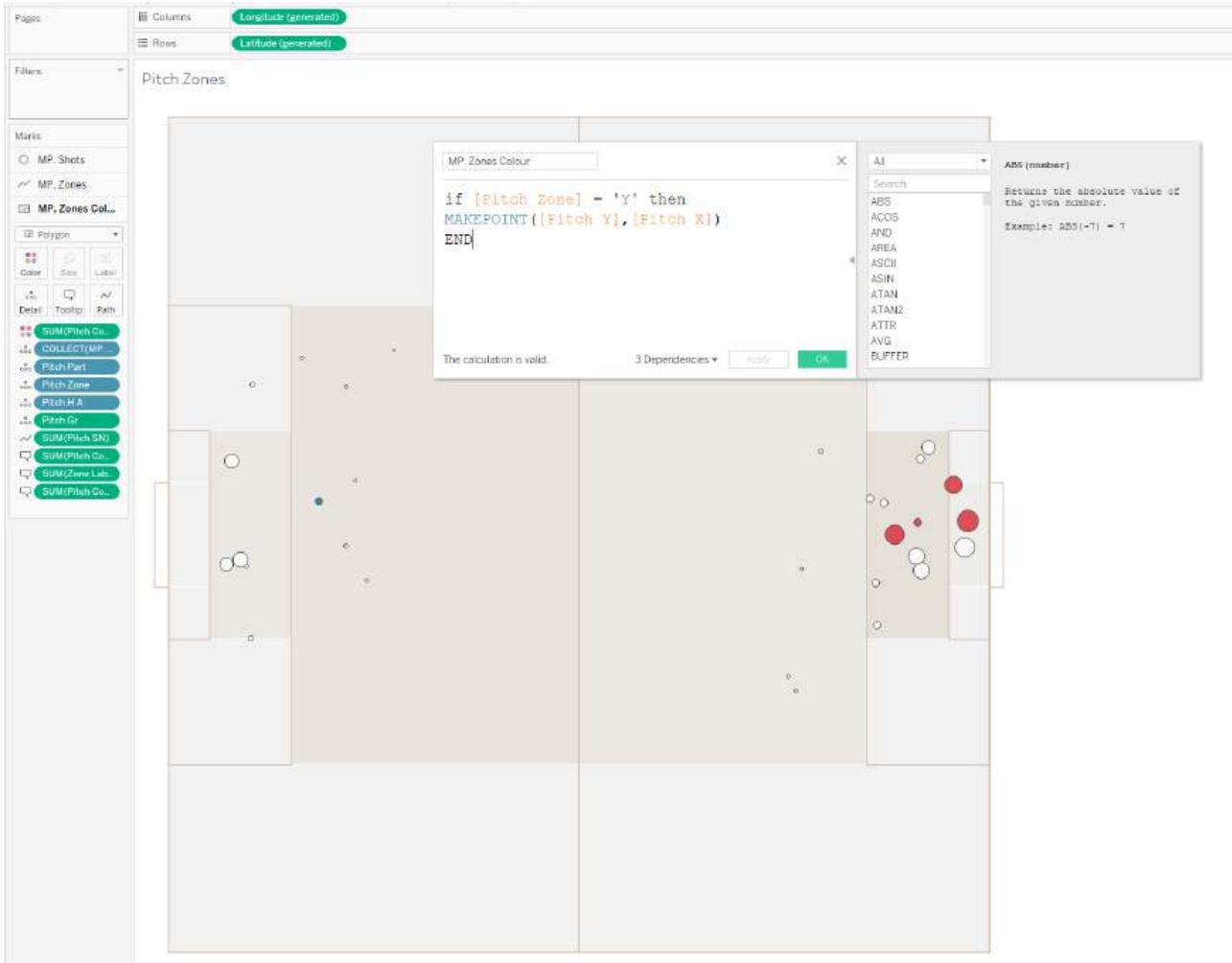
You can see how I've offset the stacking of the data. Do note also I've brought in all the pitch details too regardless if a shot wasn't taken in the specific zone or not



We can create makepoints of our x and y co-ordinates of the shots to plot the different teams. You will see they resemble the chosen match from Understat.



We can then use our pitch X & Y co ordinates we've created to draw the outline of the pitch. Make sure the SN (Sequence Number) goes onto the path, and you have the correct pills on the details card from the screenshot.



The final layer we want to add is our Zones Layer – Here we write an IF statement to look at the original data that is our zone polygons, we then plot these points in a similar way, again, with SN on path and the required dimensions on detail. By adding our aggregated count to colour we can see which zones were most frequently shot from. And that's it! The toughest part to this was really creating the initial dataset for the zones but once that's done it is fairly smooth sailing. GOING FURTHER:

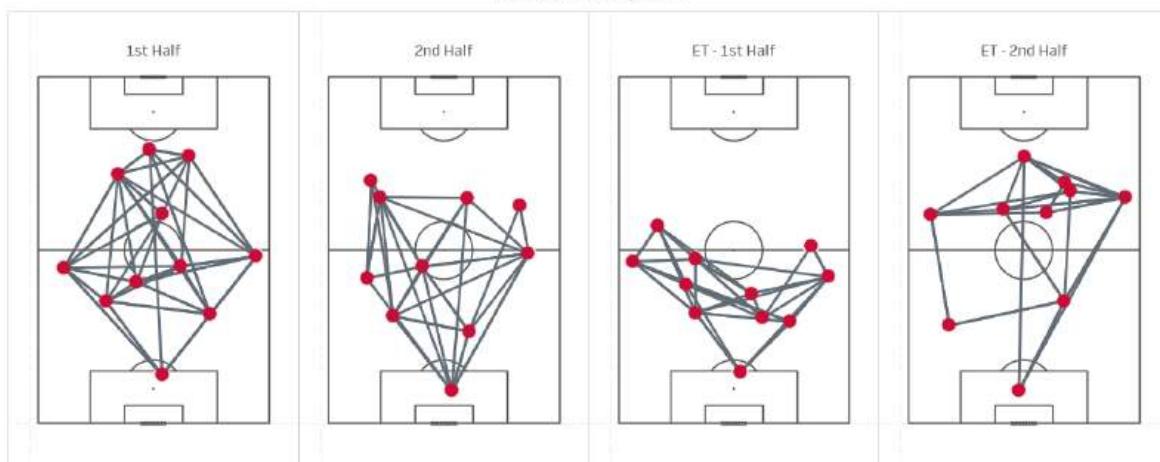
- Why not change the zones you look at?
- Try map all of one teams games as a small multiple
- Try adding more layers with details around the score, team and xG
- Try other dimensions on colour – perhaps xG from each zone may be interesting?

Message me on LinkedIn or Twitter if you have any questions – as always the resources can be found on Github! Take care & **LOGGING OFF, CJ**

TIMEFRAME PASS NETWORK IN TABLEAU PT2

Hi all, Hope everyone is doing well. Todays blog is a bit of a mix between technical and thought piece. I've previously written about how to create pass networks using alteryx and tableau [here](#).

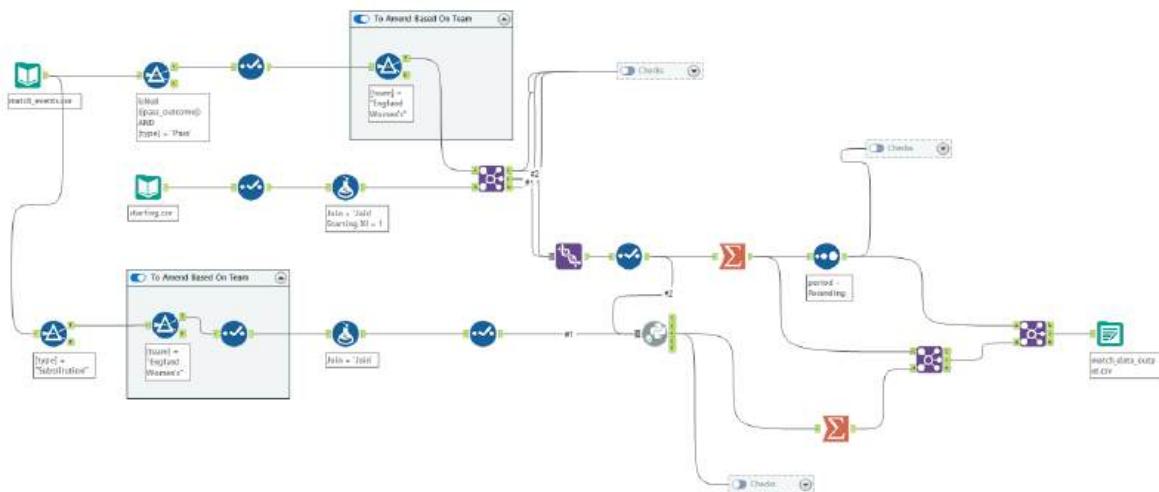
Women's Euro Final | England Pass Network
 31 July 2022, 17:00 England 2-1 Germany
 Resources on CJ-Mayes.com



Please read that one before moving on as i'm going to skip over the bits of data collection and talk mainly about the data processing in the Alteryx workflow. As a brief refresher, all of todays files can be found in the GitRepo (under the header) We will then transform the data using Alteryx, and finally we will export or data and showcase how it can be used to create a pass network in Tableau. We won't refresh how to do the Tableau build as the other blog covers it off, and I followed it to make my charts for this week! For those that want to relive the moment of England beating Germany in the Euros Final in 2022, you can [watch it here.](#) Okay? lets begin! Once you've run the code you will end up with the following outputs for the Womens Euro Final from 2022:

- match_events.csv – All match events associated to the chosen match ID (Where we will get all our pass information from)
- starting.csv – The starting line up for each team
- fig.png – the background football pitch

So with little transformation I thought best to do the rest of the prep in Alteryx. This is where my workflow has changed from previous as we now want to build the same chart multiple times over. **I wanted to now look at networks during the game given certain time frames. For this example I will look at the first half, second half, and then extra time 1st and second half.** let's talk through the process of how the Alteryx flow works



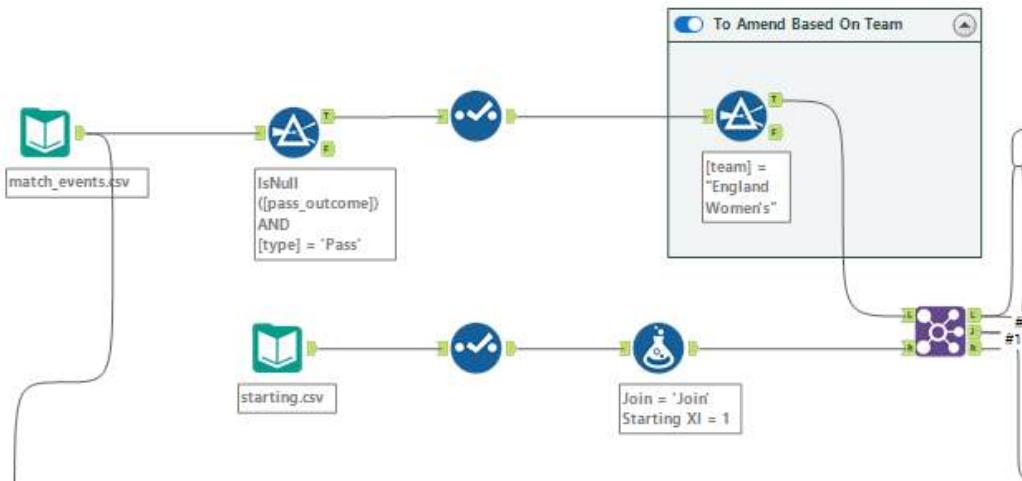
The top part of the flow connects to the raw match event data. We end up filtering for all successful passes (where the outcome is blank) and where the type of event is a pass. I use the select tool just to pick out the columns from the data that are most appropriate and filter this data just to have the England teams information in.

minute	second	pass_recipient	period	player	team	location_x	location_y	pass_end_location_x	pass_end_location_y
0	0	Millie Bright	1	Bethany Mead	England Women's	61	40.1	35.7	45.3
0	10	Rachel Daly	1	Lauren Hemp	England Women's	92.1	7.3	78.5	6.3
0	13	Keira Walsh	1	Rachel Daly	England Women's	76.6	6	71.1	15.0
0	15	Leah Williamson	1	Keira Walsh	England Women's	69.7	16.6	56.7	9.2
0	18	Mary Alexandra Earps	1	Leah Williamson	England Women's	57	9.4	30.1	36.3
0	23	Leah Williamson	1	Mary Alexandra Earps	England Women's	36.5	39.2	36.8	21.4
0	26	Millie Bright	1	Leah Williamson	England Women's	36.8	21.4	35.4	58.4
0	31	Bethany Mead	1	Millie Bright	England Women's	43.9	64.2	84.6	77.0

The second input takes our Starting XI and joins it to our main dataset, flagging who was in our original Starting XI. We don't use this metric, but I think it could be useful in general to have.

minute	second	pass_recipient	period	player	team	location_x	location_y	pass_end_location_x	pass_end_location_y	Right_Join	Starting XI
0	0	Millie Bright	1	Bethany Mead	England Women's	61	40.1	35.7	45.3	Join	True
18	8	Lucy Bronze	1	Bethany Mead	England Women's	120	0.1	105.9	58.6	Join	True
18	34	Lucy Bronze	1	Bethany Mead	England Women's	120	0.1	111.2	48.5	Join	True
19	37	Lucy Bronze	1	Bethany Mead	England Women's	120	0.1	109.1	52.3	Join	True
27	32	Lauren Hemp	1	Bethany Mead	England Women's	84.5	61.3	94.8	63.9	Join	True
35	40	Lucy Bronze	1	Bethany Mead	England Women's	73.5	73.8	69.5	73.3	Join	True
36	57	Lucy Bronze	1	Bethany Mead	England Women's	64.4	67.9	66.3	69.3	Join	True
37	61	Ellen White	1	Bethany Mead	England Women's	116.3	66.3	105.6	46.4	Join	True

So now the top part of the flow has for our match, including a flag for if they were in the starting 11, all the pass match events.



The bottom part

of the flow takes our original data but this time looks at when the substitutions happen. Now previously, I ended up just taking all events prior to the first substitution. This isn't really a practical solution for when a sub happens early on in the game so decided to revisit the idea. We can cross reference our sub data externally, [here](#). This time round, we want to look at all the



substitutions.

Record	minute	second	period	player
1	54	55	2	Francesca Kirby
2	54	59	2	Ellen White
3	62	55	2	Bethany Mead
4	87	45	2	Rachel Daly
5	87	56	2	Georgia Stanway
6	119	16	4	Lauren Hemp

So the problem that now arises is if we are to build network maps based on each period, we would have points for anyone who has been on the pitch, in this case could be up to 16 players for England in the second half. My solution is to take out the subbed off player from that time period. Its not a full solution. Because lets face it dependant when during that half you're subbed massively impacts how many times you're like to touch the ball. Perhaps if I was to do this on minute segmentation I would look at number of passes made and visualise the player of the subbed dependent on impact, but for now the solution works. Here's how that looks in alteryx:

```
Run Alteryx.help() for info about useful functions.  
i.e., Alteryx.read("#1"), Alteryx.write(df,1),  
Alteryx.getWorkflowConstant("Engine.WorkflowDirectory")
```

```
In [1]: # List all non-standard packages to be imported by your  
# script here (only missing packages will be installed)  
from ayx import Package  
#Package.installPackages(['pandas', 'numpy'])
```

```
In [2]: from ayx import Alteryx  
import pandas as pd
```

```
In [3]: df2 = Alteryx.read("#1")  
SUCCESS: reading input data "#1"
```

```
In [4]: df1 = Alteryx.read('#2')  
SUCCESS: reading input data "#2"
```

```
In [5]: Alteryx.write(df1, 1)  
SUCCESS: writing outgoing connection data 1
```

Out[5]:

	minute	second	pass_recipient	period	player	team	location_x	location_y	pass_end_x
0	64	23	Lauren Hemp	2	Alessia Russo	England Women's	96.5	54.9	
1	80	1	Ella Toone	2	Alessia Russo	England Women's	60.0	40.0	
2	92	14	Alex Greenwood	3	Alessia Russo	England Women's	62.0	14.0	
3	97	38	Keira Walsh	3	Alessia Russo	England Women's	51.1	49.3	
4	101	38	Alex Greenwood	3	Alessia Russo	England Women's	57.8	19.0	
...									

This build in python IDE reads the data we've transformed.

```
In [7]: filtered_rows = [] # To store the filtered rows
```

```
In [8]: for index, row in df1.iterrows():
    # Check if there is a match in the second dataframe
    if (row['player'] in df2['player'].values) and (row['period'] in df2['period']):
        continue # Skip the row if it matches the criteria
    else:
        filtered_rows.append(row) # Add the row to the filtered_rows list
```

```
In [9]: # Create a new dataframe with the filtered rows
filtered_df = pd.DataFrame(filtered_rows, columns=df1.columns)
```

```
In [10]: filtered_df
```

Out[10]:

	minute	second	pass_recipient	period	player	team	location_x	location_y	pass_end_L
0	64	23	Lauren Hemp	2	Alessia Russo	England Women's	96.5	54.9	
1	80	1	Ella Toone	2	Alessia Russo	England Women's	60.0	40.0	
2	92	14	Alex Greenwood	3	Alessia Russo	England Women's	62.0	14.0	
3	97	38	Keira Walsh	3	Alessia Russo	England Women's	51.1	49.3	
4	101	38	Alex Greenwood	3	Alessia Russo	England Women's	57.8	19.0	
...
352	36	38	Keira Walsh	1	Rachel Daly	England Women's	28.8	25.4	
353	36	40	Mary Alexandra Earps	1	Rachel Daly	England Women's	28.0	26.7	
354	38	55	Leah Williamson	1	Rachel Daly	England Women's	49.5	20.3	

It

then does a for loop to say that if from our subbed off players are equal to the time period and player name then remove them from the dataset. Do this for all the players and times listed in our subs dataset. I'm sure this probably could have been an iterative macro somehow.

```
In [15]: Alteryx.write(filtered_df_pr, 3)
```

```
SUCCESS: writing outgoing connection data 3
```

Out[15]:

	minute	second	pass_recipient	period	player	team	location_x	location_y	pass_end_k
0	80	1	Ella Toone	2	Alessia Russo	England Women's	60.0	40.0	
1	92	14	Alex Greenwood	3	Alessia Russo	England Women's	62.0	14.0	
2	97	38	Keira Walsh	3	Alessia Russo	England Women's	51.1	49.3	
3	101	38	Alex Greenwood	3	Alessia Russo	England Women's	57.8	19.0	
4	106	8	Chloe Kelly	4	Alessia Russo	England Women's	95.6	18.0	
...
306	36	38	Keira Walsh	1	Rachel Daly	England Women's	28.8	25.4	
307	36	40	Mary Alexandra Earps	1	Rachel Daly	England Women's	28.0	26.7	
308	38	55	Leah Williamson	1	Rachel Daly	England Women's	49.5	20.3	
309	43	30	Ellen White	1	Rachel Daly	England Women's	96.1	0.1	
310	45	56	Lauren Hemp	1	Rachel Daly	England Women's	51.1	0.1	

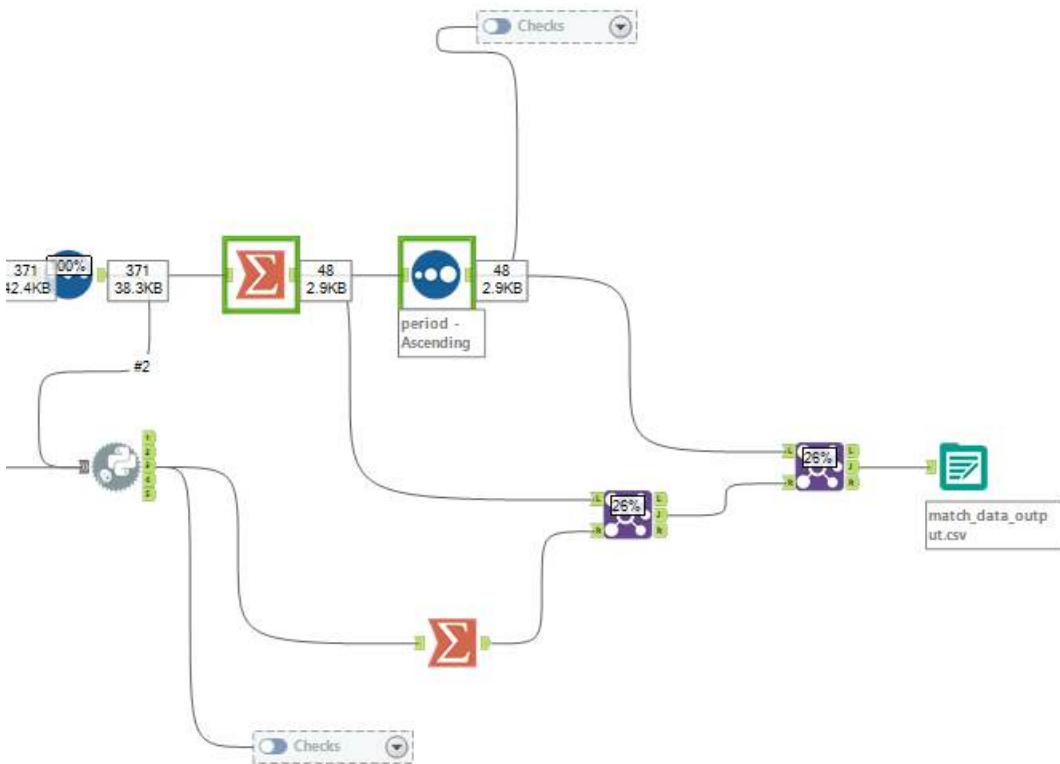
311 rows × 11 columns



In []:

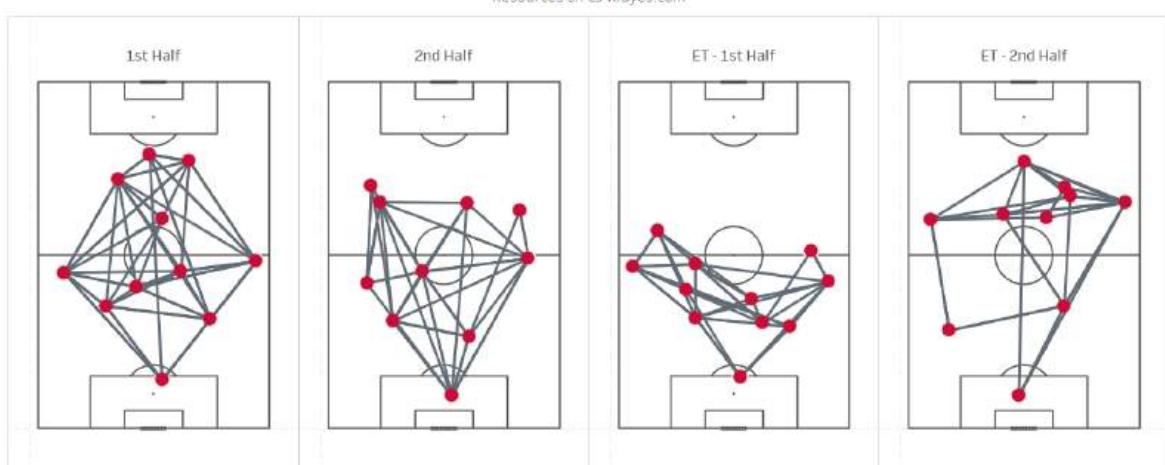
We

have to repeat the same process again, but rather than looking at player also look at pass recipient, because of course we want to remove this player from both making the passes or receiving the ball too. It then exports this information back into our flow ready to use.



The last part of the flow is all about creating average locations for each of the players. So we take their position when on the ball and take an average of those marks. Finally we glue back in data to do with the recipient in order to know which player is passing to who! We also count how many times each player combination happened. All goes well, you will end up with an output such as seen in the **match_data_output.csv**. From here we are ready to build in Tableau. For Tableau you can refer back to the previous [blog here](#). As always all the resources are found in the GitRepo as well as the dashboard is downloadable from my Tableau Public profile.

Women's Euro Final | England Pass Network
31 July 2022, 17:00 England 2-1 Germany
Resources on CJ-Mayes.com



Why not go further?

- Can you create a small multiple of all games leading up to the final?
- Can you add in jersey numbers into the code and alteryx flow?
- Can you change the design of the pitch to be different colours?
- Can you create the same workflow but split by 15 minute time segments?

I still think there is a lot of potential for building into the flow more ideas around frequency of passes, greatest interactions around players, dynamic shaping as well as more accurate represent what players are on the pitch during allotted time frames. Lots of food for thought, but perhaps you can take the workflow and make it better. Catch you in the next one. **LOGGING OFF**,

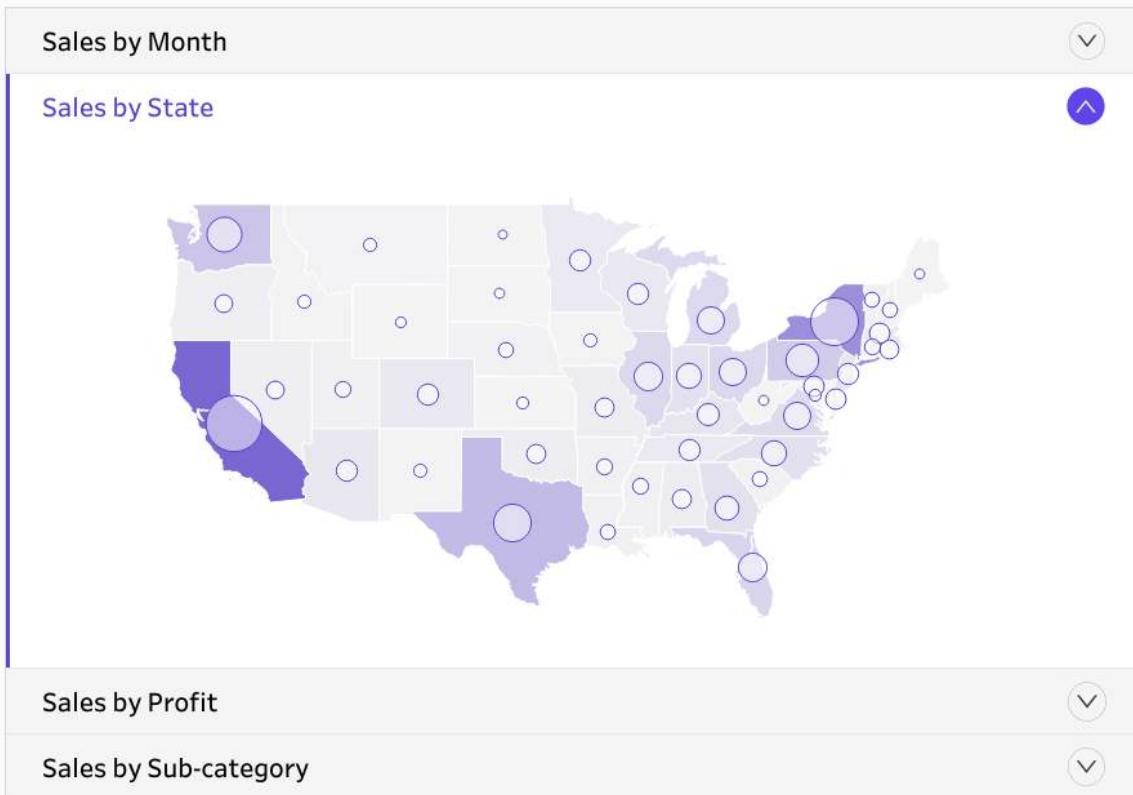
CJ

ACCORDION MENUS WITH LISA TRECOTT

Hi all, We are back with another guest blog! It's my pleasure to invite Lisa Trescott onto the site, someone in the community that many of us admire and shot to Tableau royalty winning the IronViz competition back in 2021. More recently Lisa put out a post using an accordion menu to create drop downs of different chart types and kindly agreed to share the process behind it with us all. Over to Lisa to talk us through the process.

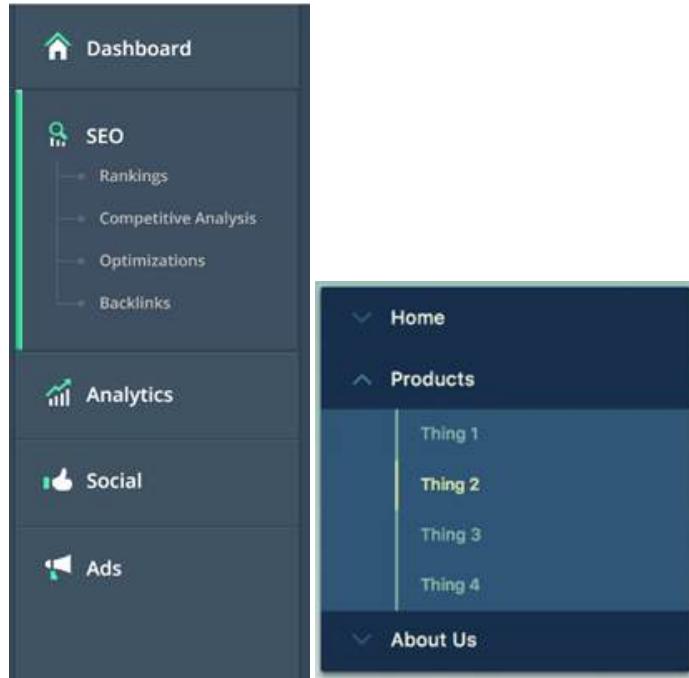
Accordion Menu

This viz utilizes dynamic zone visibility and parameter actions to create an accordion menu



Designed by Lisa Trescott

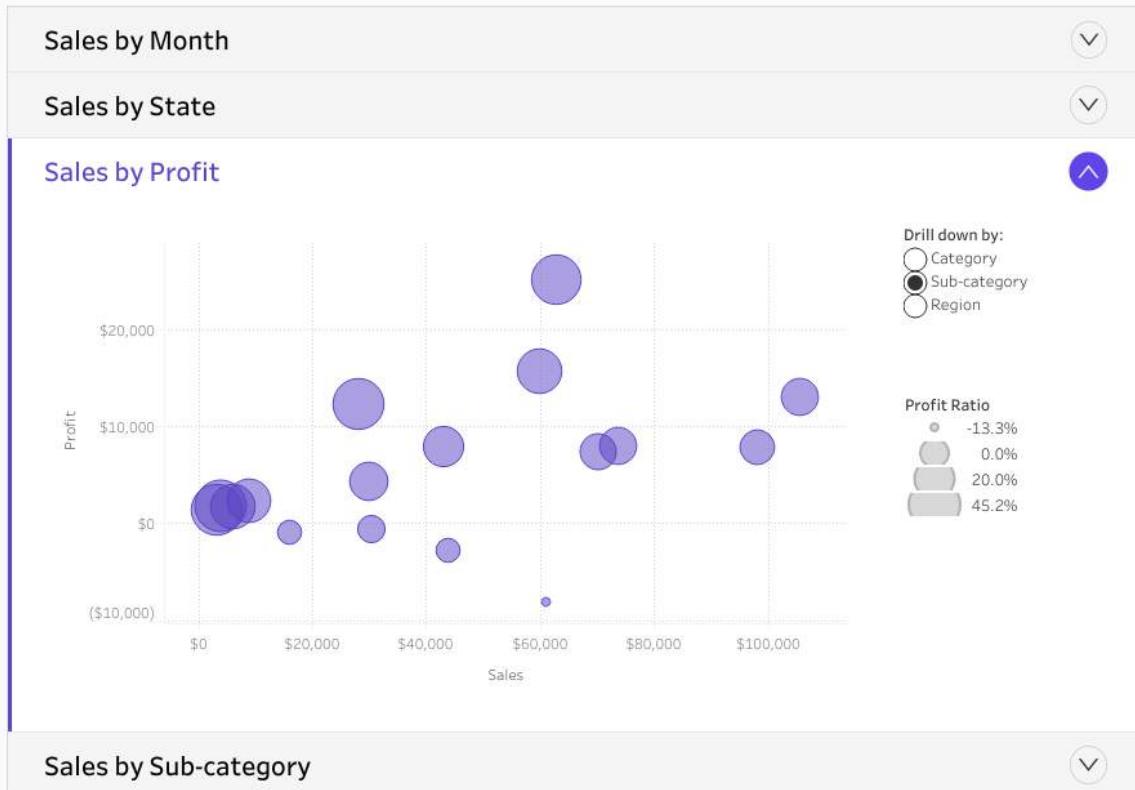
My favorite feature in Tableau is parameter actions because they allow you to do so many cool things with your user interface. Now we have dynamic zone visibility, and the two together are a match made in UI heaven! Recently, I was grappling with a way to fit multiple charts on a single dashboard without making the dashboard too large. I remembered a [viz](#) by Zak Geis in which he created an accordion menu, and I thought that could be just the ticket. He used parameter actions to show/hide single sheets in his workbook, but I needed to be able to show and hide entire containers of charts, descriptive text, and legends. Enter dynamic zone visibility (DZV). With DZV you can show and hide containers full of content, and when paired with parameter actions you can create powerful UI experiences. I started by looking at examples of accordion menus to get design ideas. Here are a few that



served as my inspiration. Fairly similar in style, both have a way of recognizing the different element you are clicked on through the use of colour and side border. And here is my version in Tableau

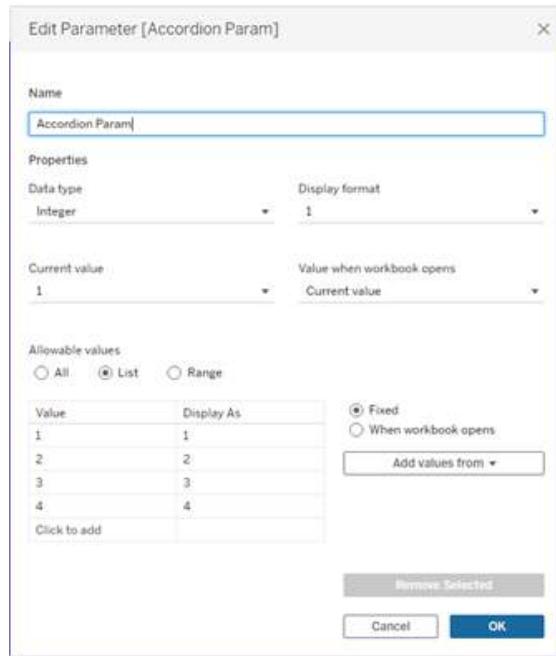
Accordion Menu

This viz utilizes dynamic zone visibility and parameter actions to create an accordion menu



Designed by Lisa Trescott

You can download a copy from the top link in the page or clicking [here](#). To start building this, create a parameter to control the sections of the accordion menu. My parameter has values 1 through 4 because my accordion has 4 sections. **If you have more**

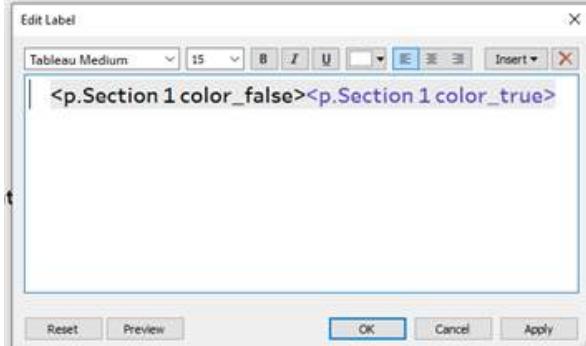


or fewer sections, you will need to add/remove values.

This parameter will control the DZV on the dashboard. Dynamic zones require a field or parameter to control them, and that field/parameter must be Boolean, a single value, or a field that returns a constant value (like a fixed LOD). I used Boolean fields and made 4 calculations that result in true/false when selected/not selected by the parameter. **p.section 1 =1 p.section 2 =2 p.section 3 =3 p.section 4 =4** **Section Headers** Each section header is a dual axis chart with MIN(1.0) as a bar and MIN(.95) as a shape. The bar chart acts as a background that changes color when the section is selected. For each header, the corresponding "p.section" calculation is placed on color.

- section = TRUE returns white (#ffffff)
- section = FALSE returns light grey (#f5f5f5)

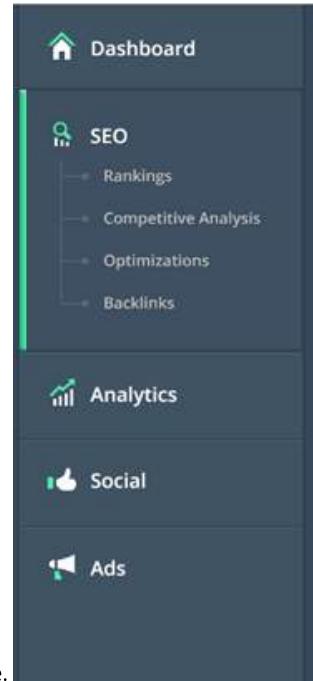
The label for each section header changes color based on two calculated fields: **p.Section 1 color_true** IF =1 THEN 'Sales by Month' ELSE " END **p.Section 1 color_false** IF <>1 THEN 'Sales by Month' ELSE " END Note: You will want to use the name of your own sections in place of "Sales by Month". Place these fields on label and format them appropriately:



The shape mark uses p.section on Shape with the following two custom shapes assigned to true and false. I used a purple up shape and a down grey shape. These header sheets will work in conjunction with parameter actions on the dashboard. This means we need the parameter value to be on Detail so it will get passed to the corresponding parameter action on the dashboard. Each section header corresponds to a number, so place the MIN of that number on Detail:

- header_section 1 = MIN(1)
- header_section 2 = MIN(2)
- header_section 3 = MIN(3)
- header_section 4 = MIN(4)

Place two dummy fields on Detail. I use dummy fields equal to '0' and '1', but these fields can be any values (as long as they are different from each other). More on the purpose of these dummy fields later. Build a header sheet for each section of your accordion. **Color Block** One of the accordion menus I found as inspiration had a colored line on the side of the section when it



was selected. I wanted to emulate that, so I created a sheet to act as that colored line bar chart using MIN(1). We have already used p.section on color, so if we use it again it will return the same grey and white color that has been previously assigned. To get around this, place two measures on color: p.section and a dummy field. We could also duplicate p.section and use the copy on color to achieve the same result. If the section is selected by the accordion parameter, the color is set to purple (#5b42c6). If it's not selected, the color is set to light grey (#f5f5f5) to match the header's background. Build these color block sheets for each section of your accordion. **Putting it Together** Each section of the accordion menu is composed of the following elements.

Header Container (remove all outer padding)

1.

Sheet: color_section

2.

Sheet: header_section

1.

Remove all padding

2.

Edit width to 3

2.

Chart Container (remove all outer padding)*Note: this whole container is controlled by dynamic zone visibility*

3.

Blank object

1.

Remove all padding

2.

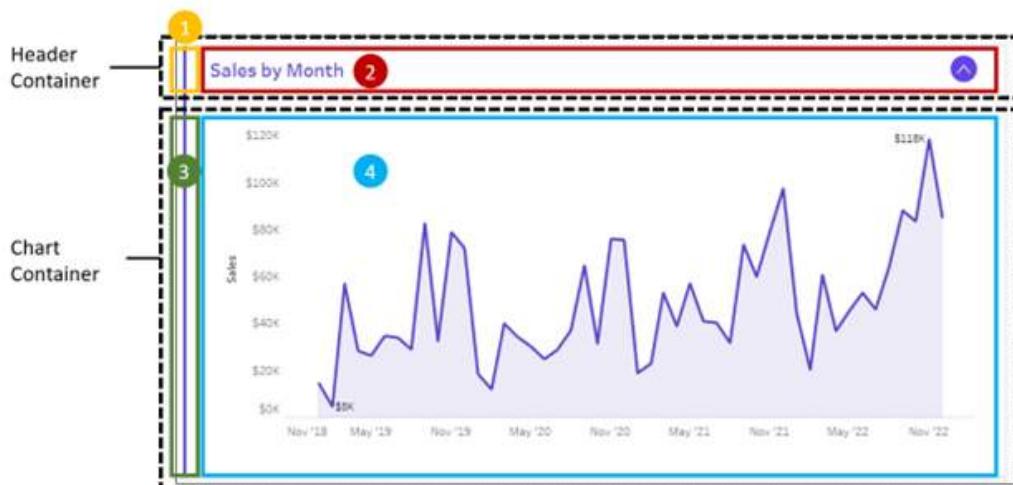
Set color to purple (#5b42c6)

3.

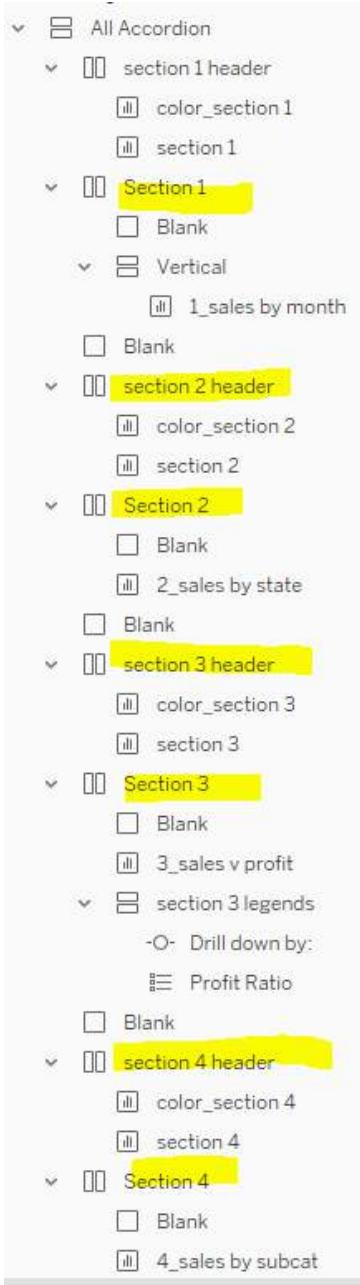
Edit width to 3

4.

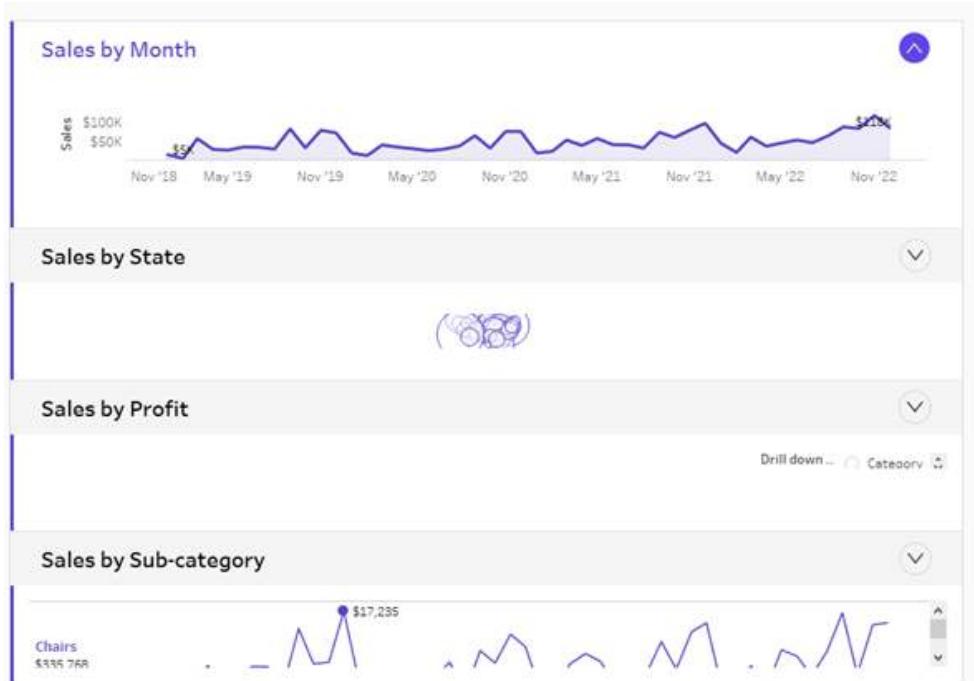
Chart



Tip: Build out all your containers before you place them into your accordion menu container. It is also helpful to name your containers in the layout pane.



Place all containers into a large master container that will house all the headers and sections. Note: I also placed a blank object between each section break (no padding, height is set to 1, color is set to grey (#e7e6eb)). The accordion menu is taking shape now, but it looks very cramped and messy. Don't worry! Once we add DZV controls it will all



fall into place.

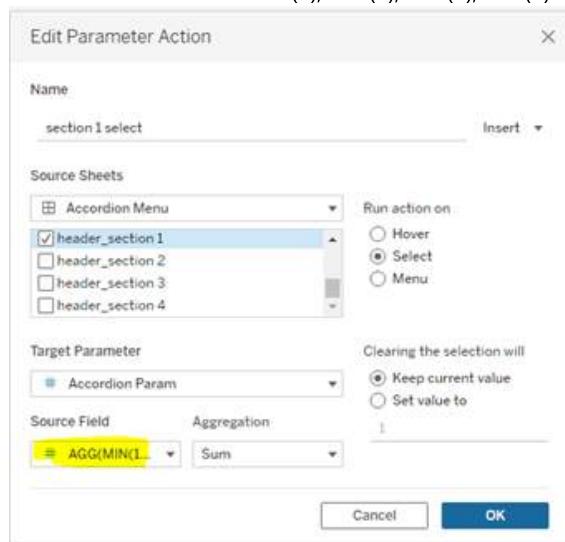
Select the

container housing your section 1 chart, navigate to the layout pane, and click "Control visibility using". Select the variable that corresponds to that section (p.section 1).

The layout pane shows the configuration for 'Section 1' under the 'Selected Item' tab. The 'Control visibility using value' checkbox is checked, and the 'p.section 1' parameter is selected in the dropdown menu. The main view shows a line chart titled 'Sales by Month' with sales values fluctuating over time from November 2018 to November 2022.

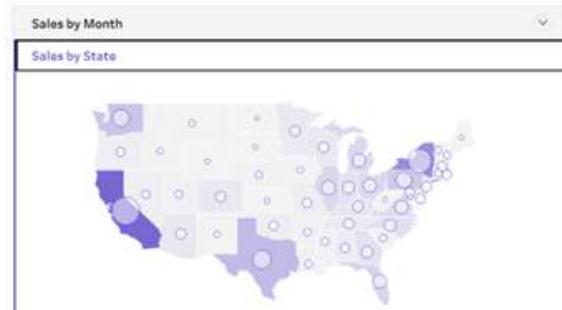
Repeat these steps for

all sections. **Parameter Actions** Each header has a parameter action associated with it. For a parameter action, you need to select the source sheet (i.e., the sheet you're clicking to change the value of the parameter), the target parameter (i.e., the parameter that is being controlled by the action), and a source field (i.e., the value that gets passed to the parameter). Note the "source field" will be MIN(1), MIN(2), MIN(3), MIN(4) to correspond to the header value in the accordion parameter.

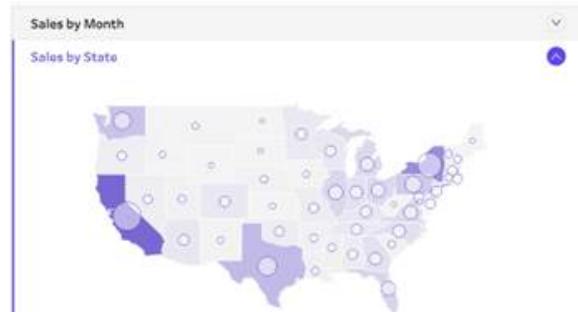


Create actions for all 4 headers. **Filter Actions** When you select

something in Tableau, it will remain selected until you click somewhere else. We want to override that, otherwise the menu will



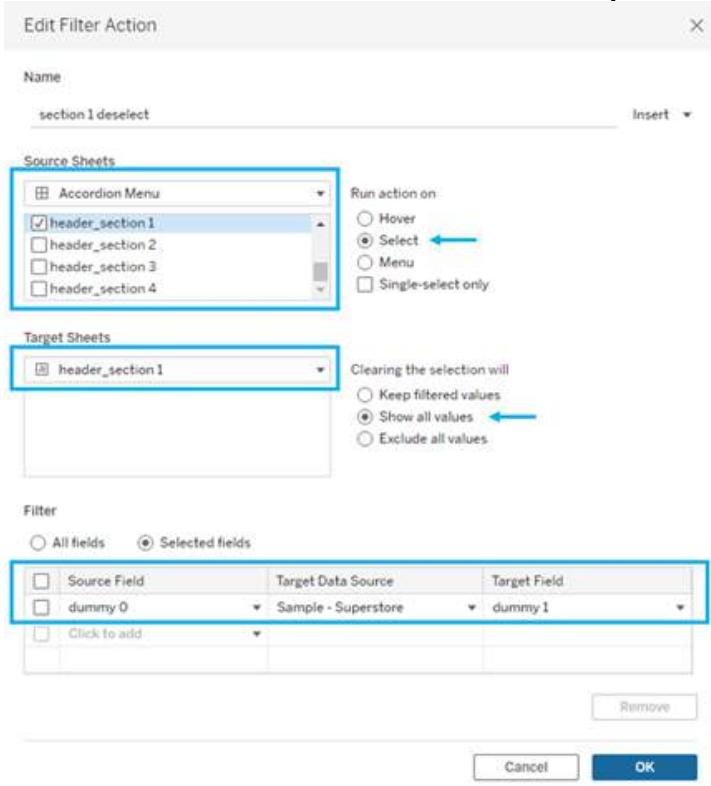
look like this upon selecting a header:



We want it to have a more

seamless interface and look like this

we put on Detail? This is where they come into play. Create a filter action that runs on select of header_section 1. For target sheet, select header_section 1. Select "Show all values" when the selection is cleared. Under "selected fields", make your source

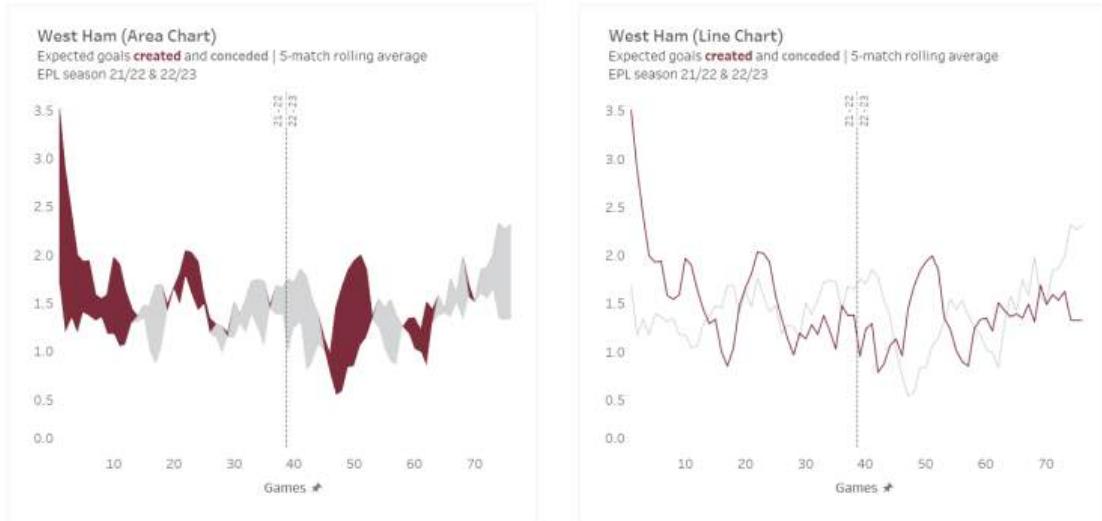


and target fields your dummy 0 and dummy 1 fields.

Create filter actions for all 4 section headers. And there you have it! An accordion menu built entirely in Tableau using parameter actions and dynamic zone visibility. I hope you have fun playing around with this concept and making other designs and variations of this menu. I'd love to see what you come up with! CJ Round-Up: Wow! How exciting! A good one for those wanting to utilise dynamic zone visibility and parameter actions in a really effective way. Be sure to check out Lisa's profile, and if you have any questions reach out via [Twitter](#). **LOGGING OFF CJ**

TEAM XG IN TABLEAU

Hi all, Short Tableau one this week on creating rolling xG charts. I've previously written about [match xG charts here](#). Todays content will look at season xG moving averages. Here is what we will create.



What is xG?

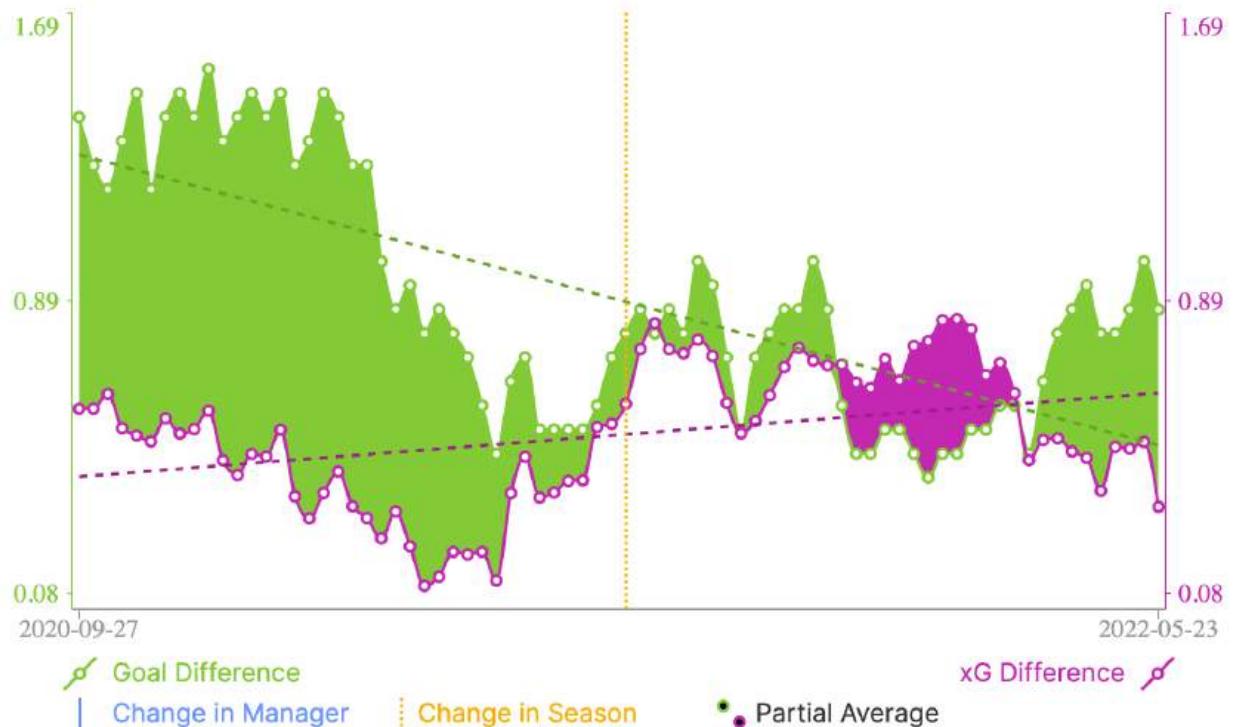
Let's take the FBRef definition. Very simply, xG (or expected goals) is the probability that a shot will result in a goal based on the characteristics of that shot and the events leading up to it. Some of these characteristics/variables include: Every shot is compared to thousands of shots with similar characteristics to determine the probability that this shot will result in a goal. That probability is the expected goal total. An xG of 0 is a certain miss, while an xG of 1 is a certain goal. An xG of .5 would indicate that if identical shots were attempted 10 times, 5 would be expected to result in a goal.

Atlético Madrid

La Liga

2020-09-27 to 2022-05-23

Trendline



StatsBomb

Without going into details, there are a number of xG models out there as to what makes up that probability of a goal being scored. For today we will just look at the calculations needed to create the chart in Tableau. A few shout outs of other cool xG tutorials: [This one from Maram using excel and powerpoint](#) [Son of a Corner xG rolling plot using python Statsbomb - how can xG be used](#)

How can xG be used?

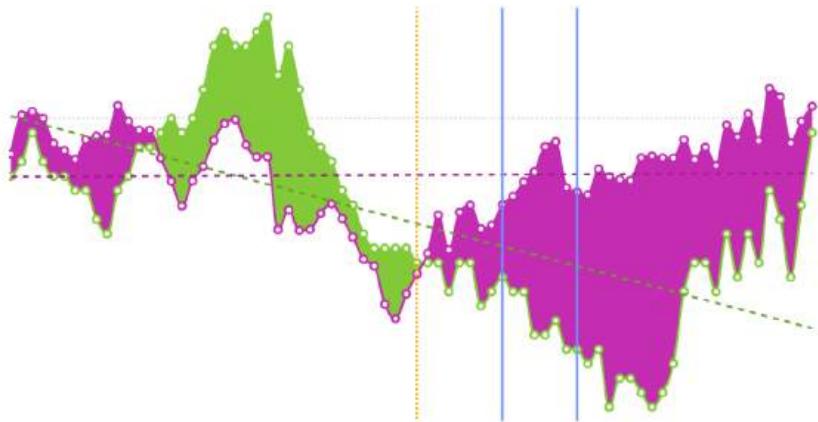
Team Analysis

xG can be used to get a better idea of underlying team quality beyond current results, highlighting teams who are over or under-performing their expected numbers and whose results may soon begin to change.

+ Player Analysis

+ Head Coach Scouting

+ Predictive Modelling



So how can we recreate this in Tableau? Let's take west ham's xG shot data for the past two seasons. You can find it in the github repo at the top of the page. First things first we need to make sure we order the data in the correct way. I do this in excel based on the match date and minute.

I then quickly create the de-duplicated version of the match IDs and rank them (they are now already in order). The main reason I do this is because there are various breaks in football and so we can't necessarily use a date field in Tableau so we have to use the equivalent of what would be an order of matches. For me those matches start at the start of the 21/22 season and end at the end of the 22/23 season. We've limited our dataset to just look at West Ham matches.

⊕- West Ham 21-23+ (Copy of West Ham)



Sort fields		Data source order											<input type="checkbox"/> Show aliases		<input type="checkbox"/> Show hi									
# Order	Match Id (Order)	# Order	A Team	Date	Atc West Ham 23	Atc West Ham 23	Atc West Ham 23	# Order	A Team	Date	Atc West Ham 23	Atc West Ham 23	# Order	A Team	Date	Atc West Ham 23	# Order	Year	# Order	Year	# Order	West Ham 23	Atc West Ham 23	X G Flag
		CRM Order	A Team	Atc West Ham 23	Date	Atc West Ham 23	H A	Atc West Ham 23	H Team	# Order	West Ham Id	Atc West Ham 23	Match Id	Atc West Ham 23	Result	# Order	West Ham 23	Atc West Ham 23	X G	Match	# Order	West Ham 23	Atc West Ham 23	X G
16384	1	West Ham	15/08/2021 13:00:00	a	Newcastle United	426219	16384	BlockedShot	1	0.044666	West Ham - Newcastle...	2021	xG Created											
16384	1	West Ham	15/08/2021 13:00:00	a	Newcastle United	426218	16384	BlockedShot	1	0.037903	West Ham - Newcastle...	2021	xG Created											
16384	1	West Ham	15/08/2021 13:00:00	h	Newcastle United	426220	16384	Goal	4	0.706295	West Ham - Newcastle...	2021	xG Conceded											
16384	1	West Ham	15/08/2021 13:00:00	a	Newcastle United	426221	16384	BlockedShot	7	0.013080	West Ham - Newcastle...	2021	xG Created											
16384	1	West Ham	15/08/2021 13:00:00	h	Newcastle United	426222	16384	BlockedShot	9	0.020707	West Ham - Newcastle...	2021	xG Conceded											
16384	1	West Ham	15/08/2021 13:00:00	a	Newcastle United	426223	16384	Goal	17	0.044954	West Ham - Newcastle...	2021	xG Created											
16384	1	West Ham	15/08/2021 13:00:00	a	Newcastle United	426224	16384	SavedShot	22	0.399756	West Ham - Newcastle...	2021	xG Created											
16384	1	West Ham	15/08/2021 13:00:00	h	Newcastle United	426225	16384	SavedShot	22	0.128386	West Ham - Newcastle...	2021	xG Conceded											
16384	1	West Ham	15/08/2021 13:00:00	h	Newcastle United	426226	16384	BlockedShot	25	0.059223	West Ham - Newcastle...	2021	xG Conceded											
16384	1	West Ham	15/08/2021 13:00:00	h	Newcastle United	426228	16384	MissedShots	26	0.130065	West Ham - Newcastle...	2021	xG Conceded											
16384	1	West Ham	15/08/2021 13:00:00	h	Newcastle United	426227	16384	BlockedShot	26	0.022003	West Ham - Newcastle...	2021	xG Conceded											
16384	1	West Ham	15/08/2021 13:00:00	a	Newcastle United	426229	16384	BlockedShot	28	0.046119	West Ham - Newcastle...	2021	xG Created											
16384	1	West Ham	15/08/2021 13:00:00	a	Newcastle United	426230	16384	BlockedShot	30	0.096596	West Ham - Newcastle...	2021	xG Created											
16384	1	West Ham	15/08/2021 13:00:00	a	Newcastle United	426231	16384	MissedShots	30	0.002020	West Ham - Newcastle...	2021	xG Created											

Here's what that will look like as data. Now we have the data its time to split out whether the action corresponds to West Ham or the opposing team.

Pages xG Flag Id

Filters: xG Rolling

Marks: Automatic, SUM({xG Flag})

Columns: xG Flag, xG Conceded, xG Created

Rows: Id

Calculation:

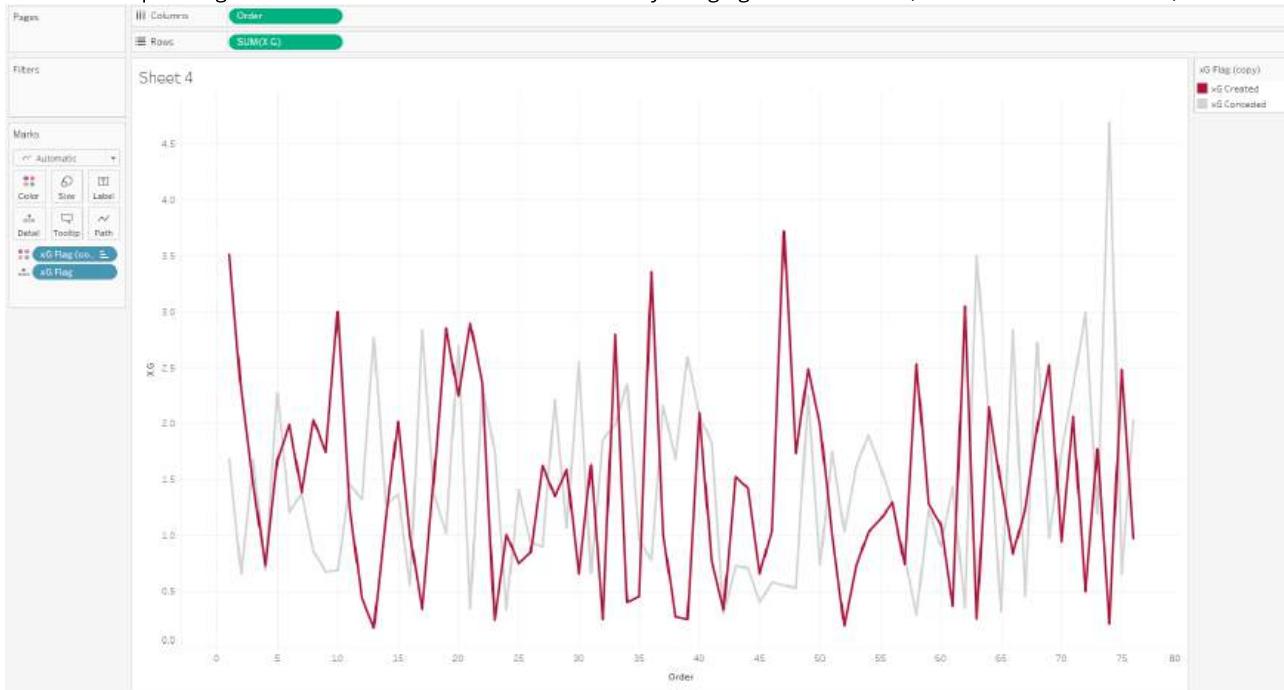
```
if [B Team] = 'West Ham' and [B A] = 'a' then 'xG Created'
ELSEIF [B Team] = 'West Ham' and [B A] = 'h' then 'xG Conceded'
ELSE 'xG Conceded'
END
```

The calculation is valid.

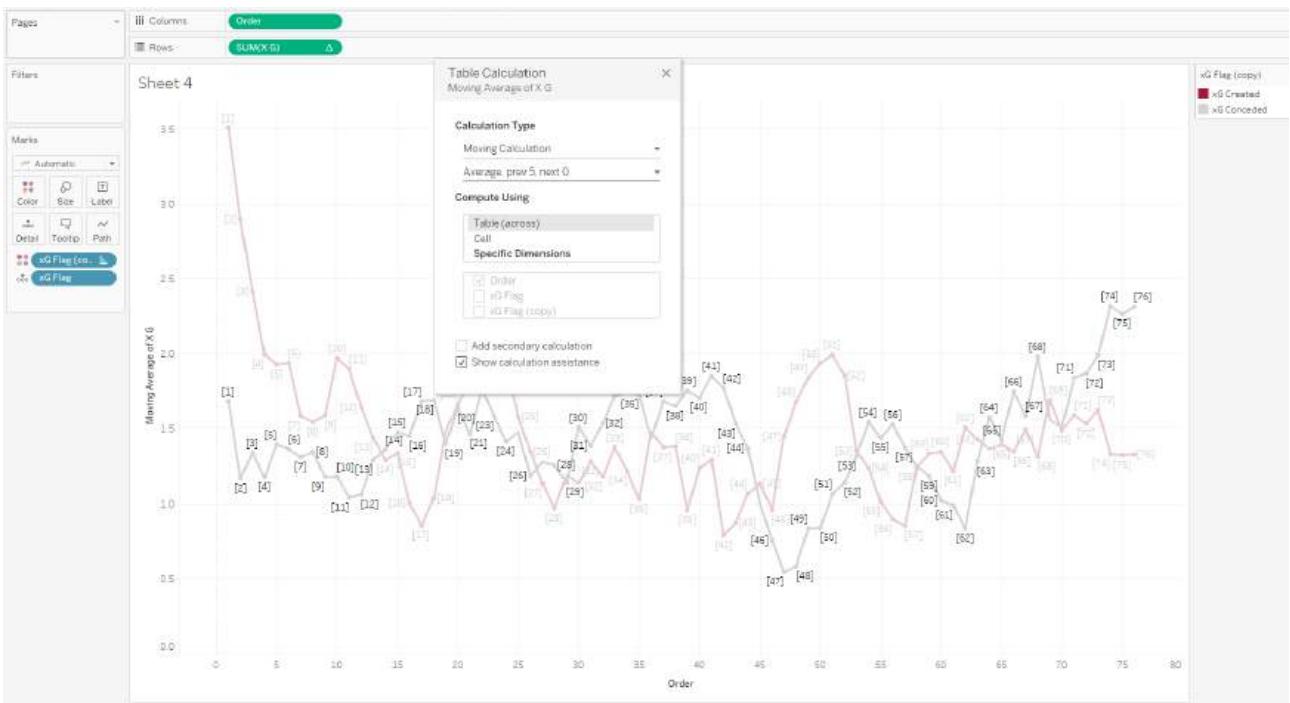
Dependencies: None

Help: ABS(number)
Returns the absolute value of the given number.
Example: ABS(-7) = 7

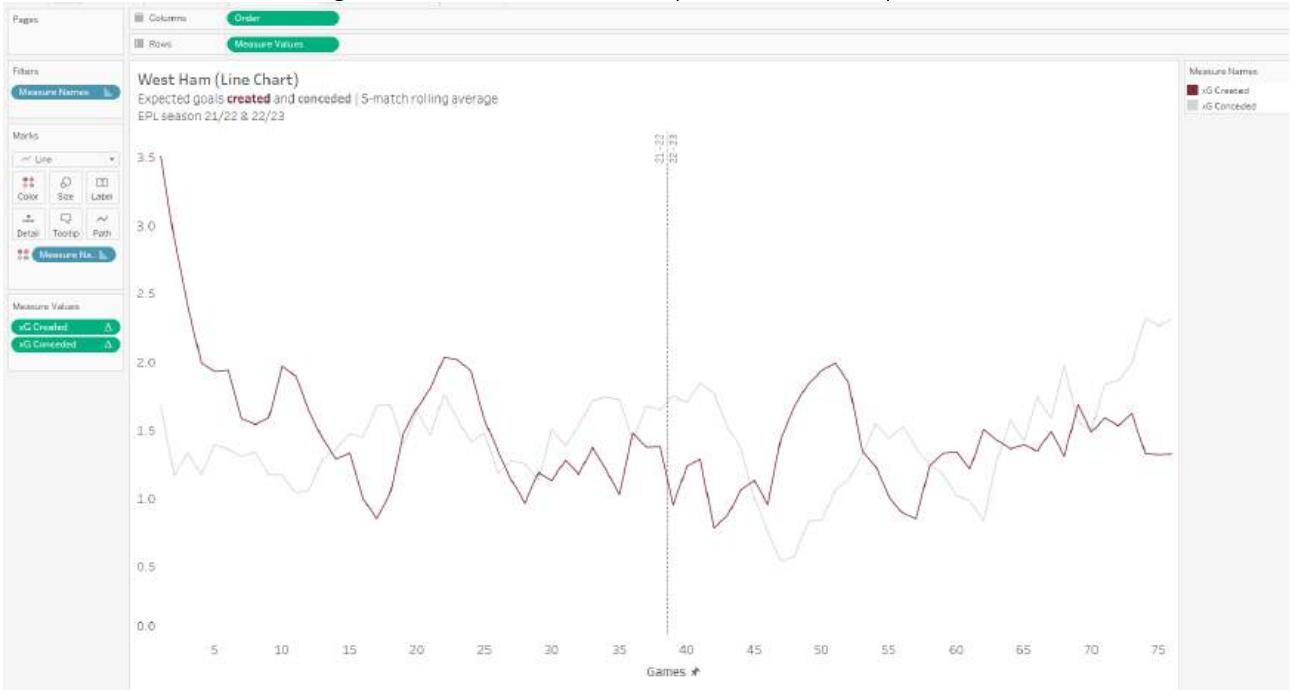
`if = 'West Ham' and = 'a' then 'xG Created' ELSEIF = 'West Ham' and = 'h' then 'xG Created' ELSE 'xG Conceded' END` The above should make sense, pretty much says if its west ham, dependent on home or away then count it as created else its a conceded expected goal. We can double check the calculation by bringing it into the view (I've left it in the workbook)



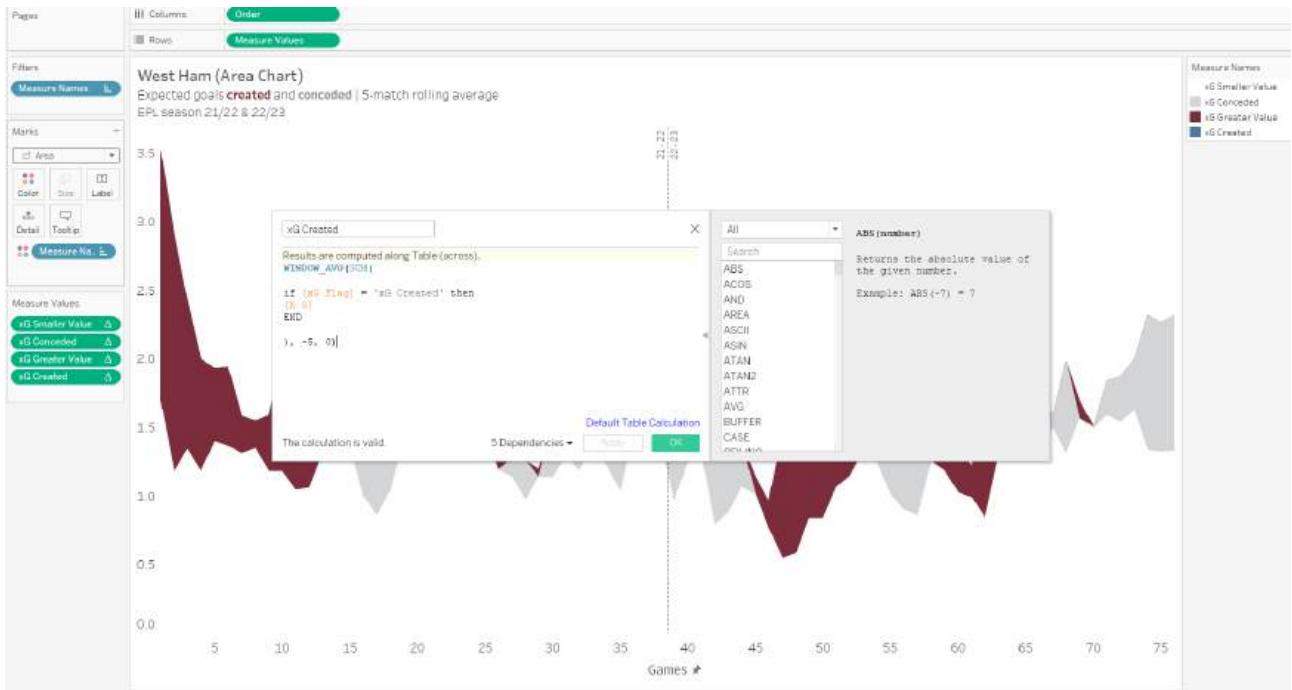
See how if we just look at the sum of xG with xG Flag on detail we have for each game (date) the flickering xG combinations – but it does feel a bit hectic on the page!



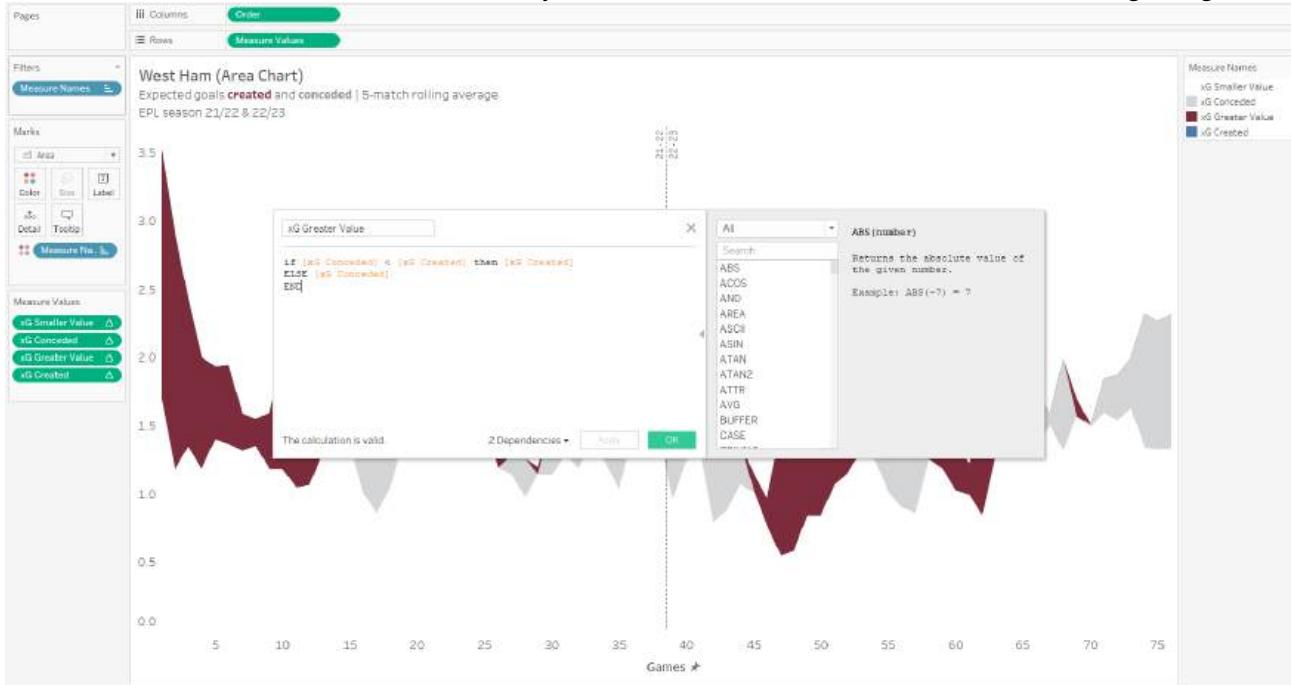
We can amend our calculation to be a moving calculation. In this example above i've chosen a moving calculation of the previous 5 games. My only hesitation with this calculation is for the first 5 games, it feels like an inaccurate 'rolling' value. Perhaps it would be nice to have the 5 games of the season before to help flatten out the xG spike at the start of our chart.



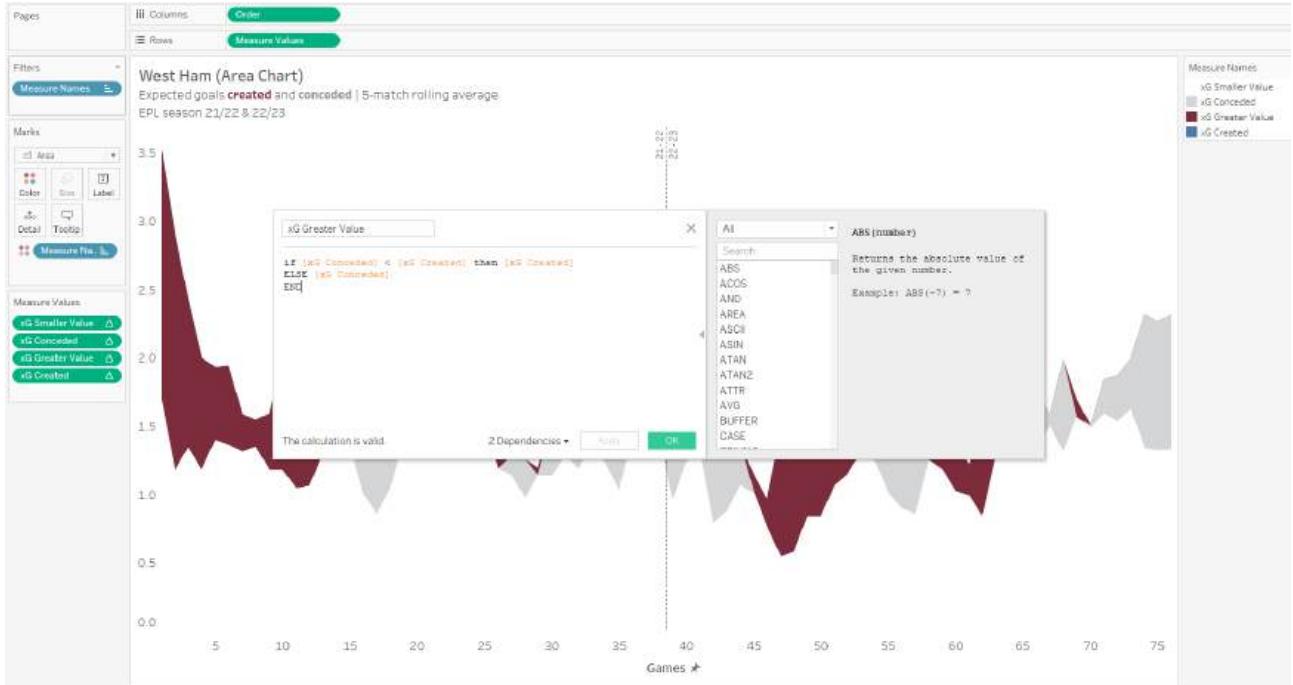
Now this is theoretically the final version of our line chart version so you could stop here. But i've seen individuals take it even further doing area charts. Create two new calculations to split out the different moving averages for West Ham and the Opponent.



xG Created WINDOW_AVG(SUM(if = 'xG Created' then END), -5, 0) and xG Conceded WINDOW_AVG(SUM(if = 'xG Conceded' then END), -5, 0) The -5 and 0 here refer to how many forwards and backwards we want to look at in our moving average.



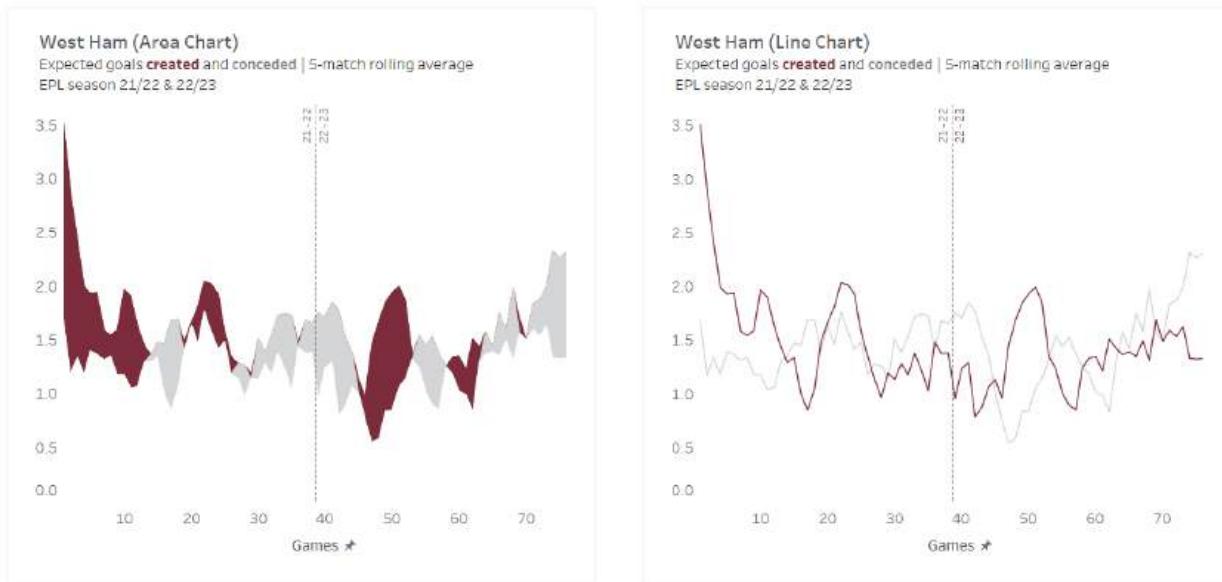
To finalise our chart we need to create a few calculations to figure out what value is greater out of the conceded and created.



By now plotting these new calculations as measure values in the rows column we can create an area chart as desired. xG Greater is as below, if < then ELSE END Of course, xG smaller is the opposite of this. Now How to build from scratch: Open a new sheet Drag Order onto columns as an unaggregated dimension. Add measure values to the rows. Add measure names to filter and only filter for measure values of:

- xG conceded, xG created, xG Greater Value, xG Smaller value

Add measure names to column. Change the marks card to area. Currently it will have it so that they are stacked. Go to Analysis stack marks and turn it off. Re-arrange your marks to have xG Smaller Value at the front, make this white in colour. xG Conceded should be the next mark in the colour range order as it will show where the opposing teams aggregated xG was higher than west hams. xG Greater value is when West Hams xG was greater than the opposition so this should be the third mark of colour. xG Created isn't necessarily needed but i left it in just so you can check your calculations are correct. (For our marks we may need to turn the stack marks off under analysis, thank you Deb Simmonds for helping me with this aspect! – I was a little confused about why the area and line version originally didnt look the same)



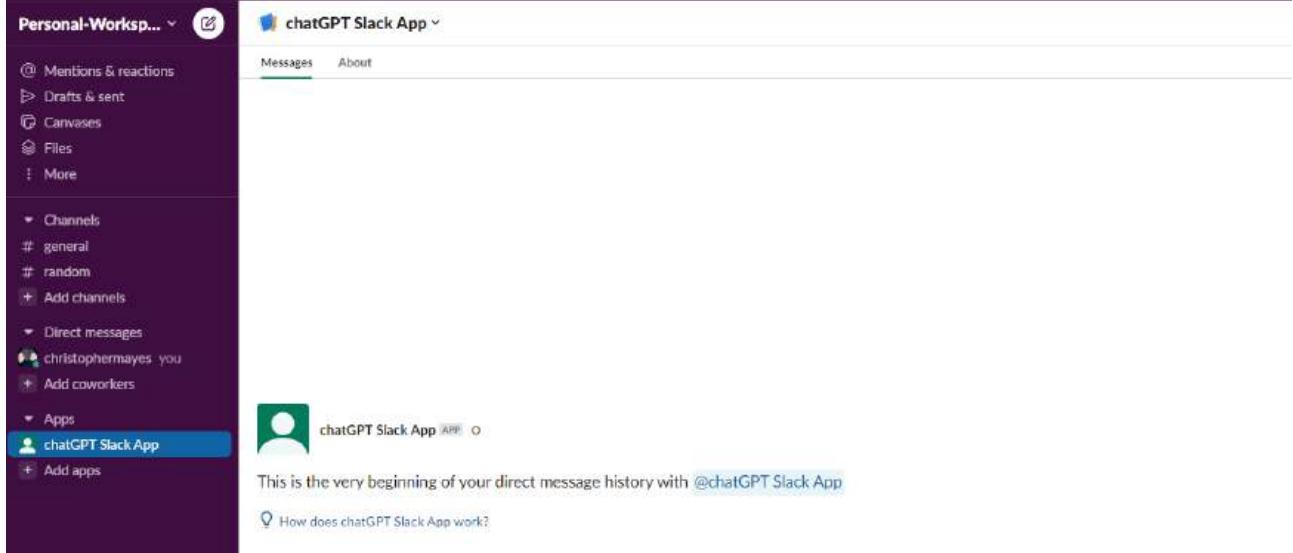
And there we have it. I think the only think I want to close out on is saying, Im unsure on how to best approach the idea of rolling averages of xG in terms of the start and end if using the idea of previous x number of games. Maybe I need more data, but this tutorial really is focusing on the calculations as oppose to your own implementation. If you think a different value is better served here by all means amend where necessary. The only other thing is, where the data has fewer marks of games there becomes natural 'cross over' points between xG created and conceded where the lines cross one another, I don't think this is the prettiest in style but unsure on alternative methods to overcome it. Final touches are cosmetic, I added some reference lines to show the split in season games, made colours indicate based on team brand, and added some padding and titles to help

indicate what the chart was showing. Remember all my resources are free to download on GitHub and Tableau Public Going further:

- Try create the chart using your own data.
 - Try look at all teams xG charts within a season
 - Add in some extra formatting to show average across the whole season
- West Ham are massive. **LOGGING OFF, C]**

CREATE A CHATGPT APP FOR SLACK

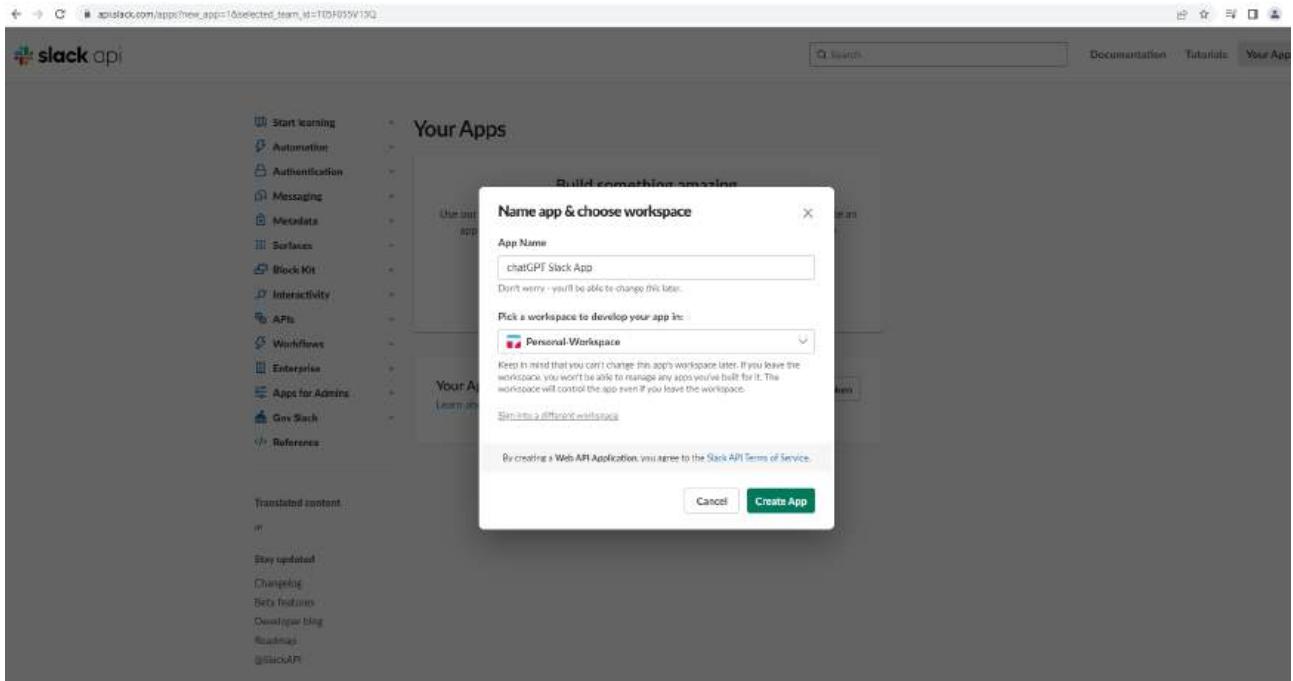
Hi all, In previous blogs we have looked at analysing CSV files, answering questions against PDF files as well as having google extensions to help with queries within excel. This week we will look at how we can integrate ChatGPT into slack as a bot. Now I must start with this isn't a how to guide on [SlackGPT](#). SlackGPT is its own product Salesforce are building out, which is there vision for generative AI in Slack. With the Salesforce SlackGPT you'll be able to instantly bring generative AI into your automated processes in Slack, without a single line of code. [If you're after that, please visit here instead.](#) This blog, however is looking at taking OpenAI and adding it in as an bot using the slack API. It'll be a blog with far less glamour around data protection, privacy and security **so in no way is a guide for workplace implementation.**



For this tutorial you will need:

- A slack workspace
- Somewhere to write your python code
- An OpenAI account and api key

Setting Up Slack Go to <https://api.slack.com/> Go to Your Apps and create new app, Create it from scratch and give your app a name & pick the workspace you want to develop in. For this tutorial I created a brand new workspace called Personal-Workspace.



Once you've created it go down to OAuth and Permissions. Add the following scopes to bot scope: chat:write chat:write:public
Scroll back up and add these to the workspace.

chat:write

Send messages as @chatGPT Slack App



chat:write.public

Send messages to channels @chatGPT Slack



App isn't a member of

what the authentication looks like.

Below is



**chatGPT Slack App is requesting permission to
access the Personal-Workspace Slack
workspace**

What will chatGPT Slack App be able to view?

👤 Content and info about you ►

What will chatGPT Slack App be able to do?

👤 Perform actions as you ►

Cancel

Allow

Once you've done that, under OAuthTokens for your workspace it will show your **bot token**, you will need to store this as we will need it later. The next token we will need is an app token. Go to basic information and scroll down to App tokens.

App-Level Tokens

App-level tokens allow your app to use platform features that apply to multiple (or all) installations—for example, the API to list event authorizations. Features have distinct scopes, so maximum

Generate an app-level token

Token Name

My App Token

Scopes to be accessed by this token

Scope

Description

[connections:write](#)

Route your app's interactions and event payloads over WebSockets



Add Scope

Cancel

Generate

Generate the

app token ready for use, and add the connections:write scope. Store the app token somewhere ready for later use in the code. Next go to socket mode, and enable socket mode.



Socket Mode

Settings

Basic Information

Collaborators

Socket Mode

Install App

Manage Distribution

Features

App Home

Org Level Apps

Incoming Webhooks

Interactivity & Shortcuts

Slash Commands

Workflow Steps

OAuth & Permissions

Receive app payloads via Websockets instead of Request URLs

Turning on Socket Mode will route your app's interactions and events over a WebSockets connection instead sending these payloads to Request URLs, which are public HTTP endpoints.

This setting is intended for internal apps that are in development or need to be deployed behind a firewall. It is not intended for widely distributed apps. Please set up Request URLs for your app before submitting to the App Directory. [Learn more](#)

Connect using Socket Mode

To start receiving payloads in Socket Mode, turn on the toggle below and call the `apps.connections.open` endpoint using an App Level Token to establish a connection.

Enable Socket Mode



Allow your app to connect via Socket Mode. You can disable this any time and revert to using any Request URL you've already defined.

The final thing we need to do to set up our app is make sure interactivity is switched on,

Interactivity

On

Any interactions with shortcuts, modals, or interactive components (such as buttons, select menus, and datepickers) will be sent to a URL you specify. [Learn more.](#)

Socket Mode is enabled. You won't need to specify a Request URL.

as well as in event subscription adding app_mention and message.im to the bot events.

Features

- App Home
- Org Level Apps
- Incoming Webhooks
- Interactivity & Shortcuts
- Slash Commands
- Workflow Steps
- OAuth & Permissions
- Event Subscriptions**
- User ID Translation
- App Manifest NEW
- Beta Features

Submit to App Directory

- Review & Submit
- Give feedback

Slack ❤️

Help

New event authorization format

Recent changes to Events API payloads
The Events API now sends information about authorized users and workspaces in a new, compact format. [Learn more.](#)

Subscribe to bot events

Apps can subscribe to receive events the bot user has access to (like new messages in a channel). If you add an event here, we'll add the necessary OAuth scope for you.

Event Name	Description	Required Scope
app_mention	Subscribe to only the message events that mention your app or bot	app_mentions:read
message.im	A message was posted in a direct message channel	im:history

Add Bot User Event

We will need to then save the changes again to our slack app.

You've changed the permission scopes your app uses. Please [reinstall your app](#) for these changes to take effect (and if your app is listed in the Slack App Directory, you'll need to resubmit it as well).

Features

- App Home
- Org Level Apps
- Incoming Webhooks
- Interactivity & Shortcuts
- Slash Commands
- Workflow Steps
- OAuth & Permissions
- Event Subscriptions**
- User ID Translation
- App Manifest NEW
- Beta Features

Submit to App Directory

- Review & Submit
- Give feedback

Slack ❤️

Help

New event authorization format

Recent changes to Events API payloads
The Events API now sends information about authorized users and workspaces in a new, compact format. [Learn more.](#)

Subscribe to bot events

Apps can subscribe to receive events the bot user has access to (like new messages in a channel). If you add an event here, we'll add the necessary OAuth scope for you.

Event Name	Description	Required Scope
app_mention	Subscribe to only the message events that mention your app or bot	app_mentions:read
message.im	A message was posted in a direct message channel	im:history

Now it is ready for use! **Generating API Key**. Part two needs us to have an API key for OpenAI. You can log into OpenAI [here](#). Within your [API key area](#), you can generate a key for use. You can see I made a new key called slack Key for this tutorial temporarily.

platform.openai.com/account/api-key

Overview Documentation API Reference Examples Playground Personal

API keys

Your secret API keys are listed below. Please note that we do not display your secret API keys again after you generate them.

Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically rotate any API key that we've found has leaked publicly.

NAME	KEY	CREATED	LAST USED
My Test Key	sk-...K3b0	Apr 30, 2023	May 20, 2023
Slack Key	sk-...uCeX	Jul 2, 2023	Never

+ Create new secret key

Default organization

If you belong to multiple organizations, this setting controls which organization is used by default when making requests with the API keys above.

Personal

Note: You can also specify which organization to use for each API request. See [Authentication](#) to learn more.

If you're using the OpenAI free trial, make sure you have enough credits left, by going to usage.

ORGANIZATION

Personal

Settings

Usage

Rate limits

Members

Billing

USER

API keys

Usage

Below you'll find a summary of API usage for your organization. All dates and times are UTC-based, and data may be delayed up to 5 minutes.

July

DAILY CUMULATIVE

Daily usage (USD)

\$0.03
\$0.02
\$0.01
\$0.00

01 Jul 02 Jul

Free trial usage

\$1.56 / \$5.00

GRANT # CREDIT GRANTED EXPIRES (UTC)

Grant 1 \$5.00 August 1, 2023

Time to Code Get up the python IDE of your choice, I'll be using pycharm. You can either set your environment variables for your slack app, bot and open ai token or hold them in a separate file. First we need to install some packages, In the terminal pip install slack_bolt, openai and langchain.

C:\Users\44771\PycharmProjects\SlackProject\chatgpt-slack-app> pip install langchain LangChain

is a powerful tool that can be used to work with Large Language Models (LLMs). LLMs are very general in nature. I would recommend reading up on this if you haven't prior. Now copy and paste the code from the repository into PyCharm. Firstly run the .py file called Tokens, after adding your tokens that we captured in the slack and open ai set-up. You will need to amend

```
main.py Tokens.py
```

```
1 import os
2
3 SLACK_BOT_TOKEN = 'redacted'
4 SLACK_APP_TOKEN = 'redacted'
5 OPENAI_API_TOKEN = 'redacted'
6 os.environ['OPENAI_API_KEY'] = OPENAI_API_TOKEN
```

lines 3-5

After that, copy and

paste into the main.py file your main code. You won't need to amend this file at all.



```
import os
from slack_bolt import App
from slack_bolt.adapter.socket_mode import SocketModeHandler
from langchain import OpenAI, LLMChain, PromptTemplate
from Tokens import SLACK_BOT_TOKEN, SLACK_APP_TOKEN, OPENAI_API_TOKEN
from langchain.memory.buffer_window import ConversationBufferWindowMemory

openai_api_key = OPENAI_API_TOKEN
# Start your app with your bot token and socket mode handler
app = App(token=SLACK_BOT_TOKEN)

# Langchain setup
template = """Assistant is a large language model trained by OpenAI.
{history}
Human: {human_input}
Assistant:"""

prompt = PromptTemplate(
    input_variables=["history", "human_input"],
    template=template
)

chatgpt_chain = LLMChain(
    llm=OpenAI(temperature=0),
    prompt=prompt,
    verbose=True,
    memory=ConversationBufferWindowMemory(k=2),
)

# Slack handler
@app.message(".*")
def message_handler(message, say, logger):
    print(message)

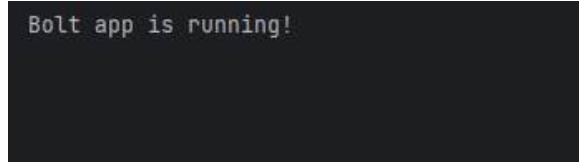
    output = chatgpt_chain.predict(human_input=message['text'])
    say(output)

# Start the app
if __name__ == "__main__":
    SocketModeHandler(app, SLACK_APP_TOKEN).start()
```

As a brief overview.

1. It imports slack bolt for building Slack apps, and specific components from the slack bolt socket mode.
2. It imports modules and classes from Langchain used for language generation and managing conversation history.
3. It imports the values of three tokens from a separate file called tokens.py
4. It initializes the OpenAI API key with the value of our open api token.
5. It creates a Slack Bolt app instance with the provided Slack bot token.
6. It sets up a language generation template that defines the structure of the conversation between a human and an AI assistant.
7. It defines a message handler for the Slack app. Any message received by the bot will trigger this handler. The message handler prints the received message and generates a response.
8. The response is sent back to the Slack channel.

If all things are working well at this stage you should be able to run your code, and it will say: Bolt App is running.



For now, this will need to be running for our app to work. If we stop this code running, the app will not work. Restart your slack. Go to your workspace and add the new app we have created.

A screenshot of a Slack workspace. On the left, the sidebar shows "Personal-Worksp..." with options like "Mentions & reactions", "Drafts & sent", "Canvases", "Files", "More", "Channels", "# general", "# random", "Add channels", "Direct messages", "christophermayes", "Add coworkers", "Apps", "chatGPT Slack App" (which is highlighted in blue), and "Add apps". The main area shows a message from "chatGPT Slack App" with the text "This is the very beginning of your direct message history with @chatGPT Slack App". Below it is a link "How does chatGPT Slack App work?"

When you now click into the app it should allow for you to write a message. If it doesn't let you write a prompt go back & double check your permissions and scopes. Try a prompt of your choosing, I've gone for: "Give me a step by step guide on how to create a slack app using the slack API"



When you click send in slack you will see in your pycharm terminal the message is triggered.



The response will show from ChatGPT in your slack window. You will also notice if you stop your terminal running then the app will stop working. Below you can see how it no longer responds.

random
+ Add channels
▼ Direct messages
christophermayes you
+ Add coworkers
▼ Apps
chatGPT Slack App
+ Add apps

christophermayes 4:04 PM
Give me a step by step guide on how to create a slack app using the slack API

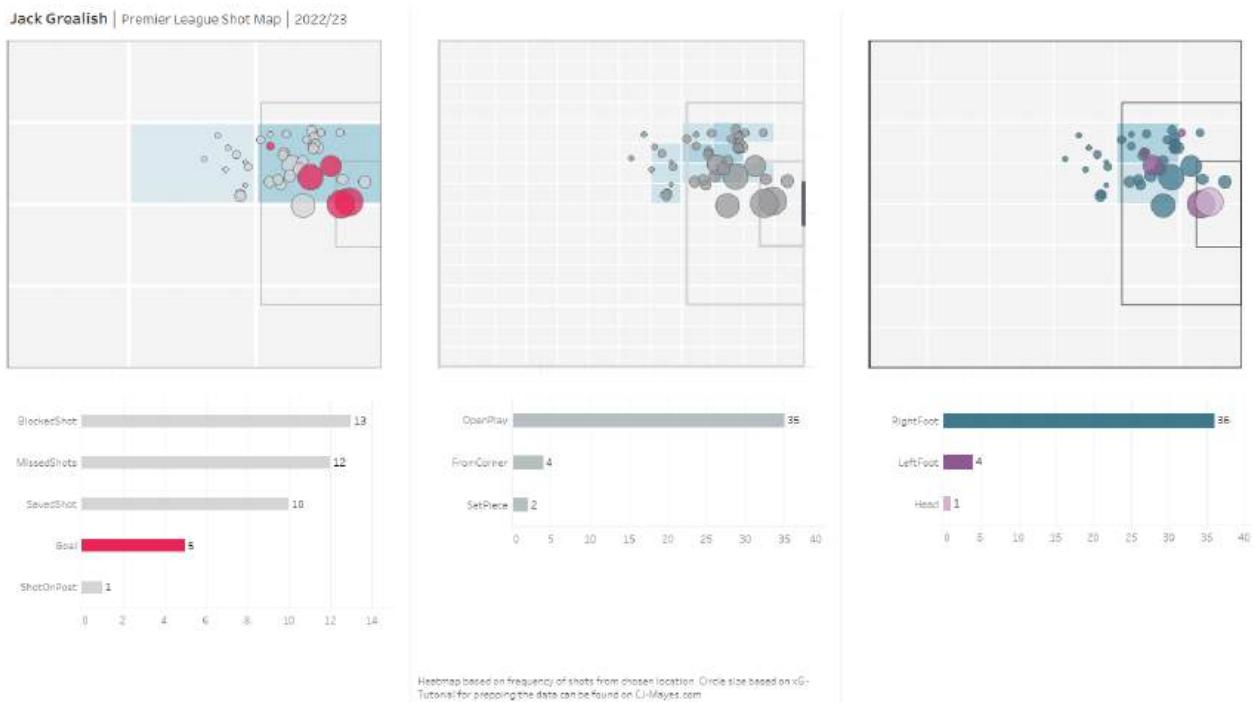
chatGPT Slack App 4:04 PM
Sure, here is a step-by-step guide on how to create a Slack app using the Slack API:
1. Create a Slack account and log in.
2. Go to the Slack API page and click "Create an App".
3. Enter the name of your app and select the workspace you want to use.
4. Click "Create App".
5. On the left-hand side of the page, click "OAuth & Permissions".
6. Click "Add a New Redirect URL" and enter the URL of your app.
7. Click "Save URLs".
8. On the left-hand side of the page, click "Bot Users".
9. Click "Add a Bot User" and enter the name of your bot.
10. Click "Add Bot User".
11. On the left-hand side of the page, click "Install App".
12. Click "Install App to Workspace" and follow the instructions.

christophermayes 4:09 PM
Will the app work if I stop my terminal running?

That's it for this week, of course lots more you can start to explore with this idea – we are only just scraping the surface. To take it further why not start playing around with temporarily storing message captures, not hosting it locally, and varying the types of prompts you ask. Chat soon & take care. **LOGGING OFF CJ**

AUTOMATED SHOT MAPS IN TABLEAU (MAKE GRID IN ALTERYX)

Hi all, Hope everyone is doing well. Last week I got to attend [Salesforce London World Tour on AI](#). So I'm still on a bit of a high from that having got to hang with Tableau friends, the wider community and hear from Mo Gawdat in the closing note from the day. In a rapidly changing environment, I find myself constantly in a position of listening and learning which I love! Anyway for today we leave AI at the door and go back to our tried and tested Tableau & Alteryx.



A week or so ago I wrote about how we can make our shot maps a little bit prettier in Tableau. Most of it was replicating design ideas seen in the Analyst website.... although I'm sure they do theirs through code? Who knows. Anyway, it got me really thinking. On **MPLsoccer** they have so many great examples of different pitch charts we can create and some are harder than others when trying to replicate it in Tableau. One of those charts is grid heat maps.

mpl soccer

Search docs

CONTENTS:

- Installation
- Pitch Basics
- Radar Charts
- Bumpy Charts
- StatsBomb

Examples

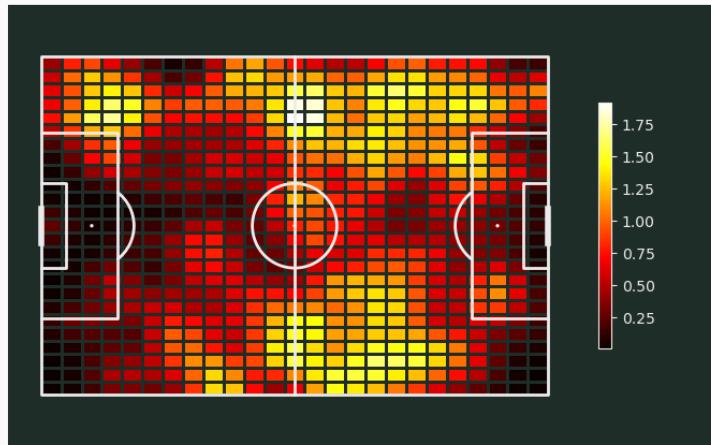
- Radar charts
- Pizza Plots
- Bumpy Charts
- Pitches
- Tutorials
- StatsBomb

Pitch setup

- Radar charts
- Pizza Plots
- Bumpy Charts
- Pitches
- Animation
- Convex Hull
- Customize colormaps
- Cyberpunk theme

Example code for "Pitch setup" included below

```
fig.set_facecolor('#22312b')
bin_statistic = pitch.bin_statistic(df_pressure.x, df_pressure.y, statistic='count', bins=(25, 25))
bin_statistic['statistic'] = gaussian_filter(bin_statistic['statistic'], 1)
pcm = pitch.heatmap(bin_statistic, ax=ax, cmap='hot', edgecolors="#22312b")
# Add the colorbar and format off-white
cbar = fig.colorbar(pcm, ax=ax, shrink=0.6)
cbar.outline.set_edgecolor('#fefefef')
cbar.ax.yaxis.set_tick_params(color='#fefefef')
ticks = plt.setp(plt.getp(cbar.ax.axes, 'yticklabels'), color='#fefefef')
```



Load some fonts, path effects, and a custom colormap

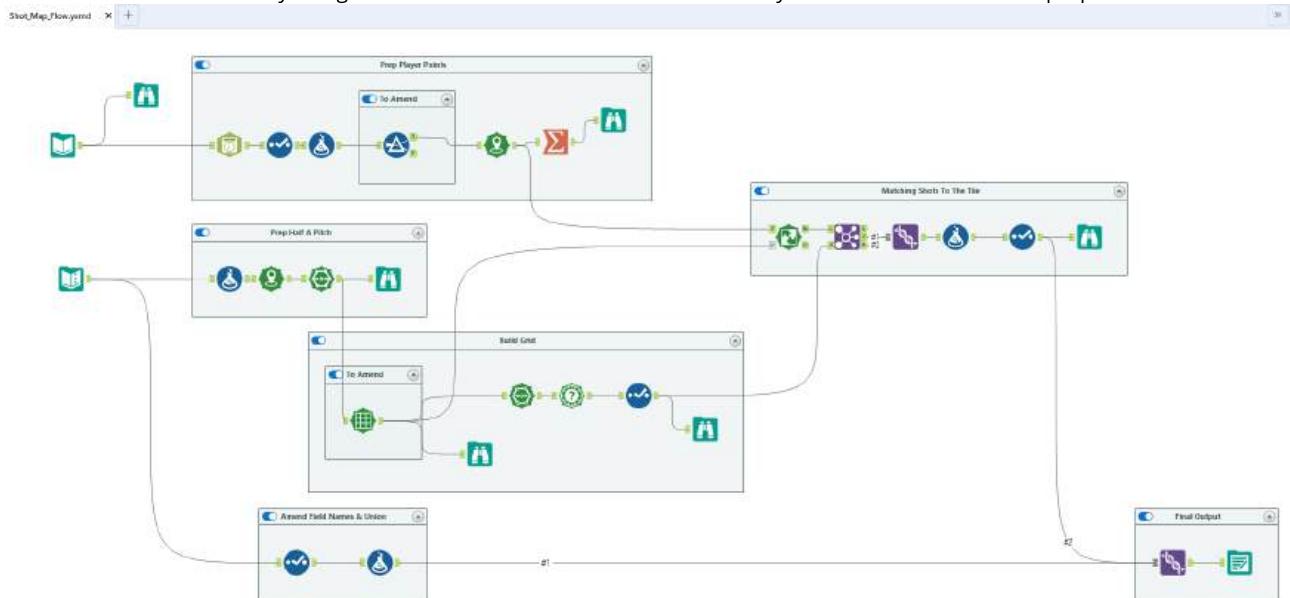
```
# fontmanager for google font (roboto)
roboto_regular = FontManager()

# path effects
path_eff = [path_effects.Stroke(linewidth=1.5, foreground='black'),
            path_effects.Normal()]

# see the custom colormaps example for more ideas on setting colormaps
pearl_earring_cmap = LinearSegmentedColormap.from_list("Pearl Earring - 10 colors",
                                                       ['#15242e', '#4393c4'], N=10)
```

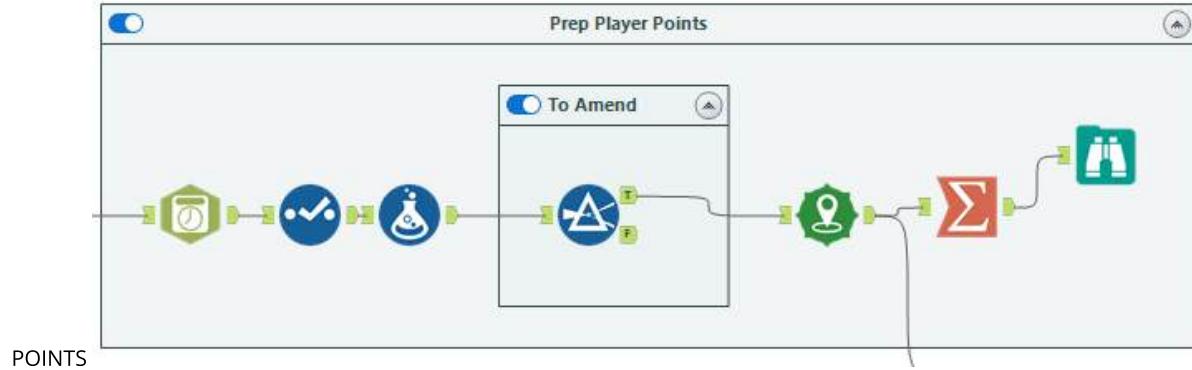
Now, I was really fortunate to be able to host back in Q1 of this year a guest blog from Fuad who produced such stunning heatmaps for shots in his Iron Viz entry for the 2023 feeder. It was a lovely workflow he shared, but the workflow did assume that the grid would have to be worked out based on the co-ordinates. I.e if you got a width of 10 then if you want 10 grids its a width of 1, 5 grids width of 2 etc. It got me thinking could we get it so we can just input an arbitrary number and that will help solve our grid issue? Well – I've come up with an alteryx solution that I hope people will enjoy. We will then visualise it in Tableau.

So Todays blog is all about the Make Grid function in Alteryx. Lets start with the data prep.



Bit of a chunky flow but we will explain each part. First things first is our dataset. We use understat data for the EPL 22/23

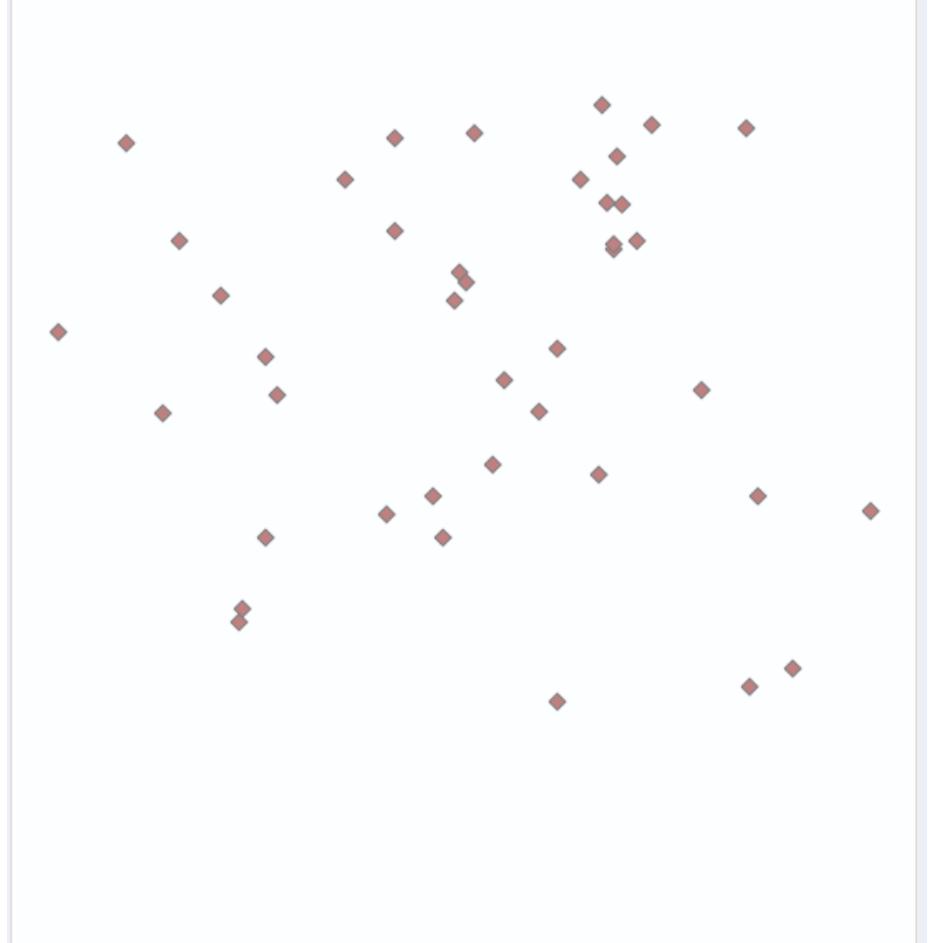
season. You can find a copy on my Github link at the top of the page, alongside the other resources to save locally. PREP PLAYER



POINTS

- The first three tools are really cleaning and transforming, nothing too crazy going on. I do end up scaling my X & Y points up drastically as I find bigger numbers easier to understand than small decimals, this is of course just preference and they get scaled back down at the end.
- The To Amend is a filter I added in, you'll see in my dashboard I end up restricting my points to just **Jack Grealish's** shots. Do update this or remove it if you want alternative players.
- The create points is a well named tool that.... makes points. We set our longitude and latitude values to our x and y coordinates.
- By using the summarise tool and combine action we are able to build this into one spatial object.

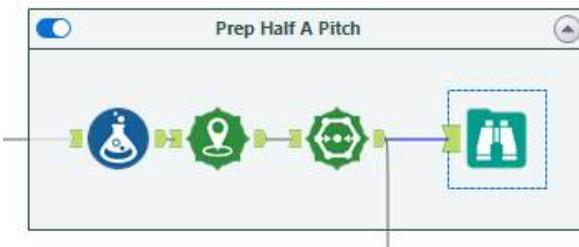
The output of that data looks like this



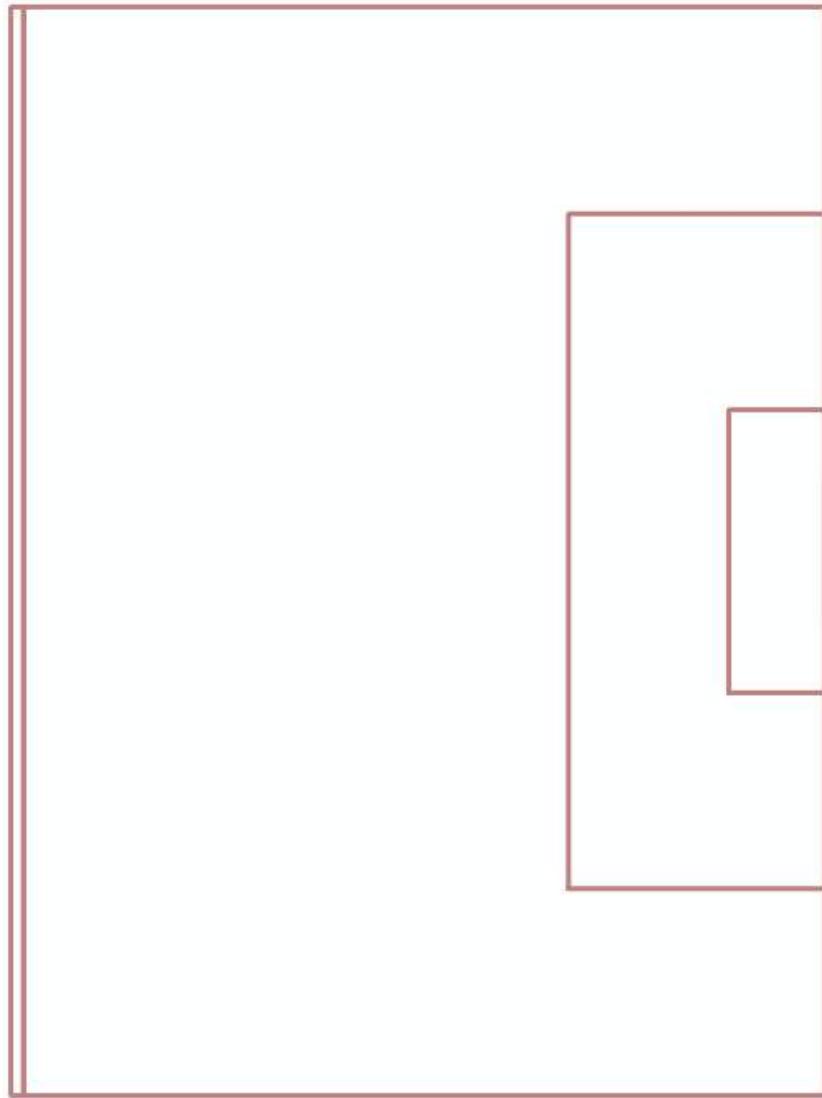
At this point you might think it's a little confusing with no context. It's always good to remember we have the understat website as a reference point.



HALF PITCH The next thing we want to do is to start building a base of points for the pitch. (Of course, we can render these as a background image in Tableau, but why not for this tutorial we give it a go creating a pitch using lines!

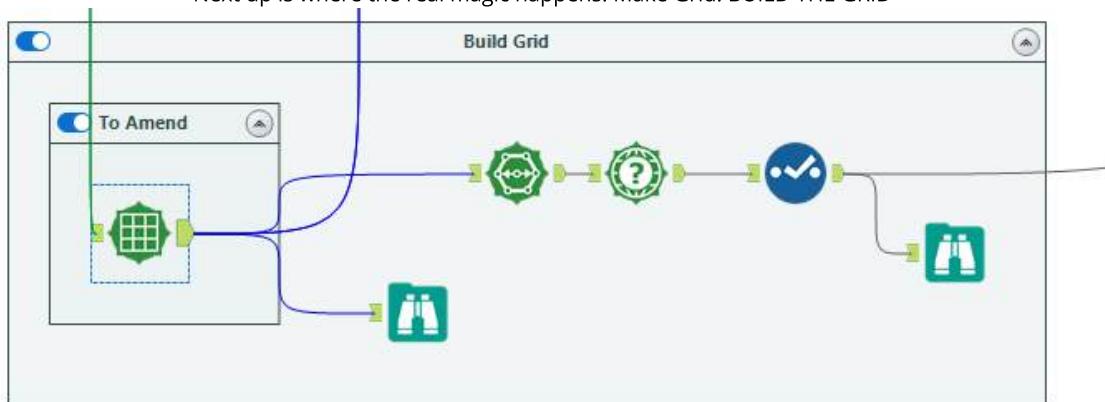


- First we scale our pitch values to that seen within the top half of the flow for convenience.
- We use the create points to identify the pitch sections including the 6 yard box, penalty box and half of the pitch.
- Next we add the Poly Build tool which helps us essentially assign a path order to our plotted points to build the frame of our pitch.

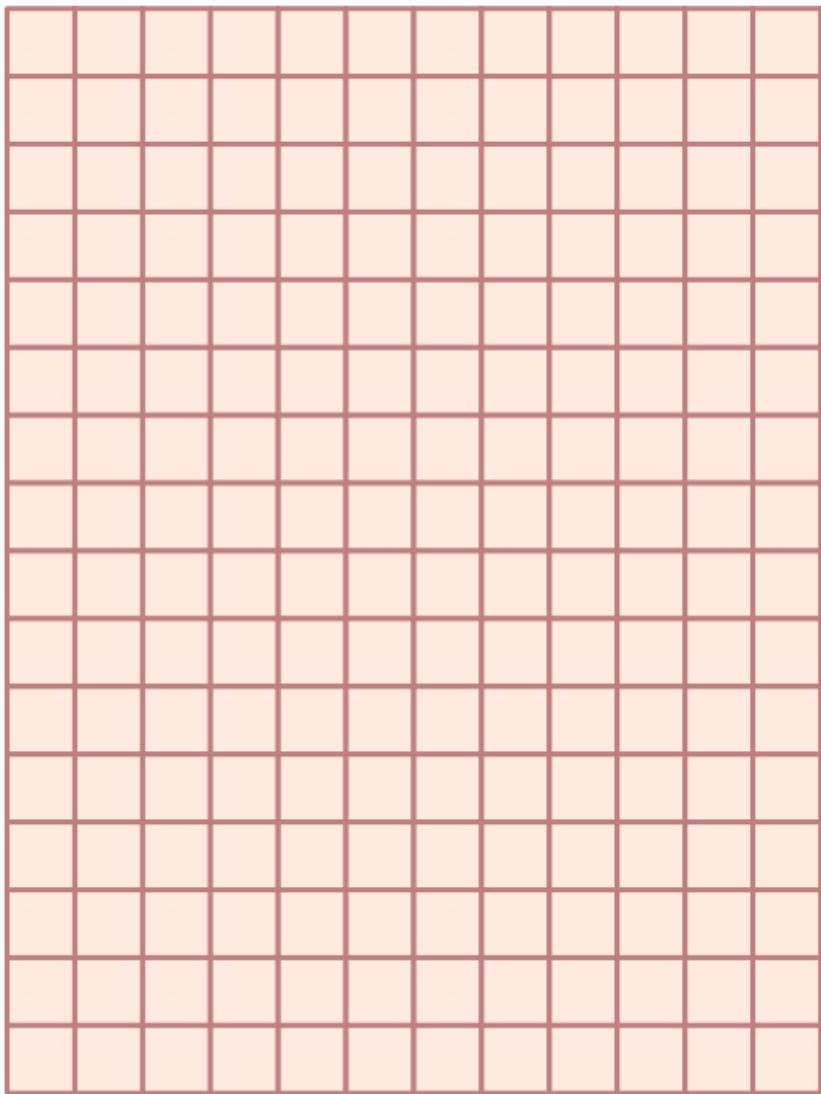


- I do recognise this is a little basic. Since drawing this i've gone back and added in some extra lines for the goal that we can bring into Tableau. The other thing to note is it is missing the semi-circle on the left, perhaps I add that in too at some point.
- You'll also see the weird stripe to the left.... reason I've added that in is because my grid wasn't aligning properly so needed to make the pitch wider to allow the next steps to fit. We can hide it when it comes to the final visual.

Next up is where the real magic happens. Make Grid! BUILD THE GRID



- Onto my magical Make Grid. We can set this value at different values to split our chart up. In my Tableau Output you will see that I have tried three different variations.
 - 1 at 1.47 which gives a 3*4 grid.
 - 1 at 0.735 which gives a 6*8 grid
 - 1 at 0.3675 which gives a 12*16 grid.



Really all this is doing is creating those grids that we will overlay on top of our pitch. But now we really want to find what the co-ordinates are for each square!

- We then use the PolySplit field to split each of the grids into points.

3 of 3 Fields | Cell Viewer | 960 records displayed | ↑ ↓

Record	GridName	Split_SpatialObj	Split_SequenceNum
1	Grid0_0	Point - View Browse Tool Map Tab	0
2	Grid0_0	Point - View Browse Tool Map Tab	1
3	Grid0_0	Point - View Browse Tool Map Tab	2
4	Grid0_0	Point - View Browse Tool Map Tab	3
5	Grid0_0	Point - View Browse Tool Map Tab	4
6	Grid1_0	Point - View Browse Tool Map Tab	0
7	Grid1_0	Point - View Browse Tool Map Tab	1
8	Grid1_0	Point - View Browse Tool Map Tab	2
9	Grid1_0	Point - View Browse Tool Map Tab	3
10	Grid1_0	Point - View Browse Tool Map Tab	4

Each grid is 5 points joined up in a

specific order to create the square.

Record	GridName	Split_SpatialObj	Split_SequenceNum	CentroidX	CentroidY
1	Grid0_0	Point - View Browse Tool Map Tab	0	0.0615	0.085
2	Grid0_0	Point - View Browse Tool Map Tab	1	0.066822	0.085
3	Grid0_0	Point - View Browse Tool Map Tab	2	0.066822	0.079678
4	Grid0_0	Point - View Browse Tool Map Tab	3	0.0615	0.079678
5	Grid0_0	Point - View Browse Tool Map Tab	4	0.0615	0.085

- By using the Spatial info tool, we can find the centroid as X and Y co-ordinates, identifying where they lie for each point. This way we will be able to create a dataset underlay for our polygons.

That's all well and good so to this point in the tutorial we have:

- Decided which players shots we are looking at
- Built a framework for the pitch to plot the shots
- Been able to create a grid to start to overlay onto the pitch.

We have one final big step to do, and that's to understand of each of the shots what grid do they land in? That brings me onto the final part of the flow. MATCHING SHOTS TO THE TILE



- We use the spatial match tool here, with the target as our centroid field (the X and Y of the players shots) and the Universe connector is then the grid field. I.e which grid does the shot take place.
- The Join tool is then used to bring in the polygon co-ordinates for each shot (yes 5 rows for each shot) This might start to explode your dataset so use with caution. You might want to rework this step otherwise.
- The last part to this is a matter of clean up.

Before I produce the final output I end up gluing in a full outer join the pitch details just so i can have this in my dataset, realistically this shouldn't be in our dataset but it was easier to have it all in one file. Then we finally export our data to a CSV.

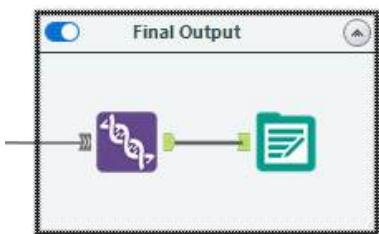
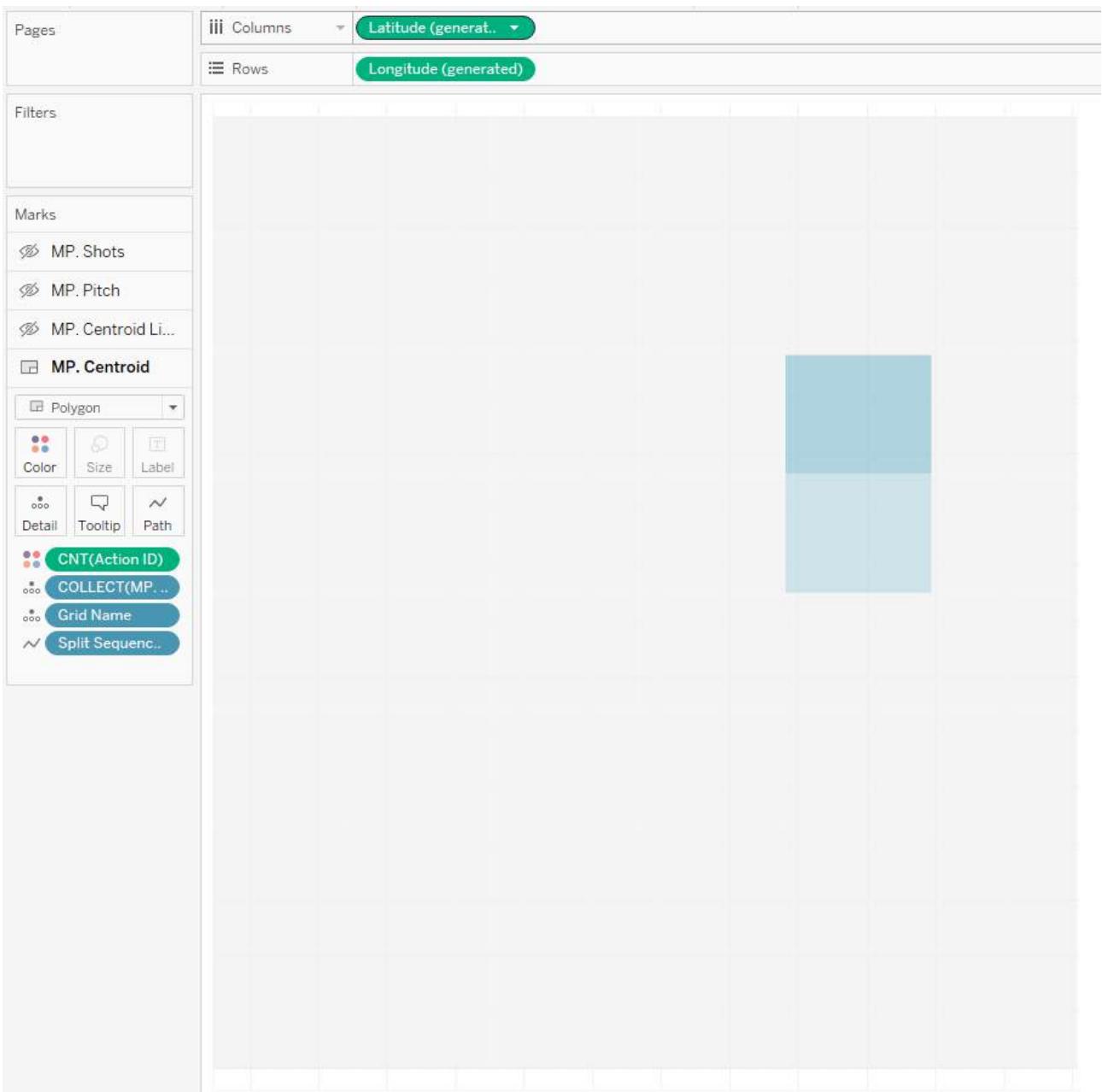
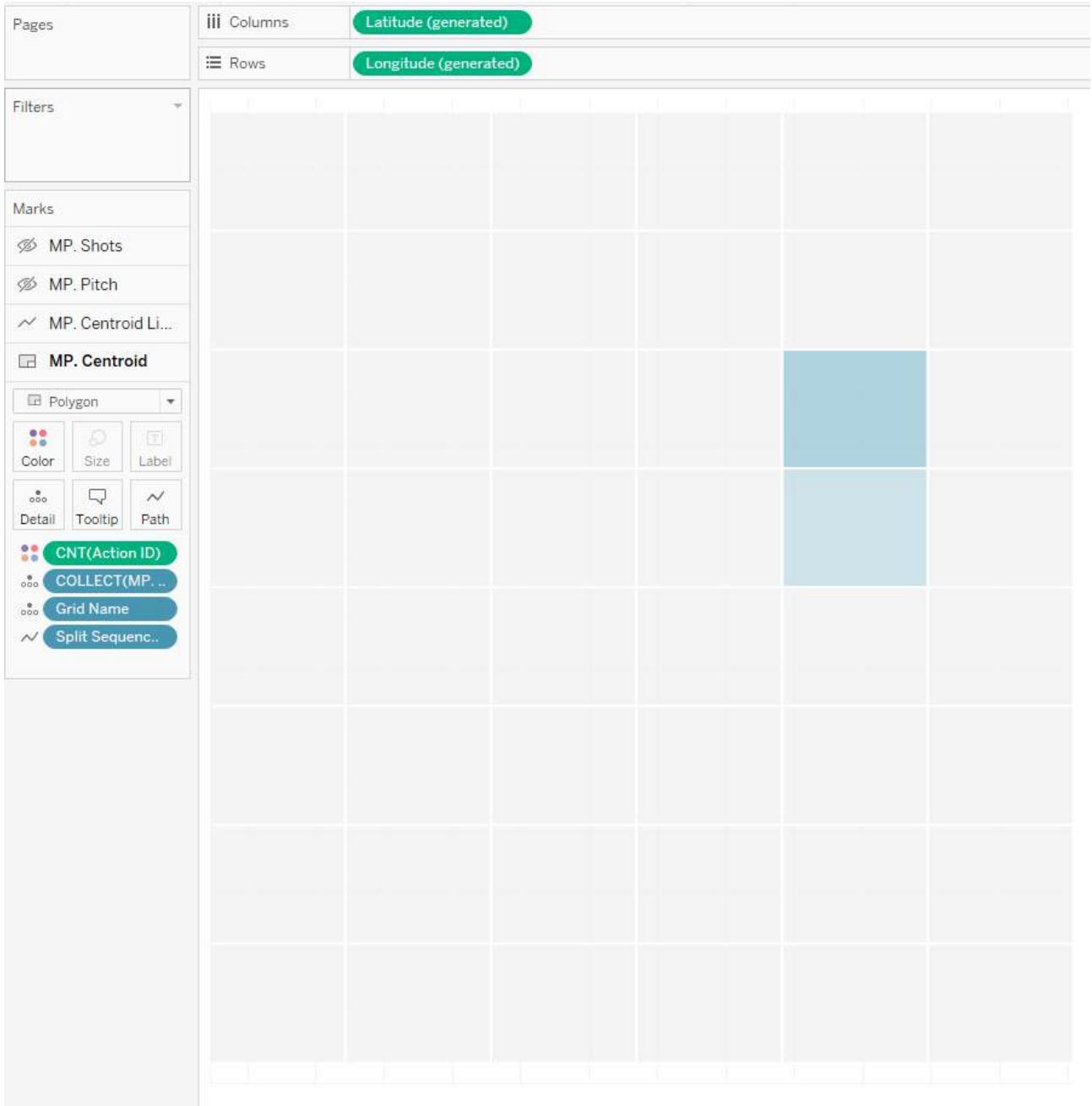


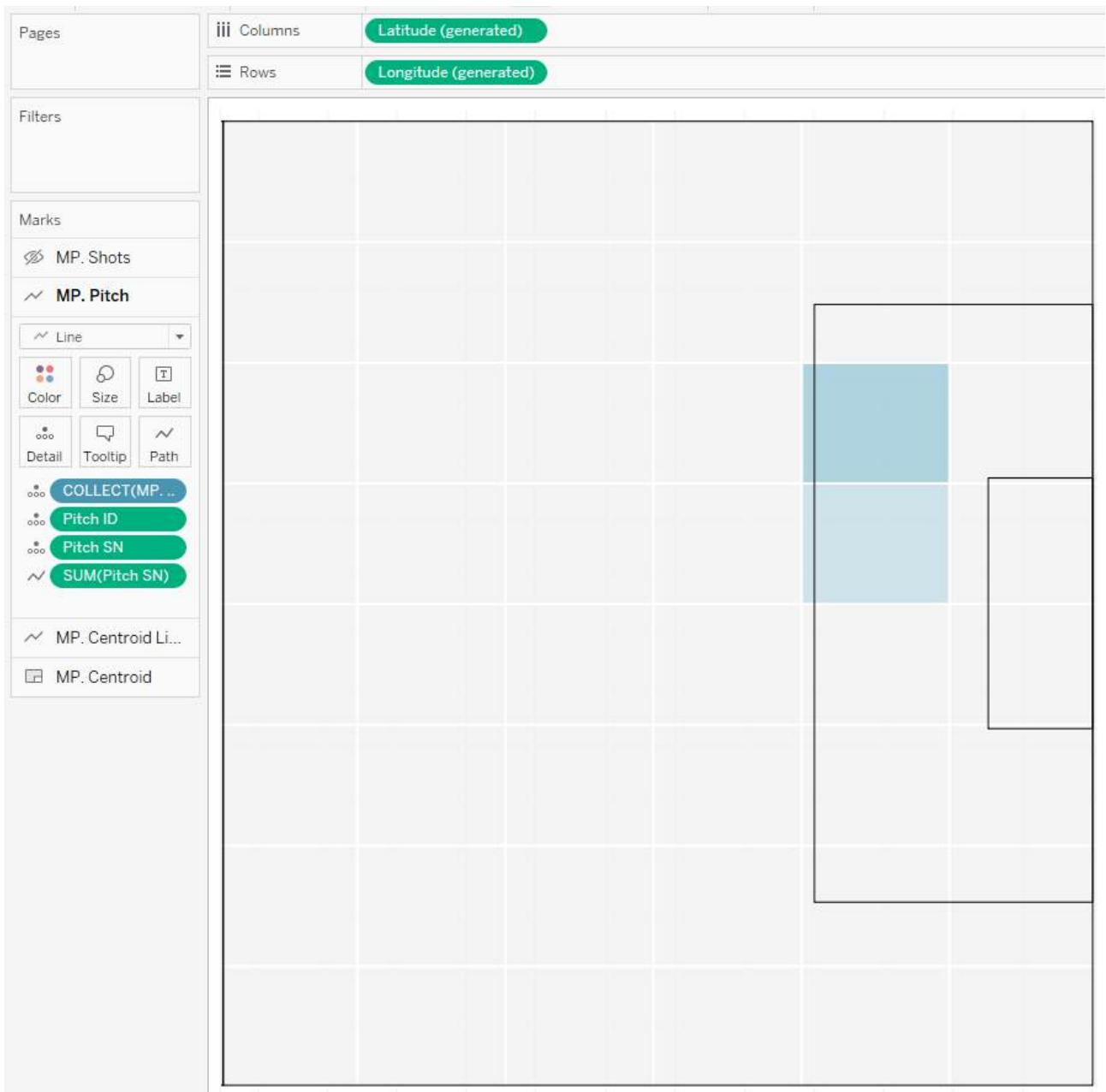
TABLEAU TIME You will find my visual attached at the top of this blog using the usual links. I don't think for the sake of this blog I will need to explain how map layers works - do just read any other tableau blog of mine if you need a little help, but lets take a little look at the different layers in action.



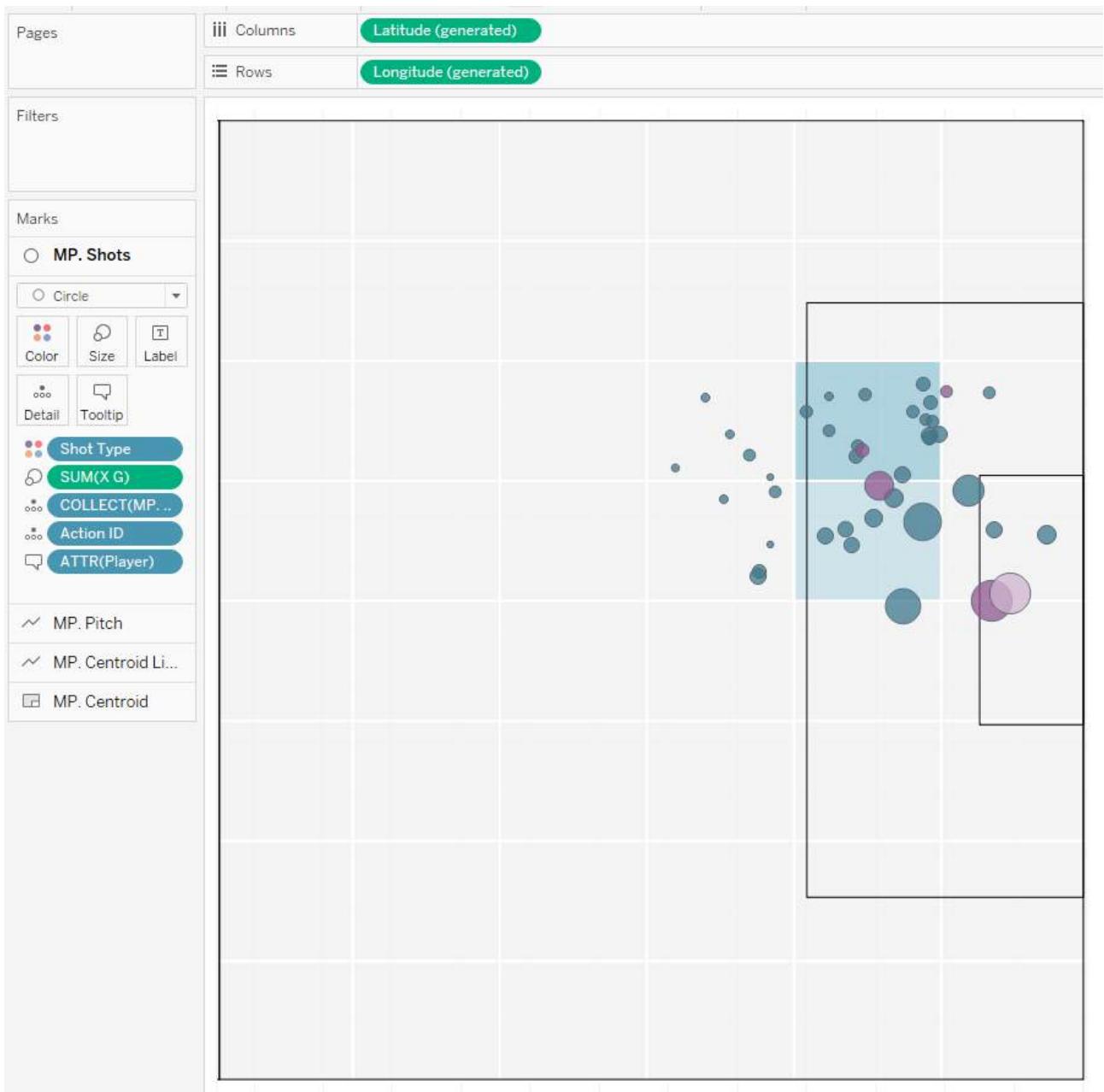
The first layer is using the grid information. We make separate grids of action, making sure the split sequence is on path to join our squares. This mark type is a Polygon. We are then able to count the number of actions in each box! More ActionID's relate to more shots taken. The next layer we start to think about bordering the polygons. Of course we could use the custom border but these are at preset levels. By adding in the same layer but making it a line, we can adjust the thickness of what this looks like.



Next we can add in the pitch frame. This data we had joined at the very end.

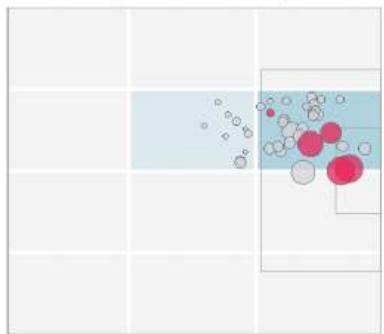


With all things lines and polygons, of course again we need to make sure we have the right things on detail and more importantly path.

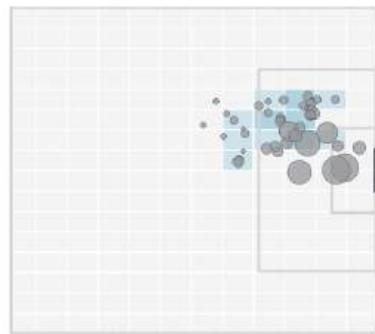


Final touches therefore is creating a new layer of just the shot co-ordinates to overlay on top. You can start to play around with grid size, colour and pitch layout to be able to transform your visuals. Going Further:

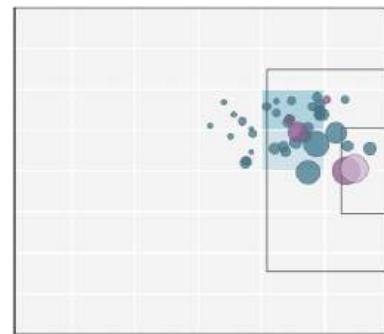
- Try plot shot maps for a whole team.
- Try hide squares that weren't shot from.
- Try adding in interactivity or tooltips into the visual.



BlockedShot	13
MissedShots	12
SavedShot	10
Goal	5
ShotOnPost	1



OpenPlay	35
FromCorner	4
SetPiece	2



RightFoot	36
LeftFoot	4
Head	1

Heatmap based on frequency of shots from chosen location. Circle size based on xG. Tutorial for prepping the data can be found on CJ-Myles.com.

Everything built in Tableau with no background maps feels quite slick in my opinion. Let me know your thoughts. **LOGGING OFF,**
CJ

REFERENCE BORDER TABLE CELLS IN TABLEAU

Hi all, Down the rabbit hole I go ever since being mind blown by a Workout Wednesday in Tableau. You will know the one I'm on about. Week 20 from this year.

Can you build a heat map with bathymetry lines?

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Accessories	3%	6%	12%	16%	22%	28%	38%	45%	60%	68%	83%	100%
Appliances	3%	7%	14%	19%	26%	33%	37%	49%	59%	67%	85%	100%
Art	3%	7%	12%	21%	29%	38%	45%	52%	65%	72%	86%	100%
Binders	6%	8%	15%	21%	26%	32%	37%	48%	66%	75%	85%	100%
Bookcases	5%	6%	13%	17%	22%	30%	38%	43%	63%	70%	90%	100%
Chairs	3%	6%	13%	19%	27%	33%	40%	46%	61%	69%	83%	100%
Copiers	3%	3%	20%	22%	34%	35%	41%	45%	52%	77%	87%	100%
Envelopes	5%	9%	19%	24%	31%	34%	41%	46%	59%	67%	85%	100%
Fasteners	1%	3%	5%	8%	9%	10%	12%	15%	20%	85%	91%	100%
Furnishings	4%	7%	14%	22%	29%	35%	43%	47%	60%	66%	83%	100%
Labels	2%	4%	11%	15%	22%	31%	44%	51%	63%	73%	88%	100%
Machines	4%	9%	27%	37%	43%	49%	51%	55%	69%	74%	92%	100%
Paper	3%	6%	14%	19%	27%	36%	41%	50%	64%	70%	86%	100%
Phones	4%	7%	16%	21%	28%	36%	43%	52%	63%	71%	88%	100%
Storage	4%	7%	13%	20%	27%	35%	41%	49%	62%	70%	86%	100%
Supplies	9%	10%	33%	46%	49%	51%	70%	72%	86%	88%	90%	100%
Tables	5%	7%	15%	20%	25%	33%	38%	46%	55%	65%	81%	100%

CHALLENGE BY : LUKE STANKE
CREDIT TO: DONNA COLES BLOG

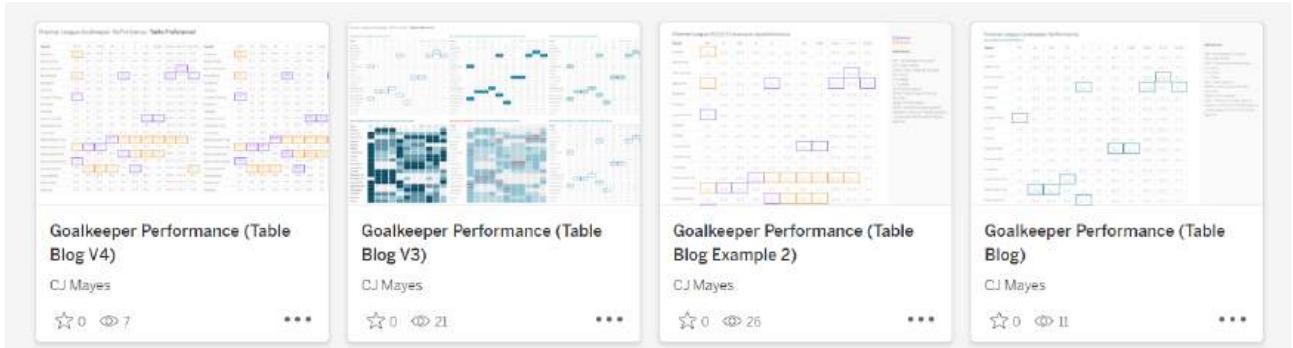
#WOW2023 | WEEK 20

RECREATED BY : CJ MAYES

Truth be told, On the most part, I feel at a point that things that are released now in the community I can generally look at and think I could have a half decent go at replicating if needed. That being said, when [this bathymetry lines](#) came out as a challenge from Luke Stanke I was genuinely lost as how to approach it. Long story short, we are fortunate enough to have individuals in the community that are kind enough to lend a hand in reverse engineering the workbook, and I leant heavily on the work of **Donna Coles** to be able to finally understand how the reference lines were working off the colour calculations. Donna is literally world class!! If you'd like to recreate the above do check out her blog [here](#). I really like her use of the lookup function, and her design choices of doing fixed width bars. Anyway, said visualisation made me start thinking about other types of tables. Today's blog will be a thought piece on tables and the use of colour.

Premier League Goalkeeper Performance - Table Preference?

I'll append the calculations needed to create one of the final tables at the end (they're simple don't worry), but by all means all my visuals are free to download from my Tableau Public.



Let's start with this table using FBREF data on goalkeeper performance from this season. I've always found the tough thing about heatmaps is the overlay of text marks. Here I have a 4 step ranked colouring.

Generic Heatmap, Stepped Colour Split - Thick White Column Borders

Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	108.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	136.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	8.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.3	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	10.0	21.0	78.0	2.1	108.0	185.0	60.5
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	60.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	33.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	18.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

In Tableau the

labels on auto don't render too well contrast wise in my eyes, as they sink into the page. I wish they were adjustable in a more flexible way. It becomes difficult when your lightest and darkest colours differ in shade values strongly. Part of me started to think okay what if we lighten some of these values, but then add in some of the logic we've learnt from the bathymetrics reference lines. I came up with this,

Max Value For Each Metric || With Additional HeatMap Background Ranking

Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	108.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	136.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	8.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.3	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	10.0	21.0	78.0	2.1	108.0	185.0	60.5
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	60.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	33.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	18.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

It started to feel like to me the narrative of what I was trying to show was changing. We've gone from showing how values vary amongst 20 teams to showing how they vary amongst the teams but also what is the maximum value of each column. Two things stuck out to me... by trying to answer two questions in one am I really answering anything? Do you feel overwhelmed looking at this visual? I did slightly. Saying that, I did quite like the use of a lighter colour palette. The rabbit hole continues. Maybe I should just be looking at the maximum values of the different metrics. It could make for a cleaner visual and I can still use my new bordering technique I learnt from the Work out Wednesday.

Max Value For Each Metric || Outline Boxed

Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	108.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	136.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	8.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.3	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	10.0	21.0	78.0	2.1	108.0	185.0	60.5
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	60.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	33.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	18.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

I quite like this.

It's more subtle, it's clear. The main positive to doing something like this, is it feels a bit more old school reporting. You'd be able to put this in a report, or print it out and you wouldn't have problems of box colour spill. (You wouldnt even have to print it in colour!) But do we lose some of the attention to those maximum areas?

Max Value For Each Metric || Outline Boxed, With Blue Text Overlay

Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	108.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	136.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	8.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.3	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	10.0	21.0	78.0	2.1	108.0	185.0	60.5
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	60.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	33.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	18.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

We can create

separate calculations for the maximum boxes and add them. Now our values also lift off the page when we highlight the maximum using text colour. A minor upgrade! But are we over complicating it?

Max Value For Each Metric || Max Highlighting, With Custom White Text Contrast

Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	106.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	156.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	6.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.5	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	21.0	78.0	2.1	106.0	185.0	60.5	
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	92.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	53.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	16.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

Max Value For Each Metric || Same Text, Dotted Box

Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	106.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	156.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	6.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.5	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	21.0	78.0	2.1	106.0	185.0	60.5	
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	92.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	53.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	16.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

Above left is block colour Tableau in its native form. Simply a colour encoding for the maximum values. It is intuitive, it draws attention and the message is clear. Maybe the reference lines aren't needed at all. Perhaps you may think the lighter blue version to the right works better, maybe a border isn't needed at all, this could be the end to the rabbit hole. But alas, Lets change the context of the visual one more time. Currently we are only looking at the maximum value. Well if we consider the metrics maximum for some of them will be a positive thing, (e.g number of shots saved, number of wins) but for other the minimum would be a good indicator of success (number of losses, goals against) Lets build that logic into our visualisation.

Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	108.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	136.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	8.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.3	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	10.0	21.0	78.0	2.1	108.0	185.0	60.5
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	60.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	33.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	18.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

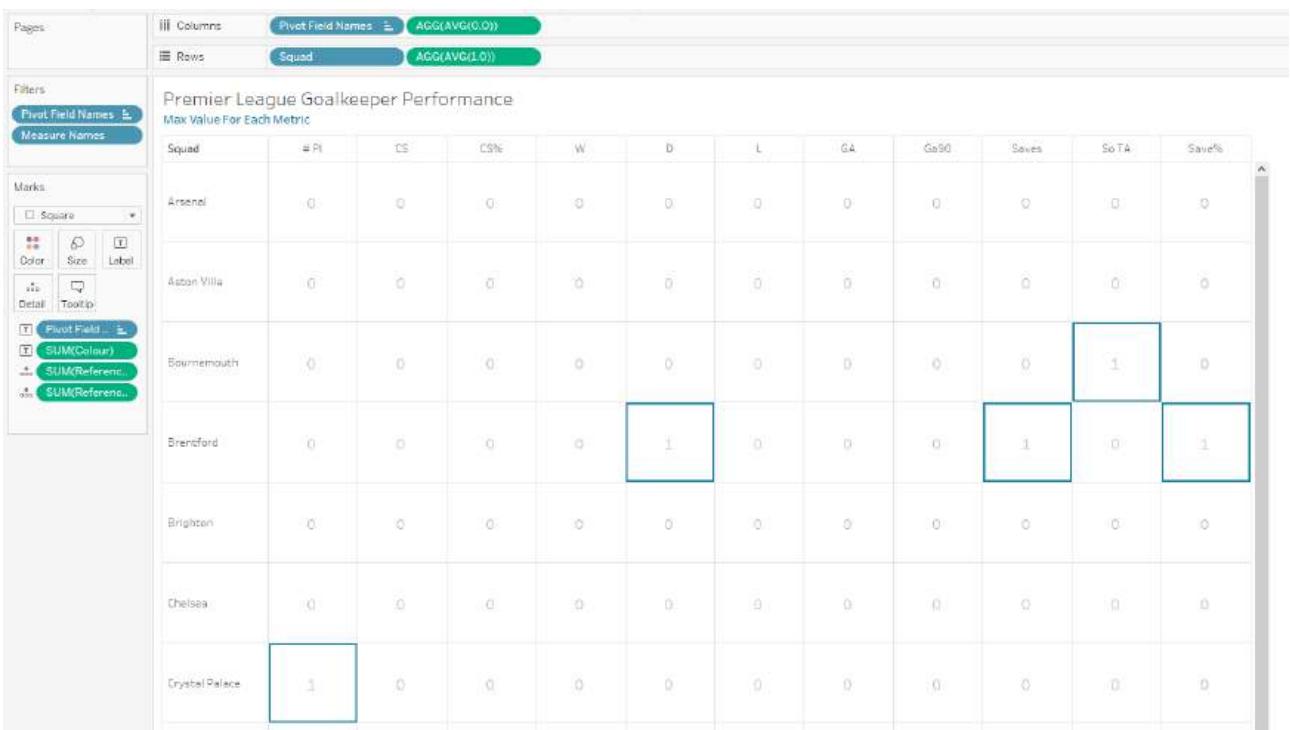
Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	108.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	136.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	8.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.3	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	10.0	21.0	78.0	2.1	108.0	185.0	60.5
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	60.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	33.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	18.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

Above are my two 'report print' visuals using the logic I learnt from the Workout Wednesday.

Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	108.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	136.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	8.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.3	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	10.0	21.0	78.0	2.1	108.0	185.0	60.5
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	60.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	33.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	18.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

Squad	# Pl	CS	CS%	W	D	L	GA	Ga90	Saves	So TA	Save%
Arsenal	1.0	14.0	36.8	26.0	6.0	6.0	43.0	1.1	94.0	136.0	70.6
Aston Villa	2.0	12.0	31.6	18.0	7.0	13.0	46.0	1.2	108.0	151.0	72.8
Bournemouth	2.0	8.0	21.1	11.0	6.0	21.0	71.0	1.9	136.0	204.0	68.1
Brentford	1.0	12.0	31.6	15.0	14.0	9.0	46.0	1.2	154.0	197.0	77.7
Brighton	2.0	12.0	31.6	18.0	8.0	12.0	53.0	1.4	82.0	131.0	64.1
Chelsea	2.0	10.0	26.3	11.0	11.0	16.0	47.0	1.2	115.0	159.0	71.1
Crystal Palace	3.0	9.0	23.7	11.0	12.0	15.0	49.0	1.3	114.0	160.0	71.9
Everton	2.0	9.0	23.7	8.0	12.0	18.0	57.0	1.5	127.0	185.0	71.9
Fulham	2.0	9.0	23.7	15.0	7.0	16.0	53.0	1.4	144.0	196.0	76.5
Leeds United	2.0	5.0	13.2	7.0	10.0	21.0	78.0	2.1	108.0	185.0	60.5
Leicester City	2.0	7.0	18.4	9.0	7.0	22.0	68.0	1.8	113.0	178.0	64.0
Liverpool	2.0	14.0	36.8	19.0	10.0	9.0	47.0	1.2	111.0	157.0	71.3
Manchester City	2.0	13.0	34.2	28.0	5.0	5.0	33.0	0.9	60.0	92.0	67.4
Manchester Utd	1.0	17.0	44.7	23.0	6.0	9.0	43.0	1.1	99.0	142.0	71.1
Newcastle Utd	2.0	14.0	36.8	19.0	14.0	5.0	33.0	0.9	94.0	127.0	75.6
Nott'ham Forest	3.0	8.0	21.1	9.0	11.0	18.0	68.0	1.8	112.0	182.0	65.9
Southampton	2.0	4.0	10.5	6.0	7.0	25.0	73.0	1.9	81.0	150.0	52.7
Tottenham	2.0	10.0	26.3	18.0	6.0	14.0	63.0	1.7	109.0	172.0	65.1
West Ham	2.0	9.0	23.7	11.0	7.0	20.0	55.0	1.5	115.0	170.0	70.0
Wolves	2.0	11.0	28.9	11.0	8.0	19.0	58.0	1.5	115.0	174.0	67.8

& Here are the two more standard Tableau colourings you'd expect to see with Fill colour. I'll leave it to you all to decide when you may want to implement one solution over another, but I have thoroughly enjoyed thinking about colour contrast, shades, the offset of text colour against backgrounds as well as what context the visual will be used in, be it print, hosted on Tableau Server or elsewhere) Finally for those that are wondering how the reference lines are created, the calculations are below: First we find the value of each metric that is the maximum. { Fixed : max()} Next we can create a colour boolean, if the value is equal to our maximum then assign it a 1. if = then 1 else 0 END



Finally, both axis will want to add a reference line at 1, and another at -1 both horizontally and vertically here is the calculations.
 $IF = 1 \text{ THEN } 1 \text{ ELSE null END}$ The other is of course, $IF = 1 \text{ THEN } -1 \text{ ELSE null END}$ (It's worth mentioning I made my boxes axis between -1 & 1 hence the above two calcs. These will vary depending what your box size is. With this method it did need 4 reference lines. That's it for this week, Message me on LinkedIn or Twitter if you have any questions – as always the resources can be found on Tableau Public. I want to close out by saying thank you to Steve Wexler who replied fairly rapidly when I threw this idea past him the last few days – someone who does a spectacular job for keeping us honest to best practice when it comes to visualisation, and a viewpoint many of us respect in the community....without telling you his thoughts, I want to hear from you first, Which version works best in your eyes? Check the [visual here](#). **LOGGING OFF, CJ**

DASHED LINES IN RADIALS (TABLEAU)

Hi all, WOOOOOO!!!! Thank you tableau. We now have dashed lines in 2023.2 meaning a new way of using the functionality to draw attention to specific parts of our charts. Now, early disclaimer. I will leave the when to use / how and why to others. There are plenty of better advocates for best practice in the community and resources at your finger tips of when to use dashes in your charts. In fact, just this week the **Flerlage Twins** released this fantastic blog on just that. It really dives into some great examples around projections, confidence intervals as well as the different activity of lines vs colour. Please check it out [here!](#)

TABLEAU'S 2023.2'S DOTTED LINES BUT WHERE SHOULD WE USE THEM?

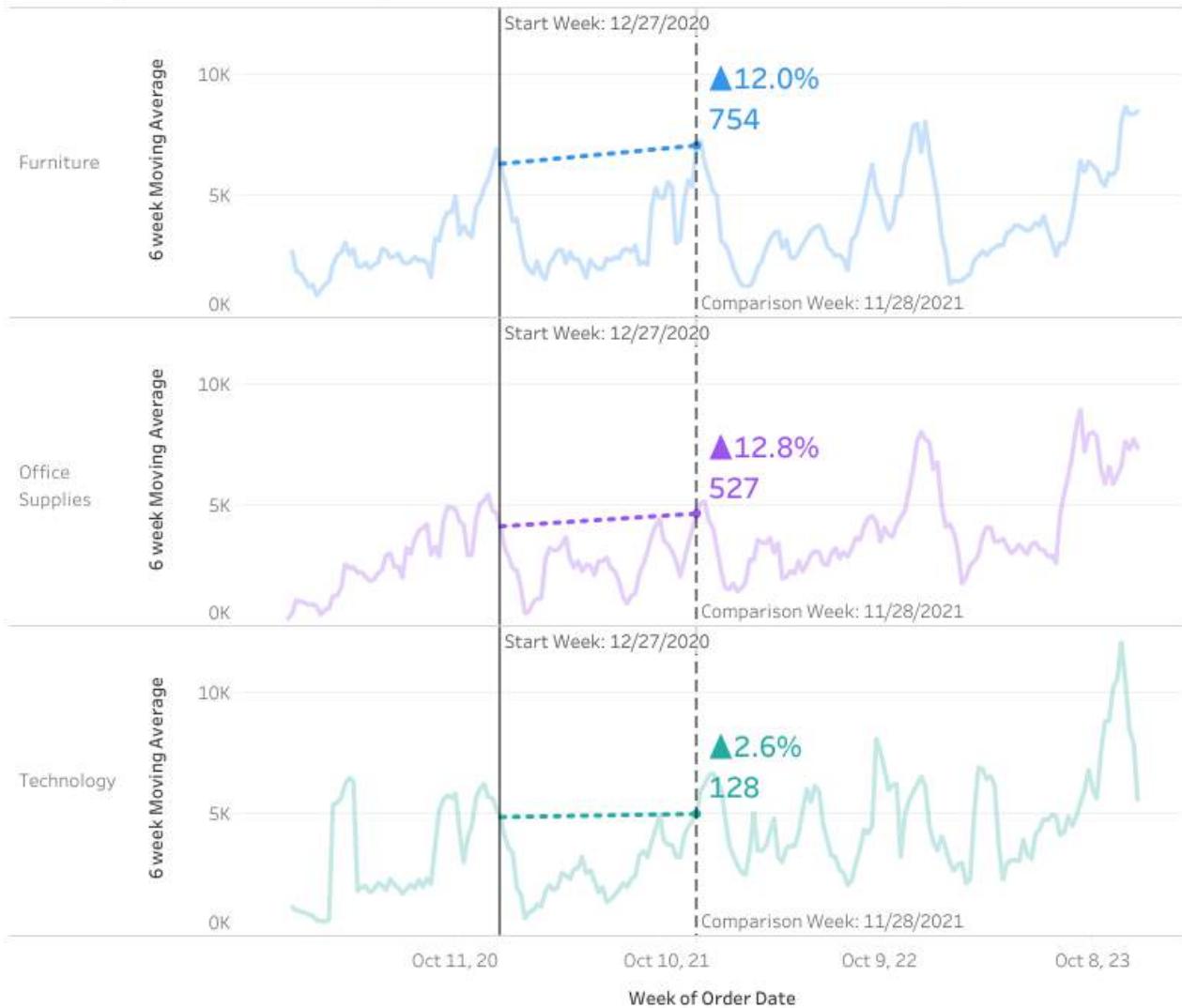
Tableau 2023.2's Dotted Lines - Where Should We Use Them?

by Kevin Flerlage · 2 DAYS AGO · 3 MINUTE READ

If you want to put your new dot lines in 2023.2 to good use, why not also have a go at the latest weeks [Workout Wednesday](#).

#WOW2023 W25

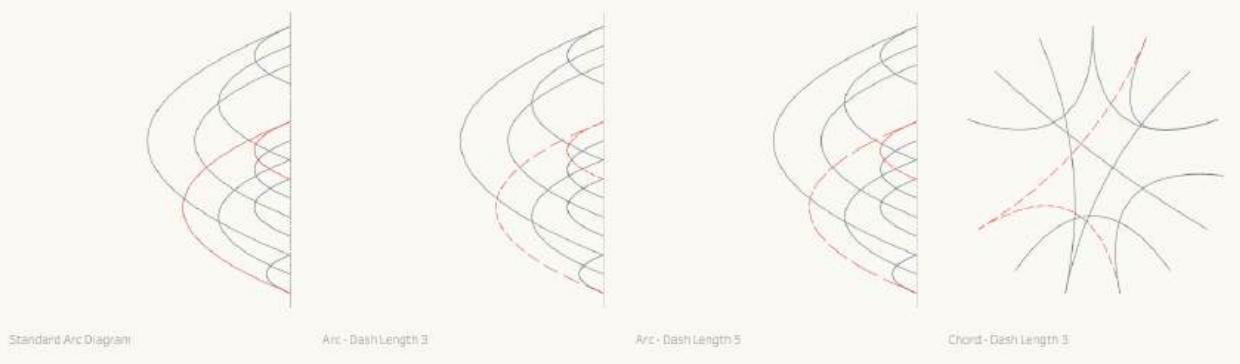
Can you calculate a percent difference between a range of dates?



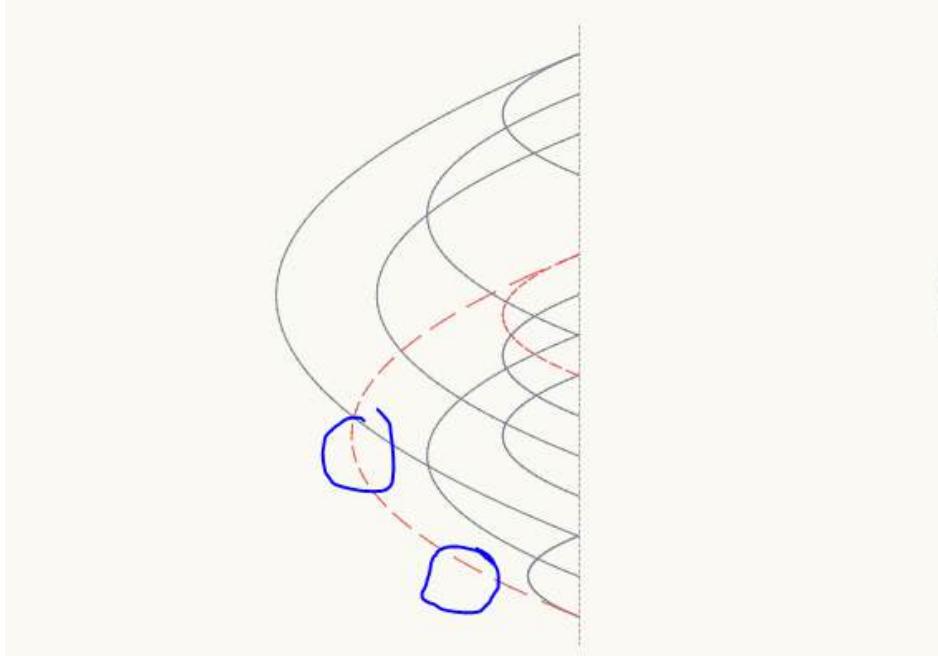
#WOW2023 W25 | Data: Superstore | Inspired By: @Jacro28 | Challenge By: @_Lorna_Brown

Anyway, for the blog this week we will look at what it means for the maths behind radial and curved designs, and the improvements it can bring. **What it was before when looking at radials?** Before the introduction of dashed lines we could create a similar effect through data scaffolding. You can check out how chord / arc charts we could create dashed effects using an extra column onto detail. This help essentially 'cut up' our data densification points splitting them out. [Here's](#) where i've done it before.

BEZIER | Development



But what you may notice with these dashed lines are they are of different lengths.



It was previously only in the case of perfect circles that the length of each fake dash would be of equal length. Which makes sense when we are essentially marking where we want our points along a curve, they won't all be equidistance if our lines start spiralling here there and everywhere. Below you can see how they work fine for radials, where I've used them in a [champions league visual](#), years ago. (Extra time had a radial of certain length vs going to penalties....)

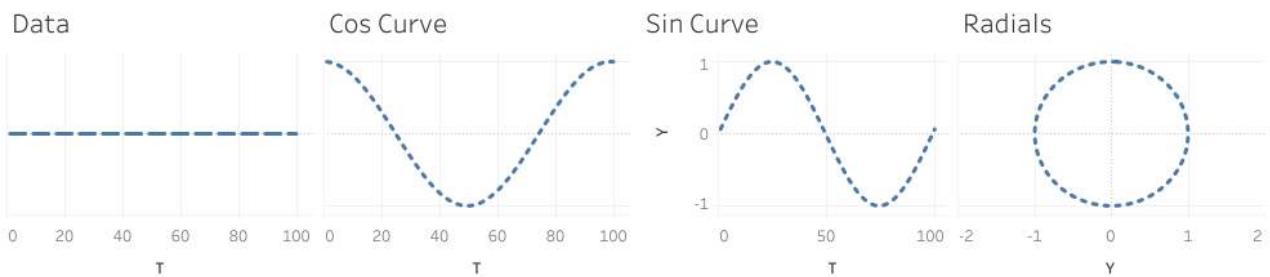
UEFA CHAMPIONS LEAGUE



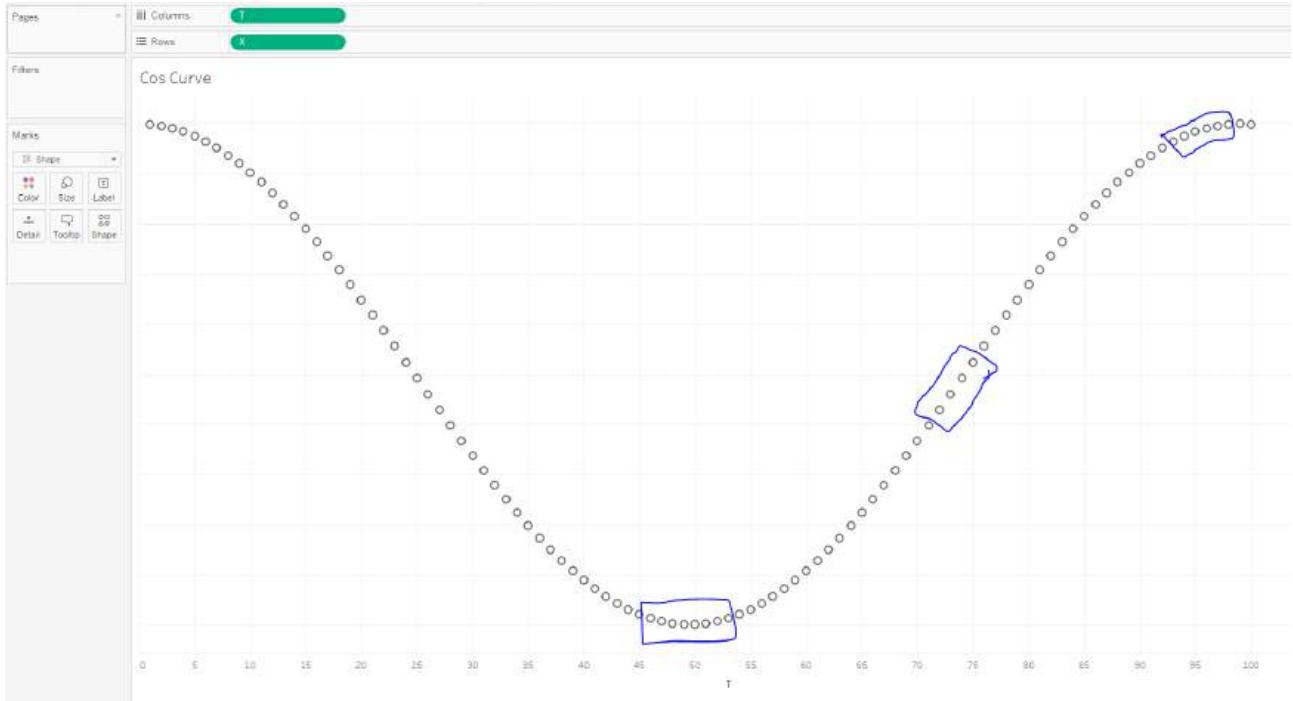
Mathematically this became problematic for non-circles because we want a point because we need our line to curve and not be jagged, but we don't want the point because it makes for unequal length arc dots. Tough, which meant previously unless looking at circles it was kind of a design no go. **New 2023.2 Update** The good news is, that problem is no more. Opening up that old Bezier curve work I had created, I can go back in and remove the detail and change the line to dashes very easily. It is under path and Line Pattern!

By all means take a copy of the dashboard and do the same. **Let's See It In Practice From Scratch** Lets go from the very beginning. Connect to a dataset with 100 rows (Column Name: T)

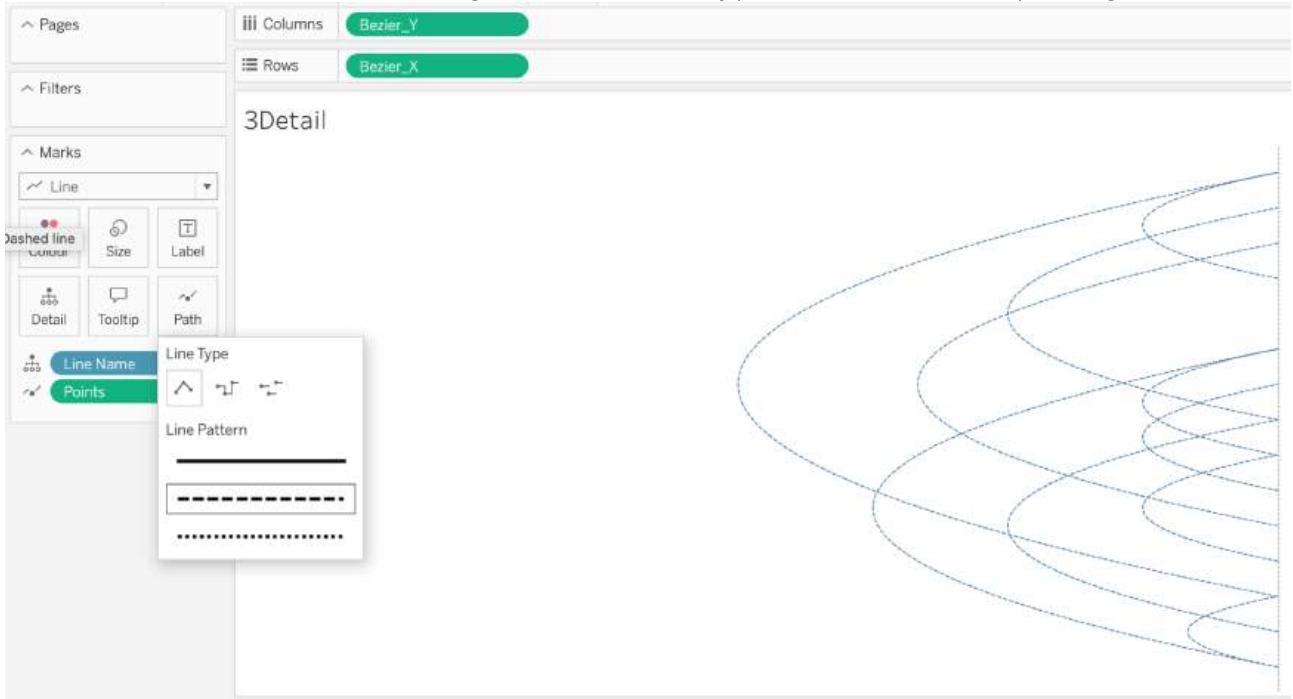
- Create a field Rank, with the value T in it.
- Create and Angle field with $360/100$ (This is the a full circle / the number of points, each angle should be the same gap)
- Create a rank angle calculation ($\text{rank} * \text{angle}$) This creates the spread of each mark to be along the curve.
- Create x field, This is $\cos(\text{radians}(\text{rankangle}))$
- Create y field, This is $\sin(\text{radians}(\text{rankangle}))$ x and y is following trigonometry to create the waves that together make a circle.
- Plot them against one another to see how they interact and make the new mark line of the path a dash. You can download my example workbook from my page [here](#).



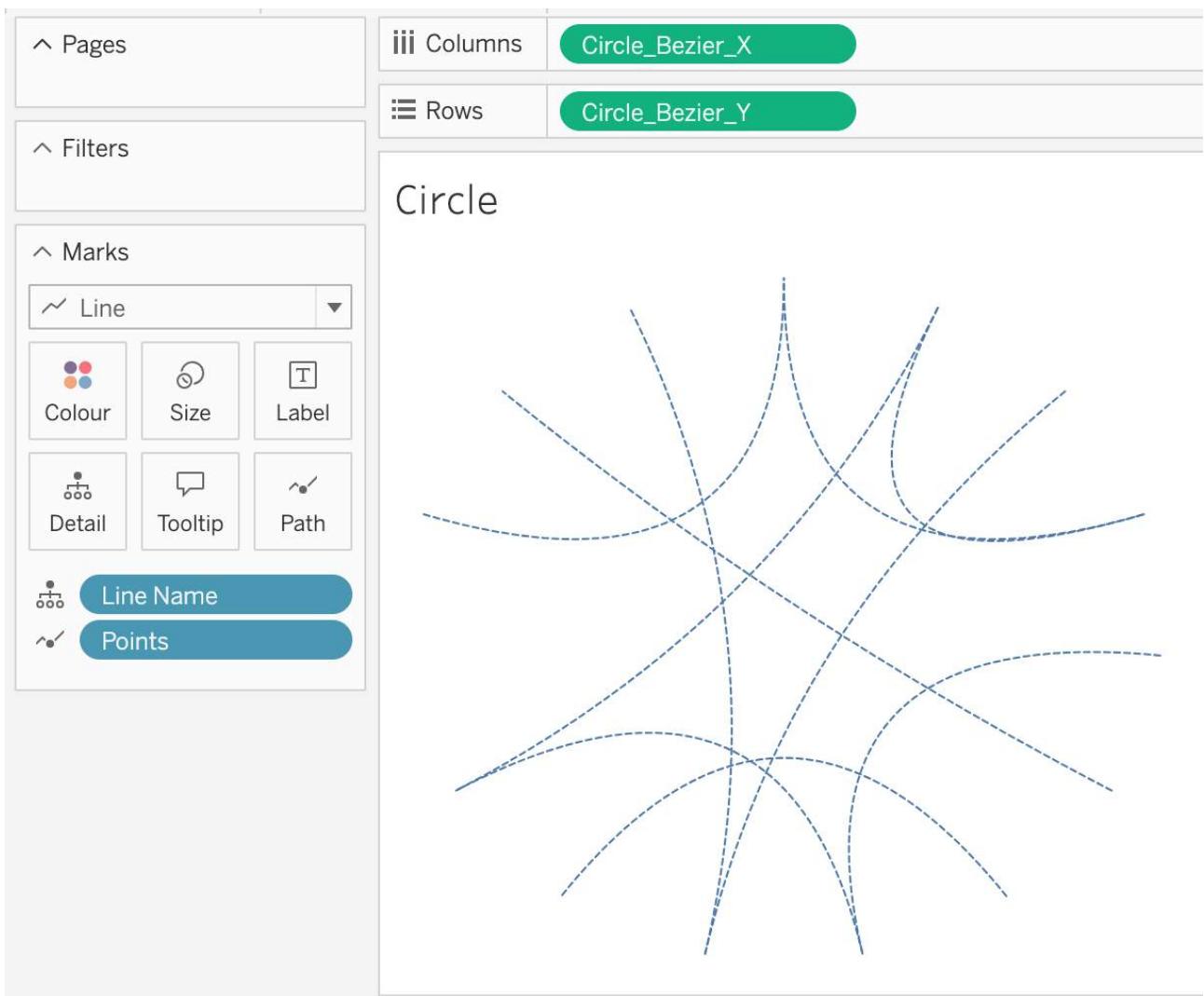
Not the most creative design or curves, but its a good way of showcasing the even 'spacing' of the dots from the dashed line ability. This is regardless of our data scaffolding bunching that you see, if we change our marks card back to circles it now doesn't matter that they no longer are equidistant.



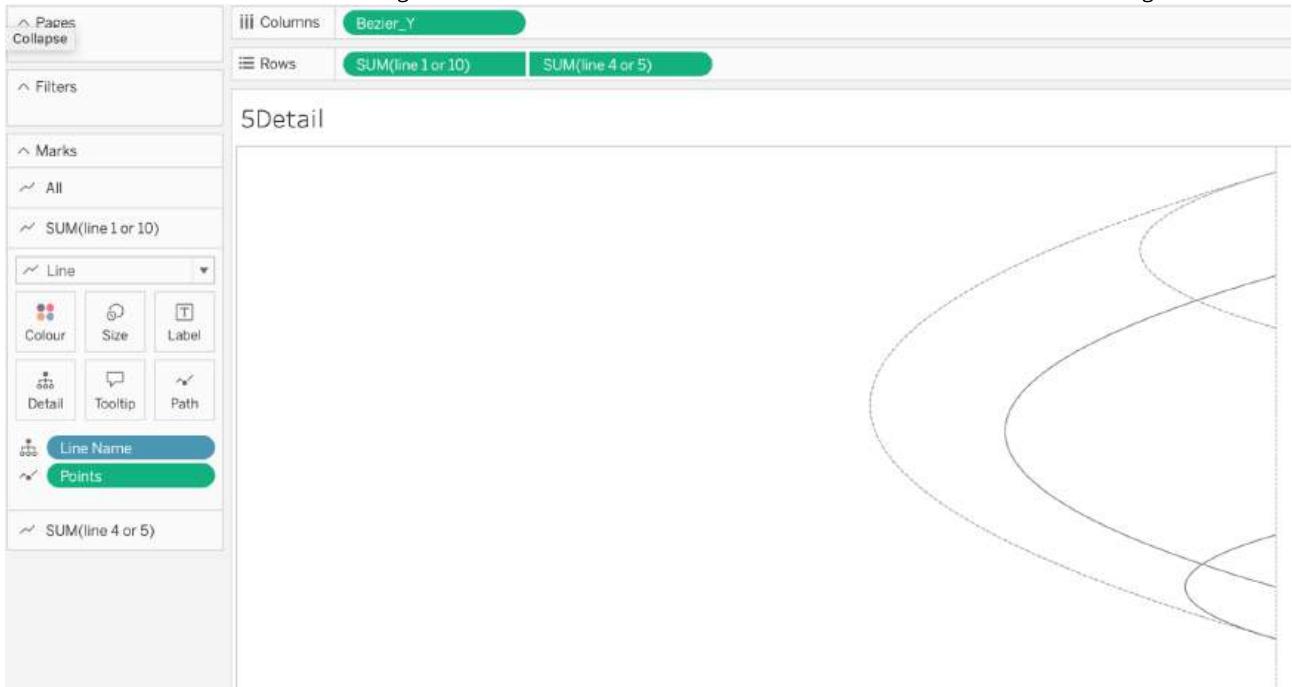
We can also check this works using the old workbook on my profile, here are a few examples using arcs



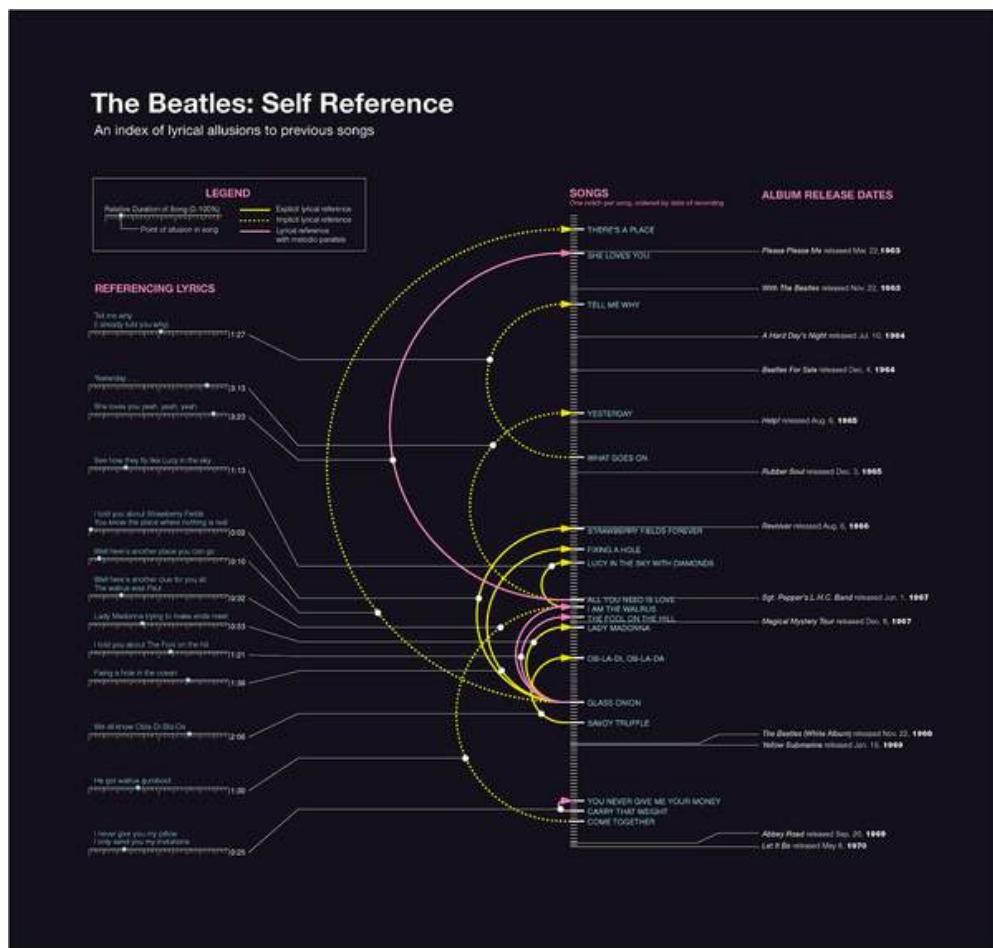
and chord diagrams.



We can even dual axis two lines against one another to have some of our lines as dashes and others as straight lines.



Are There Any Good Examples Online? Well it jogged my memory of when I was putting together my iron Viz Feeder for 2022, and the inspiration I found in this beatles visual. You can find a copy on [designinspo](#).



If you are after alternative creative ways of using dashes that isn't radial inspired, then I've also seen [Dennis](#) doing some stuff with it implemented in maps – with some amazing user cases including pedestrian/bike paths, park trails, restricted roads, planned roads, disputed boundaries.



Bike Pathways Central Ottawa

Excited to see how people utilise it in creative ways. Speak soon – time for a morning run! **LOGGING OFF, CJ**

EVEN MORE SHOT MAPS (PYTHON)

Hi all, I thought it had been a while since I've done a thought piece. I have been a huge fan of some of the content that has been released this year within the wider football community in terms of visuals, so wanted to do a piece on design tips and tricks I've seen that works well. It also offers me the opportunity to say that I've updated some previous code that stopped working last year. I want to say thank you to Adam Silverman for letting me know! My code is sloppy at the best of times. Within the new code, Main.py will create 2 datasets. The first is the summary stats for the season 2022/23 of the Premier League.

The second is shot data for every player that played in the premier league in 2022/23.
(This shot data if running the code will export every shot for that player regardless of season and league if on understat) The file I've uploaded with _EPL is a clean version of that dataset that only looks at Premier League 2022/23

Of course, you could add a line of code in to filter back to the required season and matches, but nicer for those that don't want to amend code to have everything.

```

import pandas as pd
from understatapi import UnderstatClient

understat = UnderstatClient()

league_choice = 'EPL'
season_choice = '2022'
league_player_data =
understat.league(league=league_choice).get_player_data(season=season_choice)

df = pd.DataFrame(league_player_data, columns=['id', 'player_name', 'games','time',
'goals', 'xG', 'assists',
'yellow_cards', 'red_cards', 'position',
'xA', 'shots', 'key_passes',
'team_title', 'npg', 'npxG',
'xGChain', 'xGBuildup'])
df.to_csv(r'league_player_data.csv', index=False)

player_id = df["id"]
print(f'There are {player_id.count()} players in the league for the chosen season')

player_ct = 0

# Range based on number of players
for i in range(0,554):
    df2 = understat.player(player=player_id[player_ct]).get_shot_data()
    df4 = pd.DataFrame(df2, columns=['id', 'minute', 'result', 'Goal', 'X', 'Y',
'xG',
'player', 'h_a', 'player_id', 'situation', 'season', 'shotType',
'match_id', 'h_team', 'a_team', 'h_goals', 'a_goals', 'date',
'player_assisted', 'last_action'])
    df4.to_csv('player_shot_data_all_comp.csv', mode='a', header=True)
    player_ct = player_ct + 1
    print(player_ct)

```

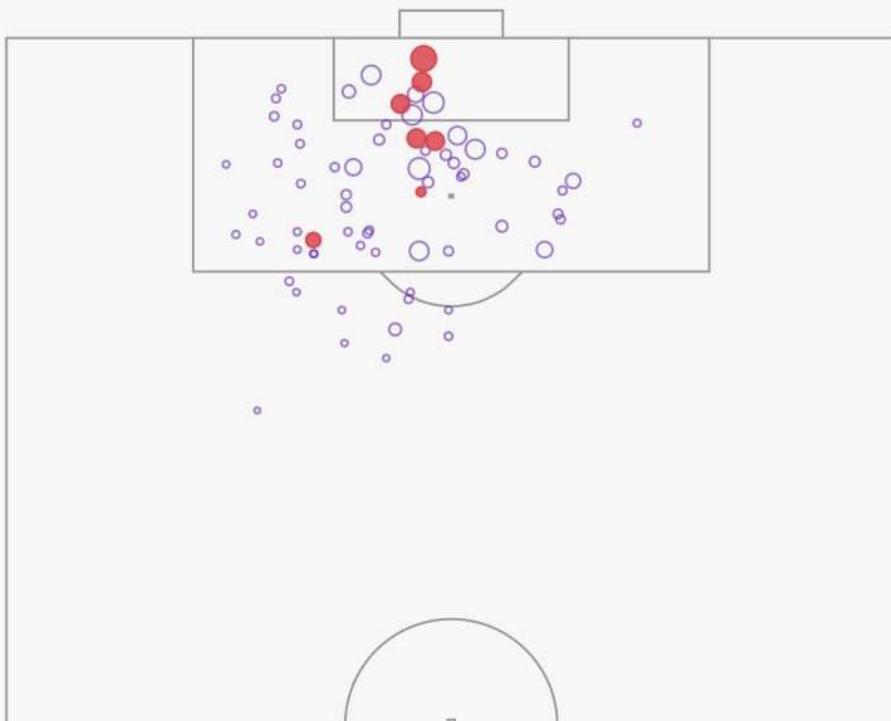
Okay so that's the code, ping me if you have any troubles.

For those that do use the dataset, make sure to cross reference it to the understat website, for example if you're doing match data you may need to flip / reflect some of the points if you want them split by team etc. But if you're just using a half pitch should be fine.

Let's dive into some inspiration that you may want to look at when using the dataset.

Where better to start than work from the Analyst?

The below image of Diogo Jota comes from the site [here](#).

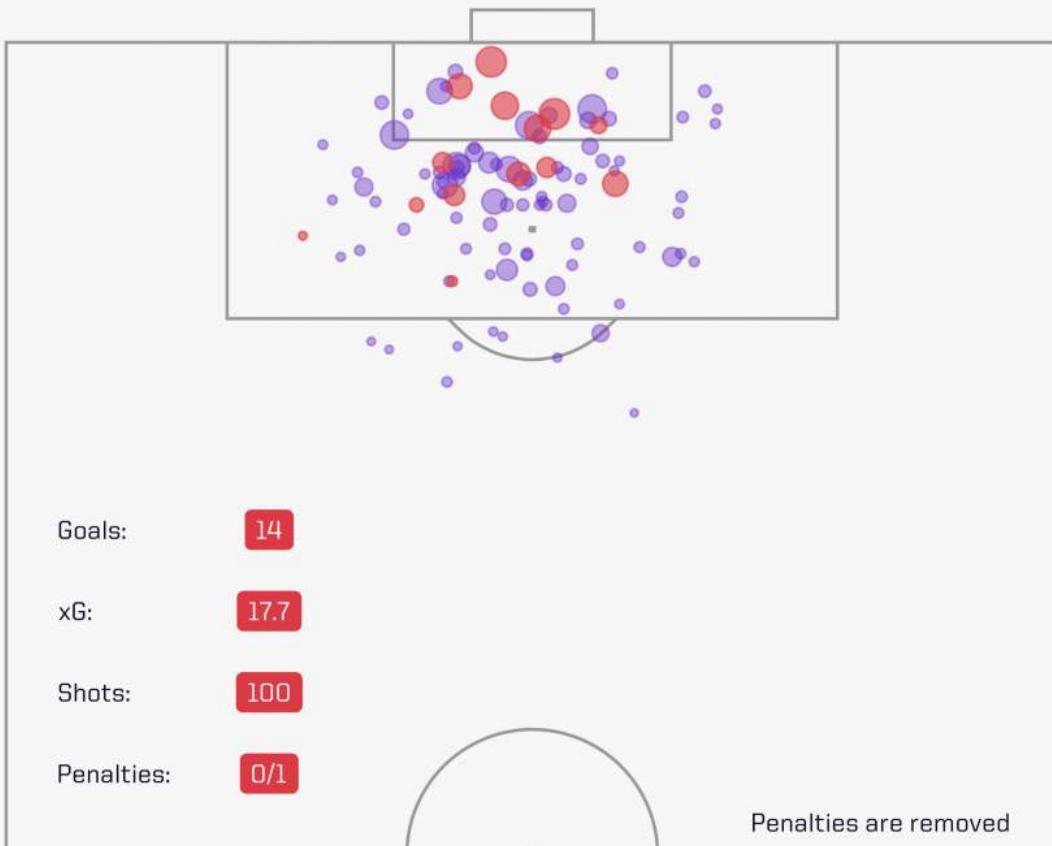


Some reasons I like this shot map.

- The pitch is lighter in colour, just a gentle grey colour, helps push the pitch framework to the background not being overpowering whilst allowing the shots to take main focus.
- Context: The visual is all about how Jota took shots worth 11.5 expected goals. With seven actual goals, he underperformed his xG total by 4.5 goals. What better way than to highlight that using a red for goals. (understandably also the analyst brand logo colours)
- Rather than have a separate legend and stats values they've integrated the two. Instantly I know shots are circles, and the shots that went in are the red. We don't have to now state this twice on the page.#

100 shots: Gabriel Jesus

Manchester City | Premier League 2019-20



Depending on positioning, you may want to frame your visual a little different. See how the Analyst here have used some of the white space to add in those key metrics where Jesus isn't shooting from range. An alternative consideration I see a lot would be to have the metrics horizontal instead. Next visual is by The Athletic.

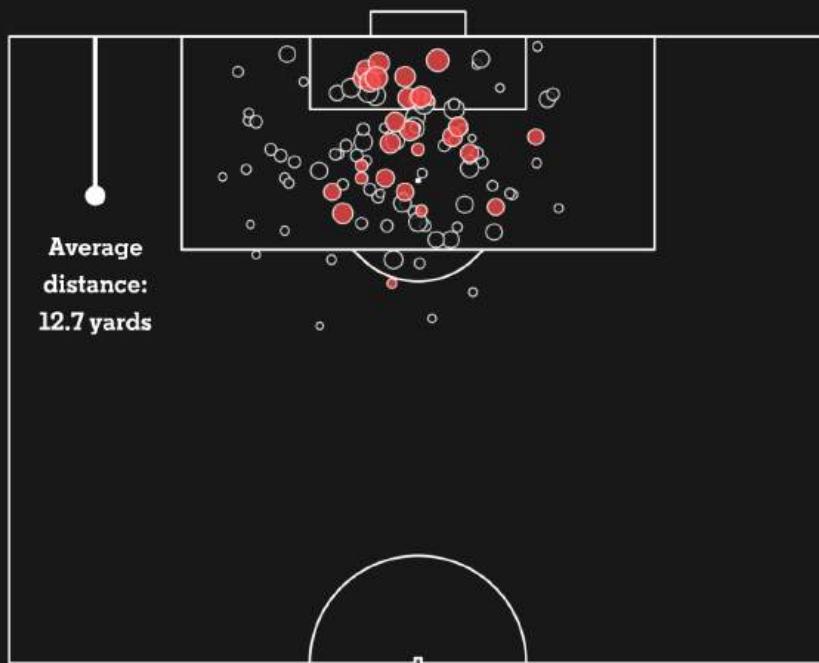
Erling Haaland

All non-penalty shots for Manchester City in Premier League in 2022-23



Low-quality chance ○ ○ ○ ○ ○ High-quality chance

● Goal ○ No Goal



Total shots

107

Right Foot

19

Left Foot

62

Head

25

Goals

35 (1.22 p90)

xG

26.8 (0.93)

xG per Shot

0.2

Other

1

The Athletic

For this one you can see the full article [here](#). For this they reverse the colours, with the dark background and white bordered pitch frame again suitable for the athletics branding colours. For me this piece really drives two main takeaways. Haaland is a machine, scoring way more than his expected average at the time of 0.19 with 35 goals, well above the 26 expected. But secondly, I love the average distance metric. I haven't really seen that on charts before and it does re-iterate danger zones or where Haalands presence is felt most.

Harry Kane

All non-penalty shots in the Premier League in 2021-22

Low Quality Chance ○ ○ ○ ○ ○ High Quality Chance

● Goal ○ No Goal



Shots
9 (2.14 p90)

Goals
0 (0)

xG
1 (0.24)

xG/Shot
0.11

The Athletic

Another one from The Athletic but this time they use the colour red to emphasize the reverse of previous. You can read the article [here](#). Sometimes the story isn't what the data is but it's the story about what it isn't. It really reminds me of this story about The Dog That Didn't Bark.

Gregory (Scotland Yard detective): "Is there any other point to which you would wish to draw my attention?"
Holmes: "To the curious incident of the dog in the night-time." Gregory: "The dog did nothing in the night-time."
Holmes: "That was the curious incident."

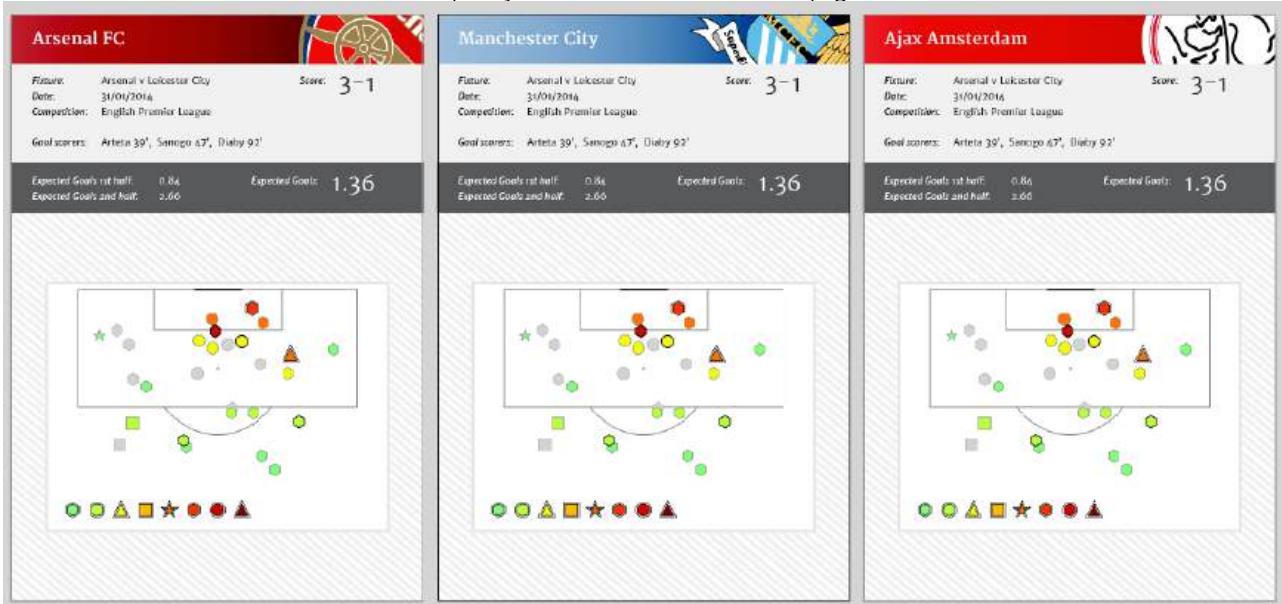
The fact that the dog did not bark when you would expect it to do so while a horse was stolen led Holmes to the conclusion that the evildoer was a not a stranger to the dog, but someone the dog recognized and thus would not cause him to bark.

Borussia Dortmund 2-2 1. FSV Mainz 05

German Bundesliga 2022-23 | 27 May 2023



This next piece is from Opta Analyst on [Twitter](#). A 2-2 tie between Dortmund and Mainz despite having 19 more shots and an expected goals of 4. Many of the compliments from the previous three shot maps also are seen in this one. The main reason I wanted to include this is for an example of framing. Opta do well to add in details of the game where there would be otherwise 'lost space'. See where the kick off circle is, they utilise this empty space where no one is shooting from the allow for some key stats. They also divide the pitch in two using the club logos, so you know which side refers to which team, without having to explicitly write it with more ink on the page.



For this one from Statsbomb, lets ignore the shot map for the time being but look at the overall design. I particularly like the framing to this visual. The banner headers are really effective in showing team. The score and xG are larger than the remaining details of fixture, date and competition in order to allow the user to see the main impact of the game before getting some of the

90s played: 12.44

Tinotenda Kadewere

non-penalty chances

Chance Map - Allsvenskan 2018

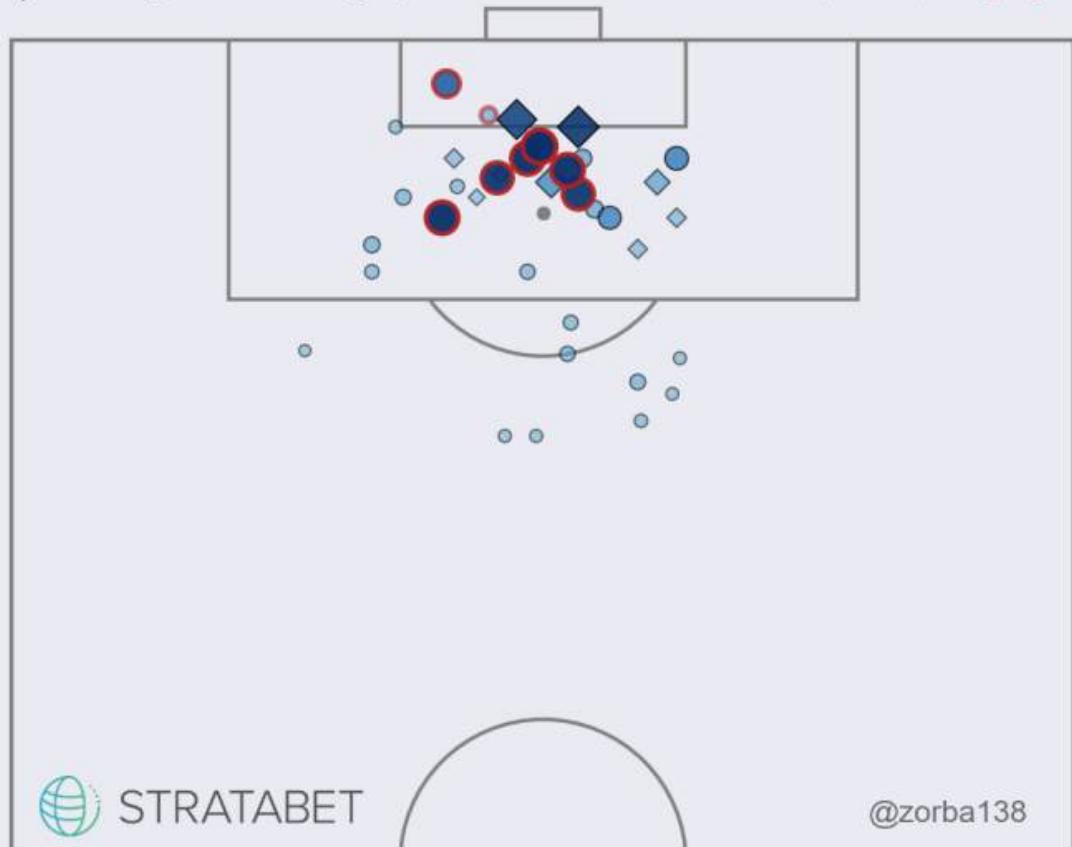
8 Goals | 5.51 xG | 36 Chances | 0.15 xG/Chance

◊ header

□ direct free kick

○ any other

size, color = xG, red edge = goal

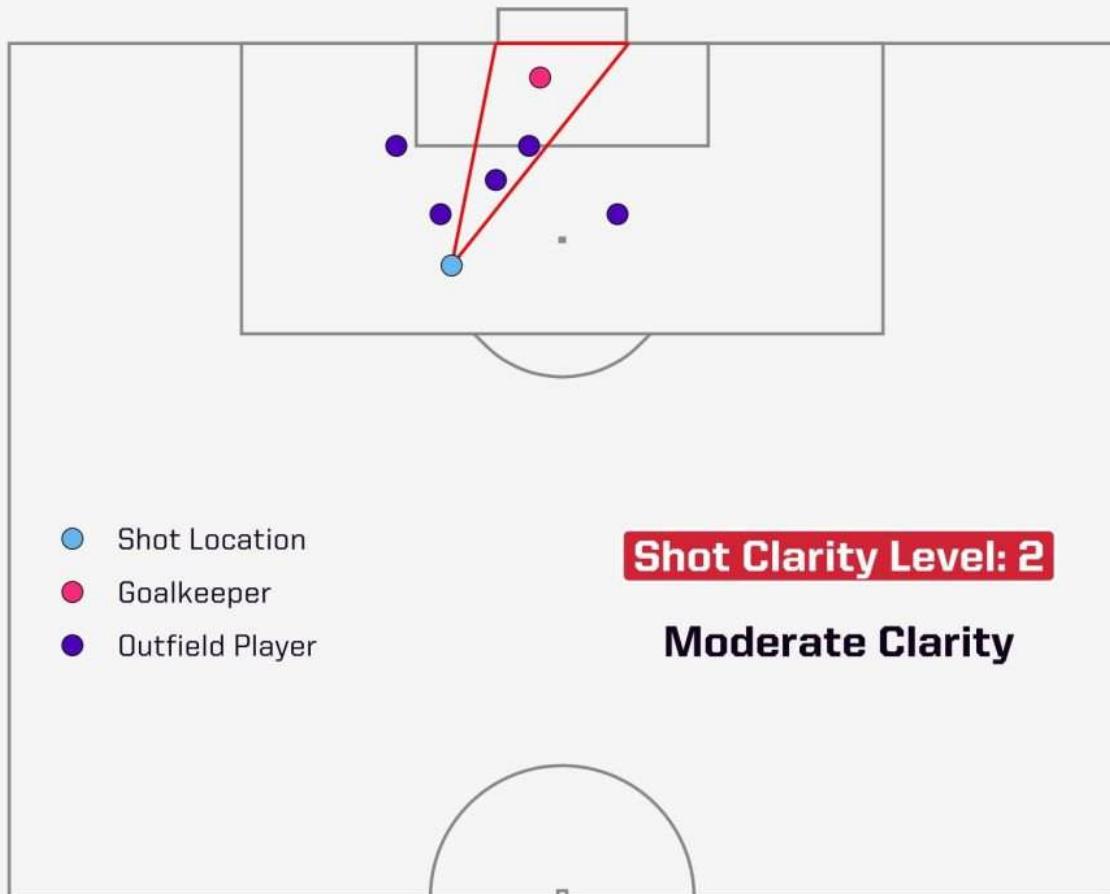


additional context.

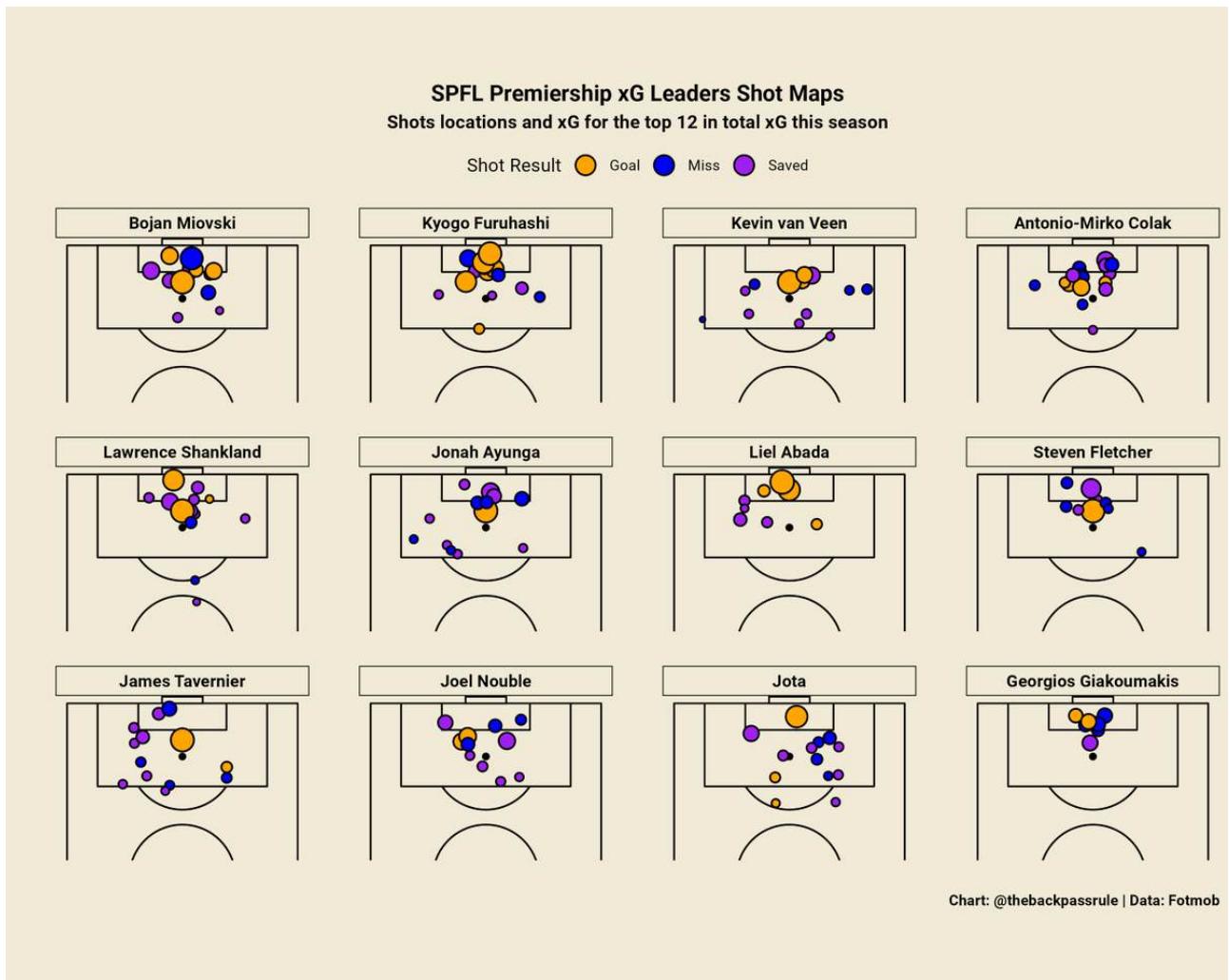
And to end the shot maps this final piece by Stratabet, Zorba. I think a challenge many of us face is how much information to put on the page or not. I include this one because I think it is the most easily replicable within Tableau of all the ones above (yay!). Stratabet do well to have various layered headings. Interestingly enough this visual will take a little longer to understand (which isn't always a bad thing) and that's because of the use of pre-attentive attributes. We have multiple things going on, with the use of size, shape and colour at the same time. It's a good reminder of thinking how these can work in unison for our story! You can read the full article [here](#). Going Further Now there is a lot of talent in the wider community but here are a few visuals that take it to the next step beyond just shot points so why not try: Shot Clarity charts (Analyst)

Shot Clarity

Shot Clarity Zone



Small Multiples (Modern Fitba)



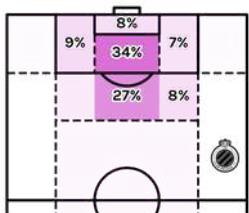
Shooting Area charts (Son of a Corner)

Shooting Areas in the Jupiler Pro League

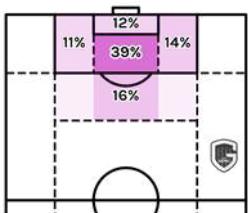
Percentage of non-penalty shots taken in each area. Viz by @sonofacorner.

Only areas with more than 5% of shots are labeled.

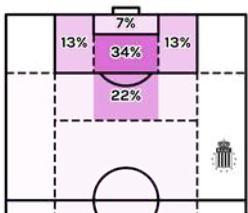
Club Brugge | shots 105



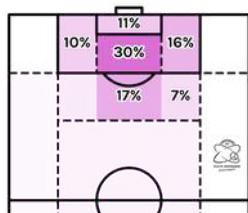
Genk | shots 93



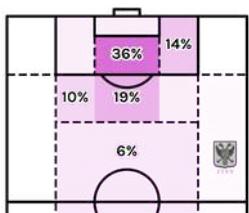
Sporting Charleroi | shots 83



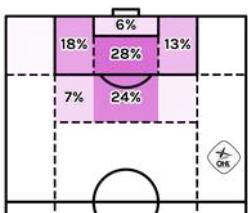
Zulte Waregem | shots 70



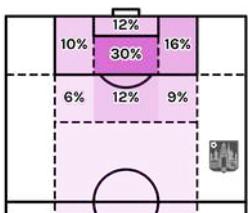
St.Truiden | shots 69



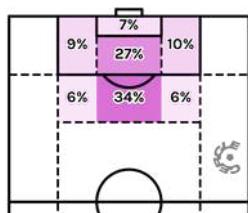
OH Leuven | shots 68



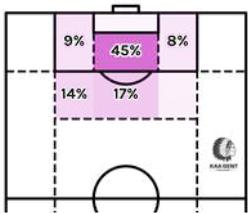
Westerlo | shots 67



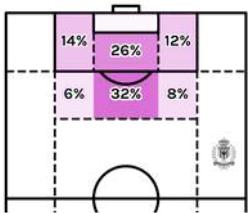
Cercle Brugge | shots 67



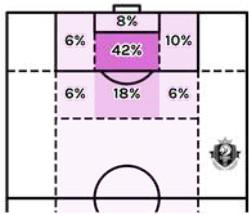
Gent | shots 66



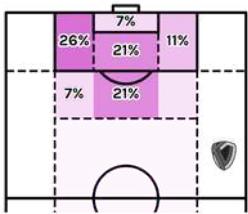
KV Mechelen | shots 65



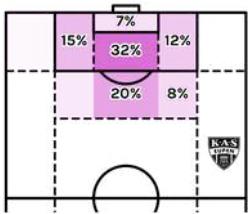
RFC Seraing | shots 62



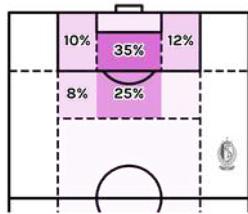
Oostende | shots 61



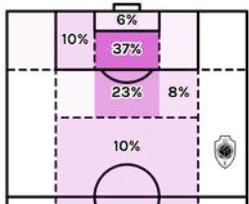
Eupen | shots 60



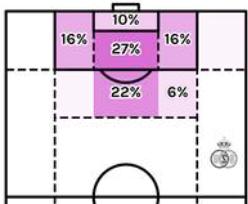
Standard Liege | shots 60



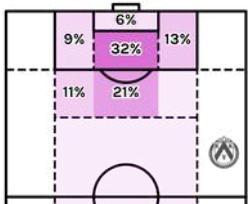
Royal Antwerp | shots 52



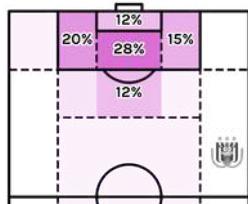
Union St.Gilloise | shots 51



Kortrijk | shots 47



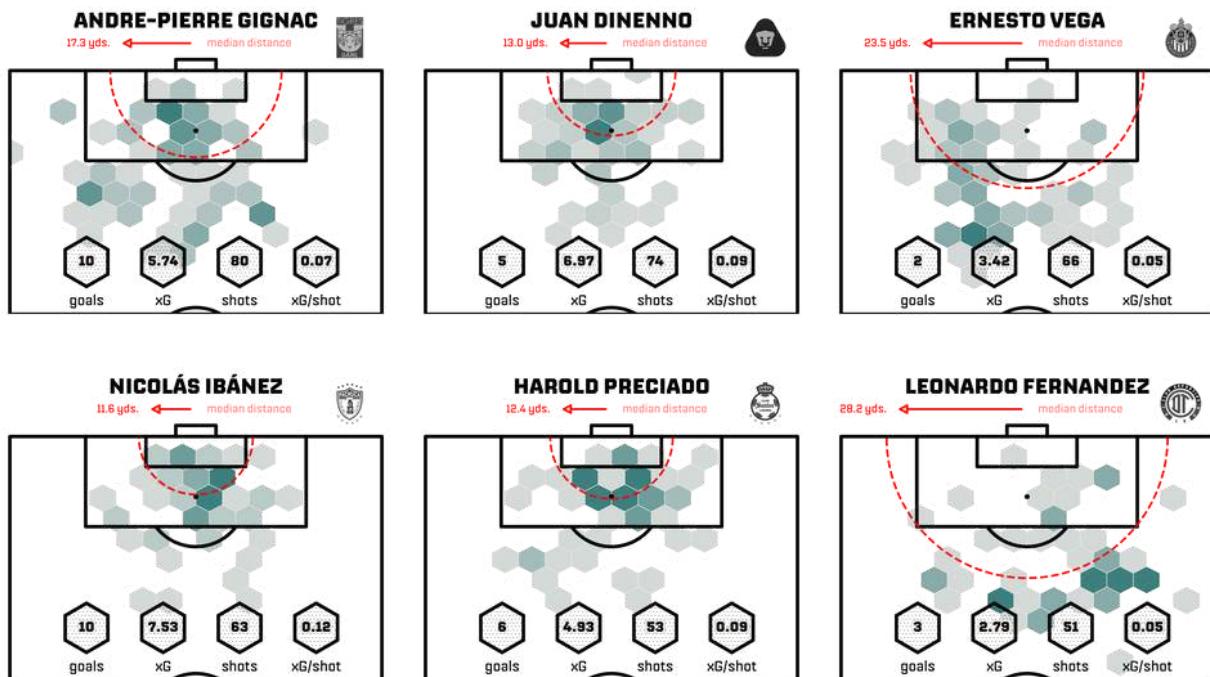
Anderlecht | shots 40



Hex shot maps Son of a Corner

LIGA MX'S TOP SHOOTERS

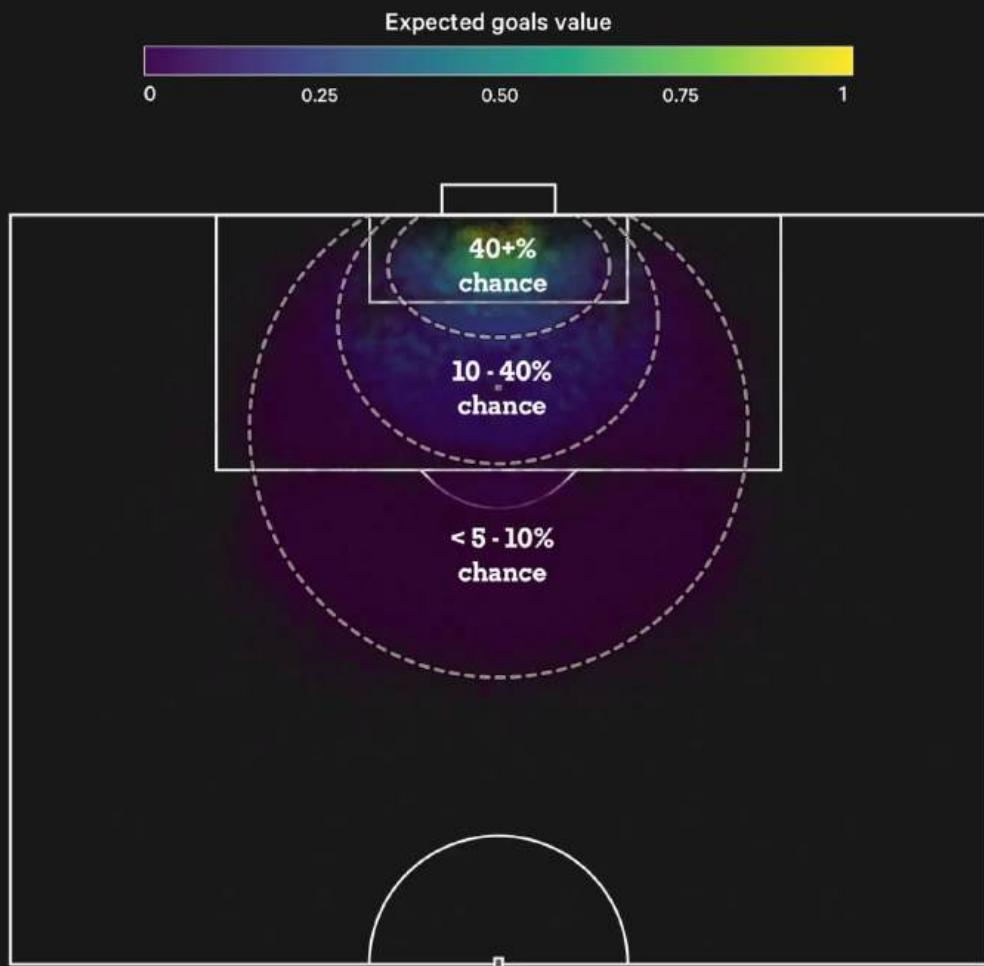
Non-penalty shot bins for Liga MX's top 6 shooters | Apertura 2022/2023 | viz by @sonofacorner, inspired by @jonollington



Goal Locations – The Athletic

How likely is a goal from different locations?

Expected goals values based on Premier League data since 2016-17 season



The Athletic

That's it for this week. Hopefully a few nice design ideas showcased within this weeks blog. Back to tutorials in weeks to come.

LOGGING OFF, CJ

CHATGPT ANALYSING PDF FORMS

Hi all, In previous blogs we looked at analysing CSV files using Langchain and ChatGPT, we've also looked at extensions that enable you to be able to use it with different tools. Today we are going to be looking at whether we can load documents (PDF's) and analyse the content within them! Now this question of analysing PDF's was asked by my best friend Arj, (*and previous flat mate for those of you that watch Iron Viz 2022, yes we have riveting conversations as he wanted to analyse Tax documents*) So to spruce it up a bit, I've decided we look at something equally as boring.... The rule handbook of tennis! You can view the ITF rulebook [here](#), as well as in the GitHub repo at the top of the page. The rulebook is 46 pages in length. We will look to refer to certain sections of it. You will need

- An API Key from OpenAI
- An IDE to write python code in (I've gone for PyCharm)

As a brief refresher, I've written previously how you can get this key through signing up to their program. (You get free \$5 worth of tokens!) You can sign up on OpenAI.com Once you've got your key, let's head on over to write our python code. The code is

stored in the **main.py** file. I also put a copy of the rulebook locally to access.

```
import os
from langchain.embeddings import OpenAIEMBEDDINGS
from langchain.vectorstores import Chroma
from langchain import OpenAI, VectorDBQA
from langchain.document_loaders import UnstructuredFileLoader
from langchain.text_splitter import CharacterTextSplitter

...
import nltk
nltk.download("punkt")
...

import pandas as pd
loader = UnstructuredFileLoader("2023-rules-of-tennis-english.pdf")
documents = loader.load()

text_splitter = CharacterTextSplitter(chunk_size=3000,chunk_overlap=0)
texts = text_splitter.split_documents(documents)

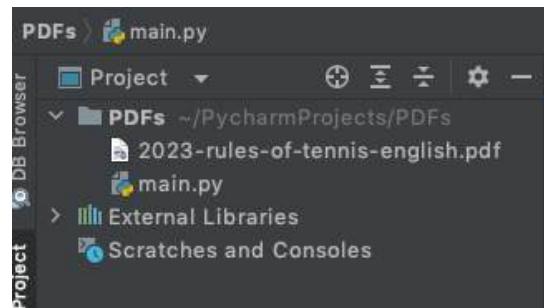
KEY = os.environ["OPENAI_API_KEY"] = " Insert Your Key Here "
embeddings = OpenAIEMBEDDINGS(openai_api_key=KEY)
doc_search = Chroma.from_documents(texts,embeddings)
chain = VectorDBQA.from_chain_type(llm=OpenAI(), chain_type="stuff",
vectorstore=doc_search)

query = "During a tie-break how many points does a player need to win the game
and set?"
# A generic numerical question
print(query)
print(chain.run(query))

query = "If a player is serving and it is their first serve, the ball hits the
net but bounces in the correct service court, what happens?"
# Open-ended question testing the rulebooks documentation
print(query)
print(chain.run(query))

query = "What is the difference in centimetres between the baseline and the
net?"
# Asking a question that is only more vaguely cited in the rulebook
print(query)
print(chain.run(query))
```

1. Importing necessary modules: The code begins by importing several modules, including `os` and `langchain`. You will realise that when your code errors, it will tell you which packages will need to be installed beyond this. These modules provide functionalities for working with files, embeddings, vector stores, document loaders, text splitting, and data manipulation.
2. Downloading NLTK data (commented out): There is a commented-out section, this doesn't need to be done more than once. NLTK stands for Natural Language Toolkit.
3. We then Initialize the `UnstructuredFileLoader`: The code creates an instance of the `UnstructuredFileLoader` class from the `Langchain` library, passing our pdf file "2023-rules-of-tennis-english.pdf" This is our file we want to load, of course you can use



one of your own. The loader is used to... load the content.

4. That loaded document is then stored in the documents variable.
5. After that we split the text into chunks – The code creates an instance of the CharacterTextSplitter class from Langchain, specifying a chunk size of 3000 characters (You can pick your own) and no overlap between chunks (again, you can amend this value).
6. The split documents method is then called on the text_splitter, passing through the documents. This breaks it up into smaller storables chunks.
7. The code sets the value of the KEY variable as the OpenAI API key, this can be re-written if you want to host your key elsewhere.
8. You will see we run three queries in my example. Each query is a question stored in the query variable each time. We print the result into the console. Perhaps you may want to make this so that they are saved somewhere as an output, if you do amend the code.

So that's the code, but what's the output and was it entirely accurate? We asked it based on the document: **During a tie-break how many points does a player need to win the game and set?** It responded: **The first player/team to win seven points wins the “Game” and “Set”, with a deciding point if the score reaches six all.** Which is correct! The answer to this was found on page 5.

b. Tie-break game

During a tie-break game, points are scored “Zero”, “1”, “2”, “3”, etc. The first player/team to win seven points wins the “Game” and “Set”, provided there is a margin of two points over the opponent(s). If necessary, the tie-break game shall continue until this margin is achieved.

The player whose turn it is to serve shall serve the first point of the tie-break game. The following two points shall be served by the opponent(s) (in doubles, the player of the opposing team due to serve next). After this, each player/team shall serve alternately for two consecutive points until the end of the tie-break game (in doubles, the rotation of service within each team shall continue in the same order as during that set).

The player/team whose turn it was to serve first in the tie-break game shall be the receiver in the first game of the following set.

Additional approved alternative scoring methods can be found in Appendix V.

We then asked it: **If a player is serving and it is their first serve, the ball hits the net but bounces in the correct service court, what happens?** It responded: **The service is a let and the server shall serve again. The whole point must be replayed.** Again, correct! The context for this question can be found on page 9 of the rulebook.

22. THE LET DURING A SERVICE

The service is a let if:

- a. The ball served touches the net, strap or band, and is otherwise good; or, after touching the net, strap or band, touches the receiver or the receiver's partner or anything they wear or carry before hitting the ground; or
- b. The ball is served when the receiver is not ready.

In the case of a service let, that particular service shall not count, and the server shall serve again, but a service let does not cancel a previous fault.

Additional approved alternative procedures can be found in Appendix V.

23. THE LET

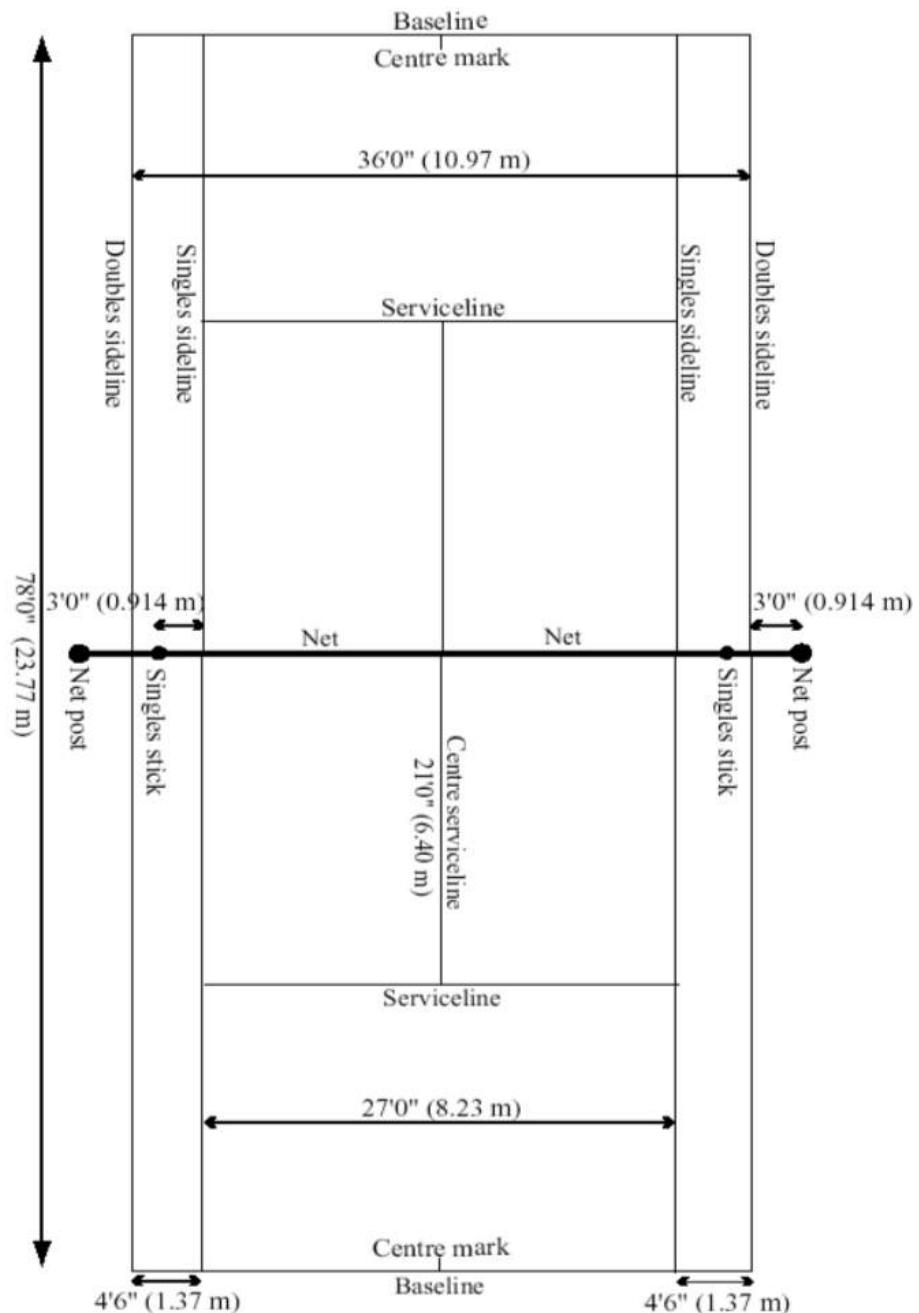
In all cases when a let is called, except when a service let is called on a second service, the whole point shall be replayed.

Case 1: When the ball is in play, another ball rolls onto court. A let is called. The server had previously served a fault. Is the server now entitled to a first service or second service?

Decision: First service. The whole point must be replayed.

APPENDIX VIII

PLAN OF THE COURT



Note: All court measurements shall be made to the outside of the lines.

It responded: The baseline is 21 feet (6.40 m) from the net. So not quite the answer, but it has managed to look at the diagram and state the difference between the **service line and the net**. I guess maybe I shouldn't ask questions that aren't directly within the PDF. This does make sense in terms of how the code operates but something to be aware of in terms of the type of questions asked and how you phrase them. So that's it for this week. A nice easy one where you can go add your own PDF's in. A few hiccups and pointers along the way to get your code working:

- Make sure you have all your necessary packages
- Amend your chunk sizes and overlap as appropriate

- Make sure you have a working API key
Going further, why not:
build a front end where you can upload a document and ask the question in a chat bot?
Use your own PDF that you want to analyse
- Play with different ways of articulating a question to see the differences in responses
Chat soon! **LOGGING OFF, CJ**

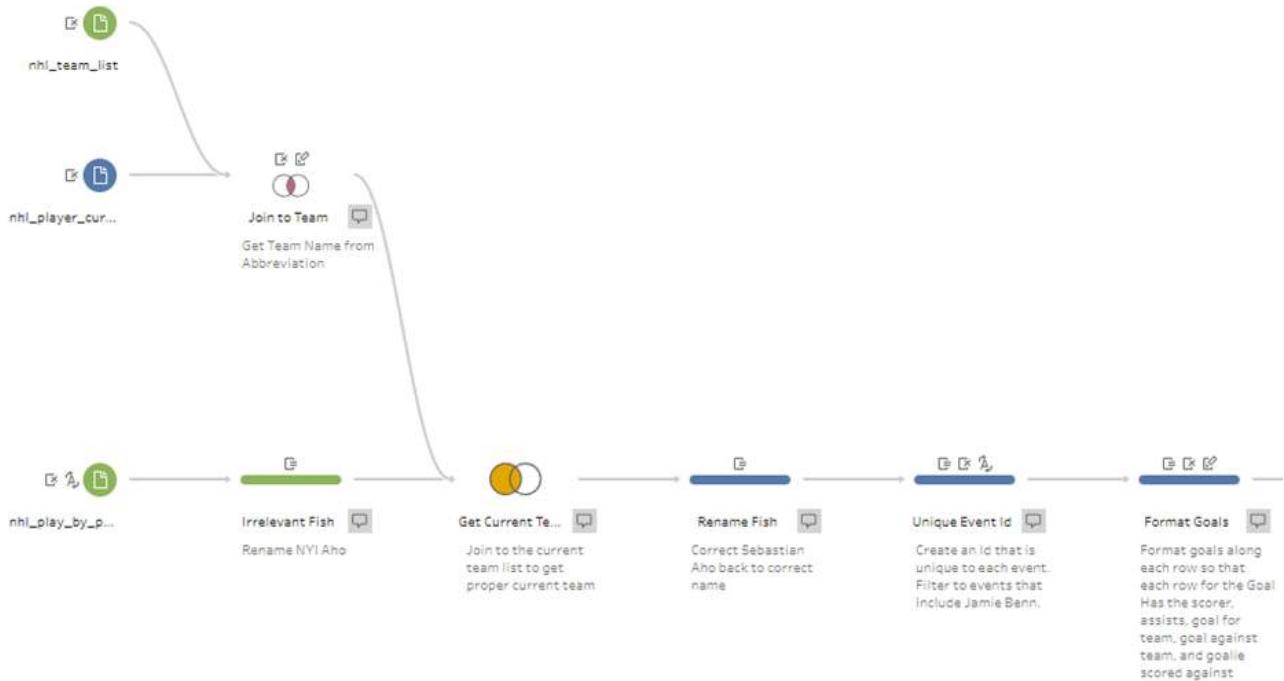
NHL PLAYER STATS WITH JARED FLORES PT2 (PREP & DESKTOP)

Hi all, Welcome back to part II of creating an E2E process for NHL Player Stats. If you missed it, earlier this week Jared joined the site to talk about how he leveraged code prompts in ChatGPT to help him access the NHL API. It's a delight to be able to follow that blog up, with Jared showcasing the final prep & build of his NHL visualisation. Jared, over to you.

Formatting the Data with Tableau Prep

Because of the formatting required, I took the output from the Python script and ran it through Prep to get my desired results. (You can see the files on Github)

Due to some performance issues and a ton of LOD calculations, I had to split up the flow into three parts. The first flow takes the play-by-play data and combines it with some tables that indicate each players current team. Players get traded throughout the year so I wanted to ensure I could show their stats based on their current team. The first flow takes the play-by-play data and combines it with some tables that indicate each players current team. Players get traded throughout the year so I wanted to ensure I could show their stats based on their current team.



After combining the data and handling for a duplicate name (There are two players in the league named Sebastian Aho that I have to account for to properly join some external pieces of data), I create a Unique Event Id.

Edit Field

Field Name

Unique Event Id

```

STR([about.dateTime])+
[result.eventCode]+str([about.eventIdx])
  
```

This allows me to ensure my LOD calculations are fixed to the proper event. Next, I start manually pivoting the data. You can see I made 25

Changes (25)

- Calculated Field
Goal Scored By Player
IF [player_type] = 'Scorer' THEN
[player_name] END
- Calculated Field
Goal Scored By Player
{ FIXED [Unique Event Id]:
MIN([Goal Scored By Player])}
- Calculated Field
Goal Scored By Team
IF [player_type] = 'Scorer' THEN
[team.name] END
- Calculated Field
Goal Scored By Team
{ FIXED [Unique Event Id]:
MIN([Goal Scored By Team])}
- Calculated Field
Assisted By Player
IF [player_type] = 'Assist' THEN
[player_name] END
- Calculated Field
Assisted By Position
IF [player_type] = 'Assist' THEN
[player_primaryPosition.name]
- Calculated Field
...
IF [player_type] = 'Goalie' and
[result.event] = 'Goal' THEN
- Calculated Field
...
{ FIXED [Unique Event Id]:
MIN([Goal Scored Against
- Calculated Field
Goal Scored Against Team
IF [player_type] = 'Goalie' and
[result.event] = 'Goal' THEN IF
- Calculated Field
Goal Scored Against Team
IF ISNULL([Goal Scored Against
Team]) and [result.event] = 'Goal'

changes just to get Goals and Assists pivoted

For the logic, I find the
data point based on the player_type

Edit Field

Field Name

Goal Scored By Player

```
IF [player_type] = 'Scorer' THEN [player_name] END
```

Now that I have the player who scored the goal, I can fix this value to every row that contains the same event id.

Edit Field

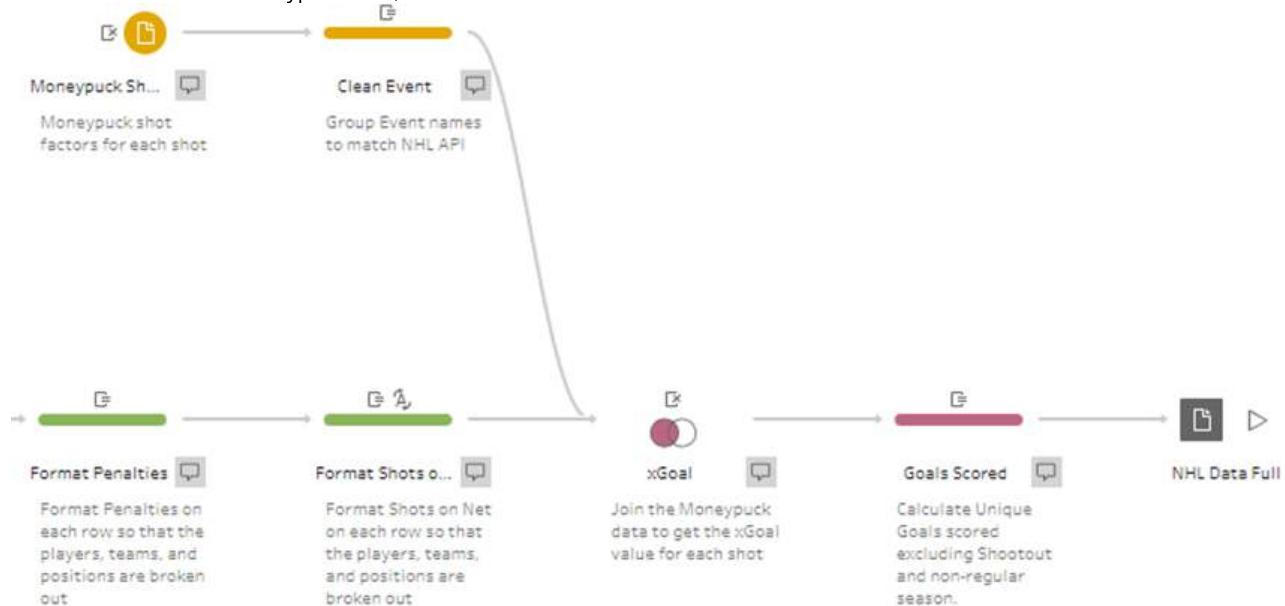
Field Name

Goal Scored By Player

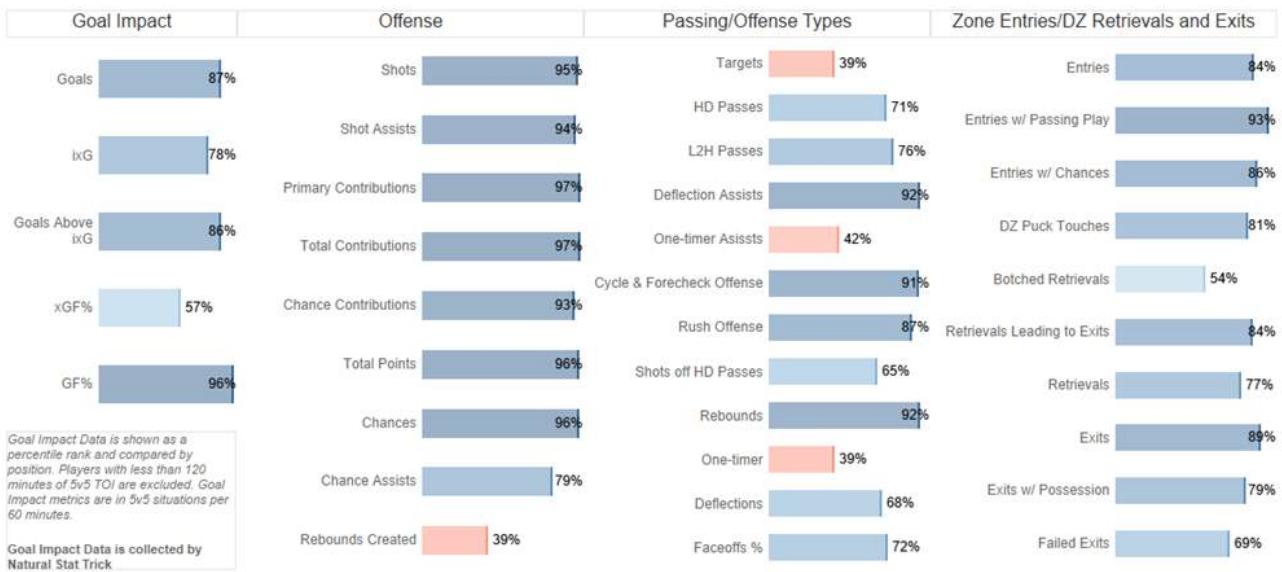
```
{ FIXED [Unique Event Id]: MIN([Goal Scored By Player]) }
```

I repeat this

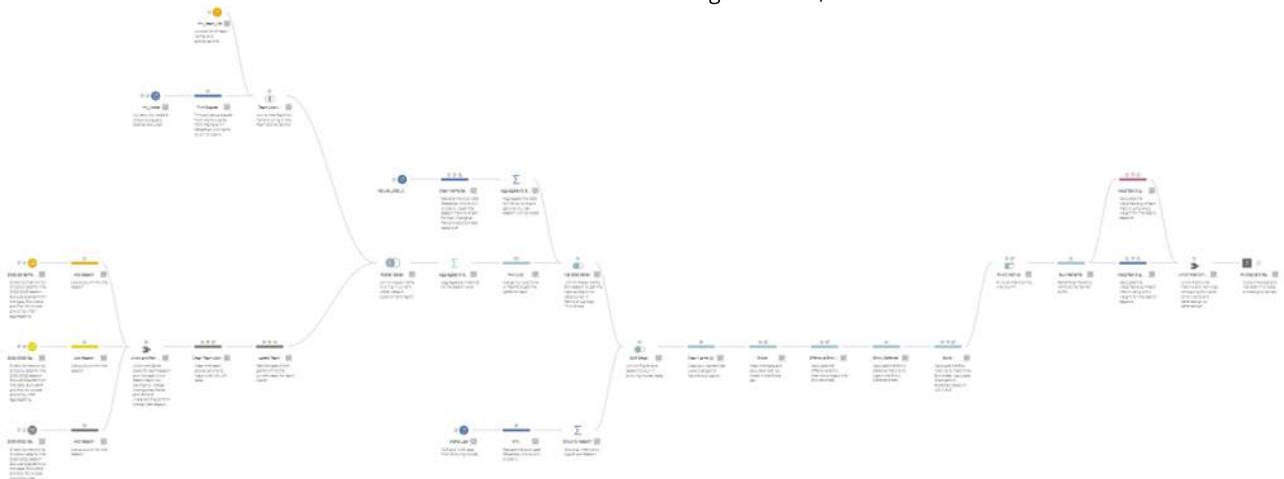
for assists, faceoff winner and loser, penalty drawn by and taken by etc. Eventually, I am able to bring in some external data from moneypuck.com, a site that tracks several metrics around the effectiveness of shots taken.



With this data, I can ensure the tooltip shows the expected goal value of every goal! While this Prep flow provides the main data set used for the majority of the dashboard, I also have some rankings listed near the top of the dashboard:



These metrics are coming from Natural Stat Trick and All Three Zones. Both projects track various player metrics to get a better understanding of how effective each player actually is outside raw goals and assists. There's a ton happening in the Prep flow that combines the various data sources so I won't go in detail, but look at this monster!



With all of this data, we're able to finally visualize each player's season and also how effective they are on ice. While I love data preparation, building Prep flows, and solving the data puzzle, building this visualization was one of my favorites. **TABLEAU** There were a few inspirations for the visualizations in the dashboard. The first was the NHL's website



JAMIE BENN | #14

LW | 6' 2" | 205 lb | Age: 33 | Dallas Stars

SUMMARY STATS NEWS AWARDS SHOP

Jamie Benn

Born: July 18, 1989

Birthplace: Victoria, BC, CAN

Shoots: Left

Draft: 2007 DAL, 5th rd, 8th pk (129th overall)

[View Player Bio +](#)

Season	GP	G	A	P	+/-	PIM	PPG	PPP	SHG	SHP	GWG	OTG	S	S%
2022-2023 Playoffs	12	2	8	10	2	34	2	4	0	0	0	0	29	6.9
Career Playoffs	78	22	39	61	10	96	7	16	2	2	3	0	203	10.8
2022-2023	82	33	45	78	23	34	13	30	2	3	4	0	190	17.4
NHL Career	1,030	362	485	847	70	768	101	233	15	27	55	11	2,741	13.2

[Follow @jamiebenn14](#)

→ Watch Jamie Benn Highlights

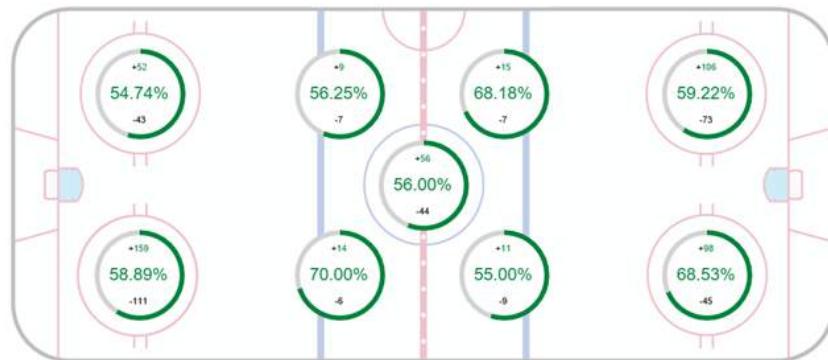
LAST 5 GAMES

Opponent	Date	G	A	P	+/-	PIM	PPG	SHG	S	Shifts	TOI
SEA	MAY 13	0	0	0	-2	2	0	0	2	23	16:57
SEA	MAY 11	0	1	1	1	0	0	0	1	24	16:51
SEA	MAY 9	1	1	2	1	4	1	0	2	22	17:16
SEA	MAY 7	0	1	1	-1	12	0	0	2	19	16:00

I wanted to take this view and enhance the information provided. I was also inspired by Meghan Chayka from Stathletes (@MeghanChayka) to incorporate the Faceoff chart:

Faceoff Wins

Choose Faceoff View
Ice Location



[SIGN UP TODAY](#)

Since I was using a rather large table and incorporating Image Roles into the viz, I wanted to ensure I wasn't including too many rows at once so I included the classic table pagination from **Luke Stanke**:

Game Logs

		G	A	P	+/-	PIM	PPG	PPP	SHG	SHP	GWG	OTG	S	TOI	Shifts	S%	FO%
Nov 3	@ Arizona Coyotes	1	2	3	2	0	0	1	0	0	0	0	5	13:44	22	20.00%	37.50%
Nov 5	@ Edmonton Oilers	3	0	3	1	0	2	2	0	0	0	0	4	12:55	22	75.00%	72.73%
Nov 8	@ Winnipeg Jets	0	0	0	0	0	0	0	0	0	0	0	0	15:15	21	0.00%	40.00%
Nov 11	vs San Jose Sharks	2	0	2	2	5	0	0	0	0	0	0	3	15:57	19	66.67%	75.00%
Nov 13	@ Philadelphia Flyers	0	1	1	0	0	0	1	0	0	0	0	1	13:39	20	0.00%	75.00%
Nov 15	@ Tampa Bay Lightning	0	1	1	-2	0	0	0	1	0	0	0	0	16:21	25	0.00%	11.11%
Nov 17	@ Florida Panthers	1	1	2	1	0	0	0	0	0	0	0	3	16:04	23	33.33%	50.00%
Nov 19	vs New York Islanders	1	2	3	3	0	0	0	0	0	1	0	3	20:02	26	33.33%	76.92%
Nov 21	vs Colorado Avalanche	0	0	0	1	5	0	0	0	0	0	0	0	15:48	23	0.00%	50.00%
Nov 23	vs Chicago Blackhawks	1	1	2	0	0	1	1	0	0	0	0	3	12:21	22	33.33%	62.50%

While there is a lot of information packed into the dashboard, each section essentially functions the same, so let's dig into the summary and goals sections.

NHL 2022-2023 Season Stats



Jamie Benn | #14

LW | 6' 2" | 205 lb | Dallas Stars

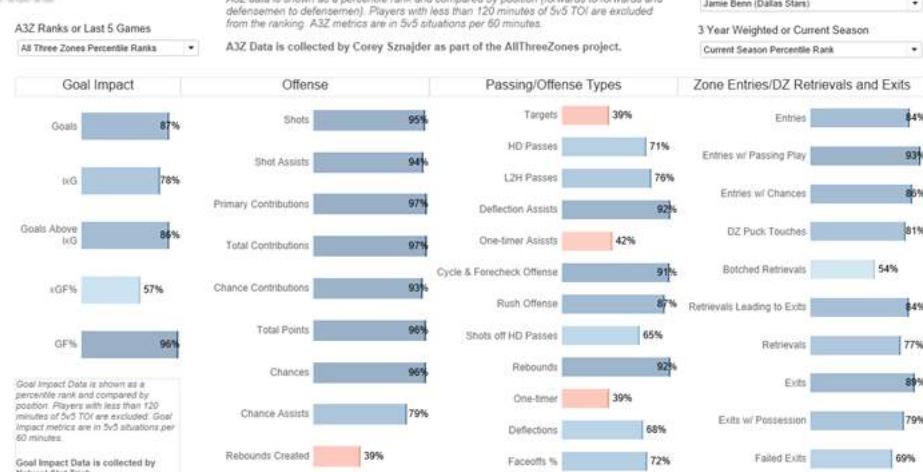
Born: July 18, 1989

Birthplace: Victoria, BC, CAN

Shoots: Left

Cap: \$9.50M x 8

Team Cap %: 13%



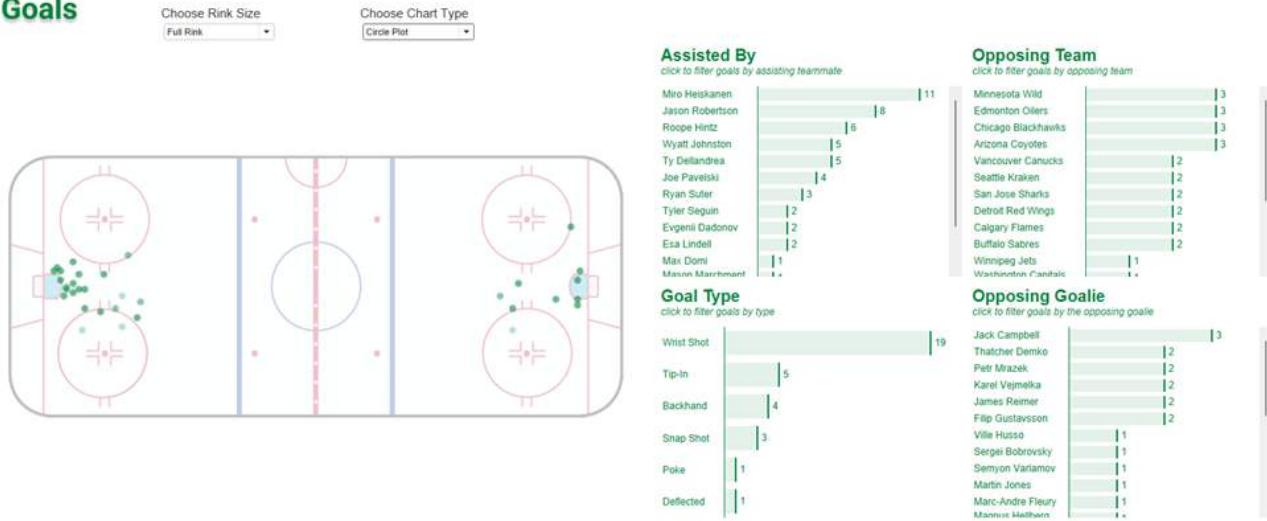
When you select a player from the top right corner, their image from the NHL website pops up along with some basic information about them (Name, Position, Height/Weight, Current Team, Birth Date and Place, and Cap Hit). You are able to see their base stats for the 2022-2023 season, similar to what we saw from the NHL website, but the percentile rankings also give a better picture into how the player actually performs compared to the league.

Goals



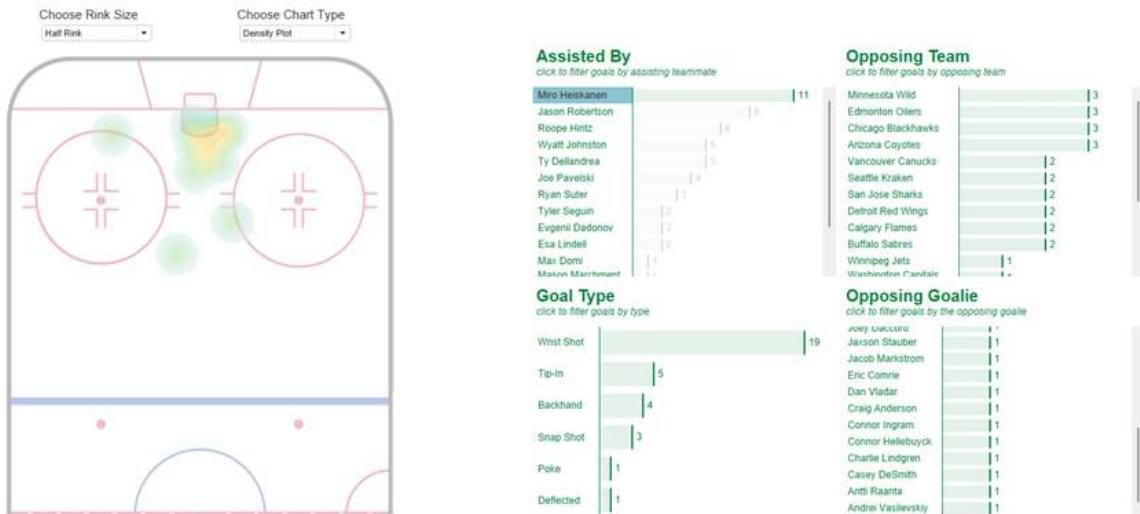
You're given a few options on how to view the on-ice events. You can choose between half or full rink and density plot or circle plot. Here's an example of the full rink and circle plot being selected:

Goals

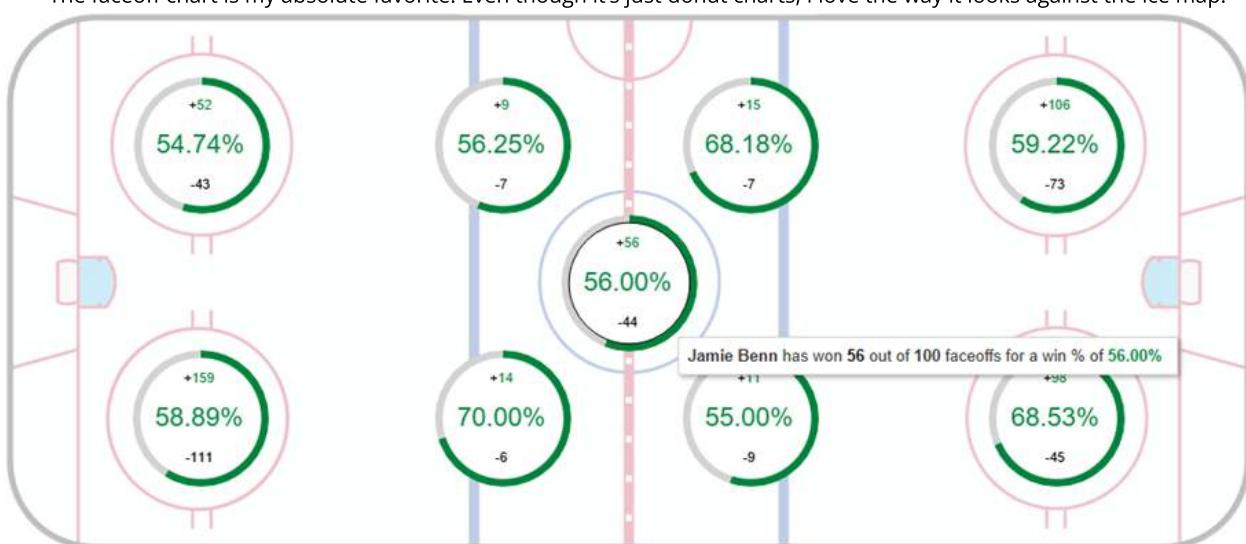


Since we manually pivoted all of the values in the Prep flow, we're able to see for the selected player who is assisting them the most, what type of goals they're scoring, and who they're scoring against. We can even filter the events by any of the charts:

Goals



The remaining sections of the dashboard all operate in the same way so you can dig into a player's assists, penalties, and hits. The faceoff chart is my absolute favorite. Even though it's just donut charts, I love the way it looks against the ice map.



At the top of the dashboard, you can also take a look at their last five games:

NHL 2022-2023 Season Stats



Jamie Benn | #14

LW | 6' 2" | 205 lb | Dallas Stars

Born: July 18, 1989

Birthplace: Victoria, BC, CAN

Shoots: Left

Cap: \$9.50M x 8

Team Cap %: 13%

A3Z Ranks or Last 5 Games
Last 5 Games

Select a Player
Jamie Benn (Dallas Stars)

Last 5 Games

		G	A	P	+/-	PIM	PPG	PPP	SHG	SHP	GWG	OTG	S	TOI
	St. Louis Blues	APR 13	0	1	1	1	0	0	0	0	0	0	3	12:27
	St. Louis Blues	APR 12	0	2	2	1	0	0	1	0	0	0	1	15:40
	Detroit Red Wings	APR 10	1	0	1	1	2	0	0	0	0	0	2	14:31
	Vegas Golden Knights	APR 8	0	0	0	0	0	0	0	0	0	0	2	15:13
	Philadelphia Flyers	APR 6	0	1	1	1	0	0	1	0	0	0	2	15:27

And there you have it! ChatGPT, Python, Tableau Prep, and Tableau Desktop all working together to create something with so much information packed in. I can't wait to continue improving the dashboard! CJ Round-up: I really want to thank Jared for taking the time to piece this together, it's been on the wish list to have Jared discuss his NHL data prep for a while. Blogs like these just go to show the breadth of skills and tools now being used in data, and what we see online as finished visuals, we sometimes forget the leg work put in to shape the data in the way we need. Can't wait to see what's next for Jared, and love that he is carrying the torch not just for Tableau Prep but also making real headway in NHL analytics. **LOGGING OFF, CJ**

NHL PLAYER STATS WITH JARED FLORES PT1 (PYTHON)

Hi all, It's a delight to be able to welcome Jared Flores onto the blog today. Jared is in fact the first person to make multiple guest appearances on the site! Pretty good going huh. Of course, when I'm not convincing Jared to write content to host on this channel you can find him on his own [Youtube](#), [Tableau Public](#), [website](#) and [Twitter](#). Jared will look to cover off using Tableau Prep, ChatGPT, and Python to create an interactive and in-depth Tableau Dashboard around your favorite NHL stars. The blog will be split into two sections. The first using ChatGPT to help with coding. The second will look at data cleansing and Tableau. All code and files as usual can be found through the Github link! Over to Jared for todays content.

ChatGPT has been all the buzz lately. From writing quirky jokes to negotiating lower prices with customer service chats, there have been so many use cases uncovered since it's public release. A few months ago, I wanted to see if I could use it to assist me in writing some Python to pull data from the NHL's public API so I could create a dashboard with the data. The process has evolved over time but the results were amazing. I'm going to show you from beginning to end what it took to create the NHL Player Stats Dashboard.

ChatGPT and Python

The First thing I did was find the link to the NHL API:

<https://statsapi.web.nhl.com/api/v1/configurations>

This link lists the main endpoints which contain various pieces of information. I started with a simple prompt:

JA

Write a python script that pulls a list of all teams from the following api

<https://statsapi.web.nhl.com/api/v1/configurations>



Certainly! Here's a Python script that uses the `requests` library to make a GET request to the specified API and retrieves a list of all teams:

```
python
import requests

# Make a GET request to the API
response = requests.get('https://statsapi.web.nhl.com/api/v1/configurations')

# Check if the request was successful
if response.status_code == 200:
    data = response.json()
    teams = data['teams']

    # Print the list of teams
    for team in teams:
        print(team['name'])
else:
    print('Failed to retrieve team data. Status code:', response.status_code)
```

[Regenerate response](#)

While this wrote a general python script to request the data, ChatGPT didn't have all the information it needed to actually pull any relevant information. I explored some of the links listed in the configurations endpoint and ended up using the following:

<https://statsapi.web.nhl.com/api/v1/teams> – This endpoint contains data about all teams

<https://statsapi.web.nhl.com/api/v1/schedule> – This endpoint contains the schedule for a specified date range

<https://statsapi.web.nhl.com/api/v1/people/{ID}> – This endpoint returns data for a specific player. You must enter their ID for valid results

<https://statsapi.web.nhl.com/api/v1/game/{ID}/feed/live> – This endpoint returns the play-by-play data for a specified game, provided you have the game id

Drew Hynes has created a publicly accessible documentation site around the most common uses for the API which can be found here: <https://gitlab.com/dword4/nhlapi>
Once I got my bearings with the API, I had a few main pieces of data I wanted to grab:

- Play-by-play data for every game in the current regular season
- A list of all current rosters for all teams
- Player specified metrics for all games (Time on Ice, Shots Taken, Goals, Faceoffs etc.)
- Final Score

I ended up creating a python script for each of these data sets with the assistance of ChatGPT. What I found was GPT helped me get started, but never really provided a

100% working code without a lot of re-prompting. It definitely helped speed up the entire process, but it still required some familiarity with Python and also an understanding of the JSON format of the data. Let's dig in to what the main bit of code does.

Accessing the NHL API

Firstly, we get the team IDs. The base URL for the NHL API is <https://statsapi.web.nhl.com/api/v1>. We then append /teams to this URL to get the endpoint for teams. Using the `requests.get()` function, we make a GET request to this endpoint, parse the JSON response into a Python dictionary, and extract the team IDs.

Fetching Specific Game Data

Next, we use these team IDs to fetch the games each team has played since a specific date. We construct the URL for the endpoint using the base URL, the date range, and the team ID.

```
44 # Iterate over the team ids in the team_id_list
45 for team_id in team_id_list:
46     # Set the endpoint for the specific team
47     team_endpoint = f"/schedule?startDate=2022-10-07&endDate={yesterday}&teamId={team_id}"
48
```

For each team, we extract the gamePk values, which are unique identifiers for each game.

```
1 import requests
2 import pandas as pd
3 import datetime
4 from datetime import timedelta
5
6 # Get yesterday's date
7 yesterday = datetime.date.today() - timedelta(1)
8
9 # Get today's date
10 today = datetime.date.today()
11
12 # Set the base URL for the NHL API
13 base_url = "https://statsapi.web.nhl.com/api/v1"
14
15 # Set the endpoint for the teams
16 teams_endpoint = "/teams"
17
18 # Send a GET request to the NHL API
19 response = requests.get(base_url + teams_endpoint)
20
21 # Parse the response as a dictionary
22 response_dict = response.json()
23
24 # Create an empty list to store the team ids
25 team_id_list = []
26
27 # Iterate over the teams in the response
28 for team in response_dict['teams']:
29     # Extract the team id
30     team_id = team['id']
31
32     # Append the team id to the team_id_list
33     team_id_list.append(team_id)
```

```

55     # Iterate over the games in the response
56     for game in response_dict['dates']:
57         # Extract the gamePk value
58         gamePk = game['games'][0]['gamePk']
59
60         # Append the gamePk value to the gamePk_list
61         gamePk_list.append(gamePk)

```

Extracting Detailed Game Data

Using the gamePk values, we then fetch the detailed data for each game. This includes information about the game type, season, teams, venue, and plays. We extract the data we're interested in and store it in a temporary Pandas DataFrame.

```

90     # Get the game type from the response dictionary
91     game_type = response_dict['gameData']['game']['type']
92
93     # Get the season from the response dictionary
94     season = response_dict['gameData']['game']['season']
95
96     # Get the away team from the response dictionary
97     away_team = response_dict['gameData']['teams']['away']['name']
98
99     # Get the home team from the response dictionary
100    home_team = response_dict['gameData']['teams']['home']['name']
101
102    # Get the venue played from the response dictionary
103    venue = response_dict['gameData']['teams']['home']['venue']['name']
104
105    # Create a temporary Pandas dataframe from the plays data
106    temp_plays_df = pd.json_normalize(response_dict['liveData']['plays']['allPlays'])
107
108    # Add the game type and gamePk columns to temp_plays_df
109    temp_plays_df['game_type'] = game_type
110    temp_plays_df['season'] = season
111    temp_plays_df['gamePk'] = gamePk
112    temp_plays_df['home_team'] = home_team
113    temp_plays_df['away_team'] = away_team
114    temp_plays_df['venue_played'] = venue

```

Extracting Player Data

An additional layer of complexity is the player data nested within each play. We define a function `extract_players` to handle this. The function checks if the players column is a list, and if so, creates a DataFrame to store the player data. It then iterates over the players in the list, extracts the player's data, and appends it to the DataFrame.

```

116     # Define a function to extract the players data to a new dataframe
117     def extract_players(row):
118         # Check if the players column is a list
119         if not isinstance(row['players'], list):
120             # Return an empty dataframe
121             return pd.DataFrame()
122
123         # Create an empty dataframe to store the player data
124         player_data = pd.DataFrame(columns=['about.eventIdx', 'player_id', 'player_name', 'player_type'])
125
126         # Iterate over the players in the list
127         for player in row['players']:
128             # Extract the player's data
129             eventIdx = row['about.eventIdx']
130             player_id = player['player']['id']
131             player_name = player['player']['fullName']
132             player_type = player['playerType']
133
134             # Append a row containing the player's data to the player_data dataframe
135             player_data = player_data.append({
136                 'about.eventIdx': eventIdx,
137                 'player_id': player_id,
138                 'player_name': player_name,
139                 'player_type': player_type
140             }, ignore_index=True)
141
142         # Return the player_data dataframe
143         return player_data

```

Merging and Cleaning Data

Next, we merge our play-by-play data with player data and clean it up, dropping rows without a result.eventCode. We then append this cleaned data to our main DataFrame, which stores data from all games.

```

190     # Merge the player data with players_df
191     temp_plays_df = temp_plays_df.merge(players_df, on=['player_id'], how='outer')
192
193     # Drop all rows in temp_plays_df that don't have a result.eventCode
194     temp_plays_df = temp_plays_df.dropna(subset=['result.eventCode'])
195
196     # Append the data from the current game to the plays_df dataframe
197     plays_df = plays_df.append(temp_plays_df)

```

Exporting Data to CSV

Finally, we export the data to a CSV file, ready for further analysis or visualization in a tool like Tableau.

```

199     # Export the data to a CSV file
200     plays_df.to_csv('nhl_play_by_play_data_all_games.csv')

```

I ended up with a dataset like seen in the GitHub Repo.

With the help of ChatGPT and Python, we have a complete script that fetches, processes, and exports NHL game data for any given time period.

There was a slight problem though. The way the NHL captures events, there are multiple rows associated with each unique event. For example, when a player scores a goal, there is a row for each of the following:

- The goal scorer
- The primary assister
- The secondary assister

The goalie scored against

Jason Robertson

Roope Hintz

Matt Johnston

Jamie Benn wristed a shot into the net against Jack Campbell on 11/5/2022 in a win against the Edmonton Oilers during the 2022-2023 season

Assisted by Miro Heiskanen and Roope Hintz

xGoal: 0.7937

Time of Goal: 04:01 In the 3rd Period



Each goal in the map needed to have a clear description as to who scored, assisted, and who was in net. I also have charts that break down goal types, opposing teams, opposing goalies etc. so to capture all of this, I needed each row to contain all of these

data points instead of having a row for each. You'd think it was a simple pivot that needed to happen, but some of these events are conditional. That's where Tableau

Prep Came in!

Join us in the next few days for Part 2, where I look to go through Tableau Prep & Desktop to creating my NHL visuals. CJ Round-Up: Jared raises some really interesting points above around API end points, but rightly describes how ChatGPT can be leveraged as developers to be able to 'fill the gaps' on writing code to help us. I am so excited for Part 2 in coming days where we get our hands on some prep & desktop to recreate elements within Jared's visual that you can find here. **LOGGING OFF, CJ**

STRAVA MAPS PT 2 – PREP YOUR GPX FILES (PYTHON)

Hi all, A few weeks back, I wrote a blog on how to manually extract your strava data, create the csv version of your gpx file and then be able to plot that map in Tableau. Well this week we will take one small part of that process and make it quicker easier through code. As a brief reminder. To download the data you will need to navigate to "["My Account"](#)

My Profile
My Account
My Performance
Display Preferences
Privacy Controls
Data Permissions
Email Notifications
My Gear
My Apps
Partner Integrations
My Badges
My API Application

Membership

Member Since: January 7, 2020

Membership Status: **Free User**

[Start Free Trial](#)

Subscribe to access premium features

- Personalized Route Suggestions
- Advanced Workout Analysis
- Segment Leaderboards

ALSO INCLUDED:

- [Recover Athletics](#)
Prehab and Injury Prevention Workouts
- [FATMAP NEW](#)
3D Maps and Guides for the Outdoors

[See All Premium Features](#)

Download or Delete Your Account

Here's where you can get an archive of your Strava data for your records, or begin the process of deleting your account. [Learn more](#)

[Get Started](#)

Click to Get started on download your stats. Once you go to download them it will send you a link to your email associated with the account

activities	✓	10/04/2023 11:24	File folder
clubs	✓	10/04/2023 11:24	File folder
media	✓	10/04/2023 11:24	File folder
routes	✓	10/04/2023 08:48	File folder
activities.csv	✓	10/04/2023 11:24	Microsoft Excel C... 278 KB
applications.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
bikes.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
blocks.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
categories_of_personal_information_we_... ...	✓	10/04/2023 11:24	Adobe Acrobat D... 52 KB
clubs.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
comments.csv	✓	10/04/2023 11:24	Microsoft Excel C... 26 KB
components.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
connected_apps.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
contacts.csv	✓	10/04/2023 11:24	Microsoft Excel C... 33 KB
email_preferences.csv	✓	10/04/2023 11:24	Microsoft Excel C... 2 KB
events.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
favorites.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
flags.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
followers.csv	✓	10/04/2023 11:24	Microsoft Excel C... 4 KB
following.csv	✓	10/04/2023 11:24	Microsoft Excel C... 4 KB
general_preferences.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
global_challenges.csv	✓	10/04/2023 11:24	Microsoft Excel C... 17 KB
goals.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
group_challenges.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
information_we_disclose_for_a_business_... ...	✓	10/04/2023 11:24	Adobe Acrobat D... 73 KB
local_legend_segments.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
logins.csv	✓	10/04/2023 11:24	Microsoft Excel C... 2 KB
media.csv	✓	10/04/2023 11:24	Microsoft Excel C... 12 KB
memberships.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
metering.csv	✓	10/04/2023 11:24	Microsoft Excel C... 128 KB
mobile_device_identifiers.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
orders.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
partner_opt_outs.csv	✓	10/04/2023 11:24	Microsoft Excel C... 2 KB
posts.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
privacy_zones.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
profile.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
profile.jpg	✓	10/04/2023 11:24	JPG File 24 KB
reactions.csv	✓	10/04/2023 11:24	Microsoft Excel C... 121 KB

Within the activities.csv you will see an associated gpx file. These files are essentially the mappings for our different activities.

2989833057.gpx	✓	10/04/2023 11:24	GPX File	107 KB
3003632713.gpx	✓	10/04/2023 11:24	GPX File	93 KB
3010269791.gpx	✓	10/04/2023 11:24	GPX File	127 KB
3050395336.gpx	✓	10/04/2023 11:24	GPX File	138 KB
3070693531.gpx	✓	10/04/2023 11:24	GPX File	120 KB
3090440097.gpx	✓	10/04/2023 11:24	GPX File	104 KB
3110426251.gpx	✓	10/04/2023 11:24	GPX File	118 KB
3130586991.gpx	✓	10/04/2023 11:24	GPX File	223 KB
3151109678.gpx	✓	10/04/2023 11:24	GPX File	246 KB
3209712232.gpx	✓	10/04/2023 11:24	GPX File	239 KB
3212677120.gpx	✓	10/04/2023 11:24	GPX File	206 KB
3219783138.gpx	✓	10/04/2023 11:24	GPX File	253 KB
3234246540.gpx	✓	10/04/2023 11:24	GPX File	100 KB
3241629424.gpx	✓	10/04/2023 11:24	GPX File	134 KB
3252466640.gpx	✓	10/04/2023 11:24	GPX File	361 KB
3256995695.gpx	✓	10/04/2023 11:24	GPX File	123 KB
3257140484.gpx	✓	10/04/2023 11:24	GPX File	86 KB
3266098044.gpx	✓	10/04/2023 11:24	GPX File	419 KB
3270484548.gpx	✓	10/04/2023 11:24	GPX File	111 KB
3274964275.gpx	✓	10/04/2023 11:24	GPX File	163 KB
3284054261.gpx	✓	10/04/2023 11:24	GPX File	731 KB
3296972879.gpx	✓	10/04/2023 11:24	GPX File	257 KB
3307769727.gpx	✓	10/04/2023 11:24	GPX File	483 KB
3312495731.gpx	✓	10/04/2023 11:24	GPX File	458 KB
3329901744.gpx	✓	10/04/2023 11:24	GPX File	102 KB
3379095210.gpx	✓	10/04/2023 11:24	GPX File	32 KB
3405802731.gpx	✓	10/04/2023 11:24	GPX File	67 KB
3417103983.gpx	✓	10/04/2023 11:24	GPX File	159 KB
3428012328.gpx	✓	10/04/2023 11:24	GPX File	104 KB
3448363481.gpx	✓	10/04/2023 11:24	GPX File	402 KB
3464051804.gpx	✓	10/04/2023 11:24	GPX File	115 KB
3475138379.gpx	✓	10/04/2023 11:24	GPX File	98 KB
3487702528.gpx	✓	10/04/2023 11:24	GPX File	119 KB
3540167750.gpx	✓	10/04/2023 11:24	GPX File	98 KB
3552943056.gpx	✓	10/04/2023 11:24	GPX File	137 KB
3566437977.gpx	✓	10/04/2023 11:24	GPX File	71 KB
3570812732.gpx	✓	10/04/2023 11:24	GPX File	118 KB
3576196503.gpx	✓	10/04/2023 11:24	GPX File	428 KB

The

only problem we had was... Tableau doesn't allow for gpx files, so we need to convert them into records of data.

Tableau

X



This file is not in a recognizable format. See the Help documentation for a list of supported file types.

OK

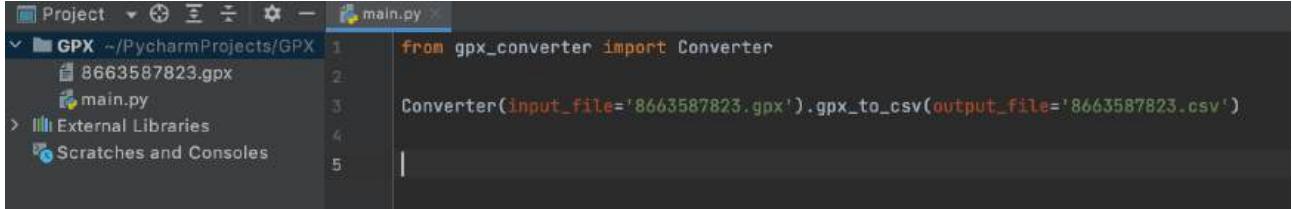
Now if you don't want to use python then by

all means visit my previous blog, or navigate to [GPS Visualiser website](#). But this blog is all about how to more easily do this

through code. Lets take my previous gpx file from when I ran the Paris Half back in March. You will find it in the Github repository at the top of the page. (8663587823.gpx) Open up a python IDE of your choice in my case I'm using pycharm. Go to your terminal and add pip install -U gpx-converter

```
(PycharmProjects) christophers-MacBook-Pro:GPX christophermayes$ pip install -U gpx-converter
Collecting gpx-converter
  Downloading gpx_converter-2.1.0-py2.py3-none-any.whl (8.1 kB)
Collecting gpxpy>=1.4.0 (from gpx-converter)
  Downloading gpxpy-1.5.0.tar.gz (111 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━ 111.6/111.6 kB 1.6 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: numpy>=1.18.1 in /Users/christophermayes/PycharmProjects/lib/python3.8/site-packages (from gpx-converter) (1.24.3)
Requirement already satisfied: pandas>=1.0.3 in /Users/christophermayes/PycharmProjects/lib/python3.8/site-packages (from gpx-converter) (2.0.1)
Requirement already satisfied: python-dateutil>=2.8.2 in /Users/christophermayes/PycharmProjects/lib/python3.8/site-packages (from pandas>=1.0.3->gpx-co
```

Next thing, add the gpx file to your repository area.



```
Project ~/PycharmProjects/GPX main.py
└ GPX ~/PycharmProjects/GPX
    └ 8663587823.gpx
        └ main.py
    > External Libraries
    > Scratches and Consoles
```

```
from gpx_converter import Converter
Converter(input_file='8663587823.gpx').gpx_to_csv(output_file='8663587823.csv')
```

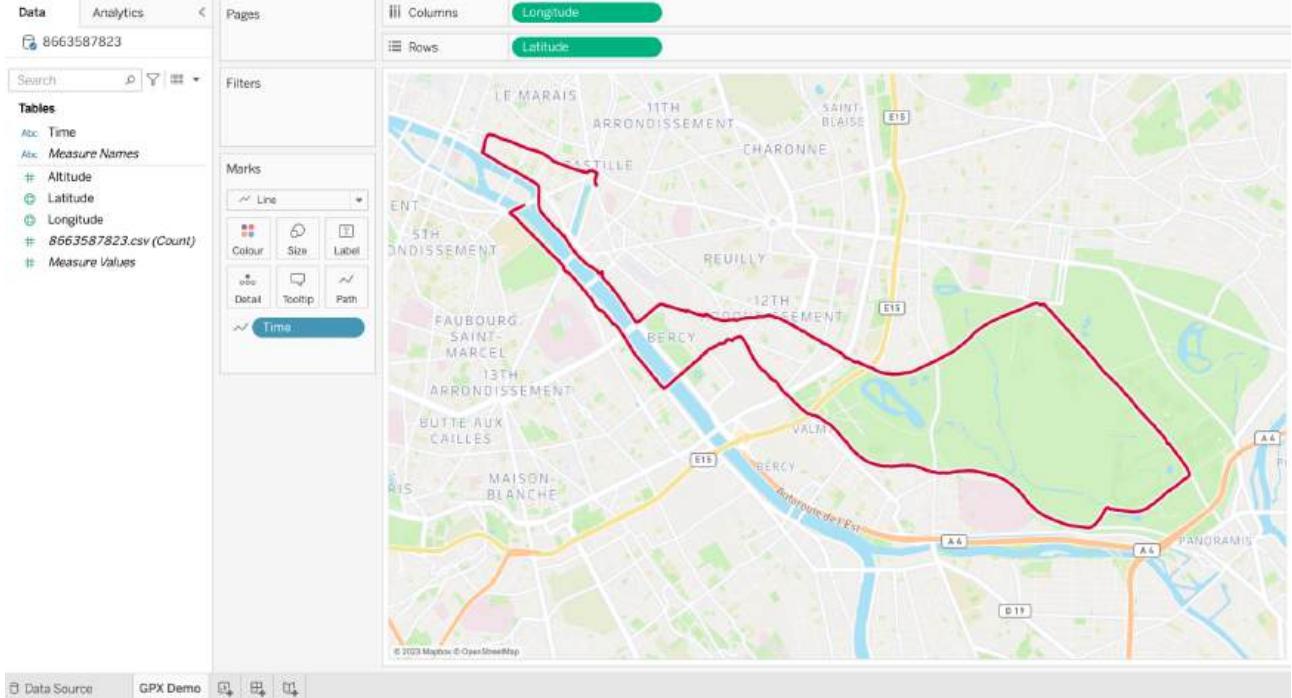
All we need to do is update our code with the correct input file name and output file name and we are ready to hit run. For me, I've left them as the original strava activity numbers from my Strava page.



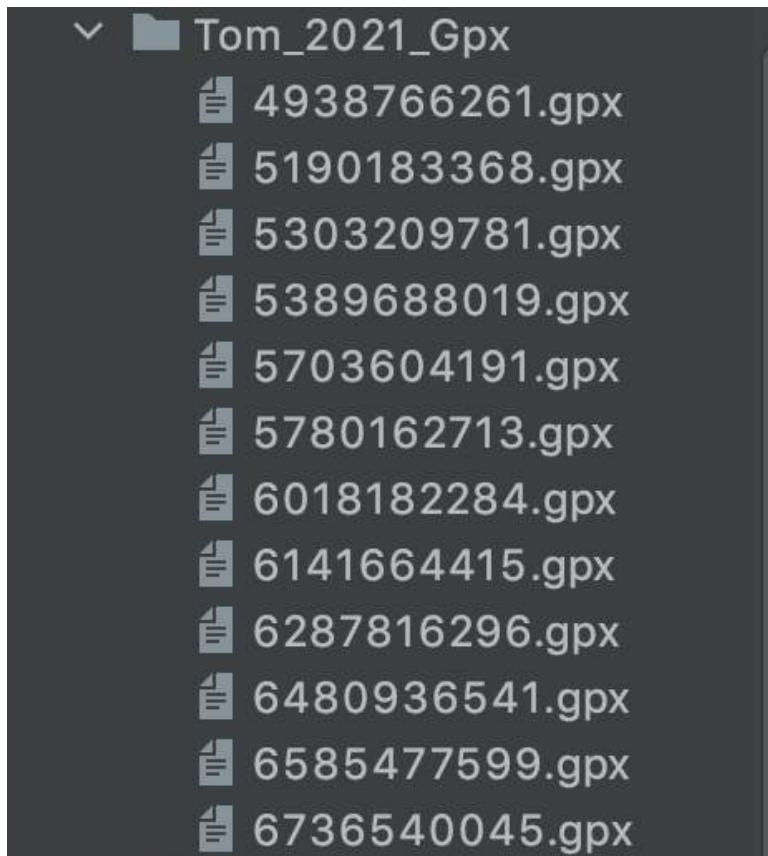
```
Project ~/PycharmProjects/GPX main.py
└ GPX ~/PycharmProjects/GPX
    └ 8663587823.csv
    └ 8663587823.gpx
        └ main.py
    > External Libraries
    > Scratches and Consoles
```

```
from gpx_converter import Converter
Converter(input_file='8663587823.gpx').gpx_to_csv(output_file='8663587823.csv')
```

When we click run, you will see the csv version appear. Let's go to Tableau to see if it has worked.



Looks pretty spot on. Do note, we need Time on path for Tableau to know in which order to draw the path for our longitude and latitude points. Fantastic, this way we have completely eliminated the need for using an online converter! Now this is okay but it only works for one gpx file at a time. So really we will need to create the equivalent for each of our files. LOOKING AT MULTIPLE FILES This next bit of code of course you can go away and amend. I've written it so that we get a csv for each of our gpx files, if you only want one final csv file perhaps you rewrite this code but I wanted individuals to have both available. For this bit of the tutorial I will be gluing together 10 gpx files, courtesy of my good friend Tom Smith. Tom actually got me into running in the first place, and I managed to dig out a few of his 10K badges from back in 2021 (before he moved home) that inspired me to take it



more seriously.

First step is put your gpx files into a

folder, like above.

```
# import required module
import os
# assign directory
directory = 'Tom_2021_Gpx'
for filename in os.listdir(directory):
    f = os.path.join(directory, filename)
    # checking if it is a file
    if os.path.isfile(f):
        df = Converter(input_file=f).gpx_to_dataframe()
        df['filename'] = filename
        df.to_csv(f'OutputFiles/{filename}.csv')
```

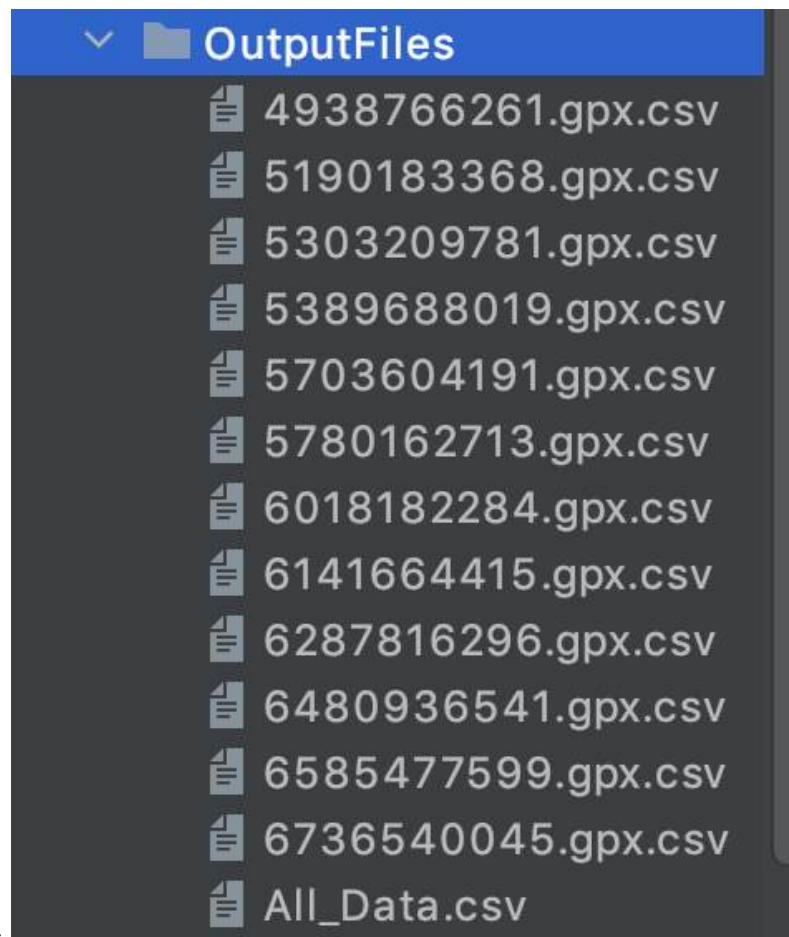
This code snippet will convert the gpx file to a dataframe and add a column with the name of the gpx file. It will then save these files down to a chosen location as csvs. In this example I save them to a folder I created called OutputFiles. This means you will have a csv for each activity.

```
import glob
import pandas as pd
os.chdir("/Users/christophermayes/PycharmProjects/GPX/OutputFiles")

extension = 'csv'
all_filenames = [i for i in glob.glob('*.{}'.format(extension))]

#combine all files in the list
combined_csv = pd.concat([pd.read_csv(f) for f in all_filenames])
#export to csv
combined_csv.to_csv("All_Data.csv", index=False, encoding='utf-8-sig')
```

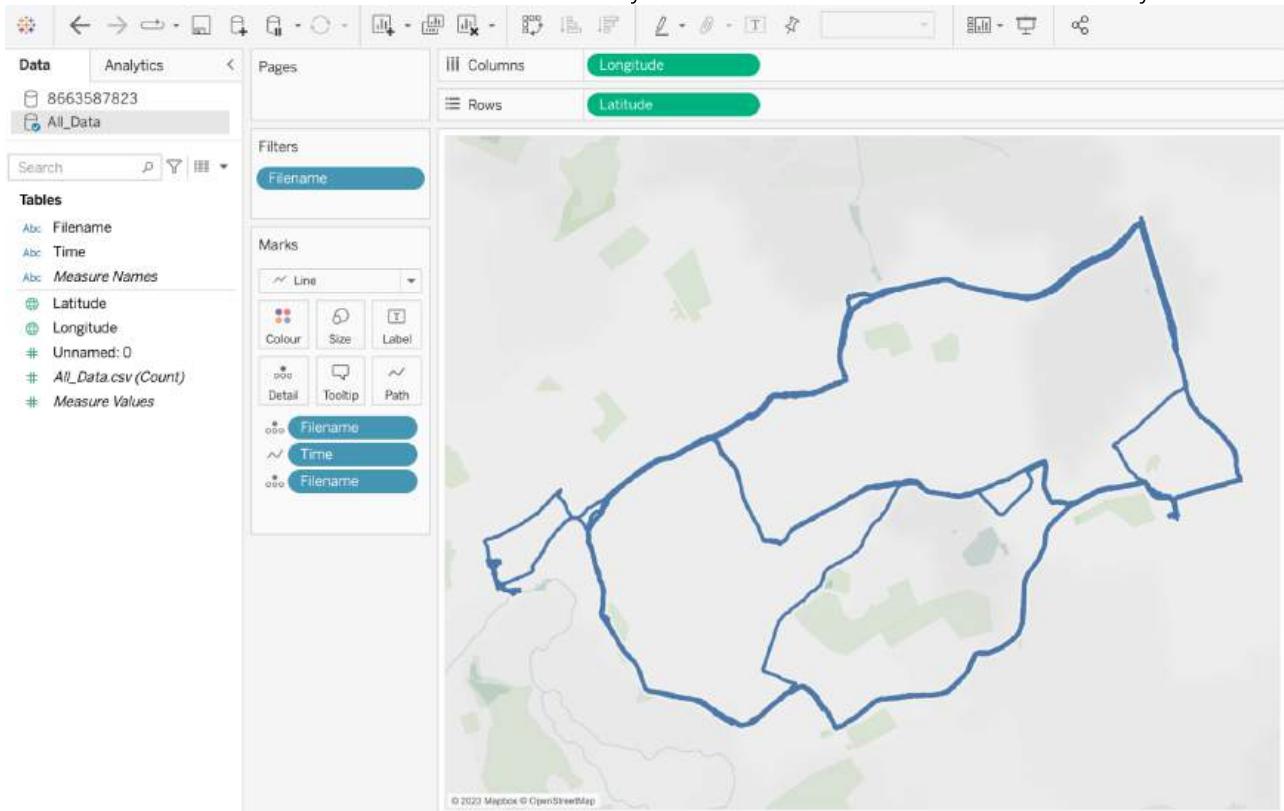
The next bit of code looks to locate the output files and then glue them together. It's save this final document down as



All_Data.csv You'll see it appear in the folder.

We

can finally sense check that our data is correct. We see that the number of rows adds up to those of our 12 files. We can also throw it into Tableau to check that it works the way we intend. Here are some of those runs as overlays.



That should be everything needed in terms of automating the gpx part of the process. If you get stuck I recommend checking out the documentation [here](#). My plan down the line is to be able to build full automated workflows for Strava dashboards for those that join my online fitness club. You can join [here](#). Going further:

- Chart all your different activities that have maps
- Create the same automation with a coding language of your choice
- Improve the performance of the python code
- Healthy body, healthy mind. **LOGGING OFF, CJ**

CHATGPT GOOGLE SHEETS EXTENSION

Hi all, In previous blogs we looked at analysing CSV files using Langchain and ChatGPT. This time we will look at some new future tools that have been released that may help analysts around those pesky csv files.

Use ChatGPT in Google Sheets and Docs

Supports all models: GPT-4 (if you have access), ChatGPT (gpt-3.5-turbo) and GPT-3

- ✓ Clean lists, extract entities, convert formats
- ✓ Edit, summarize, translate, classify
- ✓ Generate ad creative, taglines, subject lines, outlines, blogs, emails

Our GPT functions are perfect for short text in Sheets.

Our Docs integration and presets are great for long form content.

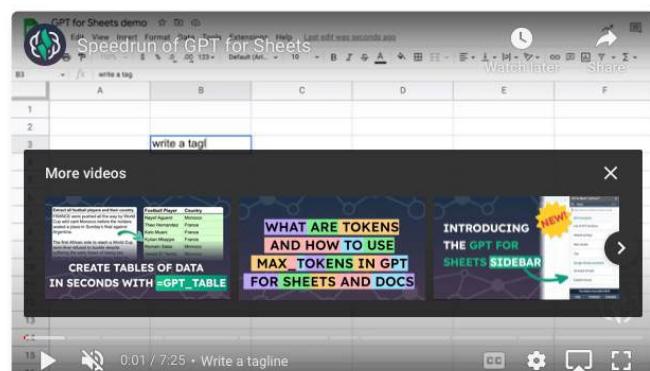
Works with all base and fine-tuned models.

 4.6/5 · 2M+ installations

[Install to get started](#)

[Video tutorials](#)

The add-on is free! [The only cost is OpenAI.](#)



GPT for Sheets and Docs:
one add-on that works with both

 Google Sheets
 Google Docs

Marc Andreessen  @pmarca · [Follow](#)

Killer app!

 Q-Cap  @qcapital2020
Oh god , I wanna throw up.

We will look just at the top 25 records in the Sample Superstore data from Tableau. How to get set up? Well first you will need an Open API key, I've written previously how you can get this key through signing up to their program. (You get free \$5 worth of tokens!) You can sign up on OpenAI.com We will want to also install the GPT for google sheets add on. You can access it through the link [here.](#)



GPT for Sheets™ and ...

[Install](#)

ChatGPT in Google Sheets™ and Docs™. Get the full power of AI for inspiration data cleaning, classification, extraction, translation, edition, summarization, writing directly in you...

By: [Talarian](#)

Listing updated: May 4, 2023

When you install, you will need to sign into your google account associated to Google sheets, and allow the access for chatGPT

This will allow **GPT for Sheets and Docs** to:

- View and manage documents that this application has been installed in (i)
- View and manage spreadsheets that this application has been installed in (i)
- ⊕ Connect to an external service (i)
- ⊕ Display and run third-party web content in prompts and sidebars inside Google applications (i)

to run against your file.

Click through and the set up is

 GPT for Sheets™ and Docs™ has X
been installed!

2 of 2 Additional setup

GPT for Sheets™ and Docs™ requires additional setup before your users can start using it. It is recommended that you complete this setup now.

[Complete additional setup now](#)

Opens in a new browser tab/window 

[PREV](#) [DONE](#)

complete.

We can then load up a data file. Here I'm looking at

the Sample Superstore Orders file – I ended up just reducing the file to 25 rows because I didn't want to take up a lot of run time against my free tokens.

The screenshot shows a Google Sheets interface with the title "Sample - Superstore_Orders". The Extensions menu is open, and the "GPT for Sheets™ and Docs™" option is selected. A dropdown menu from this option lists various functions:

- Add-ons
- Macros
- Apps Script
- GPT for Sheets™ and Docs™**
- Launch & Enable functions
- Set API key
- Manage cache
- Replace GPT formulas with values in selection
- Replace GPT formulas with values in entire sheet
- Import data to Google Sheets
- Mail merge from Google Sheets
- Help

The main spreadsheet area displays a table of data with columns: Category, City, Country/Region, Customer, Postal Code, Product Name, Region, and Segment.

	A	B	C	D	H	I	J	K
	Category	City	Country/Region	Customer	Postal Code	Product Name	Region	Segment
1	Office Supplies	Houston	United States	Darren	0380C	77095 Message Book	\ Central	Consumer
2	Office Supplies	Naperville	United States	Phillina	12326	60540 GBC Standard P	Central	Home Office
3	Office Supplies	Naperville	United States	Phillina				Home Office
4	Office Supplies	Naperville	United States	Phillina				Home Office
5	Office Supplies	Naperville	United States	Phillina				Home Office
6	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05/01/2019	US-2019-	Consumer
7	Furniture	Henderson	United States	Maria Etezadi	Global	06/01/2019	US-2019-	Home Office
8	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06/01/2019	US-2019-	Home Office
9	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06/01/2019	US-2019-	Corporate
10	Office Supplies	Henderson	United States	Maria Etezadi	Ibico	06/01/2019	US-2019-	Home Office
11	Office Supplies	Henderson	United States	Maria Etezadi	Alliance	06/01/2019	US-2019-	Home Office
12	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06/01/2019	US-2019-	Home Office
13	Office Supplies	Los Angeles	United States	Lycoris Saunders	Xerox	06/01/2019	US-2019-	Consumer
14	Technology	Henderson	United States	Maria Etezadi	Other	06/01/2019	US-2019-	Home Office
15	Technology	Henderson	United States	Maria Etezadi	GE	06/01/2019	US-2019-	Home Office
16	Furniture	Huntsville	United States	Vivek Sundaresa	Howard Miller	07/01/2019	US-2019-	Consumer
17	Office Supplies	Huntsville	United States	Vivek Sundaresa	Acco	07/01/2019	US-2019-	Consumer
18	Office Supplies	Laredo	United States	Melanie Seitz	Newell	09/01/2019	US-2019-	Consumer
19	Technology	Laredo	United States	Melanie Seitz	Memorex	09/01/2019	US-2019-	Consumer
20	Furniture	Springfield	United States	Anthony Jacobs	Howard Miller	10/01/2019	US-2019-14902C	22153 Howard Miller 11 South

To enable the ability for GPT for sheets to work, go to extensions and find it in the drop down. Click launch and enable functions.

Below are some examples of just some of the things you can do while in your google sheet!

This add-on brings ChatGPT and GPT-3's AI power to Google Sheets™ by exposing many simple custom functions:

GPT to get the result of ChatGPT in a single cell

GPT_LIST to get multiple result in a column (one item per cell)

GPT_TABLE to get a table of items from a prompt

GPT_FILL to fill a range from examples

GPT_FORMAT to sanitize your spreadsheet data into the same format

GPT_EXTRACT to extract entities from your Sheets™ data

GPT_EDIT to edit your spreadsheet content

GPT_SUMMARIZE to summarize your spreadsheet content

GPT_CLASSIFY to classify spreadsheet content into a single category

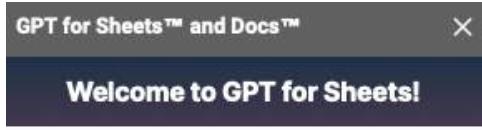
GPT_TAG to apply tags to your spreadsheet content

GPT_TRANSLATE to translate your spreadsheet content

GPT_CONVERT to convert from table to csv to html to json to xml, etc

GPT_MAP to map values from two columns

Here it is when it will ask you for your OpenAI API Key. Enter it and click next and we will get through the prompts page.



To start using GPT for Sheets, you first need to set your OpenAI API key. If you don't have one, [create it on the OpenAI dashboard](#).

Enter your OpenAI API key

Next

ⓘ [How to create an OpenAI API key](#)

Here are a list of just some of the different functions that we can do using GPT.

GPT for Sheets™ and Docs™	
← List of GPT functions	
BETA	
GPT	▼
GPT_LIST	▼
GPT_HLIST	▼
GPT_SPLIT	▼
GPT_HSPLIT	▼
GPT_FILL	▼
GPT_TABLE	▼
GPT_MAP	▼
GPT_FORMAT	▼
GPT_EDIT	▼
GPT_TAG	▼
GPT_CLASSIFY	▼
GPT_EXTRACT	▼
GPT_SUMMARIZE	▼
GPT_TRANSLATE	▼
GPT_CONVERT	▼
GPT_CODE	▼
GPT_CREATE_PROMPT	▼

Lets do a few basic functions. Dates.... They are always a tricky one.

```
GPT_FORMAT(input, target_format,
[source_format], [temperature],
[max_tokens], [model])
```

=GPT_FORMAT(L2, "DD month YY","DD/MM/YYYY",0.9)

formula will do, is looks at the respective column, asks for the target format, and then what the source format column will be. (This would be optional if you have lots of different date formats). Temperature is a measure of accuracy (Between 0-1) and tokens is to ensure you don't use all your token credits up.

What this

U	V
GPT Format Ship Date	GPT Format Ship Date 2
01 July 19	01 Monday July 2019
01 Aug 19	01 Thursday August 2019
01 Aug 19	01 Monday August 2019
01 Aug 19	01 Thursday August 2019
01 Dec 19	01 Monday December 2019
01 Oct 19	01 Wednesday October 2019
01 Oct 19	01 Wednesday October 2019
01 Jul 19	01 Tuesday July 2019
01 Oct 19	01 Wednesday October 2019
01 Oct 19	01 Wednesday October 2019
01 Oct 19	01 Wednesday October 2019
01 Aug 19	01 Tuesday August 2019
01 Oct 19	01 Wednesday October 2019
01 Oct 19	01 Wednesday October 2019
13 Jan 19	13 Monday January 2019
13 Jan 19	13 Monday January 2019
15 Jan 19	15th day January 2019
15 Jan 19	15 Monday January 2019
14 Jan 19	14 Monday January 2019
18 Jan 19	18 Wednesday January 2019
16 Jan 19	16 Wednesday January 2019
18 Jan 19	18 Wednesday January 2019

The right hand column is =GPT_FORMAT(U2,

"DD day Month Year", "DD MMM YY", 0.9) This may not wow you but it may add some productivity gains in day to day work for those using google sheets. Lets look at some other examples. Can we find the code required to extract the first name from a full name field. =GPT_CODE("extract the first name from a customer name field with a full name in", "python", 0.8) Here is the prompt I put into excel.

W	X	Y
GPT Code Python	GPT Code SQL CREATE FUNCTION getFirstName(fullName VARCHAR(50)) RETURNS VARCHAR(50) BEGIN DECLARE firstName VARCHAR(50); SET firstName = (SELECT SUBSTRING_INDEX(fullName, ' ', 1)); RETURN firstName; END;	GPT Code Tableau LEFT([Customer Name], FIND(' ', [Customer Name]))

Well lets check.

```

def extract_first_name(full_name):
    return full_name.split()[0]

full_name = "Christopher James Mayes"
print(extract_first_name(full_name))

```

Run: main
 /Users/christophermayes/PycharmProjects/bin/python /Users/christophermayes/PycharmProjects/ChatGPT/main.py
 Christopher
 Process finished with exit code 0

I've copied out the function and tested it against my name. Seems to work fine! What about changing the word Python to Tableau? Well not entirely correct in this case, but I can imagine it is more accurate for actual coding languages. You'll see by comparing the prompt answer given above with my answer below the find tool had the customer name and the blank string field the wrong way round!

Customer Name

Aaron Bergman	Aaron
Aaron Hawkins	Aaron
Aaron Smayling	Aaron
Adam Bellavance	Adam
Adam Hart	Adam
Adam Shillingsburg	Adam
Adrian Barton	Adrian
Adrian Hane	Adrian
Adrian Shami	Adrian
Almee Bixby	Aimee
Alan Barnes	Alan
Alan Dominguez	Alan
Alan Haines	Alan
Alan Hwang	Alan
Alan Schoenberger	Alan
Alan Shonely	Alan
Alejandro Ballentine	Alejand..
Alejandro Grove	Alejand..
Alejandro Savy	Alejand..
Aleksandra Gannaway	Aleksan..
Alex Avila	Alex
Alex Grayson	Alex
Alex Russell	Alex
Alice McCarthy	Alice
Allen Arnold	Allen
Allen Goldenen	Allen

First Name

LEFT([Customer Name], FIND([Customer Name], " "))

The calculation is valid. 1 Dependency ▾ Apply OK

Still, not bad going and easy to amend. Side note, I hadn't really ever used FIND in Tableau?? Was pleasantly surprised by its positioning power. Okay lets look at one final simple example of the ChatGPT Sheets Extension. Can we generate a formula in excel? I want to take the discount price and apply it to the sales value. Below I write in my prompt.

Generate formula



Describe your goal

Use discount as a % and find the value of the new sales price

Generate formula

=A1*(1-B1) A1 = Original
Sales Price B1 = Discount

Insert in active cell

Copy

By then adding a sales column with the new formula it has created our

P	Q	R	S	T
Discount	Profit	Quantity	Sales	Sales Price
0.2	5.5512	2	16.448	13.1584
0.8	-5.487	2	3.54	0.708
0.2	4.2717	3	11.784	9.4272
0.2	-64.7748	3	272.736	218.1888
0.2	4.884	3	19.536	15.6288
0	746.4078	9	2573.82	2573.82
0	1.4796	2	5.48	5.48
0	5.2398	3	12.78	12.78
0	274.491	2	609.98	609.98
0	0.3112	4	31.12	31.12
0	3.0084	1	6.54	6.54
0	9.3312	3	19.44	19.44
0	204.1092	4	755.96	755.96
0	113.6742	2	391.98	391.98
0.6	-53.7096	3	76.728	30.6912
0.8	-18.2525	7	10.43	2.086
0.2	1.168	2	9.344	7.4752
0.2	9.75	3	31.2	24.96
0	21.2954	1	51.94	51.94
0	1.3583	1	2.89	2.89
0	3.0814	2	9.94	9.94
0.15	3.9294	3	333.999	283.89915
0	87.3504	6	545.94	545.94
	0	6.567	5	19.9
discounted prices!				

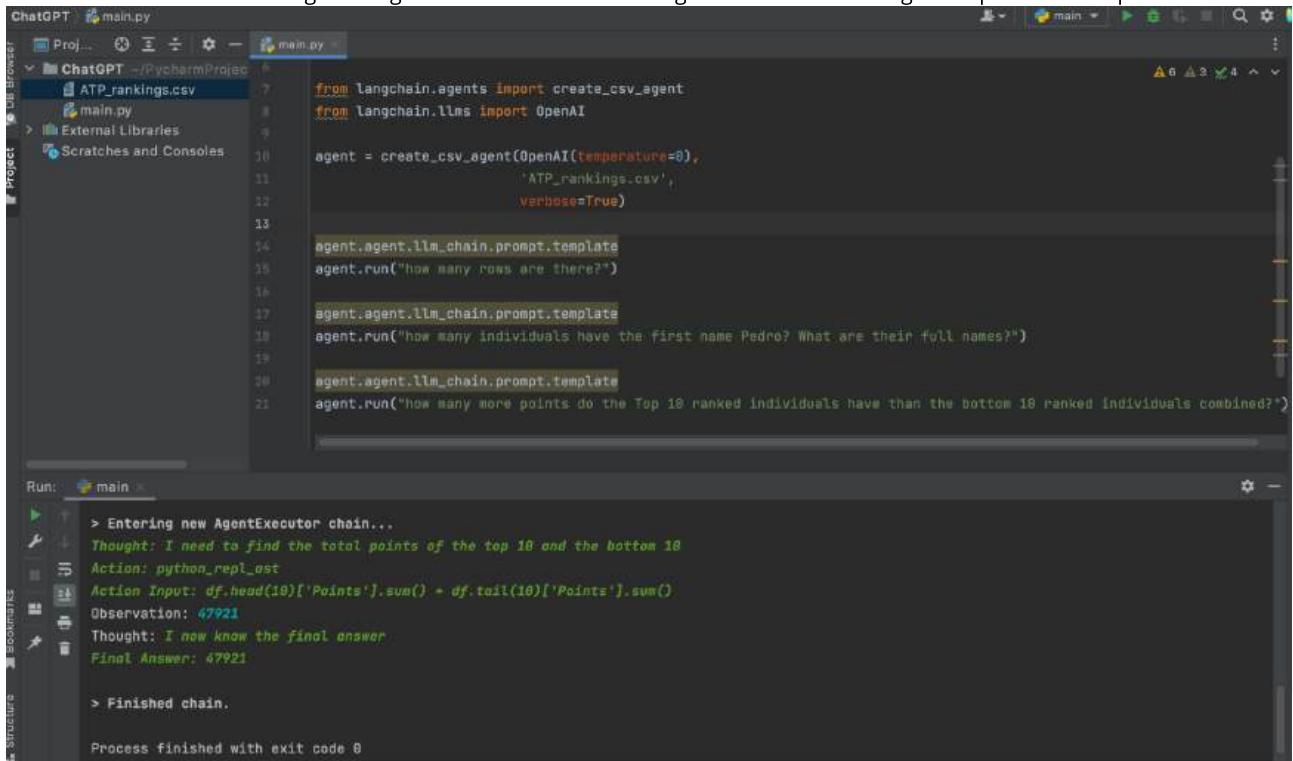
Lets end with a quick final translation example. =GPT_TRANSLATE(A2, B2)

A	B	C
Text	Language	Translation
I love Tableau	English	I love Tableau.
I love Tableau	Spanish	¡Me encanta Tableau!
I love Tableau	French	J'adore Tableau
I love Tableau	German	Ich liebe Tableau.
I love Tableau	Dutch	Ik hou van Tableau.

What are my overall thoughts? With over 15 different functions at a basic level this will really help business analysts be able to do their job in a more efficient way. I think it adds a good middle ground for formatting information, creating subsets and analysing data that you may have had to either create more tricky excel formulas for or written code for. Whether in ChatGPT or within this add on, it certainly has helped me debug code a lot quicker than searching online for answers. It does however come with limitations, whilst we are in such a growth phase of AI – the field will be littered with mistakes and excel document responses from chatGPT will be no different from that – and some of the other functions this extension offered I found more challenging to adopt. Overall, the future is bright and I suspect that this small demo of getting this extension added and running a few queries is a drop in the ocean of what's to come. Check it out if you have some spare time and let me know if your organisation has started to use AI advancements in your job roles. Would love to learn more from you. That's it for this week, catch you soon. **LOGGING OFF CJ**

ANALYSING CSV FILES WITH LANGCHAIN (CHATGPT)

Hi all, Can we get OpenAI to answer our questions based on a csv input? We are back with another coding snippet this week. This week focussing on Langchain and how we can autogenerate answers using the capabilities of OpenAI.



```

from langchain.agents import create_csv_agent
from langchain.llms import OpenAI

agent = create_csv_agent(OpenAI(temperature=0),
                        'ATP_rankings.csv',
                        verbose=True)

agent.agent.llm_chain.prompt.template
agent.run("how many rows are there?")

agent.agent.llm_chain.prompt.template
agent.run("how many individuals have the first name Pedro? What are their full names?")

agent.agent.llm_chain.prompt.template
agent.run("how many more points do the Top 10 ranked individuals have than the bottom 10 ranked individuals combined?")

```

Run: main

```

> Entering new AgentExecutor chain...
Thought: I need to find the total points of the top 10 and the bottom 10
Action: python_repl_ost
Action Input: df.head(10)['Points'].sum() + df.tail(10)['Points'].sum()
Observation: 47921
Thought: I now know the final answer
Final Answer: 47921

> Finished chain.

Process finished with exit code 0

```

The above printscreen isn't the prettiest end output, but is what we will be striving towards today and I am very excited to showcase it. This tutorial will look to show how we can use the OpenAI package and langchain, to look at a csv file and ask it questions about the file and the agent will send back a response. Today we will look at LLMs. Large language models (LLMs) are emerging as a transformative technology, enabling developers to build applications that they previously could not. However, using these LLMs in isolation is often insufficient for creating a truly powerful app – the real power comes when you can combine them with other sources of computation or knowledge. Read more about it [here](#) Firstly, let's look at how we can get set up. We will need

- A Python IDE (I choose to use PyCharm)
- An OpenAI account and an API key generated.
- A dataset to analyse (I Choose tennis)

As a brief refresher, here is where you can access your API token. It is stored under User, where you can create a secret key.

The screenshot shows the 'API keys' section of the OpenAI platform. On the left, there's a sidebar with 'ORGANIZATION' and 'USER' sections. Under 'USER', 'API keys' is selected and highlighted in green. The main content area is titled 'API keys' and contains a note about not displaying secret API keys again after generation. It also cautions against sharing the key with others. A table lists the single key: 'My Test Key' with key value 'sk-...', created on '30 Apr 2023' and last used on '30 Apr 2023'. Below the table is a link '+ Create new secret key'. A section titled 'Default organization' allows selecting the default organization for requests, currently set to 'Personal'. A note at the bottom says 'Note: You can also specify which organization to use for each API request. See [Authentication](#) to learn more.'

You can see your amount of usage of your free credits under the usage section.

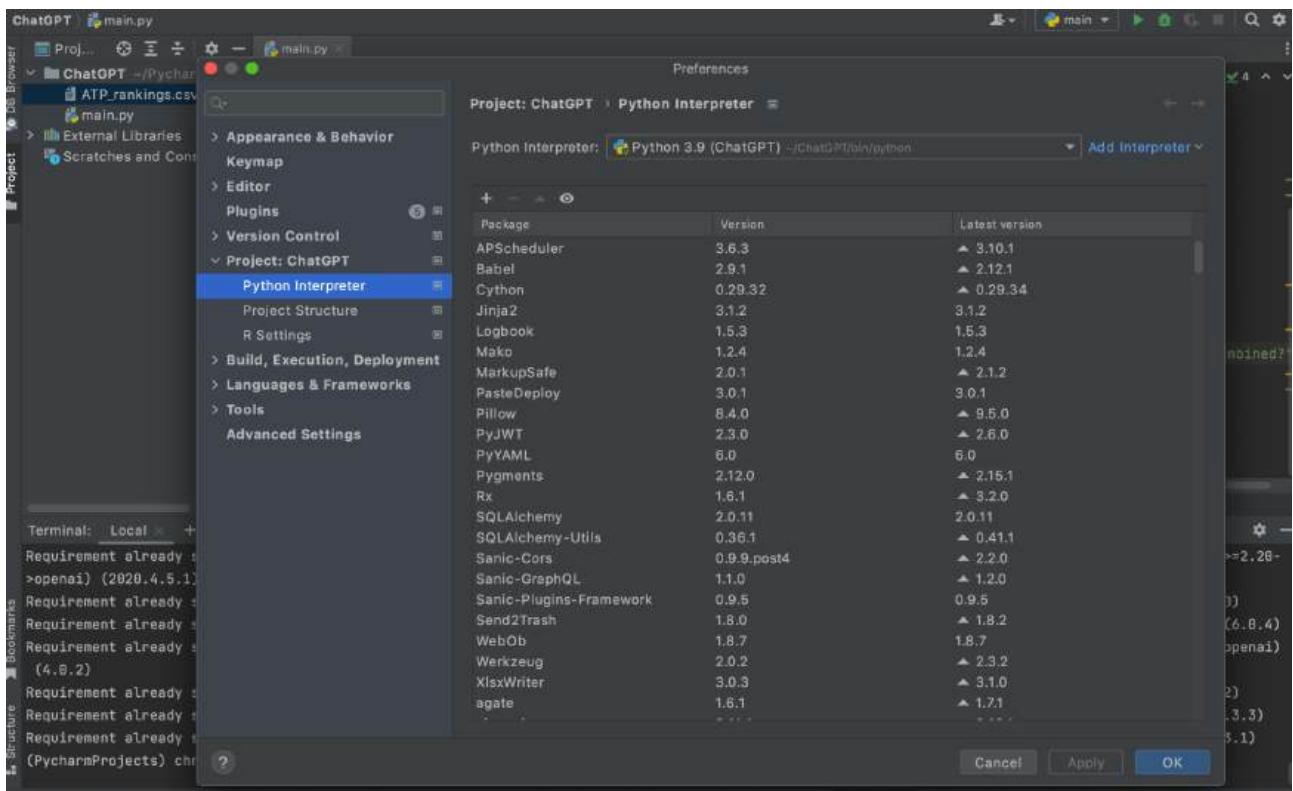
The screenshot shows the 'Usage' section of the OpenAI platform. The sidebar has 'Settings' and 'Usage' selected. The main area displays a chart titled 'Daily usage (USD)' for April. The y-axis ranges from \$0.00 to \$0.006. The x-axis shows dates from 01 Apr to 28 Apr. A single bar for April 28 shows a usage of approximately \$0.0048. Below the chart is a section titled 'Free trial usage' with a progress bar showing '\$0.01 / \$5.00'. A table below lists a grant: 'Grant 1' with credit granted '\$5.00' and expiration '1 August 2023'.

By now, you will know my IDE tool of choice tends to be PyCharm. Copy and paste the code snippet into your own main.py file. Make sure to add in your api key into the code. Download the ATP Rankings dataset from the repo. I chose this dataset from a previous tutorial I wrote – I felt like it was a nice easy dataset to interpret showing the rankings of the top 500 mens players....

back in November-ish time. So for this to work you will need to be on python 3.9 – I was previously on 3.8 and it was not compatible and was throwing up errors. Also you will need to pip install your packages. For this we need OpenAI, langchain, os and pandas.

```
(PycharmProjects) christophers-MacBook-Pro:ChatGPT christophermayes$ pip install langchain
Collecting langchain
  Downloading langchain-0.0.153-py3-none-any.whl (696 kB)
    696.4/696.4 kB 2.2 MB/s eta 0:00:00
Requirement already satisfied: PyYAML>=5.4.1 in /Users/christophermayes/PycharmProjects/lib/python3.8/site-packages (from langchain) (6.0)
Collecting SQLAlchemy<3,>1.4 (from langchain)
  Downloading SQLAlchemy-2.0.11-cp38-cp38-macosx_10_9_x86_64.whl (2.8 kB)
    2.8/2.0 kB 2.8 MB/s eta 0:00:00
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /Users/christophermayes/PycharmProjects/lib/python3.8/site-packages (from langchain) (3.8.4)
Requirement already satisfied: async-timeout<5.0.0,>=4.0.0 in /Users/christophermayes/PycharmProjects/lib/python3.8/site-packages (from langchain) (4.0.0)
```

You'll know this is done as there will be no underlined red in the code anymore.



Sometimes in the terminal it doesn't fully load my packages so I like to go into the interpreter to manually add them there. Once that's all done we are ready to run the code. But before we get into the responses, lets look at why we are doing this. What are agents? Agents involve an LLM making decisions about which Actions to take, taking that Action, seeing an Observation, and

repeating that until done. LangChain provides a standard interface for agents, a selection of agents to choose from, and examples of end-to-end agents. If you'd like to learn more about Langchain you can read about it [here](#). In our example we want to execute three simple questions against our csv. Firstly, how many rows are there? Secondly, how many individuals have the first name Pedro? What are there full names? Thirdly, How many more points did the Top 10 players currently have compared to the bottom 10 players? Lets look at this in action.

> Entering new AgentExecutor chain...

Thought: I need to find out how many rows are in the dataframe

Action: python_repl_ast

Action Input: len(df)

Observation: 500

Thought: I now know the final answer

Final Answer: There are 500 rows in the dataframe.

When we run our command, it translates our question into a thought. It will then take that thought and create an action, almost as if you ask an analyst how would you answer it in code. It then computes that action and gives us an observation. Of course it is no surprise our Top 500 ranking of mens tennis has 500 records in. It then has a final thought that it knows the answer, and states it to us as an answer. Nice and simple! Lets go on to do something a little more challenging and find individuals called Pedro. Now if I was to ask the agent, how many people in the dataset are called Pedro it will give us zero.

```

> Entering new AgentExecutor chain...
Thought: I need to find out how many individuals have the name Pedro in the dataframe
Action: python_repl_ast
Action Input: df[df['Name'] == 'Pedro'].shape[0]
Observation: 0
Thought: There are no individuals with the name Pedro in the dataframe
Final Answer: There are 0 individuals with the name Pedro in the dataframe.

> Finished chain.

```

So we have to be a little bit more precise.

```

agent.agent.llm_chain.prompt.template
agent.run("how many individuals have the name Pedro?")

agent.agent.llm_chain.prompt.template
agent.run("how many individuals have the first name Pedro? What are their full names?")

```

Lets ask it how many individuals have the FIRST name Pedro.

```

> Entering new AgentExecutor chain...
Thought: I need to filter the dataframe to find the individuals with the first name Pedro
Action: python_repl_ast
Action Input: df[df['Name'].str.startswith('Pedro')]
Observation:
   Rank      Name  Points
57    58  Pedro Martinez     818
58    59  Pedro Cachin     809
347  348  Pedro Bescardin Dias    138
493  494  Pedro Sousa       72
Thought: I now know the final answer
Final Answer: There are 4 individuals with the first name Pedro. Their full names are Pedro Martinez, Pedro Cachin, Pedro Bescardin Dias, and Pe

```

Well of course now it understands. It tells us the code it would use to compute the answer, it prints all those in the dataset called Pedro and gives us a written response too. Pretty impressive. So our final question. A mathematical computation question. Now I ask it to compare the Top 10 ranked individuals against the Bottom 10 and work out the number of points difference.

```

> Entering new AgentExecutor chain...
Thought: I need to find the total points of the top 10 and the bottom 10
Action: python_repl_ast
Action Input: df.head(10)['Points'].sum() + df.tail(10)['Points'].sum()
Observation: 47921
Thought: I now know the final answer
Final Answer: 47921

> Finished chain.

```

And there we are. The correct value finding the difference between them. Today we have gone over some basic examples of calculations against a csv file, but this is really only just touching the surface of what is possible. Going further:

- Try look at other types of files you can analyse including pdf's.
- Load your own CSV file in and ask that questions.
- Try debug errors in what type of questions it can't respond to.
- Try creating a log of questions answers and responses.

That's it for this week, catch you soon. **LOGGING OFF CJ**

TABLEAU DATA ANALYST EXAM REFLECTIONS (DEBORAH S & MEHRAS A)

Hi all, I'm so pleased to have two fantastic community members join the blog today to share their own reflections on the Tableau Data Analyst exam. A good chunk of planning [Deb](#) & [Mehras](#) have put into the conversation today and its really my honour to be able to host it on the site.

Mehras

@ItsMehras Follows you

Be a good human. MSDA @uhdowntown | Tableau Certified Data Analyst | Tableau Certified Desktop Specialist | Tableau K12 | Nerazzurro 🌐

📍 HTX 🌐 tabsoft.co/3K9YaiW 📅 Joined October 2013

Deborah Simmonds

@finkledeb Follows you

public.tableau.com/profile/deborasimmonds

Joined March 2013

Both have recently changed careers to work in data, although come from completely different backgrounds. They share a love of Tableau and the #datafam community. After successfully taking the Tableau Certified Analyst Exam this year, they have been kind enough to share their experience and resources to hopefully help others who are considering or already preparing to achieve this goal. **CJ: Welcome both, lets start with desire. Why did you want to do the exam?**

Tableau Certified Data Analyst

Get certified in the hottest role in data analytics.

[SCHEDULE EXAM NOW](#) [PREPARE FOR EXAM](#)

M: At the beginning of each year, we're required to set personal work goals and align them with a professional development plan. For me, I decided to take the Analyst Learning Path and began working through the e-learning courses, which took me roughly six months to complete. When I was about halfway through my courses, I told my boss that to become a certified Data Analyst, I had to pass an exam. D: Yes, we were also given the option at work to take this exam and I volunteered too. I've been working in my current role for just over a year, but I have no maths or data science background, and no MA or PhD. Having learned everything either on the job, e-learning, or self/datafam taught, I often struggle with imposter syndrome even though I've been using Tableau for some years now. It was so important to me to take this exam to try and overcome these feelings. **CJ: That's an interesting point you raise on imposter syndrome. How was it emotionally for you before the exam and after?**

D: I find it very difficult to study under pressure. I really dislike exams in general. Work paid for this and I think I kept thinking, "oh no, what if I'm not capable of doing it? What if I'm not as clever as they think I am? They're going to find out now!" Yeah. I was really worried about that. And I think that is like the imposter syndrome trap. M: Same – I felt a sense of responsibility to do well and pass it since they had invested in me. This was not the case with the [Tableau Desktop Specialist exam](#), which I personally paid for. **CJ: What were some of the ways that you tried to overcome some of those emotions?** M: I actually took the [Tableau Desktop Specialist exam](#) one month before the [Data Analyst exam](#) to familiarize myself with the online testing environment and gain a better understanding of how it worked. Overall, it was a helpful experience and prepared me for what to expect. D: I structured my revision quite well, and gave myself an hour at the end of every workday to study. That was the minimum I would do and kept this up for a number of weeks! **CJ: Did you have an accountability to help facilitate those learnings?** D: I think I'm very lucky to have a really supportive team at work. But for me, my accountability buddy was the fact that I'm trying to challenge myself to reach outside my comfort zone. M: I actually didn't have an accountability buddy, but my main motivation was to learn new things so that I can better support my department with all the technical Tableau questions. And now that more people in our team are working with Tableau, I really want to make sure I'm able to support everyone as much as possible. **CJ: Mehras, When you came to preparing for the exam, did you split some of your learnings by exam section?**

Preparing for the exam

One of the best preparations is role experience and time with the product. To be prepared, candidates are strongly encouraged to have a minimum of 6 months' experience. We also encourage you to explore some of our expert training options:



Tableau Certified Data Analyst Exam Readiness

This course is a ready-made bundle designed to help you build the skills and confidence you need to become a Tableau Certified Data Analyst. And you get two exam vouchers, just in case.

[LEARN MORE →](#)



Tableau Training Pass

Get unlimited access to the most popular Tableau training, plus exclusive pass-holder only workshops, training and office hours with certified Tableau experts.

[LEARN MORE →](#)



Analyst learning path

The Analyst learning path starts your Tableau Desktop journey with the basics and guides you through the topics you need to round out your analyst skill set.

[LEARN MORE →](#)

M: I actually took the "Prep Builder" course on Tableau eLearning specifically to prepare for Domain 1, which is all about connecting to and transforming data.



Prep Builder

This course will provide you with the tools and knowledge to clean, shape and combine data using Tableau Prep Builder. You'll learn how to go from data preparation to creating useful Tableau Prep Builder flows that output your data for analysis in Tableau Desktop and Tableau Server or Tableau Cloud. Finally, apply everything you've learned in an end-to-end capstone project using a complex, real-world scenario.

Time estimate:

14 hours

Prerequisites

For a hands-on introduction to Tableau Prep Builder, we recommend this course from our learning paths: [Getting started with Tableau Prep Builder](#).

Continuing education credit hours: 14

Learning objectives

At the end of this course, you will be able to:

- Connect to, clean and output your data for analysis in Tableau Desktop and Tableau Server or Tableau Cloud.
- Apply cleaning operations, create calculations and edit incorrect field values singly or in groups, replacing values manually or using built-in algorithms to address data issues.
- Build data prep flows to address common scenarios, such as the need to combine data using joins and unions, and reshape data using pivots and aggregations.
- Validate your data preparation work in Tableau Prep Builder and in Tableau Desktop.
- Configure and use Tableau Prep Builder's data sampling functionality.
- Design a complex flow, end-to-end, applying the range of cleaning techniques available, for a real-world scenario.

[START LEARNING](#)

[SEE SAMPLE COURSE](#)

I mainly focused on the modules from the Analyst Learning Path of eLearning for Domain 2 (Explore and Analyze Data) and Domain 3 (Create Content). But honestly, I learned a lot just by doing my work projects (I'm lucky that I get to work with Tableau a lot), participating in community projects like [Back 2 Viz Basics](#), [Preppin' Data](#), and [Makeover Monday](#), and following the #Datafam on Twitter to learn about new tricks and tips. I also watched all of [Andy Kriebel's YouTube shorts](#) (like 84 videos) and created all those vizzes, which not only helped me learn new tips but also gave me the confidence to tackle the hands-on section of the test. For Domain 4, I learned a lot from the eLearning course and got to practice everything I learned at work since we have access to Tableau Cloud. But no worries if you don't have access, you can always get a [free trial](#). **CJ: Which part of the exam were you most excited or worried about?** M: I was actually pretty excited about the Tableau Prep questions and the hands-on part of the exam. But if I had to choose, I'd say I was most worried about the Tableau Server questions. I haven't had as much experience with that part. **CJ: How does preparing for a closed book desktop exam compare to the real-world application?** D: Oh, well it doesn't, I don't know about you, but I Google everything. But I have noticed recently, since I've done it, that I've become a little bit faster. **CJ: What resources did you use? What were the best ones?** D: I did a lot of practice tests beforehand to find the holes in my knowledge, similar to how we approached the Alteryx certification at JLL, trying the test first because it's free and helps to identify where there are gaps in your knowledge. M: How did you come across those practice tests? I mostly used the e-learning and YouTube videos. I think your approach is really effective, and it reminds me of the diagnostic tests that I used to give to my students. Looking back, I wish I had tried something like that before taking exams

myself. D: Most of the resources I used were passed to me by #datafam member **Jessica Haire**, who encouraged me to try the practice tests by **Lukas Halim** and sent me this link to his [learning site](#). He provides plenty of free resources, and it is worth trying the desktop specialist practice tests too to refresh your knowledge. This site also contains a [Tableau Certified Data Analyst Study Guide](#) which is the same structure as Tableau's schedule but includes links to resources to learn each section. This makes it much quicker to find study materials when necessary. What videos did you use instead? M: About four or five days before the exam, while I was reviewing the material, I came across a YouTube channel called [Discover Analytics](#). The more videos I watched, the more comfortable I became with the content. The channel has a ["Certified Data Analyst Exam Prep Playlist"](#) with 10 great videos that cover exam domains with examples, exam tips, and some bonus test prep questions. D: Yes, I used the same video series! I was a little worried as I don't actively use Tableau Prep and although I do use tableau server I have far less administrative privileges now than in my previous job. So, I had to learn by reading around it quite a bit. I found these videos from Tableau really helpful in preparing for the exam. Note that you may need to log into Tableau to watch them. [The Tableau Prep Builder Interface New to Tableau Prep Builder? Start with these five fundamentals](#) If I'd given myself more time to prepare for the exam, I'd definitely have learned joined in with [Preppin' Data](#) community challenges. I believe you can do a Tableau Prep trial for a month. M: One tip that I found helpful while preparing for the test was to pay close attention to the steps I took in Tableau to complete a task. For example, when I needed to create a calculated field, I focused on the steps involved and tried to memorize them as a list of steps. D: **Tim Ngwena** was talking about this in his guide to the Beta version of the test, [Tableau Certified Data Analyst Certification: Exam & first impressions: Tableau Certification](#) he mentioned something that I experienced which is that if you've been using Tableau for a while, it can work against you a little bit as you tend to just do familiar tasks automatically and you don't think about the steps. I agree that preparing for this exam does make you step back and question "Why am I doing it this way?" And then you start to think about the order of operations work. M: I have also experienced that, so it is important to be mindful that sometimes there are multiple ways to perform a task correctly. For instance, I can think of three different ways to create a new calculated field. This is important because Multiple Response questions on the exam require you to select two or more answer options. Multiple Response questions are one difference between the Analyst Exam and the Desktop Specialist Exam. **CJ: Great thanks both some fantastic resources there. Do you have any tips on set-up?** M: If you have the option to use a bigger laptop, make sure you do so! The exam states how you are only allowed to use one monitor. There are tests at the start to ensure that your computer and the physical environment are properly configured. D: Some work laptops may have data loss protection settings, so I'm glad I did the checks well in advance as the first check showed that I needed to remove a Cisco product to enable the software to download. Because I had watched a video by **Tim Ngwena**, about his experience of the Data Analyst Beta Exam I didn't just do the laptop preparation once. **CJ: What about during the exam itself?** D: I was appreciative to have a invigilator on hand in case I experienced any technical issues. After the influx of virtual meetings during the last few years, I think we are all used to webcam access. I wish I was able to read the questions out loud though, as I'm use to that when problem solving. M: It's great that they have a system for exam security, making the test fair to everyone. Just try to relax and focus on doing your best on the exam. You got this! **CJ: Was the exam time intensive?** D: It did seem so. I don't know if previous exams have been easier or if they were just a bit more luxurious with the time. M: I had some extra time at the end of the Data Analyst test, even after reviewing all the questions in section three. The exam is in two parts including 45 multiple choice questions and 10 practical 'lab' questions, so you can divide your time in equal minutes as a rough guide.

TDA-C01: Tableau Certified Data Analyst

Where do you want to take your exam?

At a test center
 Online at my home or office
I have a Private Access Code

Prepare for your online exam at your home or office

Your computer
 Use a personal computer that has a reliable webcam and internet connection.
[Run system test.](#)

Your testing space
 The room should be a distraction-free, private place.
[See acceptable spaces](#) and view [permitted comfort aid list](#).

Your photo ID
 We'll verify your government-issued identification (ID) when you arrive for your exam.
[Review admission & ID policies](#)

What to expect
 Check in for your OnVUE exam 30 minutes before your appointment time.
[Watch our short video](#) to get familiar with the process.

Questions?
 Check out the [OnVUE FAQs](#) and [minimum technical requirements](#).

CJ: It's also worth noting they ask you to sign into the exam early, otherwise this eats into the exam time. Its best loads the exam up to half hour early. Don't forget your ID! **CJ: Did you use the flag feature?** M: Yes, I think it's a great feature that allows you to skip through some questions that you're unsure about and don't want to spend too much time on. Flagging questions is also useful because you might not be able to review all of them at the end of the test, but it's good to flag some so you can always

come back to them if you have extra time. **CJ: How was the hands-on lab?** M: I found the hands-on lab the easier part of the exam. They have the dashboards and sheets to amend pre-set within the Tableau file. I went back and forth several times to ensure I wasn't missing anything. Also, it's important to remember to save your progress as you go! **CJ: Which part of the exam content did you find most challenging?** D: I found some of the wording of the multiple-choice sections difficult to understand.

You had to really dig into the questions and go, right, what are they really asking? I also found the questions with multiple answers tricky, and found myself doubting my knowledge at these points. M: Some of the knowledge-based questions, especially the ones that require multiple selection and building a list of steps, can be a bit challenging. I think, it's a great idea to learn how to complete tasks in Tableau in different ways and to remember the steps because we tend to forget the steps after using the software for a while. **CJ: What about the exam content itself?**

Why learn with Tableau eLearning?



Role-based training

Tableau Learning Paths provide a clear track to proficiency with the most up-to-date Tableau training content.



Track your learning

Assessments help you evaluate where you are in your learning path and give you confidence in your new skills.



Learning at enterprise scale

Accelerate the onboarding process for new employees and help more experienced users get the most out of Tableau capabilities.

M: e-learning was perfect to prepare for the different sections of the exam, but I remember when it got to Tableau Cloud and Tableau Server I wanted to supplement it with more resources. D: Tableau Tim has some great YouTube resources that may help others in the same situation as you, such as this one. [Publishing data sources to Tableau Server & Tableau Cloud: Tableau Tutorial for Beginners](#): He has published some great beginners resources, it is well worth checking out his content to review any elements within the exam that you might be unsure of.

Publishing data sources to Tableau Server & Tableau Cloud: Tableau Tutorial for Beginners:

CJ: If you were to go back and do it again, what would you do differently in terms of preparation? D: I would have liked to have started preparing earlier! M: I would spend a bit more time on functions, especially calculated fields, and making sure that you're comfortable with FIXED LOD calculations as well. **CJ: How did it feel to pass?** M: After completing the test, I stepped out of the room and told my wife, "I'm not sure if I'll pass or not." While I felt confident about the hands-on parts, I was unsure about the rest. Approximately an hour after the test, I happened to check my email and was surprised to find that my score was already available. What a relief! D: I didn't think I'd passed either. And when I finished, my partner asked me how it went, and I told him that I couldn't call it, I could not tell if I'd passed or failed. I was so pleased. **CJ: Has it given you more confidence?** D:

Yes I think it has helped me gain a different type of confidence. I've noticed that due to the preparation and study I am completing tasks more quickly, and retaining focus a little easier, which is fantastic as these are issues that I've struggled with for

a long time. I've also felt more confident to support others with their preparation for the exam, which is something I've really enjoyed. M: For me, the journey of learning new things was more important than the actual certification and exam. Passing the test and getting certified was a validation of the effort I put in, but it was not the ultimate goal. This is how I see it. **CJ Round-Up:** Thank you both for joining the blog this week. I hope the conversation today brings others comfort around some of the type of content in the exam, some of the emotions felt both before, during and even after the exam as well as some tips around how best prepare. Take care and speak soon. **LOGGING OFF, CJ**

HOW TO LOCALLY HOST CHATGPT (GRADIO)

Hi all, ahhh ChatGPT. Some are buzzing by its endless potential, some treat it as a glorified Google search and some are sick of hearing the word and are just waiting for the mass adoption period. Well, depending where you sit on that spectrum will probably determine whether this blog is for you. Today we are going to look at how we can host a mini chatbot locally hosted instead of loading the standard chat.openai website. This is a beginner tutorial for those wanted to play around with writing python code. Below is what we will look to recreate.

The screenshot shows a local web application running at 127.0.0.1:7860. The title bar says "CJ - Demo". On the left, there's a "user_input" field containing the text "How can I get more involved in the Tableau community?". Below it are "Clear" and "Submit" buttons. On the right, under "output", is a response: "There are a few ways to get more involved in the Tableau community:
1. Join the Tableau Community: The Tableau Community is a great place to get started with Tableau. You can access it through the Tableau website, sign up for weekly newsletters, and participate in forums to ask and answer questions.
2. Attend Tableau User Groups: Tableau User Groups are local networking events that provide an opportunity to meet other Tableau users in your area. These groups host events to share tips and tricks, showcase user projects, and promote collaboration.
3. Participate in Tableau Public: Tableau Public is a great way to share your Tableau work with others. You can publish your dashboards and visualizations to Tableau Public and connect with other community members. You can also find inspiration by exploring amazing work created by others in the Tableau Public Gallery.
4. Attend Tableau conferences: Tableau hosts multiple conferences throughout the year, which are great opportunities to meet other Tableau users and enthusiasts. At these events, you can attend sessions and workshops, network with your peers, and learn new techniques from experts in the field."
At the bottom right of the output area is a "Flag" button.

As always – you can access the code in the repo at the top of the page. So what is ChatGPT?

The screenshot shows the official ChatGPT interface on chat.openai.com. The top navigation bar includes links for "About", "Help", "Log in", and "Sign up". The main area has a sidebar with a "+ New chat" button. The main content area features the "ChatGPT" logo. Below it is a summary table with three columns: "Examples", "Capabilities", and "Limitations".

Examples	Capabilities	Limitations
"Explain quantum computing in simple terms" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

At the bottom, there's a message input field with placeholder text "Send a message..." and a "Send" button.

Well, let's ask it. ChatGPT is a large language model created by OpenAI. ChatGPT is designed to generate human-like text and to respond to natural language inputs. ChatGPT can be used for a variety of applications, including language translation, question-answering, text summarization, and chatbot interactions.

What is ChatGPT?

ChatGPT is a large language model created by OpenAI. It is based on the GPT (Generative Pre-trained Transformer) architecture, specifically the GPT-3.5 variant. ChatGPT is designed to generate human-like text and to respond to natural language inputs. It has been trained on a massive amount of text data from the internet and can understand and generate text in a wide range of languages and topics. ChatGPT can be used for a variety of applications, including language translation, question-answering, text summarization, and chatbot interactions.

What does creating a front end GUI unlock in terms of potential? Just think... all the power of chatGPT but your own personalisation as a company brand. You can start to power services / chatbots / online automation whilst including all the design flexibility of the current site architecture. How do we go about creating it? Well first you will need an account (obviously) Sign up to [OpenAI](#) using your email and make a password. When you sign up you should get \$5 of credit. I think it used to be \$15 but recently reduced. This is a free trial period when it comes to using the API. We DO NOT need to add any billing to go through this tutorial. (Only if you want to go on to use additional features or run out of user credits would you need this)

platform.openai.com/account/usage

Overview Documentation API reference Examples Playground Upgrade Help Personal

Settings Usage Rate limits Members Billing

USER API keys

Below you'll find a summary of API usage for your organization. All dates and times are UTC-based, and data may be delayed up to 5 minutes.

Daily usage (USD)

DAILY CUMULATIVE

Free trial usage

\$0.01 / \$5.00

GRANT #	CREDIT GRANTED	EXPIRES (UTC)
Grant 1	\$5.00	1 August 2023

You can see your credits under Usage. Go down to user and click API Keys. Create a new key.

platform.openai.com/account/api-keys

Overview Documentation API reference Examples Playground Upgrade Help Personal

ORGANIZATION Personal Settings Usage Rate limits Members Billing

USER API keys

API keys

Your secret API keys are listed below. Please note that we do not display your secret API keys again after you generate them.

Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically rotate any API key that we've found has leaked publicly.

NAME	KEY	CREATED	LAST USED
My Test Key	sk-...KRbD	30 Apr 2023	30 Apr 2023

+ Create new secret key

Default organization

If you belong to multiple organizations, this setting controls which organization is used by default when making requests with the API keys above.

Personal

Note: You can also specify which organization to use for each API request. See [Authentication](#) to learn more.

You can see I've created a new "Test Key" This key will be added to our code to interact with ChatGPT. Now we have our key, let's look at getting set up. Open up an IDE of your choice, for me I use PyCharm. You will need to pip install openai and gradio packages in your terminal.

```
Terminal: Local + ▾
(PycharmProjects) christophers-MacBook-Pro:ChatGPT christophermayes$ pip install openai
Collecting openai
  Using cached openai-0.27.5-py3-none-any.whl (71 kB)
Collecting aiohttp
  Using cached aiohttp-3.8.4-cp38-cp38-macosx_10_9_x86_64.whl (359 kB)
Collecting tqdm
  Using cached tqdm-4.65.0-py3-none-any.whl (77 kB)
Collecting requests>=2.20
  Using cached requests-2.29.0-py3-none-any.whl (62 kB)
Requirement already satisfied: certifi>=2017.4.17 in /Library/Frameworks/Python.framework/Versions/3.8/lib/python3.8/site-packages (from requests>=2.20-)

Terminal: Local + ▾
(PycharmProjects) christophers-MacBook-Pro:ChatGPT christophermayes$ pip install gradio
Collecting gradio
  Downloading gradio-3.28.1-py3-none-any.whl (17.3 MB)
    |████████████████████████████████| 17.3 MB 2.7 MB/s
Collecting pydub
  Downloading pydub-0.25.1-py2.py3-none-any.whl (32 kB)
Collecting altair>=4.2.0
  Downloading altair-4.2.2-py3-none-any.whl (813 kB)
    |████████████████████████████████| 813 kB 2.5 MB/s
Requirement already satisfied: requests in /Users/christophermayes/PycharmProjects/lib/python3.8/site-packages (from gradio) (2.29.0)
```

Now we have the packages we can paste our code into a main.py file.



```
import openai
import gradio as gr
openai.api_key = "redacted"

messages = [{"role": "system", "content": "Test"}]

def CustomChatGPT(user_input):
    messages.append({"role": "user", "content": user_input})
    response = openai.ChatCompletion.create(
        model = "gpt-3.5-turbo",
        messages = messages
    )
    ChatGPT_reply = response["choices"][0]["message"]["content"]
    messages.append({"role": "assistant", "content": ChatGPT_reply})
    return ChatGPT_reply

demo = gr.Interface(
    fn=CustomChatGPT,
    inputs="text",
    outputs="text",
    title ="CJ - Demo"
)
demo.launch(share=False)
```

What does the code do. Well, of course we replace the word redacted with your own API key. From there you can run the code.

```

messages = [{"role": "system", "content": "Test"}]

def CustomChatGPT(user_input):
    messages.append({"role": "user", "content": user_input})
    response = openai.ChatCompletion.create(
        model = "gpt-3.5-turbo",
        messages = messages
    )
    ChatGPT_reply = response["choices"][0]["message"]["content"]
    messages.append({"role": "assistant", "content": ChatGPT_reply})
    return ChatGPT_reply

demo = gr.Interface(
    fn=CustomChatGPT,
    inputs="text",
    outputs="text",
    title ="CJ - Demo"
)
demo.launch(share=False)

```

For reference, it creates a function using the ChatGPT 3.5 turbo. Gradio is our package that creates the front end. You will see that you can pass various things through it, including a title, and various inputs and outputs. When we click run you will see the local host load.

CJ - Demo

user_input

How can I get more involved in the Tableau community?

Clear Submit

output

There are a few ways to get more involved in the Tableau community:

- Join the Tableau Community: The Tableau Community is a great place to get started with Tableau. You can access it through the Tableau website, sign up for weekly newsletters, and participate in forums to ask and answer questions.
- Attend Tableau User Groups: Tableau User Groups are local networking events that provide an opportunity to meet other Tableau users in your area. These groups host events to share tips and tricks, showcase user projects, and promote collaboration.
- Participate in Tableau Public: Tableau Public is a great way to share your Tableau work with others. You can publish your dashboards and visualizations to Tableau Public and connect with other community members. You can also find inspiration by exploring amazing work created by others in the Tableau Public Gallery.
- Attend Tableau conferences: Tableau hosts multiple conferences throughout the year, which are great opportunities to meet other Tableau users and enthusiasts. At these events, you can attend sessions and workshops, network with your peers, and learn new techniques from experts in the field.

I hope these ideas help you get started with getting more involved in the Tableau community!

Flag

Here is what that looks like. In the above example it creates an input, and then to the right will have the output. I've put in a draft question asking how to get more involved in the Tableau community? If we click flag – it will log the response in a flagged folder.

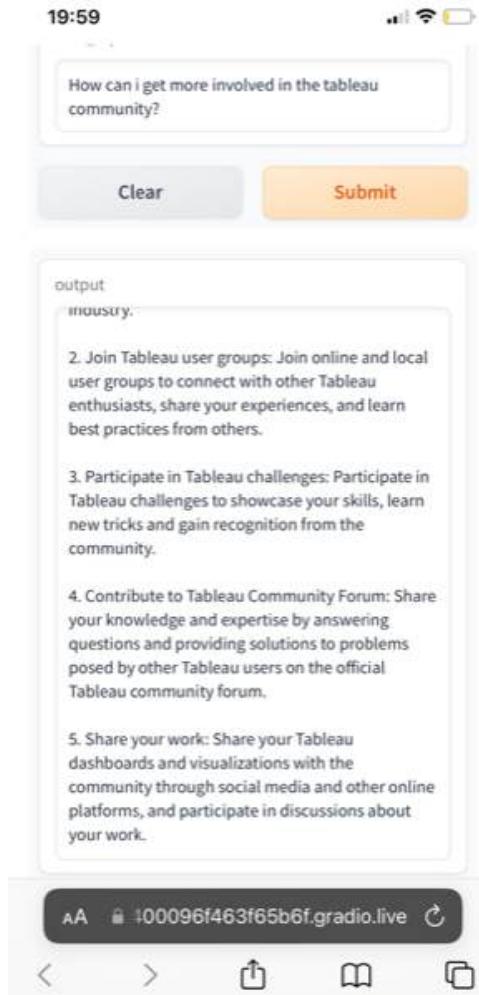
```

user_input,output,flag,username,timestamp
How can I get more involved in the Tableau community?,"There are a few ways to get more involved in the Tableau community:
1. Join the Tableau Community: The Tableau Community is a great place to get started with Tableau. You can access it through the Tableau website, sign up for weekly newsletters, and participate in forums to ask and answer questions.
2. Attend Tableau User Groups: Tableau User Groups are local networking events that provide an opportunity to meet other Tableau users in your area. These groups host events to share tips and tricks, showcase user projects, and promote collaboration.
3. Participate in Tableau Public: Tableau Public is a great way to share your Tableau work with others. You can publish your dashboards and visualizations to Tableau Public and connect with other community members. You can also find inspiration by exploring amazing work created by others in the Tableau Public Gallery.
4. Attend Tableau conferences: Tableau hosts multiple conferences throughout the year, which are great opportunities to meet other Tableau users and enthusiasts. At these events, you can attend sessions and workshops, network with your peers, and learn new techniques from experts in the field.
I hope these ideas help you get started with getting more involved in the Tableau community!","",2023-04-30 20:32:06.228

```

You will see our input, the output response from chatGPT and the time that the query was sent. Hopefully this sparks some

creativity in your own journey. A quick bonus, why not amend share to True Now we can load the site url on the phone, we are



no longer using a local host.

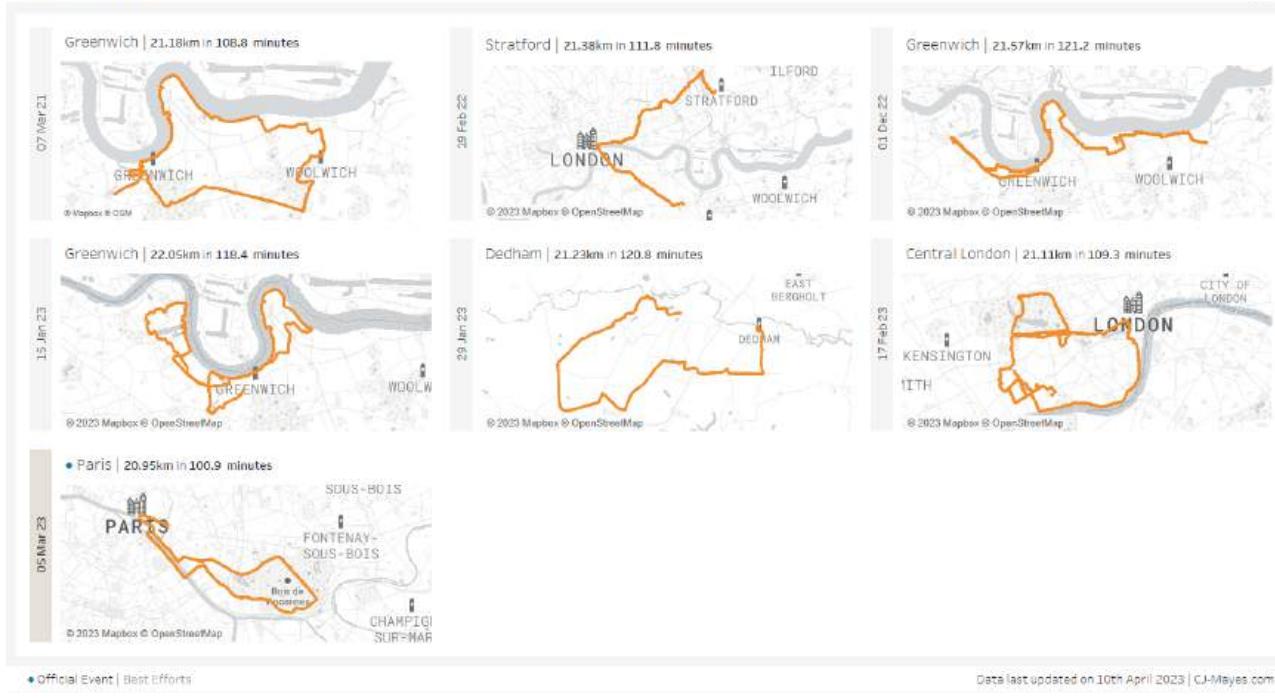
Going further:

- Try adding in some more elements to the Gradio to improve userability
- Why not create the interface using CSS?
- Try hosting the content live instead of locally
- Try making the questions unique to a specific subject content matter

LOGGING OFF, CJ

STRAVA MAPS FOR BEGINNERS

Hi all, At the start of this year I wrote a blog on how to access Strava data through the API. You can view that blog [here](#). This time round we will take the short route and download all our data from Strava itself manually, with the end outcome to look at how we can actually map our data in Tableau.



So to download the data you will need to navigate to "[My Account](#)"

- [My Profile](#)
- [My Account](#)
- [My Performance](#)
- [Display Preferences](#)
- [Privacy Controls](#)
- [Data Permissions](#)
- [Email Notifications](#)
- [My Gear](#)
- [My Apps](#)
- [Partner Integrations](#)
- [My Badges](#)
- [My API Application](#)

Membership

Member Since: January 7, 2020

Membership Status: **Free User**

[Start Free Trial](#)

Subscribe to access premium features

- Personalized Route Suggestions
- Advanced Workout Analysis
- Segment Leaderboards

ALSO INCLUDED:

- Recover Athletics
Prehab and Injury Prevention Workouts
- FATMAP **NEW**
3D Maps and Guides for the Outdoors

[See All Premium Features](#)

Download or Delete Your Account

Here's where you can get an archive of your Strava data for your records, or begin the process of deleting your account. [Learn more](#)

[Get Started](#)

Click to Get started on download your stats. Once you go to download them, it will send you a link to your email associated with the account

activities	✓	10/04/2023 11:24	File folder
clubs	✓	10/04/2023 11:24	File folder
media	✓	10/04/2023 11:24	File folder
routes	✓	10/04/2023 08:48	File folder
activities.csv	✓	10/04/2023 11:24	Microsoft Excel C... 278 KB
applications.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
bikes.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
blocks.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
categories_of_personal_information_we_...	✓	10/04/2023 11:24	Adobe Acrobat D... 52 KB
clubs.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
comments.csv	✓	10/04/2023 11:24	Microsoft Excel C... 26 KB
components.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
connected_apps.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
contacts.csv	✓	10/04/2023 11:24	Microsoft Excel C... 33 KB
email_preferences.csv	✓	10/04/2023 11:24	Microsoft Excel C... 2 KB
events.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
favorites.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
flags.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
followers.csv	✓	10/04/2023 11:24	Microsoft Excel C... 4 KB
following.csv	✓	10/04/2023 11:24	Microsoft Excel C... 4 KB
general_preferences.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
global_challenges.csv	✓	10/04/2023 11:24	Microsoft Excel C... 17 KB
goals.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
group_challenges.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
information_we_disclose_for_a_business_...	✓	10/04/2023 11:24	Adobe Acrobat D... 73 KB
local_legend_segments.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
logins.csv	✓	10/04/2023 11:24	Microsoft Excel C... 2 KB
media.csv	✓	10/04/2023 11:24	Microsoft Excel C... 12 KB
memberships.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
metering.csv	✓	10/04/2023 11:24	Microsoft Excel C... 128 KB
mobile_device_identifiers.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
orders.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
partner_opt_outs.csv	✓	10/04/2023 11:24	Microsoft Excel C... 2 KB
posts.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
privacy_zones.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
profile.csv	✓	10/04/2023 11:24	Microsoft Excel C... 1 KB
profile.jpg	✓	10/04/2023 11:24	JPG File 24 KB
reactions.csv	✓	10/04/2023 11:24	Microsoft Excel C... 121 KB

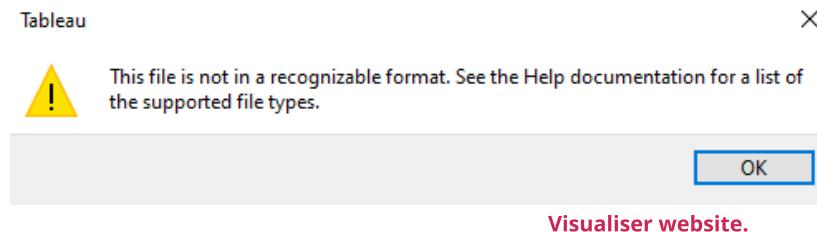
For today we will only need to look at activities and the activities folder.

Within the activities.csv you will see an associated gpx file. These files are essentially the mappings for our different activities.

2989833057.gpx	10/04/2023 11:24	GPX File	107 KB
3003632713.gpx	10/04/2023 11:24	GPX File	93 KB
3010269791.gpx	10/04/2023 11:24	GPX File	127 KB
3050395336.gpx	10/04/2023 11:24	GPX File	138 KB
3070693531.gpx	10/04/2023 11:24	GPX File	120 KB
3090440097.gpx	10/04/2023 11:24	GPX File	104 KB
3110426251.gpx	10/04/2023 11:24	GPX File	118 KB
3130586991.gpx	10/04/2023 11:24	GPX File	223 KB
3151109678.gpx	10/04/2023 11:24	GPX File	246 KB
3209712232.gpx	10/04/2023 11:24	GPX File	239 KB
3212677120.gpx	10/04/2023 11:24	GPX File	206 KB
3219783138.gpx	10/04/2023 11:24	GPX File	253 KB
3234246540.gpx	10/04/2023 11:24	GPX File	100 KB
3241629424.gpx	10/04/2023 11:24	GPX File	134 KB
3252466640.gpx	10/04/2023 11:24	GPX File	361 KB
3256995695.gpx	10/04/2023 11:24	GPX File	123 KB
3257140484.gpx	10/04/2023 11:24	GPX File	86 KB
3266098044.gpx	10/04/2023 11:24	GPX File	419 KB
3270484548.gpx	10/04/2023 11:24	GPX File	111 KB
3274964275.gpx	10/04/2023 11:24	GPX File	163 KB
3284054261.gpx	10/04/2023 11:24	GPX File	731 KB
3296972879.gpx	10/04/2023 11:24	GPX File	257 KB
3307769727.gpx	10/04/2023 11:24	GPX File	483 KB
3312495731.gpx	10/04/2023 11:24	GPX File	458 KB
3329901744.gpx	10/04/2023 11:24	GPX File	102 KB
3379095210.gpx	10/04/2023 11:24	GPX File	32 KB
3405802731.gpx	10/04/2023 11:24	GPX File	67 KB
3417103983.gpx	10/04/2023 11:24	GPX File	159 KB
3428012328.gpx	10/04/2023 11:24	GPX File	104 KB
3448363481.gpx	10/04/2023 11:24	GPX File	402 KB
3464051804.gpx	10/04/2023 11:24	GPX File	115 KB
3475138379.gpx	10/04/2023 11:24	GPX File	98 KB
3487702528.gpx	10/04/2023 11:24	GPX File	119 KB
3540167750.gpx	10/04/2023 11:24	GPX File	98 KB
3552943056.gpx	10/04/2023 11:24	GPX File	137 KB
3566437977.gpx	10/04/2023 11:24	GPX File	71 KB
3570812732.gpx	10/04/2023 11:24	GPX File	118 KB
3576196503.gpx	10/04/2023 11:24	GPX File	428 KB

So

here is some example gpx files that are associated to my account. Of course we will have some common fields to be able to join these together for example the date of the event as well as the activity ID of the gpx files. The only problem is.... Tableau doesn't allow for gpx files, so we need to convert them into records of data.



One website that is quite popular is this [GPS Visualiser website](#).

Convert a GPS file to plain text or GPX

This form reads a tracklog or waypoint file (in a recognized format) or plain-text tabular data, and

- **Addresses:** If you want to find the coordinates of a list of street addresses, it may be easier to use the [Google Earth KML](#) form.
- **Google Earth:** If you want to generate a KML or KMZ file for Google Earth, use the [Google Earth KML](#) form.
- **Leaflet/Google Maps:** To generate an HTML map, use the [Leaflet](#) or [Google Maps](#) form.
- **Non-compatible formats:** If this conversion program cannot read your file, it's possible that

Output format: Plain text GPX [Google Earth KML](#)

Upload your files here: (10 MB max. total size, .zip/.gz is supported)

File #1	Choose files	4904533511.gpx	X	Convert
File #2	Choose files	6705160233.gpx	X	
File #3	Choose files	8193216201.gpx	X	
File #4	Choose files	8396510684.gpx	X	
File #5	Choose files	8471353658.gpx	X	
File #6	Choose files	8575489308.gpx	X	
File #7	Choose files	8663587823.gpx	X	
File #8	Choose files	No file chosen	X	
File #9	Choose files	No file chosen	X	
File #10	Choose files	No file chosen	X	

[Show fewer file boxes](#)

Or paste your data here:

`name,desc,latitude,longitude`

Force text data to be this type:

Or provide the URL of a file on the Web:

Plain text delimiter: Plain text output units:

Add estimated fields: speed heading slope (%) distance VMG pace

Add DEM elevation data:

[Save these settings](#) • [Load from saved](#)

We will need to output these files to plain text. I will upload an example of the output into the GitRepo. You'll see in the printscreens above, I chose all the gpx files of where I ran at least 20km's. In the GitRepo here are the files: Strava_Workflow.ymd – an Alteryx flow that glues to activities to the output from the GPX conversion. Strava_Maps.hyper the final dataset that I use within the visual data.txt – The converted GPX data CJ_Activities – My original csv activities dataset 8663587823.gpx – the original gpx file for my Paris half mara Then all we need to do is glue our gpx file to our original dataset. Personally I used Alteryx to do this, however this could be done solely through the joins in Tableau.

Makepoint Role:

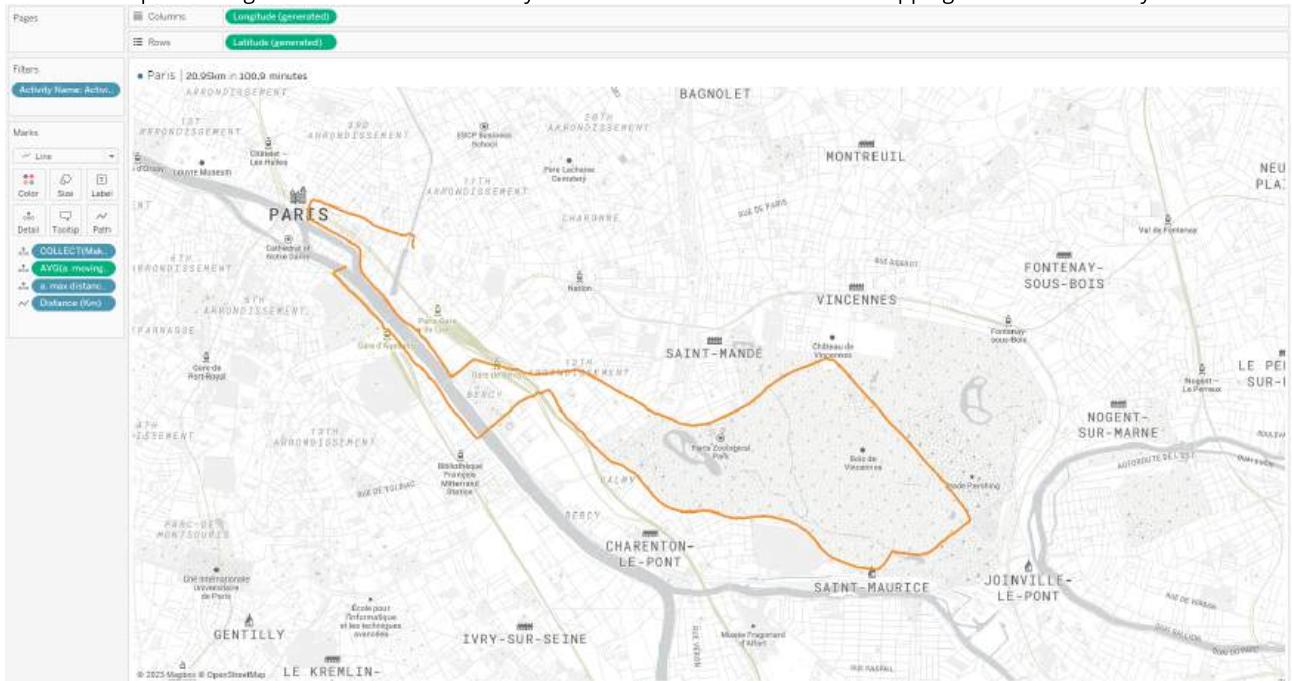
`MAKEPOINT(float([Latitude]),float([Longitude]))|`

The calculation is valid.

8 Dependencies

To create the

map, I like to use the lat and long co-ordinates from the exported version of the gpx data that is now a text file – I added a few extra calculations for the view to be able to see the distance travelled and total time taken. See how my watch cut me short 60m on my Paris run! Sigh. Final thing to do is purely cosmetics. I add the distance km to line path as it joins up the data for each of the distance logs, and amended it to be our Strava orange colour. I dialled the background colour back just using an out the box mapbox background but there is definitely a lot more we can do with these mappings in terms of the layers!



Would love to see where people take it. By all means check out my tutorial dashboard using the TP link at the top of this blog.

[Strava Metrics](#) | Half Marathon Distances



Going further:

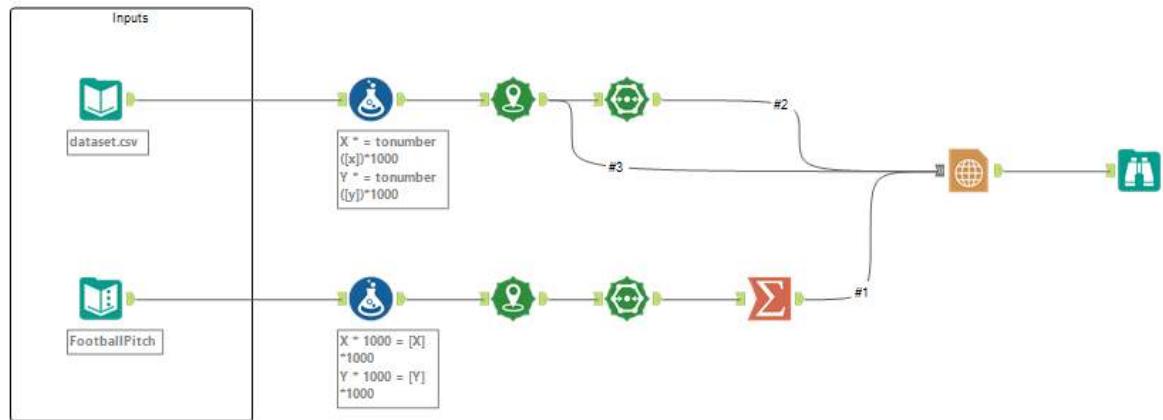
- Add a different MapBox background to your visual
- Chart all your different activities that have maps
- Using alternative map projections in Tableau: [blog](#) by [Ken Flerlage](#)
- Mapping your Strava app data in Tableau: [blog](#) by [Andy Cotreave](#) & [Andy Kriebel](#)

LOGGING OFF, CJ

CONVEX HULLS / TERRITORY MAPS IN ALTERYX

Hi all, A couple on the bounce now in terms of football/soccer content in Alteryx. Truth be told it's been a fun little series creating a few of the same charts now in Python, Tableau and now Alteryx. Hopefully this unlocks doors for people in the industry in

terms of the different tools they can utilise. I've previously written about convex hulls, [here](#). In fact it was literally a year ago now. The walk through explains what the convex hull is, what it is useful for and how to retrieve the Statsbomb data using python and then visualising it in Tableau. Well to make life simple we will take our original dataset and add it into Alteryx. Then recreate the same chart as seen in the tutorial. As always, all material can be found in the Github repository at the top of the page.



What does our workflow do? Well the initial input at the top of the flow is our data for Crystal Alyssia Dunn Soubrier from the final of the Womens World Cup for 2020/21. The data is from Statsbomb so we don't need to do anything in terms of transforming our data points. You can access my original code [here](#). We do however scale the points in equal size to make it easier to plot on a chart. The next thing we do is use the create points tool to plot these as x and y co-ordinates. The top part of the flow will be the co-ordinates of each of Crystal's passes.

Poly-Build (87) - Configuration

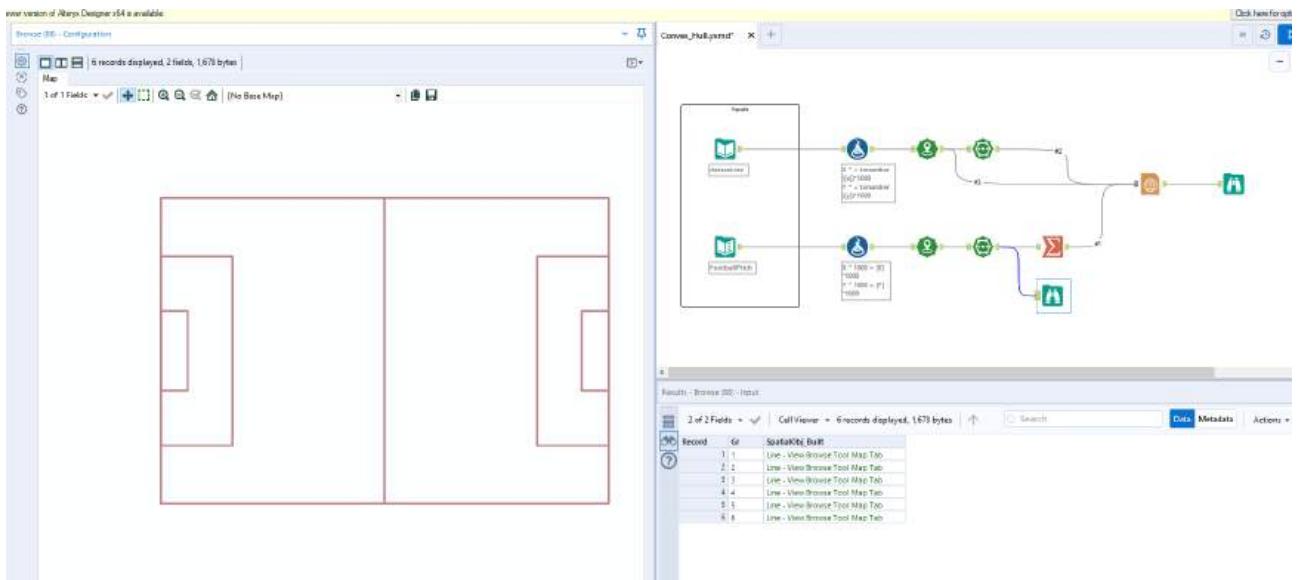
Build Method:
 Convex Hull

Source Field:
 Centroid

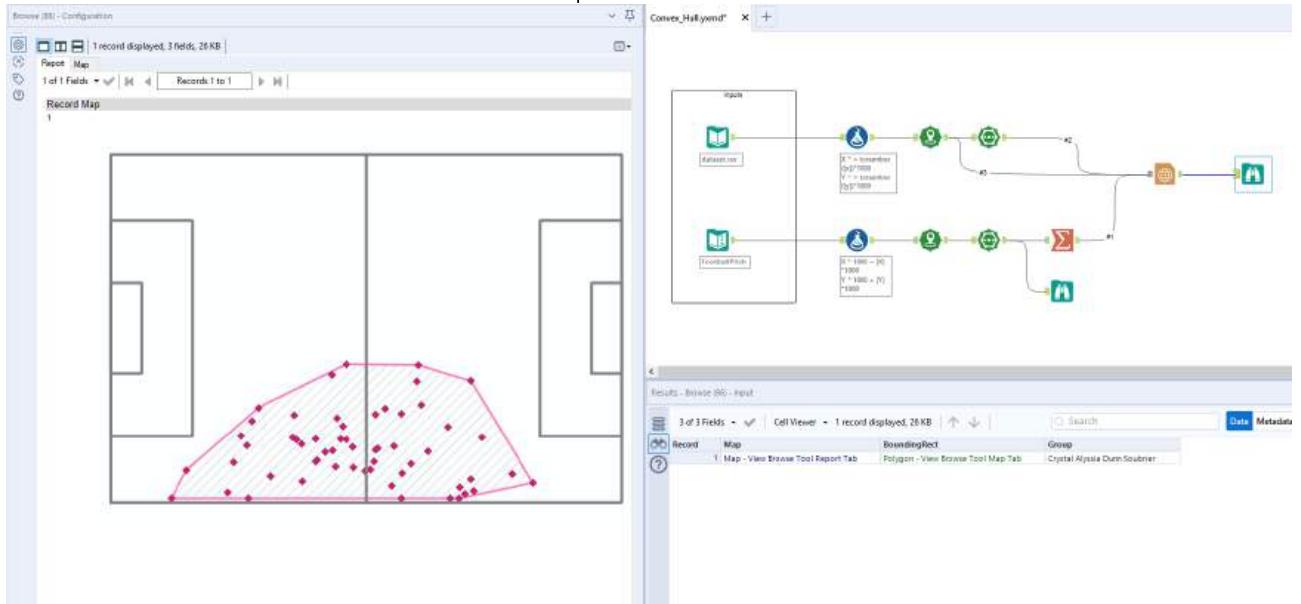
Group Field:
 player

Sequence Field:

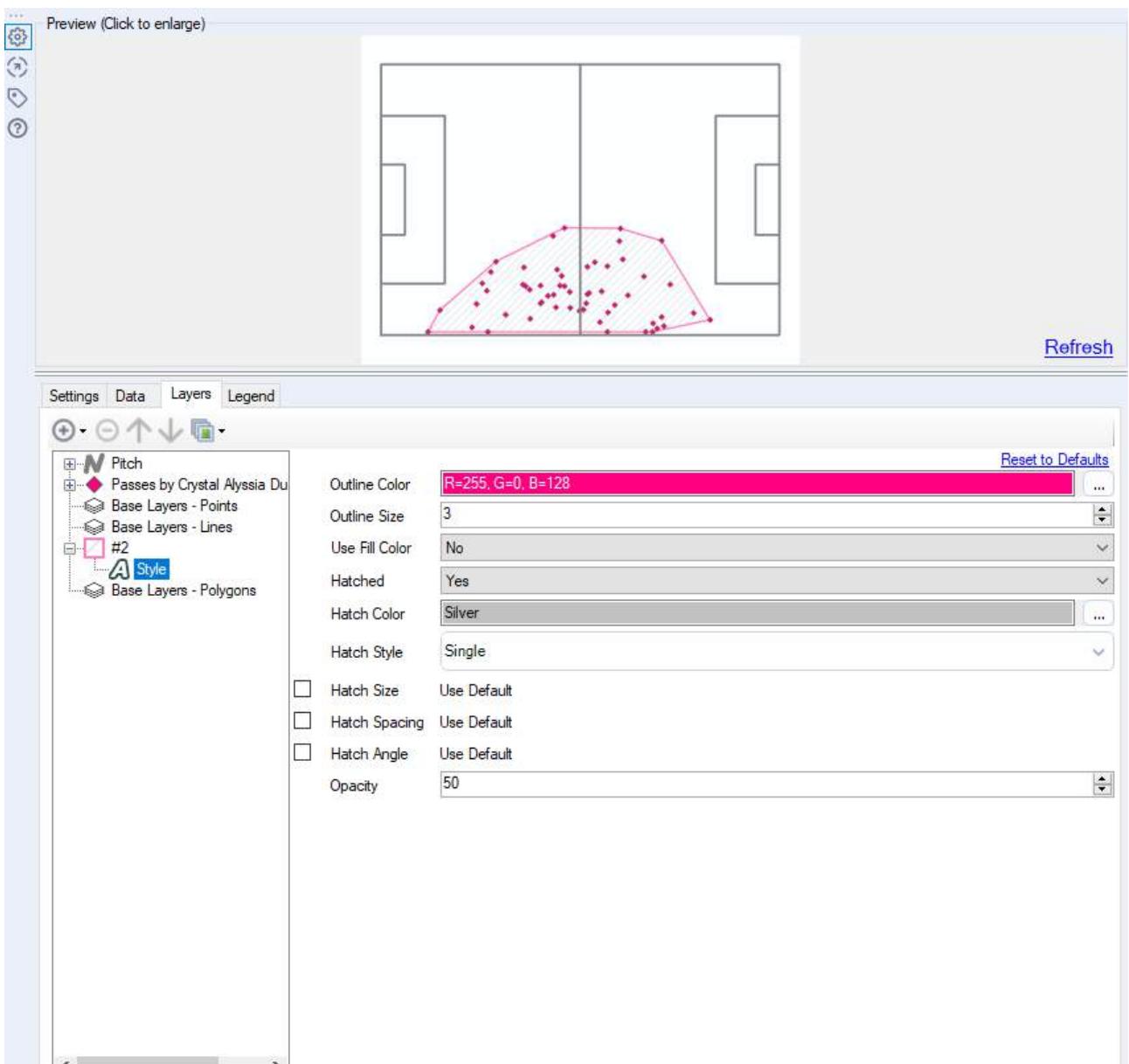

The next part of the flow uses the convex hull function based on these x and y points. In terms of grouping we group on player, this would become more useful if our dataset contained more than one player. In terms of the bottom half of the workflow, it is creating our baseline pitch, again also using the plot points tool.



The final thing to do is to use the report map feature to be able to showcase this visually. #1 is our football pitch which we configure. #2 is our convex hull. #3 is the original points, I also joined this into the final output so we can see all the passes even if it isn't part of the convex hull!



Normally I've found the design quite challenging in Alteryx, but I just LOVE that you can add a hatch into the design as well as play around with base colours and opacity in this chart.



Short blog this week, hope those who use Alteryx have enjoyed the mini series. Don't forget if you want to check out my Tableau version of this chart you can find the tutorial [here](#). Wishing those off to Tableau Conference in Vegas this week the best of times. I'm going to get my head down personally and steer clear of the social channels for a week or so because of FOMO but I hope it's everything and more for those attending! Take care and chat soon. **LOGGING OFF, CJ**

EXPANDING YOUR TOOLKIT WITH WILL SUTTON

Hi all, This week we have something really quite special. [Will Sutton](#) & I sat down to talk all about supplementary skills that will elevate you on your data journey. I think I speak for both Will & I that we may be more 'known' for Tableau related concepts, but

we both were developers at one time! For those needing the introduction – Will is a Tableau visionary, won the IronViz competition last year and will be judging this year! So if you are out in Vegas in May at TC conference do make time to say hi. Onto the chat today – lets start with a big one. **Q. Recently we've seen some huge advancements in tech, in particular with AI where chatGPT can write code after a short prompt, is it still worth learning to code?** W: Is the juice worth the squeeze?

When I began my career with data I knew Excel and was just learning Tableau. Regularly I avoided anything that would involve learning to code. I had these great no-code tools, why should I spend hours learning a new piece of tech when I already can do something similar with the tools I already have? As much as I tried to avoid coding, one day I had to face it. The call came in "Oh hey Will, Ricky's put together this awesome script in R to pull in data from the US Securities Exchange but there's an issue with it – could you take a look at it please." So I was sent this R file and asked to get it working. I download R and R Studio and was plunged into the deep end. When they talk about coding as a language they aren't joking, I looked over this script and it might as well have been written in a foreign dialect. I knew I couldn't figure it out immediately but could I get some of it to work?

Line by line I went through this script, I looked up everything, Stack Overflow, tutorials, forums, it took an age but eventually I worked out what the script was doing and fixed the error. It was a slow process but one I don't forget because it opened my eyes to what could be achieved with coding. I clicked a button and suddenly I had a dataset ready to bring into Tableau. A task that would have normally taken me hours was done in a few minutes. Amazing! I was forced out of my comfort zone and it expanded my skillset as an analyst. I was used to visualising and analysing data but now I could automate and transform new data sets

too. While chatGPT is an awesome tool I don't think the need for coding skills is going away anytime soon. We will still need people to build and implement this technology in the future. If anything chatGPT has reduced the barriers to learning coding. It's like having an on-demand expert you can call up any time day or night to ask questions. CJ: I love that idea of considering it reducing barriers to entry! It's funny how we also have a spectrum of code to no code, and tools also exist somewhere in between - your likes of Alteryx & Tableau prep, which I would say is like a front end GUI to coding. ChatGPT also will unlock everything along that spectrum. **CJ: One of the stumbling blocks is knowing which language to start with, what do you recommend for this?**

The screenshot shows the homepage of the 8 Week SQL Challenge. At the top, there's a navigation bar with links for "Data With Danny", "Case Studies", "Getting Started", "Resources", and "About". A logo for "DATA WITH DANNY" is on the left. Below the navigation, the title "Case Study #1 - Danny's Diner" is displayed, followed by the date "May 1, 2021". The main content area features a large graphic for "CASE STUDY #1" with the "DANNY'S DINER" logo, which is a stylized fork and knife inside a bowl. The text "THE TASTE OF SUCCESS" is at the bottom of the graphic. The footer contains the URL "DATAWITHDANNY.COM".

Oh absolutely, the debate of which is better Python or R still goes on. It's a question worth asking too, as you want to invest your time wisely and see a return for your efforts. Hiring managers will tell you the language shouldn't matter but what you can do with it. This is true in part but doesn't consider integrating this person into a team. For example, I worked alongside a very skilled developer that coded applications in Delphi and could build great systems but no one else on the team knows Delphi, so sharing code and knowledge becomes a challenge. Pick a language that's in regular use for your role or the role you want. For data analysts generally, it's a choice of SQL, Python and R. Aim to develop skills in one of these three. Additional languages are easier to pick up when you have some working knowledge of coding. For deciding on the right coding language consider the people around you. It's hard to learn a new language without support. I developed skills in R at the BBC purely because it was a commonly used language within the team, I would have the support of colleagues that had attempted problems before and the code they had produced I could work from. SQL skills I found incredibly valuable when working with Tableau when pulling data from databases but it's difficult to gain these skills outside of a work environment. After teaching SQL courses I recommend [Danny Ma's 8 Week SQL challenge](#), it's all about understanding the functionality in SQL and solving problems they'll face at work. [I've put together solutions and tutorials for the first few challenges on GitHub](#).

The screenshot shows a GitHub repository page for "wjsutton / data_with_danny_8_week_sql_challenge". The repository has 15 commits, 4 stars, and 1 watch. The README.md file contains the following text:

```

# Data with Danny's 8 Week SQL Challenge

Solving Danny Ma's 8-Week SQL Challenge with SQL in Snowflake 🌐

[about] [issues] [pull requests] [code] [license] [net]

Twitter | LinkedIn | GitHub | YouTube

## About

Learning SQL helped shape and develop my career as a data analyst. Danny Ma has put together a series of SQL challenges that will improve your SQL problem-solving skills. This repo shares my solutions to the tasks. Please note there are multiple ways of answering these questions.

For these challenges I have loaded the data into Snowflake to get more familiar with that platform, some of my solutions may not work across different SQL platforms due to differences in function naming and availability.

```

What would be a good starting point for picking up coding? Direction is often a challenge. There are so many paths to take with coding. I recall [Ann Jackson](#) speaking about having "T-shaped skills" where you can have a particular strength, e.g. Data

visualisation, and then add a skill to broaden into a new area or improve efficiency, which is where coding comes in. For example, I can reduce my time spent analysing data by running my dataset through a Python package called "["ydata-profiling"](#)" to return an overview report of the data I'm working with. In the example below I take a dataset of Call of Duty Players and generate a html report providing exploratory data analysis. I especially like the correlation plots at the end of the report.

```
import pandas as pd
from ydata_profiling import ProfileReport

# Load csv
df =
pd.read_csv('https://raw.githubusercontent.com/wjsutton/games_night_viz/main/chall-
enges/5_montages/cod_mw_players.csv')

# ProfileReport will build a report of your dataset and write it as an html file
profile = ProfileReport(df, title="Call of Duty Modern Warfare Players",
explorative=True)
profile.to_file("cod_report.html")
```

Or I could use code to build a dataset for me. I was having issues with my home broadband speeds and wanted to check if it was a one-off or an issue I needed fixing. So I coded up a [script in R](#) using the package "speedtest" that would check my broadband speed and write it to a file.

```
library(speedtest)

report_time <- Sys.time()

config <- spd_config()

servers <- spd_servers(config = config)

servers <- spd_closest_servers(servers, config)

best <- spd_best_servers(servers, config, max = 3)

down <- spd_download_test(best, config, FALSE, timeout = 5)

up <- spd_upload_test(best, config, FALSE, timeout = 10)

entry <- data.frame(test_time=report_time

                      ,download_speed=max(down$bw)

                      ,upload_speed=max(up$bw)

                      ,stringsAsFactors = FALSE)

if(file.exists("internet_speed_results.csv")){

  cat(paste0(Sys.time()," Appending results to existing file...\n"))

  all_results <- read.csv("internet_speed_results.csv",stringsAsFactors = FALSE)

  all_results$test_time <- as.POSIXct(all_results$test_time)

  all_results <- rbind(all_results,entry)

  write.csv(all_results,"internet_speed_results.csv",row.names = FALSE)

}

if(!file.exists("internet_speed_results.csv")){

  cat(paste0(Sys.time()," Writing results to new file...\n"))

  write.csv(entry,"internet_speed_results.csv",row.names = FALSE)

}
```

I then set the script to run every hour on my PC using Windows Task Scheduler and visualised the results in Tableau. This is a nice double win for the portfolio too, as you can talk about the code and [the visualisation](#) to illustrate a range of skills in one go.

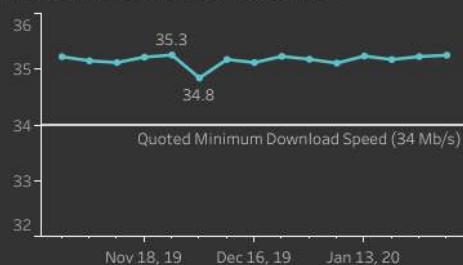
Am I Getting The Home Broadband Speed I Paid For?

We are becoming more dependant on broadband to stay in touch with the rest world, and with this reliable broadband speeds are becoming increasingly important.

To monitor my broadband at home I've coded a script (https://github.com/wjsutton/internet_speed_test) to check and record my internet speed scheduled to run every 15 minutes.

My provider, Plusnet, have quoted the minimum download speed for my area as 34 Mb/s.

AVERAGE WEEKLY DOWNLOAD SPEED (Mb/s)
Click a point to filter the charts below by week.



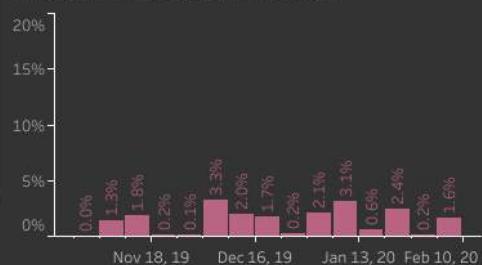
VERDICT

YES

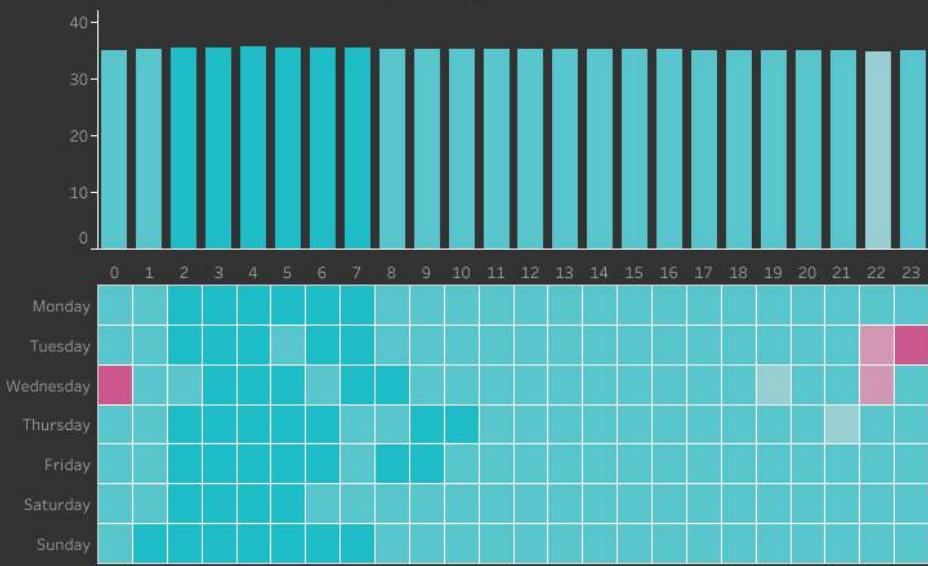
34.8 Mb/s
the worst average weekly download speed

Download Speed (Mb/s) Legend
32 [red] 36 [blue]

WEEKLY PERCENTAGE OF READINGS BELOW 34 Mb/s
Click a bar to filter the charts below by week.



DOWNLOAD SPEED READINGS BY HOUR AND DAY, 10/28/2019 - 2/9/2020



Author: Will Sutton | Source: 15 minute readings of Broadband Speed | Last updated: 9 Feb 2020

Internet Provider: Plusnet | Script: https://github.com/wjsutton/internet_speed_test

KEY TAKEAWAYS

- 99% of readings are above the minimum quoted download speed of 34 Mb/s
- Best time for download speeds is Monday 4:00 with speeds of 35.7 MB/s
- Worst time for download speeds is Wednesday 0:00 with speeds of 32.8 MB/s

CJ: Could coding help with the visual elements with Tableau? If you're familiar with polygons or lines in Tableau they allow you to draw shapes by giving x, y coordinates and a path. Knowing this you can code up a script to generate these points for you which is what I did in 2021 with my [Snowflake viz](#).

GEOMETRIC SNOWFLAKES

Snowflakes are tiny flakes of snow, usually only millimetres in size, but incredibly beautiful and iconic for the winter season.

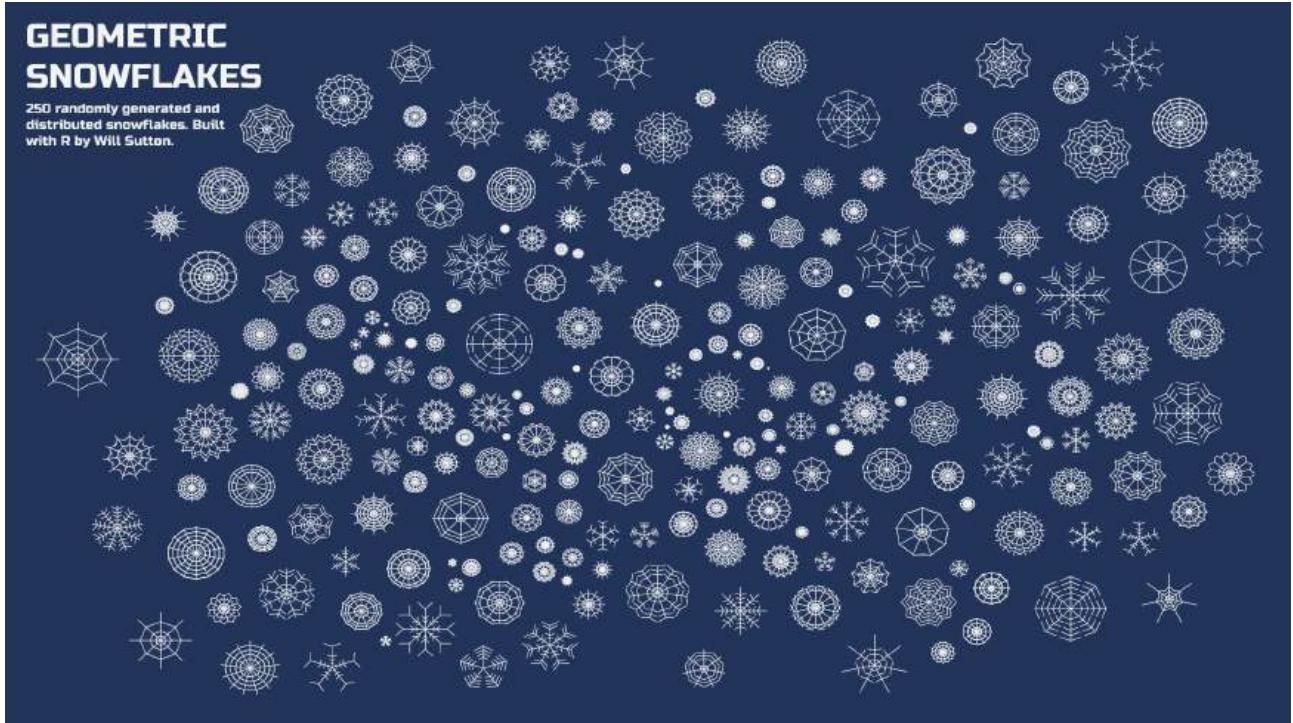
Each snowflake is unique and has six-fold symmetry, meaning you can rotate the snowflake 60° without changing its shape.

Using R and a random number generator I have built 100 snowflakes shown below. You can read more about the project on GitHub →



Author: Will Sutton | Source: Github [wjsutton/snowflake_generator](https://github.com/wjsutton/snowflake_generator)

The [script](#) itself draws lines of random length and angles to create snowflake shapes, which I've then distributed in Tableau. In 2022 I wanted to distribute the snowflakes in a more random arrangement. I found the packed circles layout to be just what I wanted if I could replace the circles with snowflakes. After some work, I was able to hack the co-ordinates out of the `packdbubbles` library in R and distribute my snowflakes in [Tableau](#).



And there are plenty of other folks using code to push Tableau further, just look at the [Advanced Visualization Generator](#) [Tristan Guillevin](#) has put together or some of the vizes from [Nick Gerend](#) CJ: What I enjoy about Tableau Public is that each piece of work goes to creating your own portfolio which is valuable when showcasing your skills. How would you demonstrate your coding skills?

[Github!](#) It's another tool to learn but very useful for displaying your work and collaborating with others. GitHub is a document technology that will monitor a file folder, called a "repo", for any changes, allowing you to push changes to files in a centralised location, or roll back any modifications. This way teams can be sure they're working with the latest version of a file or script. For the portfolio side, GitHub displays a readme which is a webpage-style document that shows up when someone accesses the code "repo". Typically you can use the readme to explain the code, introduce it to a new user, giving an overview of some of the functions. You can use this for more than just project write-ups however, think about blog posts, challenge solutions, we even used it to start hosting the [#GamesNightViz community project](#).



A

[**profile readme**](#) is a great landing page to show off your work and projects, for example on my page I've used images to highlight certain projects I've written up so it has a similar effect to clicking a viz for Tableau Public. Lastly, GitHub lets you host a website via GitHub pages if you want to branch into web development or would like a more customisable portfolio space. Mine needs a bit of update but this website was developed using R Markdown and Jekyll. Jekyll handles the themes and the blog functionality, Markdown is code that formats the documents, it takes a bit of learning but it's free to host the site and there are some great examples available to start working from.

215 contributions in the last year

Less More More

CJ Round-up: I have to admit I too am a huge fan of Github and we recently copied/stole the idea from Will to move our SportsVizSunday initiative to Github but it allows us to add code, add datasets and have essentially a one stop shop for all things data and tools. Truth be told it also saves us a whole bunch on memory and cost by not having to host it on the website. I want to thank Will for making time in his busy schedule to share some of his journey and tool stack that he has adapted over the years to be able to help people see and understand data. Please do check out Wills [Tableau Public page](#), and his [Github](#) to start your own journey with coding. Catch you all soon. **LOGGING OFF, CJ**

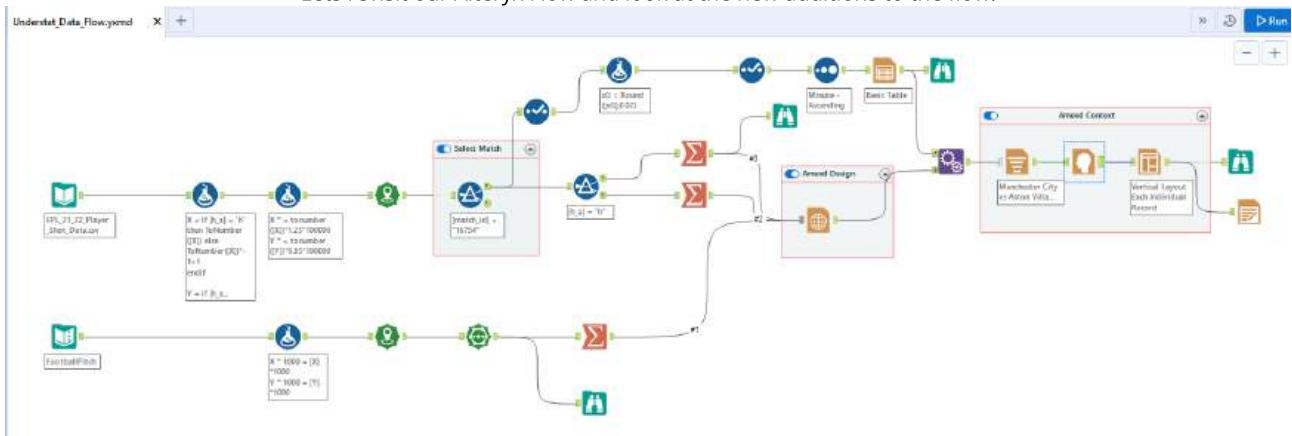
SOCcer Match Reports in Alteryx HTML

Hi all, Last week we looked at an Alteryx challenge that looked at mapping Messi & Ronaldo's goals, ([Shot Maps In Alteryx](#))

Using PolyBuild) Previously I've also blogged about how to obtain shot data through an [understat package](#). Today we will look to blend the two ideas together. I recommend you read both blogs before continuing into this one. This week;

- We will add in last years EPL data
- See if we can plot a shot map for a map, split by team
- Create a report to be exported into HTML format

Lets revisit our Alteryx Flow and look at the new additions to the flow.



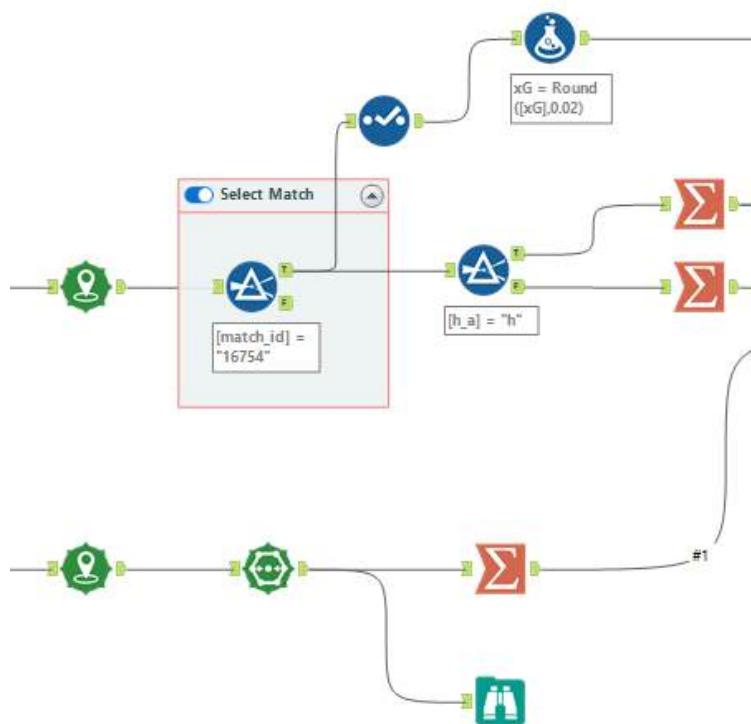
On the left hand side in the top stream. We have our understat data – this is all shots from the Premier League from last year.

Output Column	Data Preview
X	0.905999984741211
fx	If [h_a] = 'h' then ToNumber([X]) else ToNumber([X])*-1+1 endif
Y	0.420999984741211
fx	if [h_a] = 'h' then ToNumber([Y]) else ToNumber([Y])*-1+1 endif

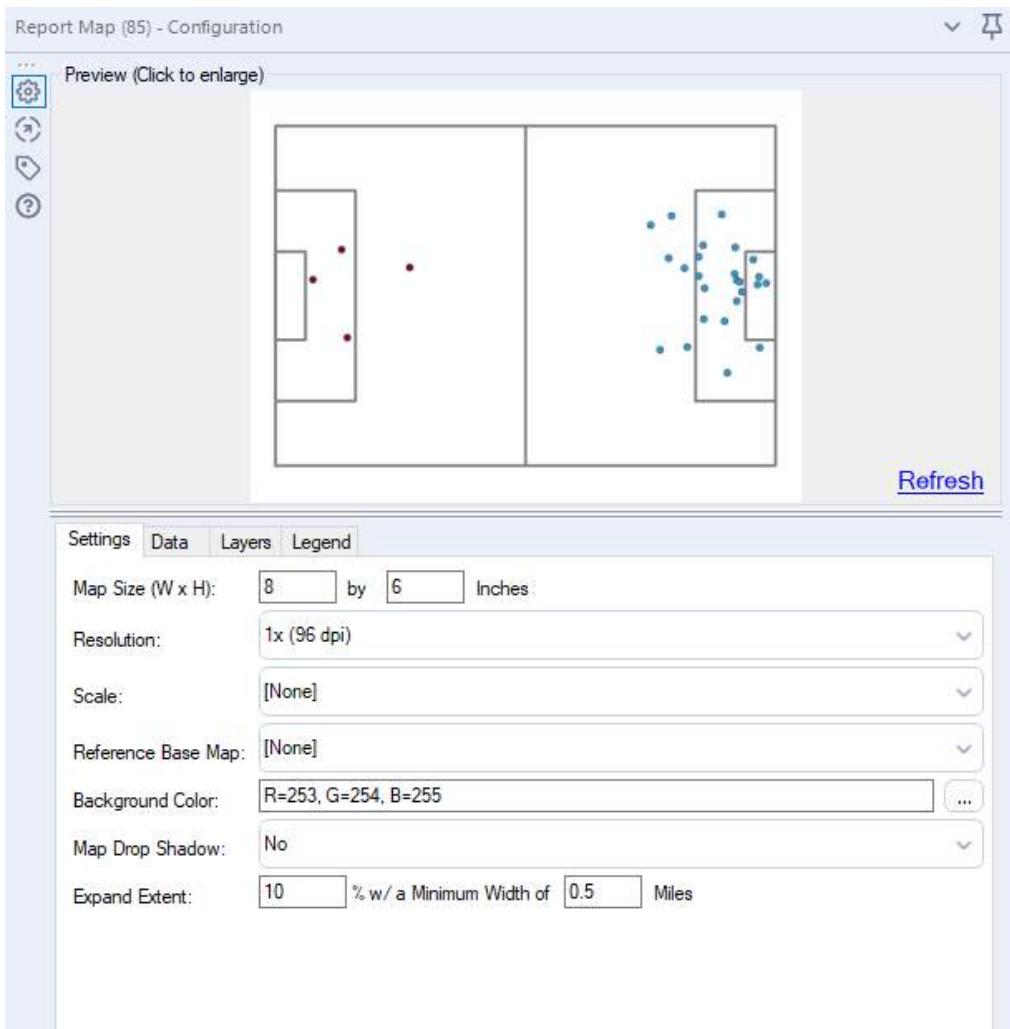
The first thing to do with our X & Y co-ordinates is to mirror the co-ordinates based on if they are in the home or away team. Reason for this is we are trying to mirror what is seen on the [Understat website](#). Currently our data only plots all the data in one side of the pitch.



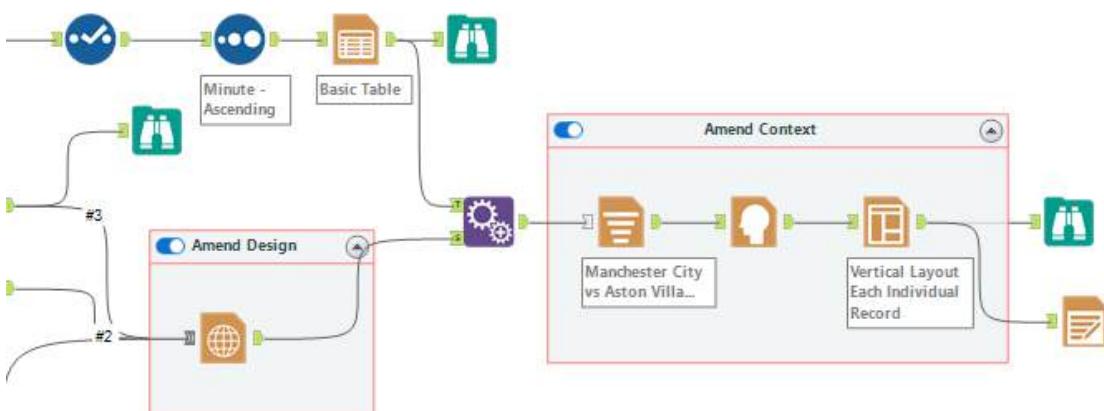
Next we do some re-scaling that was seen in the previous blog, just to allow the plot points and pitch to be easily rendered.



The middle part of the flow is the exact same as last weeks, we look to create the points and using the polybuild tool for the football pitch. The only difference here is rather than filter based on Messi & Ronaldo, we are splitting our data out on a specific match_id. For this I have chosen the Manchester City 3-2 Aston Villa game – it was one of the final games last season. You'll see my choice to split the data out as home and away to then be able to create separate points for each team.



So far so good, we have successfully been able to re-align our data to a pitch. But lets try take this a little further. We have no context as to the game. Lets take it just a few steps further by adding in a title, header, text context and maybe a table summary of the shots – we can then render this how we like.



The basic table runs along the top of the flow, It needs minimal impact, the table is mainly aesthetics and formatting that you can change, and I took the opportunity to make sure my data was sorted on by minute asc.

Bruyne (106) - Configuration

Report | 1 record displayed, 1 field, 5,377 bytes | Records 1 to 1

table

match_Id	date	Player	Minute	Result	xG	Situation	Home Or Away Team	Player ID	Shot Type	Home Team	Away Team
16754	2022-05-22	Fernandinho	2	MissedShots	0.06	SetPiece	h	614	Head	Manchester City	Aston Villa
16754	2022-05-22	João Pedro Cancelo	4	BlockedShot	0.06	OpenPlay	h	2379	RightFoot	Manchester City	Aston Villa
16754	2022-05-22	Phil Foden	23	BlockedShot	0.1	OpenPlay	h	6055	RightFoot	Manchester City	Aston Villa
16754	2022-05-22	João Pedro Cancelo	27	BlockedShot	0.04	OpenPlay	h	2379	LeftFoot	Manchester City	Aston Villa
16754	2022-05-22	Gabriel Jesus	30	MissedShots	0.06	OpenPlay	h	5543	RightFoot	Manchester City	Aston Villa
16754	2022-05-22	Matthew Cash	36	Goal	0.06	OpenPlay	a	8864	Head	Manchester City	Aston Villa
16754	2022-05-22	Ollie Watkins	40	BlockedShot	0.08	OpenPlay	a	8865	RightFoot	Manchester City	Aston Villa
16754	2022-05-22	Philippe Coutinho	41	MissedShots	0.02	OpenPlay	a	488	RightFoot	Manchester City	Aston Villa
16754	2022-05-22	Kevin De Bruyne	43	BlockedShot	0.06	OpenPlay	h	447	LeftFoot	Manchester City	Aston Villa
16754	2022-	João Pedro	44	BlockedShot	0.02	OpenPlay	h	2379	RightFoot	Manchester	Aston

Next I look to add a little text for the report,

Report Text (101) - Configuration

Text Mode

Create new field for this text
Field Name:

Attach text to existing field
Field Name:

Position:

Expert Mode: Treat text as Raw PCXML
 Validate PCXML

Text Data

B I U A

Available Fields ▾ Special Tags ▾

Manchester City vs Aston Villa Shot Map

EPL 2021/2022

This report is an example output from one of the final matches in the English Premier League Season 21/22. The game ended with a score of 3-2 to [Manchester City](#). You can see the full match report on Understat using match ID [16754](#)

Perhaps this section would be better if I could automate the values passing through it but for now I am happy as I just want to test the features. This will need manually updating for now, based on which match you choose to design.

Report Header (100) - Configuration

Report Header

The Report Header macro will append the created header to your supplied data, this new field can then be used in a Render tool as the Header for your report.

Report Title

Created on CJ-Mayes.com

Report Date

Include the date in your report header.

Report Image

Include an image (logo) on your header.
 Use the Alteryx logo
 Specify custom image:

Browse to the image to use in the report header:
C:\Users\Cj.Mayes\OneDrive - JLL\Desktop\Frame 1.png 

The size of the Alteryx logo is 150 by 55 (W x H) pixels, if you use an image of a different size or ratio, you may encounter layout issues. You may need to resize your image outside of Alteryx.

This next step is super unnecessary but it's me and I like to put my logo on things. Here I've added a header just with some website details and custom images.

Layout (104) - Configuration

Layout Mode: Each Individual Record

Group By:

Include Source Fields in Output

Layout Configuration

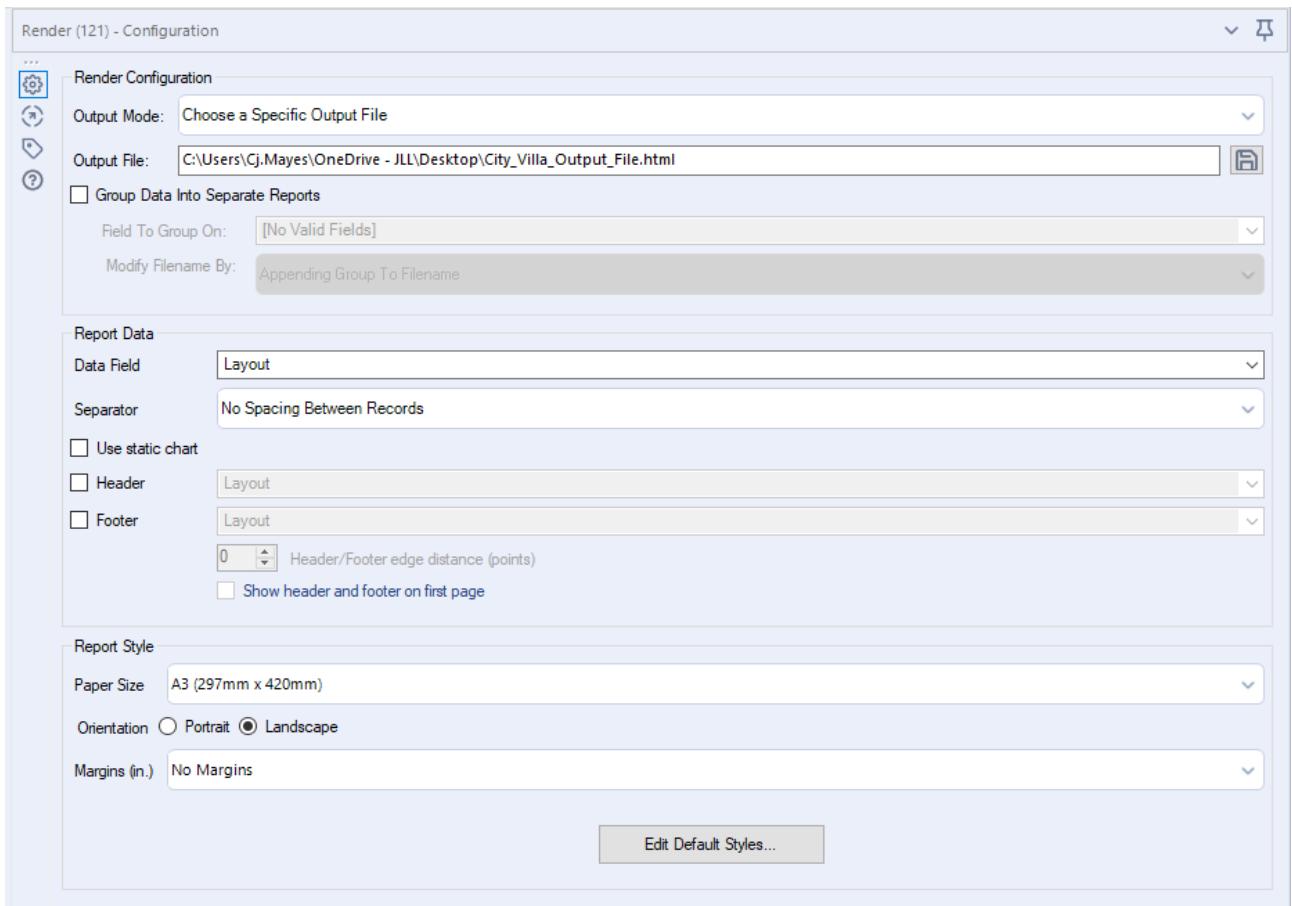
Orientation: Vertical
Layout Width: Automatic
Layout Height: Automatic
 Border
 Separator
Cell Padding: 10 pixels

Per Row Configuration

Header, Text, Map, Table

Height: Automatic
Alignment (V): Middle
Alignment (H): Center
 Fill Color

Finally, we can assign an order to how we want to lay these elements out. It takes a bit of back and forth testing different sizes, scales and layouts to find something that may feel best.



Finally I look to export the file, in this case to HTML, but of course you can export to PPT, Image amongst other things.

All	0 Errors	0 Conv Errors	0 Warnings	10 Messages	6 Files
Browse (96)	1 records				
Summarize (98)	4 records were summarized to 1 groups				
Browse (106)	1 records				
Text Input (52)	27 records were output				
Poly-Build (72)	27 points were built into 6 polys				
Summarize (75)	6 records were summarized to 1 groups				
Report Map (85)	Maps generated: 1				
Append Fields (114)	1 target records generated 1 output records				
Browse (103)	1 records				
Render (121)	C:\Users\Cj.Mayes\OneDrive - JLL\Desktop\City_Villa_Output_File.html				
Browse (91)	6 records				
Text Input (95)	27 records were output				
Designer x64	Finished running Understat_Data_Flow.ymd in 0.7 seconds				

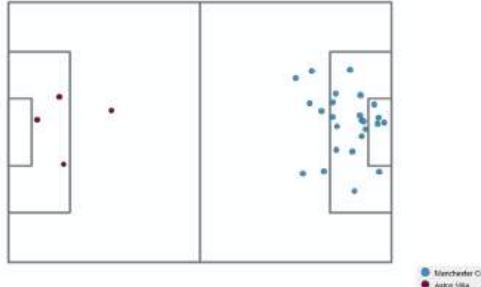
If you run the flow you will see your render save (as well as export if you haven't chosen temporary storage) By clicking the render you can see how it appears!



Manchester City vs Aston Villa Shot Map

EPL 2021/2022

This report is an example output from one of the final matches in the English Premier League Season 21/22.
The game ended with a score of 3-2 to Manchester City. You can see the full match report on Understat using match ID 36754.



match_id	date	Player	Minute	Result	xG	Situation	Home Or Away Team	Player ID	Shot Type	Home Team	Away Team	Home Goals	Away Goals	Player Assisted	Last Action
16754	2022-05-22	Fernandinho	2	MissedShot	0.06	SetPiece	h	614	Head	Manchester City	Aston Villa	3	2	Kevin De Bruyne	Cross
16754	2022-05-22	João Pedro	4	BlockedShot	0.06	OpenPlay	h	2879	RightFoot	Manchester City	Aston Villa	3	2	Phil Foden	Pass
16754	2022-05-22	Phil Foden	23	BlockedShot	0.1	OpenPlay	h	8055	RightFoot	Manchester City	Aston Villa	3	2	Gabriel Jesus	Pass
16754	2022-05-22	João Pedro	27	BlockedShot	0.04	OpenPlay	h	2379	LeftFoot	Manchester City	Aston Villa	3	2	John Stones	Chipped
16754	2022-05-22	Gabriel Jesus	30	MissedShot	0.06	OpenPlay	h	5543	RightFoot	Manchester City	Aston Villa	3	2	Phil Foden	Ball Recovery
16754	2022-05-22	Matthew Cash	36	Goal	0.06	OpenPlay	a	8864	Head	Manchester City	Aston Villa	3	2	Lucas Digne	Aerial
16754	2022-05-22	Ollie Watkins	40	BlockedShot	0.08	OpenPlay	a	4885	RightFoot	Manchester City	Aston Villa	3	2	Philippe Coutinho	Pass
16754	2022-05-22	Philippe Coutinho	41	MissedShot	0.02	OpenPlay	a	488	RightFoot	Manchester City	Aston Villa	3	2	Ollie Watkins	None
16754	2022-05-22	Kevin De Bruyne	43	BlockedShot	0.06	OpenPlay	h	447	LeftFoot	Manchester City	Aston Villa	3	2	Gabriel Jesus	None
16754	2022-05-22	João Pedro	44	BlockedShot	0.02	OpenPlay	h	2379	RightFoot	Manchester City	Aston Villa	3	2	Bernardo Silva	Pass
16754	2022-05-22	João Pedro	46	BlockedShot	0.02	OpenPlay	h	2379	LeftFoot	Manchester City	Aston Villa	3	2	Bernardo Silva	Pass
16754	2022-05-22	Kevin De Bruyne	47	MissedShot	0.06	OpenPlay	h	447	RightFoot	Manchester City	Aston Villa	3	2		Aerial
16754	2022-05-22	Aymeric Laporte	47	SavedShot	0.06	FromCorner	h	2498	Head	Manchester City	Aston Villa	3	2	Riyad Mahrez	Cross
16754	2022-05-22	Riyad Mahrez	49	BlockedShot	0.14	OpenPlay	h	730	LeftFoot	Manchester City	Aston Villa	3	2	Oleksandr Zinchenko	Pass
16754	2022-05-22	Gabriel Jesus	49	MissedShot	0.58	OpenPlay	h	3543	RightFoot	Manchester City	Aston Villa	3	2	João Pedro	Cross
16754	2022-05-22	Rodri	50	MissedShot	0.02	OpenPlay	h	2490	RightFoot	Manchester City	Aston Villa	3	2	Riyad Mahrez	Pass
16754	2022-05-22	John Stones	53	BlockedShot	0.06	FromCorner	h	586	Head	Manchester City	Aston Villa	3	2	Kevin De Bruyne	Aerial
16754	2022-05-22	Gabriel Jesus	66	BlockedShot	0.06	OpenPlay	h	5543	LeftFoot	Manchester City	Aston Villa	3	2	Phil Foden	Cross
16754	2022-05-22	Kevin De Bruyne	67	MissedShot	0.06	DirectFreekick	h	447	RightFoot	Manchester City	Aston Villa	3	2		Standard
16754	2022-05-22	Philippe Coutinho	68	Goal	0.08	OpenPlay	a	488	RightFoot	Manchester City	Aston Villa	3	2	Ollie Watkins	HeadPass
16754	2022-05-22	Phil Foden	70	MissedShot	0.1	OpenPlay	h	8055	RightFoot	Manchester City	Aston Villa	3	2	Rodri	TakeOn
16754	2022-05-22	Kevin De Bruyne	72	MissedShot	0.14	OpenPlay	h	447	RightFoot	Manchester City	Aston Villa	3	2	Raheem Sterling	Pass
16754	2022-05-22	Riley Gårdstrand	73	MissedShot	0.06	SetPiece	h	514	Head	Manchester City	Aston Villa	3	2	Kevin De Bruyne	Aerial
16754	2022-05-22	Riley Gårdstrand	75	Goal	0.5	OpenPlay	h	814	Head	Manchester City	Aston Villa	3	2	Raheem Sterling	Cross

The above two screenshots are how it renders in google chrome. Overall I'm pretty happy with it for a basic report. Of course there is lots more we can go back and add – maybe we can revisit this down the line. Last thing to note is if you export your files to html, it will also save down all the images needed to render. If these aren't stored on your computer they wont render in the final version. I.e in this case the logo would disappear and so would actually the shot map as the circles end up displaying as images. **Taking It Further**

- Take a look at a game of your choice
- Try and automate more of the flow
- Add in the choice of picking a game rather than manually updating a match ID
- Amending the layout of the visual
- Choose a different export type

LOGGING OFF, CJ

SHOT MAPS IN ALTERYX USING POLYBUILD

Hi all, Recently there has been a few blogs I've created on different ways to showcase soccer pitch co-ordinate data. This blog will be no exception, however we will just use the one tool today – Alteryx! Our region at work sat down to complete a weekly Alteryx challenge as part of our learning and development sessions, this time hosted by **Faris**. The **challenge** chosen was 346, and 347.

The idea is to Map Messi and Ronaldo's Goals, using the input provided.

Challenge #346: Map Messi and Ronaldo's Goals (Part 1)

Options ▾



ACE Qiu
19 - Altair

11-21-2022 06:20 AM

...

A solution to last week's challenge can be found [here](#).

This challenge comes to us from [@Qiu](#) who was inspired after reading the following article:
<https://medium.com/analytics-vidhya/using-spatial-and-activities-data-with-python-to-understand-how-....>
Thank you for your contribution!

This week and next week, you will use datasets to analyze how Lionel Messi and Cristiano Ronaldo do their magic.

To complete Part 1 of this challenge:

- Use the Players dataset to filter information about Messi and Ronaldo. You must include the first name, last name, role name, short name, and ID (wyid or playerid).
- Use the events_Spain_Goal dataset to retrieve information for the same 2 players. You must include the eventName, matchID, playerId, position_0_x, position_0_y, subEventID, subEventName, and tags_0_id.
- Join your 2 findings in 1 dataset. It will be used in Part 2 of the challenge.



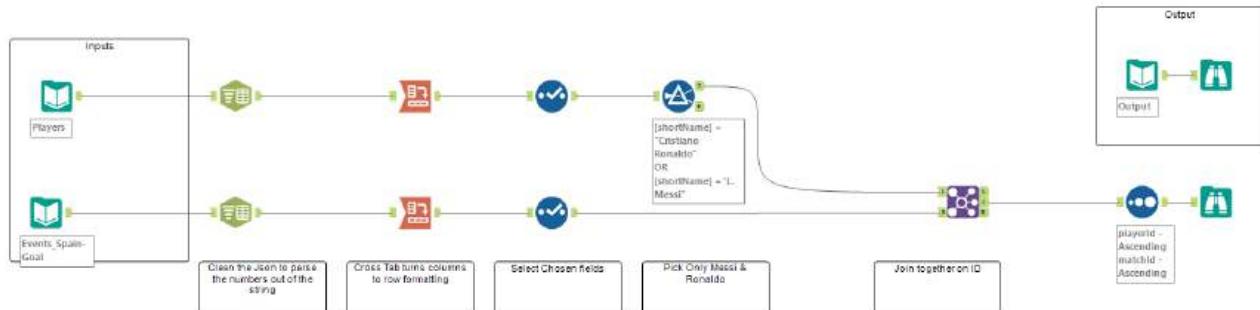
[challenge_346_start_file.yxzp](#)



[challenge_346_solution.yxzp](#)



Todays blog will be a run through of how we got on as a group, and also where you could start applying this logic to the understat datasets from previous blogs! To complete Part 1 of this challenge, we need to:
- Use the Players dataset to filter information about Messi and Ronaldo. Include the first name, last name, role name, short name, and ID (wyid or playerid).
- Use the Events_Spain_Goal dataset to retrieve information for the same 2 players. You must include the eventName, matchID, playerId, position_0_x, position_0_y, subEventID, subEventName, and tags_0_id.
- Join your 2 findings in 1 dataset. It will be used in Part 2 of the challenge.
So lets have a look like what that looks like in terms of the data prep of the flow.



The first part to the challenge is a little more simple. You can find the solution to the flow in the GitHub repository. We parse out the values within the Json, convert them from columns to rows and then select the fields and limit to just Messi & Ronaldo – finally joining the two data streams together to build our final output. The final output for this part of the challenge includes the X & Y co-ordinates of the shot locations.

Record	firstName	lastName	role_name	shortName	wyid	eventName	matchId	playerId	position_0_x	positions_0_y	subEventId	subEventName	tags_0_id
1	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505619	3322	91	00	100	Shot	101
2	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505673	3322	95	46	100	Shot	101
3	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505695	3322	88	62	100	Shot	101
4	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Free Kick	2505695	3322	90	50	35	Penalty	101
5	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505738	3322	97	45	100	Shot	101
6	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505738	3322	92	51	100	Shot	101
7	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Free Kick	2505749	3322	90	50	35	Penalty	101
8	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Free Kick	2505749	3322	90	50	35	Penalty	101
9	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505772	3322	91	48	100	Shot	101
10	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505772	3322	89	58	100	Shot	101
11	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505772	3322	98	41	100	Shot	101
12	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505780	3322	90	32	100	Shot	101
13	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505796	3322	88	50	100	Shot	101
14	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505796	3322	88	35	100	Shot	101
15	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505812	3322	86	52	100	Shot	101
16	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505812	3322	97	45	100	Shot	101
17	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505818	3322	88	42	100	Shot	101
18	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505818	3322	94	50	100	Shot	101
19	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505836	3322	87	46	100	Shot	101
20	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505836	3322	89	51	100	Shot	101
21	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505836	3322	96	48	100	Shot	101
22	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505836	3322	88	53	100	Shot	101
23	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505852	3322	91	62	100	Shot	101
24	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505876	3322	94	48	100	Shot	101
25	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505907	3322	99	51	100	Shot	101
26	Cristiano Ronaldo	dos Santos Aveiro	Forward	Cristiano Ronaldo	3322	Shot	2505927	3322	93	48	100	Shot	101

Above is what that data structure looks like. Challenge 347 is when things really start to get exciting.

Challenge #347: Map Messi and Ronaldo's Goals (Part 2)

Options ▾



11-28-2022 06:12 AM

...

A solution to last week's challenge can be found [here](#).

This challenge comes to us from [@Qiu](#) who was inspired after reading the following article: <https://medium.com/analytics-vidhya/using-spatial-and-activities-data-with-python-to-understand-how-...>. Thank you for your contribution!

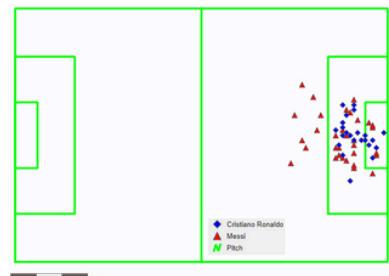
This week, we will continue using datasets to analyze how Messi and Ronaldo do their magic.

For Part 2 of this challenge, you will:

- Use the Part_2_input dataset to create the points where Messi and Ronaldo score their goals.
- Use the FootballPitch dataset to create the Football Pitch.
- Finally, joint the two previous findings to layout the Football Pitch and the goals from Messi and Ronaldo.

Hints

- The x and y coordinates are always in the range [0, 100] there, x position shall be multiplied by 1.25 and Y by 0.85 to get actual scale.
- Give the relatively small scale of a football pitch, it is better to multiply 1000 to both x, y positions of player and football pitch for visibility.



Check out our newly rebranded Spatial course.

<https://community.alteryx.com/t5/Interactive-Lessons/tkb-p/interactive-lessons/label-name/Spatial%20...>



For Part 2 of this challenge, we will need to:

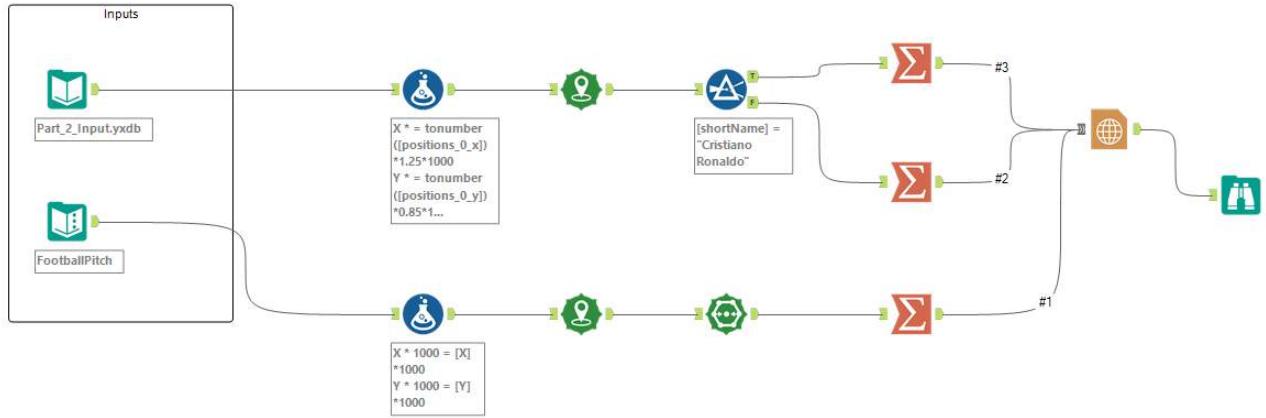
- Use the Part_2_input dataset to create the points where Messi and Ronaldo score their goals.
- Use the FootballPitch dataset to create the Football Pitch.
- Finally, joint the two previous findings to layout the

	X	Y	Gr	SN	...
1	0	0	1	1	
2	0	85	1	2	
3	125	85	1	3	
4	125	0	1	4	
5	0	0	1	5	
6	0	16.15	2	1	
7	20	16.15	2	2	
8	20	68.85	2	3	
9	0	68.85	2	4	
10	0	16.15	2	5	
11	0	31.45	3	1	
12	0	53.55	3	2	
13	7.5	53.55	3	3	
14	7.5	31.45	3	4	
15	0	31.45	3	5	
16	62.5	0	4	1	
17	62.5	85	4	2	
18	105	16.15	5	1	
19	125	16.15	5	2	
20	125	68.85	5	3	
21	105	68.85	5	4	
22	105	16.15	5	5	
23	117.5	31.45	6	1	
24	125	31.45	6	2	
25	125	53.55	6	3	
26	117.5	53.55	6	4	
27	117.5	31.45	6	5	
					...

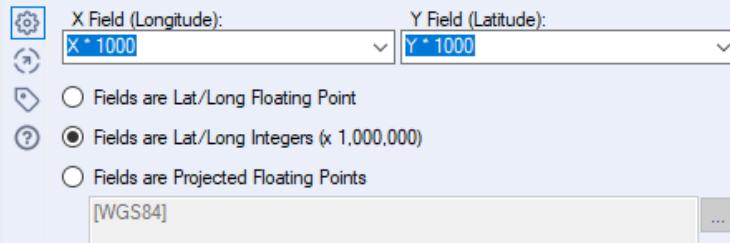
Football Pitch and the goals from Messi and Ronaldo.

So this input file is the co-ordinates

of the football pitch for the given data. Of course if you are using some external data like from opta or understat you could readjust and scale these points up and down for your own data. The group number identifies the polygon, and the SN will represent the order you join the points up in.



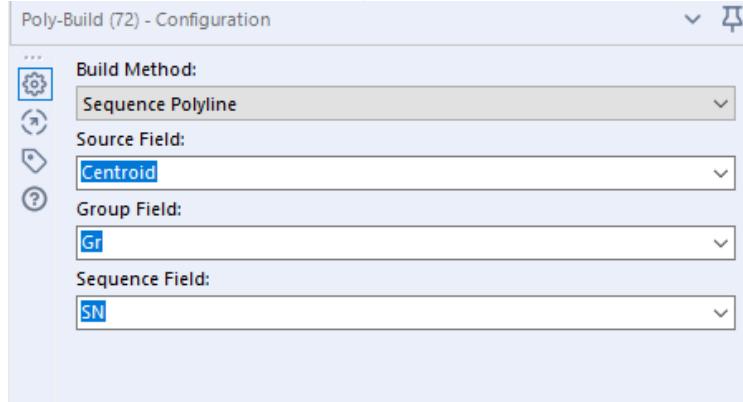
Above is the flow we will use to create our final shot map. Lets take it step by step. First we rescale our points to make easier to



visualise.

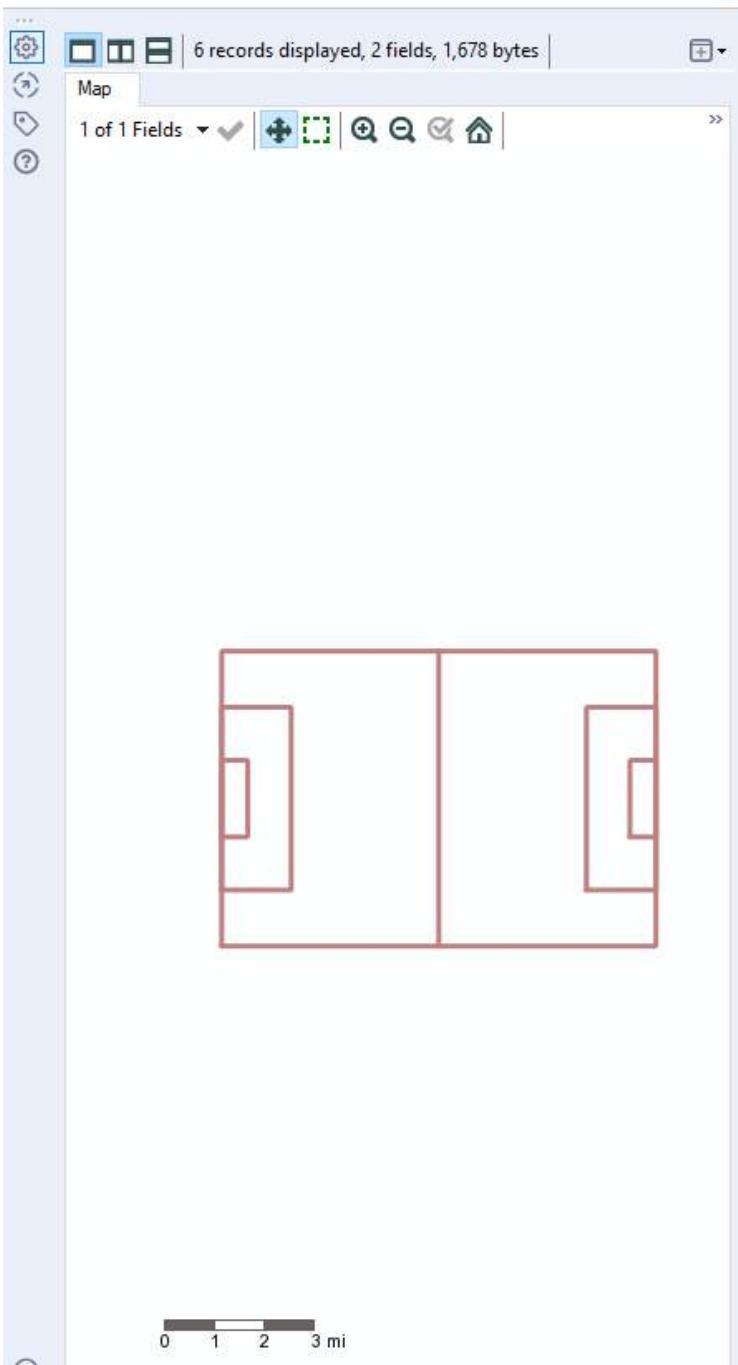
We then use the create points functionality

to be able to plot the latitude and longitude. Essentially our x and y values. We make sure to use the second option on the drop down to allow for x and y points that go beyond a typical maps limitations. For the top stream these will be unique points for each shot. For the bottom part of the stream these will be identifying points to draw our football pitch.



The polybuild function helps us then draw the pitch.

We take our input of the centroid (i.e our mapped x and y co-ordinates) and draw a sequence polyline, grouping on the group field and joining those lines up in the order of the SN field. It is important to remember to make the build method a polyline instead of polygon or convex hull! By adding in the browse we can see how this creates our pitch.



We can essentially flatten these lines into one using

Summarize (84) - Configuration

Fields:

	Field	Type
▶	firstName	V_WString
	lastName	V_WString
	role_name	V_WString
	shortName	V_WString
	wyld	V_WString
	eventName	V_WString
	matchId	V_WString
	playerId	V_WString
	positions_0_x	V_WString
	positions_0_y	V_WString
	subEventId	V_WString
	subEventName	V_WString
	tags_0_id	V_WString
▼ *		Double

Actions:

	Field	Action	Output Field Name
▶	Centroid	Combine	Messi

the combine field.

The final thing we want to do in the

top half of the screen is to split the data out, one for Messi and one for Ronaldo. Finally we will look to use the report map configuration to be able to showcase our final shot map! Things to note include:

- Adding in the incoming data fields. This will include Messi, Ronaldo shots and the background pitch.
- Adding in Layers, You'll notice that we can change the shape for each player, given we split them out as well as assign them their own colour encoding.

Settings Data Layers Legend

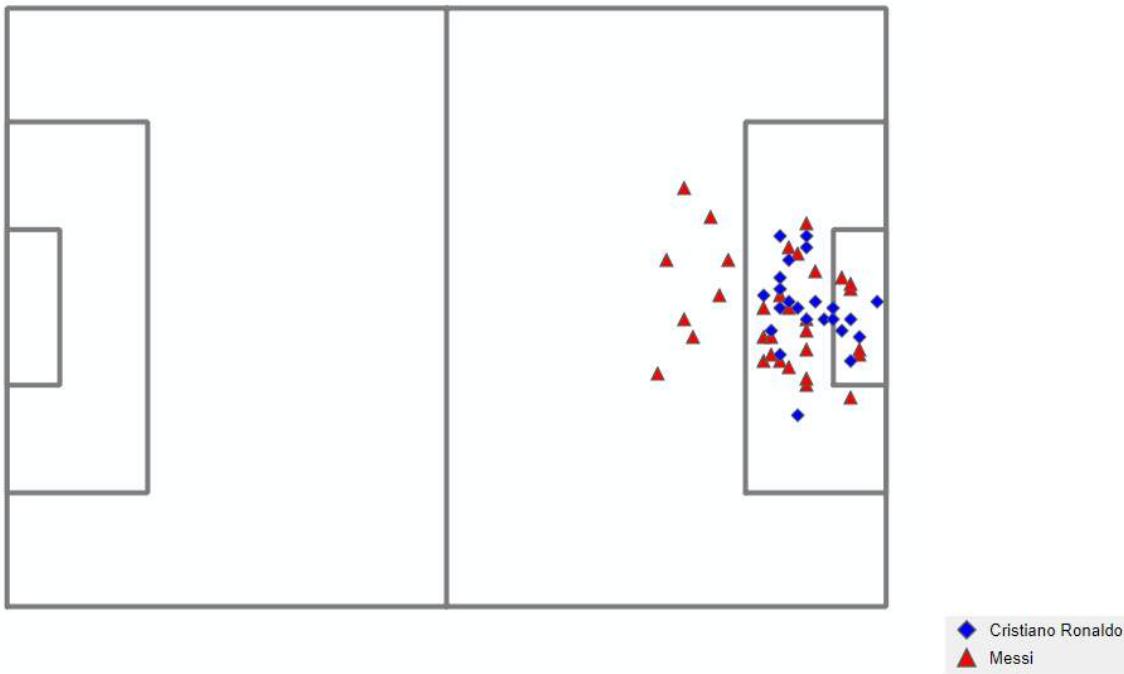
Layers:

- Cristiano Ronaldo
- Messi
- Pitch
- Style
- Base Layers - Points
- Base Layers - Lines
- Base Layers - Polygons

Layer Properties:

- Layer Name: Pitch
- Data Connection: #1
- Show in Legend:
- Zoom to Layer:
- Disable:

And there we have, if we run our workflow and add the browse tool we can see these marks drawn in Alteryx.



Maybe not the most aesthetically appealing design we have ever seen, but a good challenge to complete all within Alteryx. Maybe we will look if we can emulate some of the other blogs within Alteryx too! Once again, a massive thank you to Faris for leading the session and the team for their engagement with this challenge. We had some great inputs from [Filip Rutkowski](#) too on configuring the final mapping tools. **Going Further:**

- Apply the logic to your own dataset
- Create a different type of football pitch metrics chosen from MPLSoccer website
- Play around with the colours, layers and pitch details
- Try showcase a match with shots on both sides of the pitch.

LOGGING OFF, CJ

TRELLIS CHARTS WITH NIR SMILGA

Hi all, April is upon us! I am loving the lighter mornings and start to longer days here in the UK – It does wonders for my mental. Today, I am so pleased to be able to invite Nir Smilga onto the site. Its such an honour when people want to share there content on my site – very selfless of them. Nir has produced fantastic public work for years now, I'd urge you to check out [his profile](#). Most recently what caught my eye was a design technique of having small multiples, but rather than split each line losing the ability to cross compare, instead have them all showing and highlight each specific dimension. It just ticks the box for me between functionality and design. Over to Nir to discuss more at length...

Share Prices by Country

for Feb'23 and Change from Jan'18
Largest Increase: South Africa **▲36.4%** | Largest Decrease: Spain **▼-12.6%**

*2015 as a reference (2015=100)

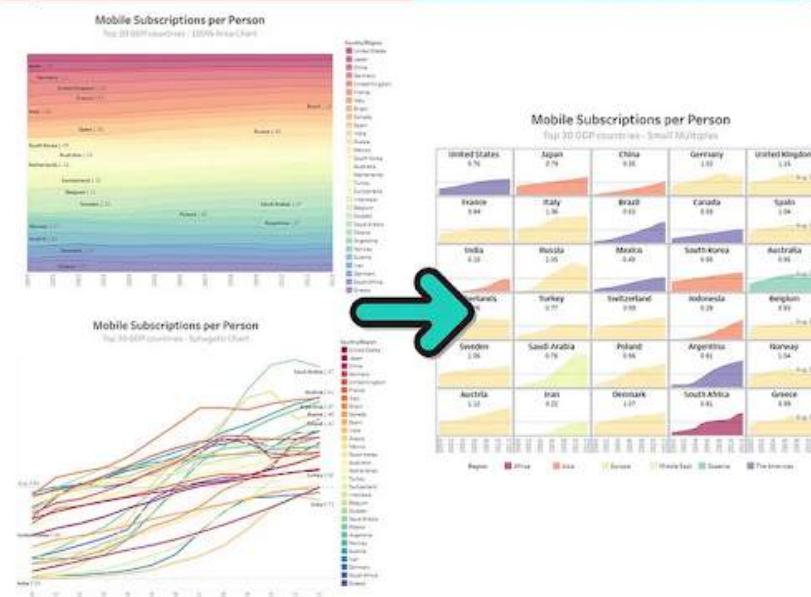
Share price indices are calculated from the prices of common shares of companies traded on national or foreign stock exchanges. They are usually determined by the stock exchange, using the closing daily values for the monthly data, and normally expressed as simple arithmetic averages of the daily data. A share price index measures how the value of the stocks in the index is changing; a share return index tells the investor what their "return" is, meaning how much money they would make as a result of investing in that basket of shares. Source: OECD Data

x



Nir: I love small-multiples (a.k.a. trellis charts). They are one of my favorite data viz techniques. It allows each group to be shown without the clutter of spaghetti charts or stacked area charts. I wrote about it [here](#) including the "how to".

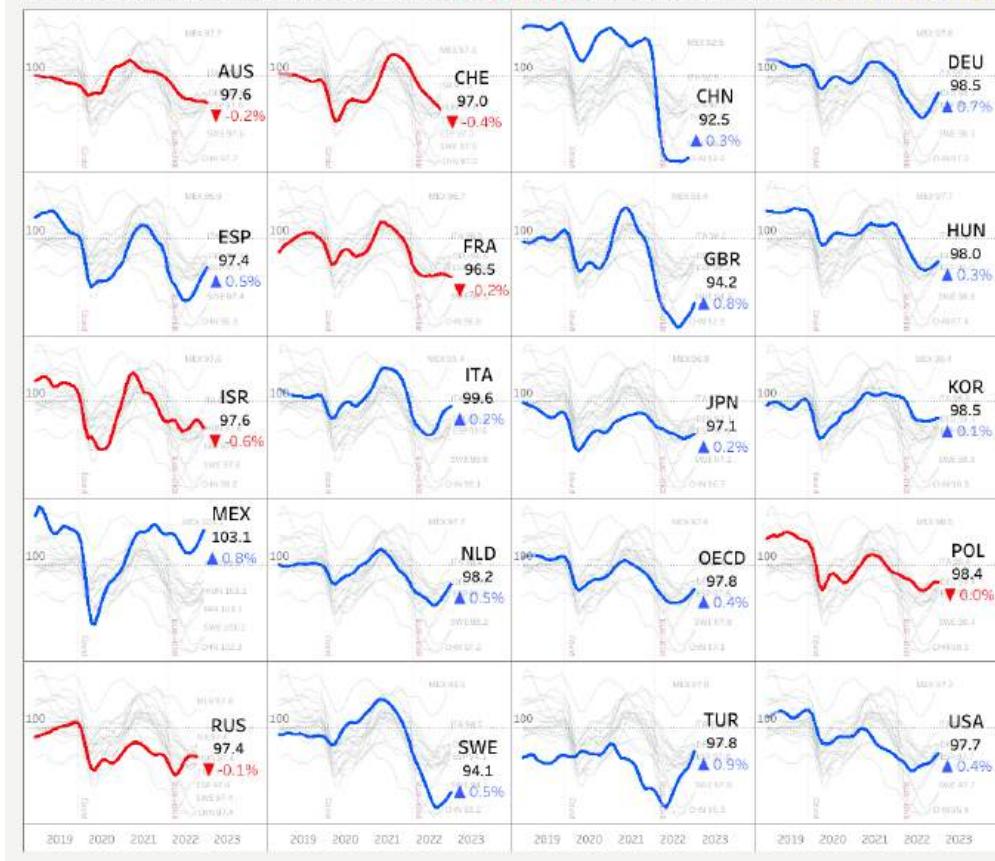
Avoid **Suggested**



But small-multiples lack the context of the other categories that makes comparisons easier. Small-multiples charts include many data points. In this example each country is shown with a specific measure over time. It is difficult to look at one and compare it to the other countries, especially if they are not on the same row. To mitigate that I added a reference line, but still, it is not easy to compare. A neat solution is to use small multiples with all categories grayed out. Seeing [Yan Holtz](#) do it in r gave me the push to finally try it in Tableau. It took me quite a few iterations but eventually I did it! Here is [the first iteration](#):

Consumer confidence index Jan'23

Change vs pre month: Largest Decrease: Israel ▼0.6% | Largest Increase: Turkey ▲0.9%



This consumer confidence indicator provides an indication of future developments of households' consumption and saving, based upon answers regarding their expected financial situation, their sentiment about the general economic situation, unemployment and capability of savings. An indicator above 100 signals a boost in the consumers' confidence towards the future economic situation, as a consequence of which they are less prone to save, and more inclined to spend money on major purchases in the next 12 months. Values below 100 indicate a pessimistic attitude towards future developments in the economy, possibly resulting in a tendency to save more and consume less.

Source: OECD Data

of Columns
4

Choose time period:
1/1/2019 to 12/31/23

Choose Countries
Multiple values

Nir Simlgi
in

And then I could not resist and created an enhanced version using map layers

Share Prices by Country for Feb'23 and Change from Jan'18

Largest Increase: South Africa **▲36.4%** | Largest Decrease: Spain **▼-12.6%**

*2015 as a reference (2015=100)

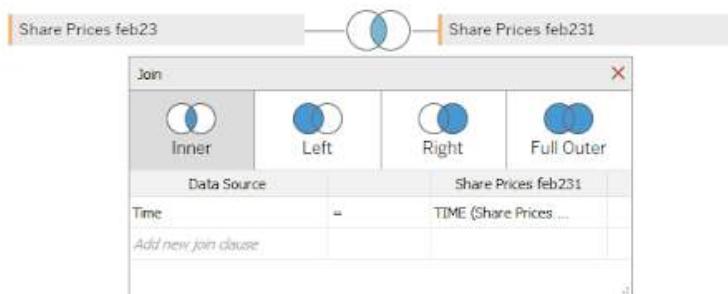
Share price indices are calculated from the prices of common shares of companies traded on national or foreign stock exchanges. They are usually determined by the stock exchange, using the closing daily values for the monthly data, and normally expressed as simple arithmetic averages of the daily data. A share price index measures how the value of the stocks in the index is changing; a share return index tells the investor what their "return" is, meaning how much money they would make as a result of investing in that basket of shares. Source: OECD Data

x



This blog will focus on how I did it in Tableau. I am not about to get into the bits and bytes of a “regular” (without the grayed out categories) small multiples chart in Tableau. I wrote about it in the blog I mentioned above. Another helpful resource [Here](#) is a very detailed tutorial by @sqlbelle explaining the logic and math behind this. Small-multiples chart in Tableau is based on table calculations that create a grid or a matrix. It is determined by the number of categories of a specific dimension, in this case the country. So in order to create the grid we will have to tell tableau to compute it for each one of the countries. But here is the challenge: how to tell tableau to show for each category the highlighted country and all of the other countries at the same time? The example I am about to demonstrate is a mix of the two versions presented above. I will demonstrate a rather simple version using dual-axes: one axis for grayed out lines and the other for the highlighted countries. I will use the data of the second example (Share Prices from the OECD) and show the highlighted countries as areas. **Data Prep** We need to break the grid by country but still need to show all the countries at the same time. After thinking about it for a while, I had an idea of joining the data to itself (self join) on the date field (which is the x-axis of each of the categories (countries))

Share Prices feb23 is made of 2 tables. ①



So for each country we get all the other

Location	LOCATION (Share Prices fe..	
AUS	AUS	Abc
	BEL	Abc
	CHN	Abc
	DEU	Abc
	ESP	Abc
	FRA	Abc
	GBR	Abc
	ISR	Abc
	ITA	Abc
	JPN	Abc
	KOR	Abc
	MEX	Abc
	NLD	Abc
	POL	Abc
	USA	Abc
	ZAF	Abc
BEL	AUS	Abc
	BEL	Abc
	CHN	Abc
	DEU	Abc
	ESP	Abc
	FRA	Abc

countries as well

First, we would like to filter the data, and

allow users to remove and add countries. As we have two country fields (as we applied self join) and we want the user to select or remove a country once and not twice, I created a **set** from the **Location** field and then created a mutual filter:

Location Set Filter

```
[LOCATION (Share Prices feb231)] in [Location Set]
and
[Location] in [Location Set]
```

Now users can add and remove countries from the set and it will appear on both **Location** and **LOCATION (Share Prices feb231)** (self joined Location) I also filtered the data to apply to the latest 6 years. **Layout** We need to tell Tableau to create the grid by the **Location**, and add to the details the self joined Location- **LOCATION (Share Prices feb231)**. Let's start building the grid based on x and y coordinates (read the blog and watch the video to understand the logic behind these calcs)

Let's add the date to the columns

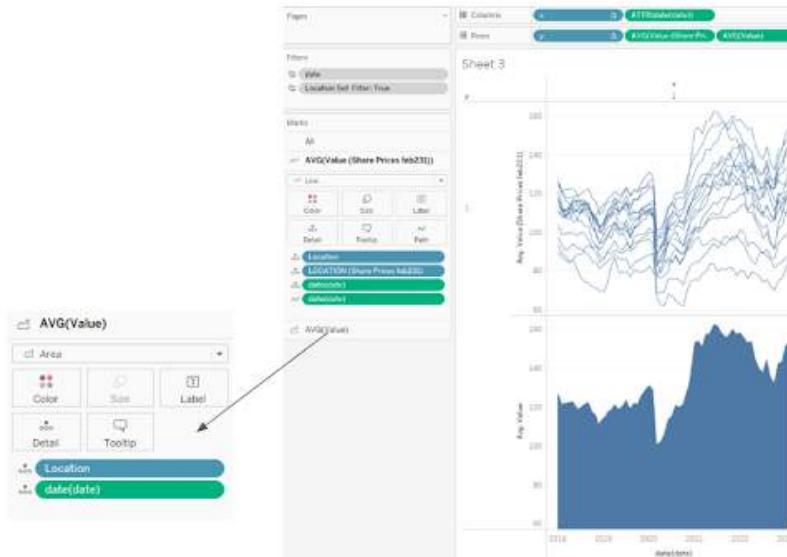
(wrapped by ATTR), the Values and again values, this time from the self joined table

iii Columns	x	Δ	ATTR(date(date))
iii Rows	y	Δ	AVG(Value (Share Pri..) AVG(Value)

Grayed Out lines **Value (Share Prices**

feb231): After adding the self-joined **Value (Share Prices feb231)** to the rows, Let's choose line chart type, and add **Location**, **LOCATION (Share Prices feb231)** -which is the self joined country, and **date** (make sure it is Exact Date and not aggregated) to

the details, and **date** again to the path mark to make continuous lines. Highlighted Country Area Chart: On **Value (Share Prices feb231)** (creating another y axis) let's choose area chart type, and add **date** and **Location** on to the details



Now we need to tell Tableau to compute the table calcs of **x** and **y** by the **Location**, **Date**, and **All LOCATION (Share Prices feb231)** fields, by editing its table

Table Calculation

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**

Location

date(date)

LOCATION (Share Prices feb231)

At the level Location

Restarting every

Sort order Specific Dimensions

Show calculation assistance

Table Calculation

y

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**

Location

date(date)

LOCATION (Share Prices feb231)

At the level Location

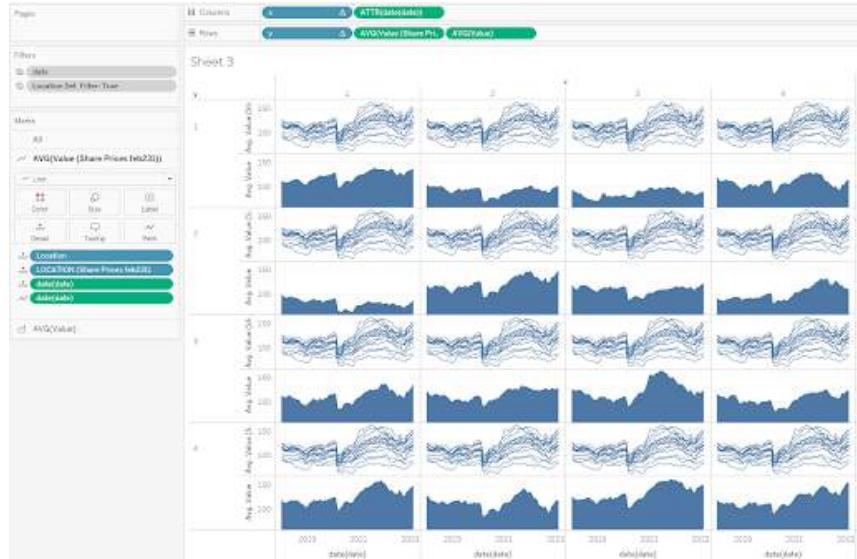
Restarting every

Sort order Specific Dimensions

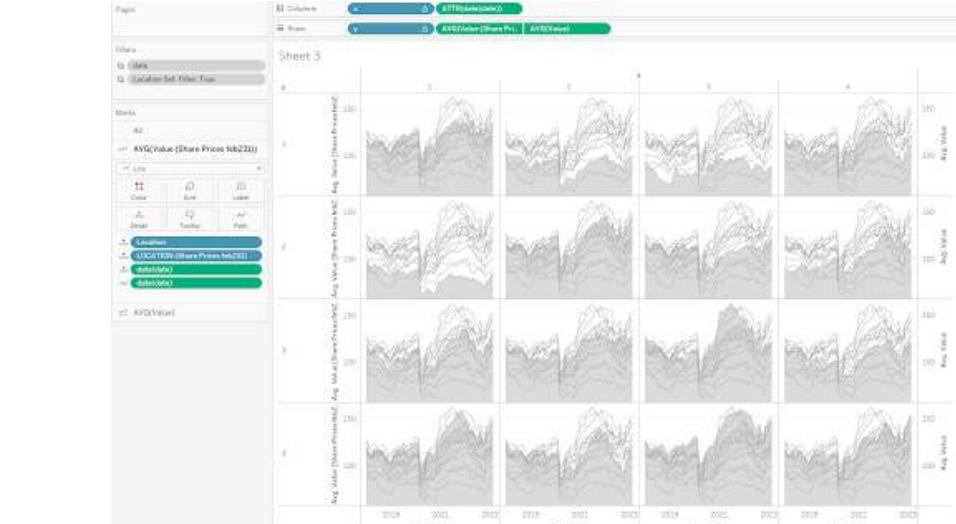
Show calculation assistance

calculations (the order matters!)

We should have something like that. I suggest you don't force the y axis to start from zero as we are showing trends\change vs 100 and not size comparison (which will require the y axis to start from 0)



Design Now, once the basic structure is set, we can focus on adding more context and on design. **Dual Axis** to combine the views, gray color with lower opacity for the



area chart

Space for country and value

labels on both sides of the charts For that we will use **reference lines** with these calculated fields

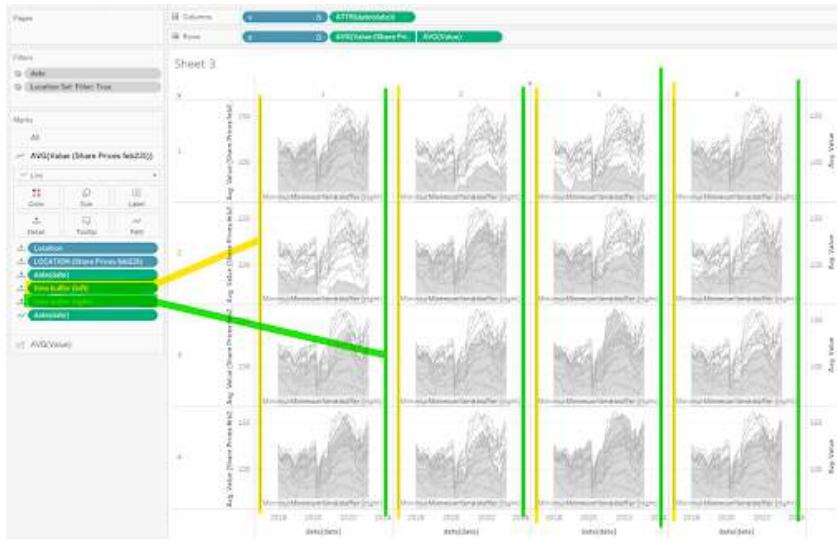
time buffer (left)

```
DATEADD('month', int( DATEDIFF('month', min([date]), max([date]))*-0.05*# of Columns ), (min([date])) )
```

time buffer (right)

```
DATEADD('month', int( DATEDIFF('month', min([date]), max([date]))*0.05*# of Columns ), (max([date])) )
```

We want that space to dynamically change upon different layouts (**# of columns**) and different **date** selections. Basically, this formula adds 5% to the left and right of the time period and multiplies by the number of columns.(make sure to hide the lines and the labels- we only need that for space)



Labels for grayed out countries

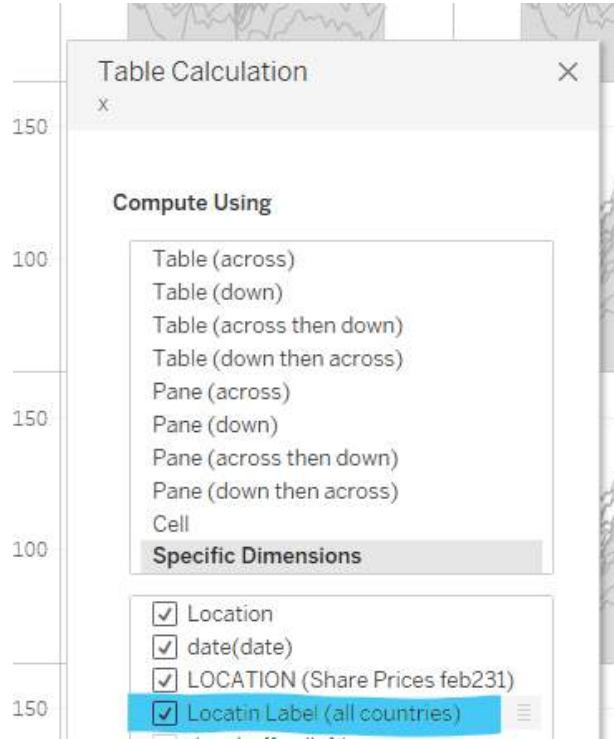
Let's create a calculated field to show all the countries besides the highlighted country (we will treat that separately next)

Location Label (all countries)

```
if [Location] != [LOCATION (Share Prices feb231)] THEN [LOCATION (Share Prices feb231)] end
```

* As we add another dimension, we will

need to tell tableau to take it into account in the table calculations for **x** and **y**. We will continue to add all the newly created dimensions to the table calculation under the "Compute Using" section



Value label for all countries except highlighted

Location Value (all countries)

```
if [Location Label (all countries)] != [Location] THEN [Value (Share Prices feb231)] END
```

Note that the value is from the self-joined table Now we will add these next to each other in the Edit Label section

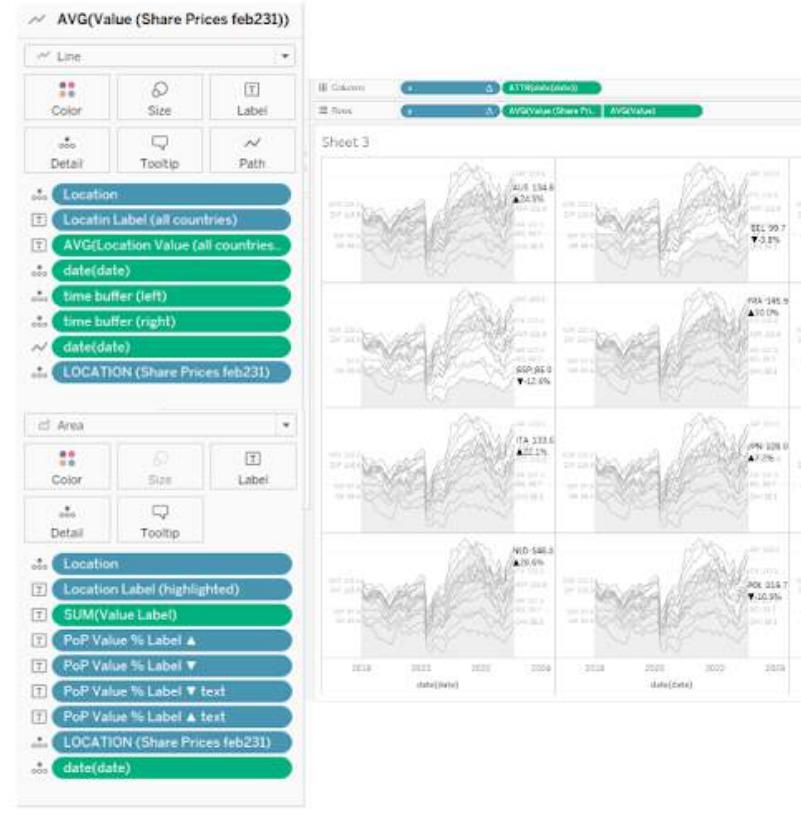
Edit Label



Labels for grayed out countries This is

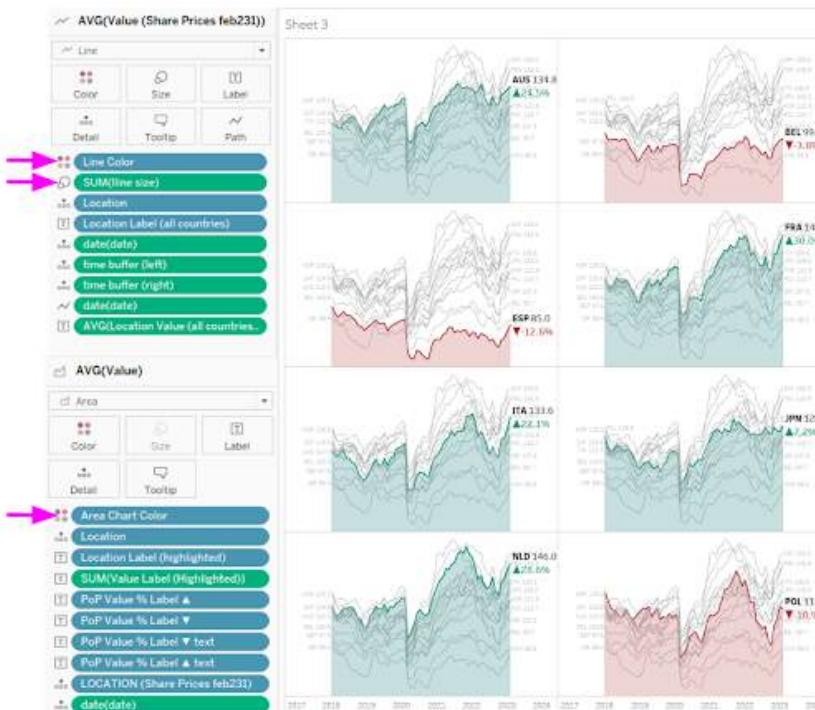
for the area chart, where labels are less customizable than on line charts. We want the labels to appear to the right of the area chart and show the country name, the value and change vs the min date value. In order to push the labels to the right I needed (after many tries and errors) It worked by adding the **LOCATION (Share Prices feb231)** field onto the details (and also to the **x**, **y** table calculations). To be honest, I am not sure I can explain here why it worked 😊 Ok, now we will need to create a bunch of calculated fields You can copy the calc text from the downloadable viz [here](#) They are called

- Location Label
- Value Label
- PoP Value % Label
- and finally the ▼▲ signs, under Pop Value % ▲



* Remember to add all the new dimensions to the table calculation "compute by"! * **Let's add some colors shall we?**

- **Line Color** – assigns green to the highlighted countries with positive change, red to negative change and gray for the grayed out countries
- **Area Chart Color** – colors the area for the highlighted countries Note: Change its opacity to 1%
- Let's make the highlighted country line thicker by distinguishing it using a calc field.



* Remember to add all the new dimensions to the table calculation compute! *Labels Color (for the highlighted countries)

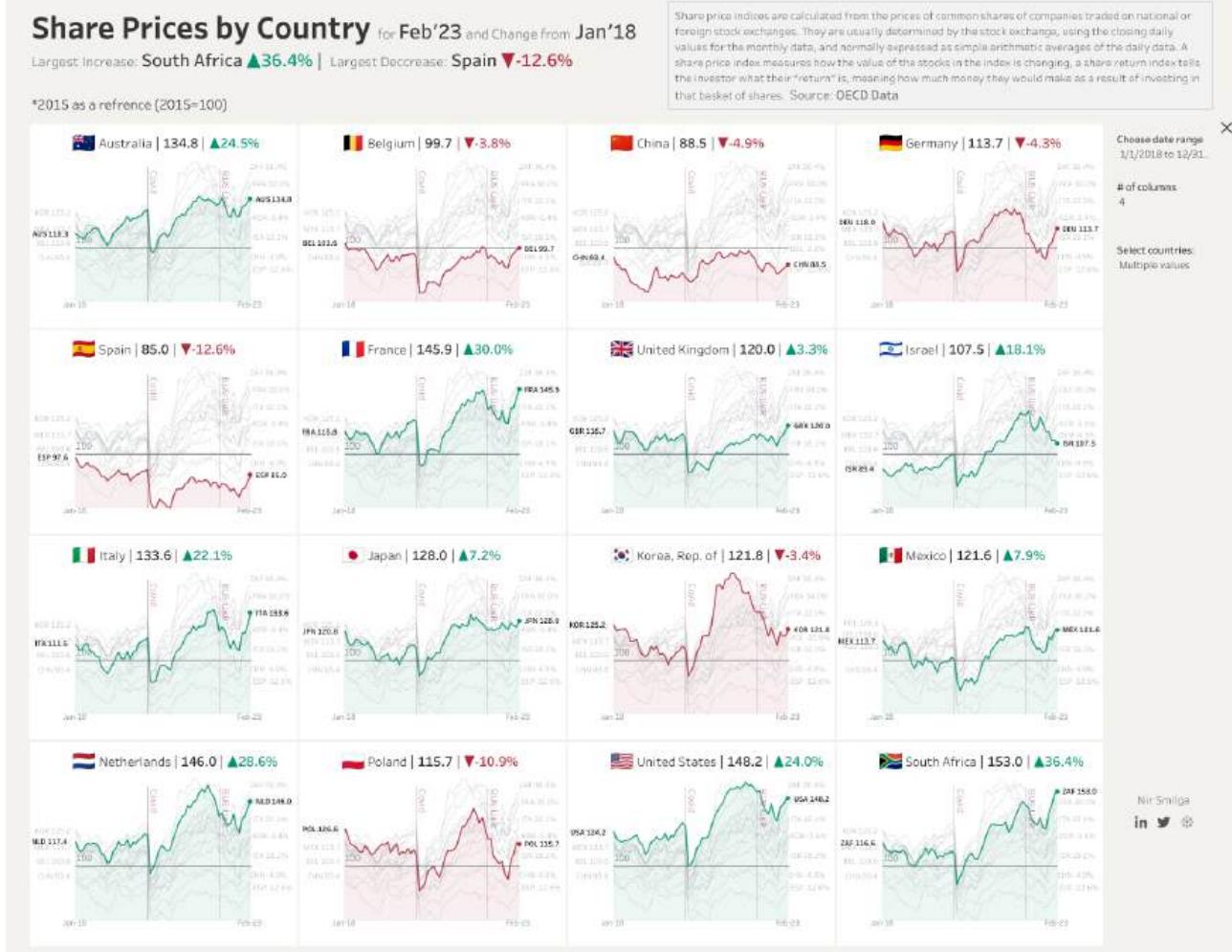
Edit Label

Tableau Bold 11 B U Insert X

<Location Label (highlighted) > <SUM(Value Label (Highlighted))>
<PoP Value % Label ▼><PoP Value % Label ▼ text><PoP Value % Label ▲ text><PoP Value % Label ▲>

I finalise my visualisation by adding more context to allow better comparisons. I would like to also see how different countries responded to major exogenous events such as COVID19 and Russia's invasion of Ukraine. Also, the values here represent changes from 2015 prices, when all values were normalized to 100, so let's add this one as well. Russia invaded Ukraine on Feb-24th 2022 but since the date granularity it monthly I moved it to Mar-01-2022*

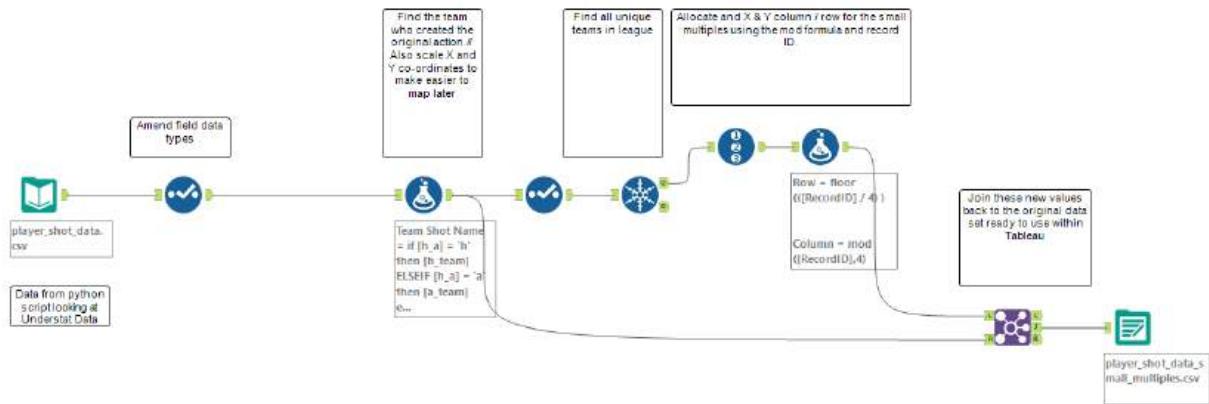
A few final touches to title and labels, and there we have it! Downloadable the full viz [here](#)



In the published viz there is an additional sheet with a line chart version, that allows hovering over all the grayed out categories (you can get to it once you edit the viz) I hope you find it useful – Share your thoughts! **CJ Round-Up:** A lengthy guest blog today jam packed with information and takeaways. We have everything from how and why trellis charts can be used through to additional tips and tricks when formatting. I love Nir's ability to create the thicker line for the country that is being looked at. The padding Nir creates either size of the viz with labels really brings that next level of context to the visual. Can't wait to see individuals emulate some of these design practices in their own small multiples moving forward. Thanks Nir! **LOGGING OFF, CJ**

SOCCKER SHOT MAP SMALL MULTIPLES

Hi all, A warm welcome back. This week we are going to look at small multiples. They can be a little fiddly at times due to various reasons, mainly because of null fields. We will look to explain some of these nuances as well as plot our all important football shot map charts. We will further our work by making our small multiples use more than one layer, allowing greater flexibility in build. **Where to start?** Well we will be using the 2021/22 season premier league data, that you can find a copy under the Github repo, or run your own version of the data from main.py that uses the understatapi package to collate the data. I have [previously blogged](#) about this data collection here. **What transformations do we need?** Well our small multiples doesn't like null values, so we can't go down our traditional route of following [Andy K's](#) video tutorial. (Yes I revisit this blog every time I need to copy and paste his calculations!) This means that our grid function needs essentially hardcoded the values this time round. You can read up more on the impact of nulls on small multiples in this blog by [Gwilym](#). So I've made a solution in Alteryx that can be applied to small multiples to always allow for updating datasets. You can find it in the repo.



What does the flow do? We take the original dataset and firstly identify whether the shot was taken by a home or attacking player. (of course we could have done this in tableau, but while we are here – why not. I also look to scale the X and Y co-ordinates up slightly for convenience. We then take a unique list of all the teams. (Of course, 20 records, given the number of teams in the Premier League) Next we allocating them a RecordID, 0 through 19. We introduce two calculations, the equivalent to the column and row headers seen in both Gwelyn's blog that hardcodes the values and Andy's blog that uses index and sizing calculations. My calculations use a floor calc, and a modular calculation.

Output Column		Data Preview	
Row			
<code>floor(([RecordID] / 4))</code>			
<code>X</code>			
<code>mod([RecordID],4)</code>			
<code>X</code>			
<code>Column</code>			
<code>Int16</code>			
<code>Size: 8</code>			
<code>Size: 2</code>			

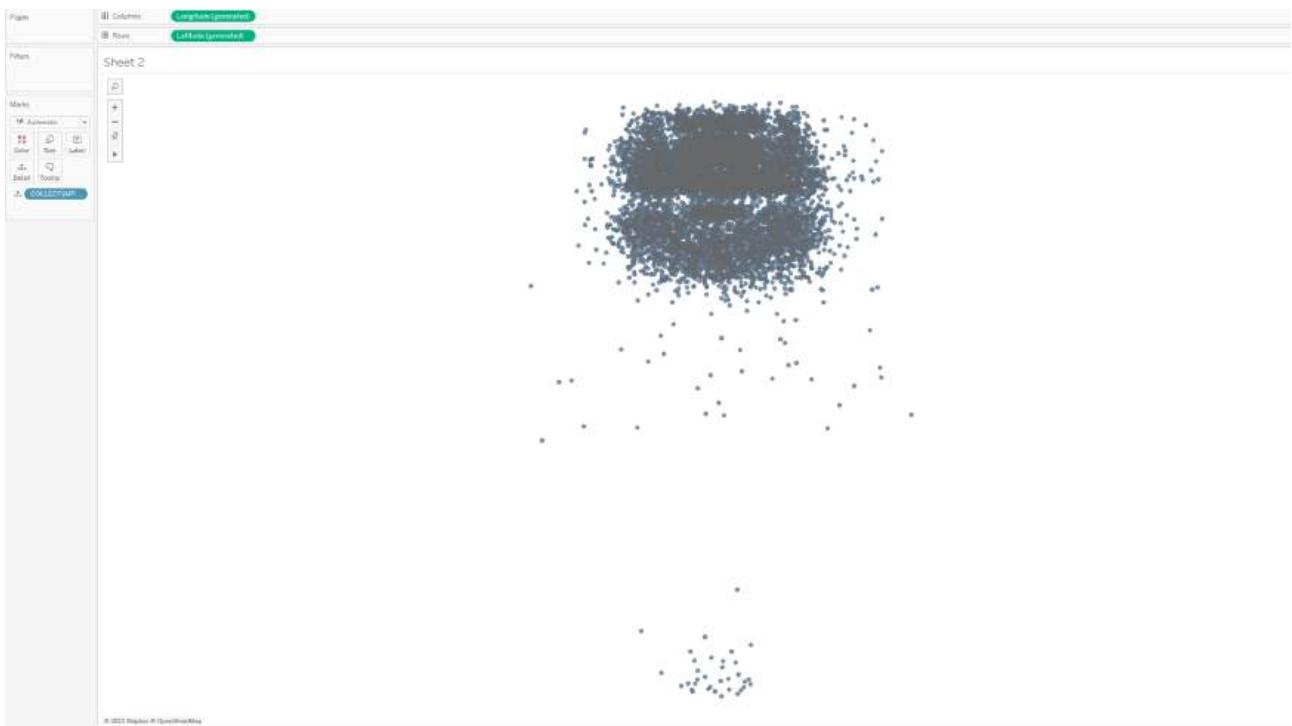
You can see

how this builds out the different column and row dividers we will need.

RecordID	Team Shot Name	Row	Column
0	Arsenal	0	0
1	Aston Villa	0	1
2	Brentford	0	2
3	Brighton	0	3
4	Burnley	1	0
5	Chelsea	1	1
6	Crystal Palace	1	2
7	Everton	1	3
8	Leeds	2	0
9	Leicester	2	1
10	Liverpool	2	2
11	Manchester City	2	3

Finally, we glue it back to our original dataset on the team

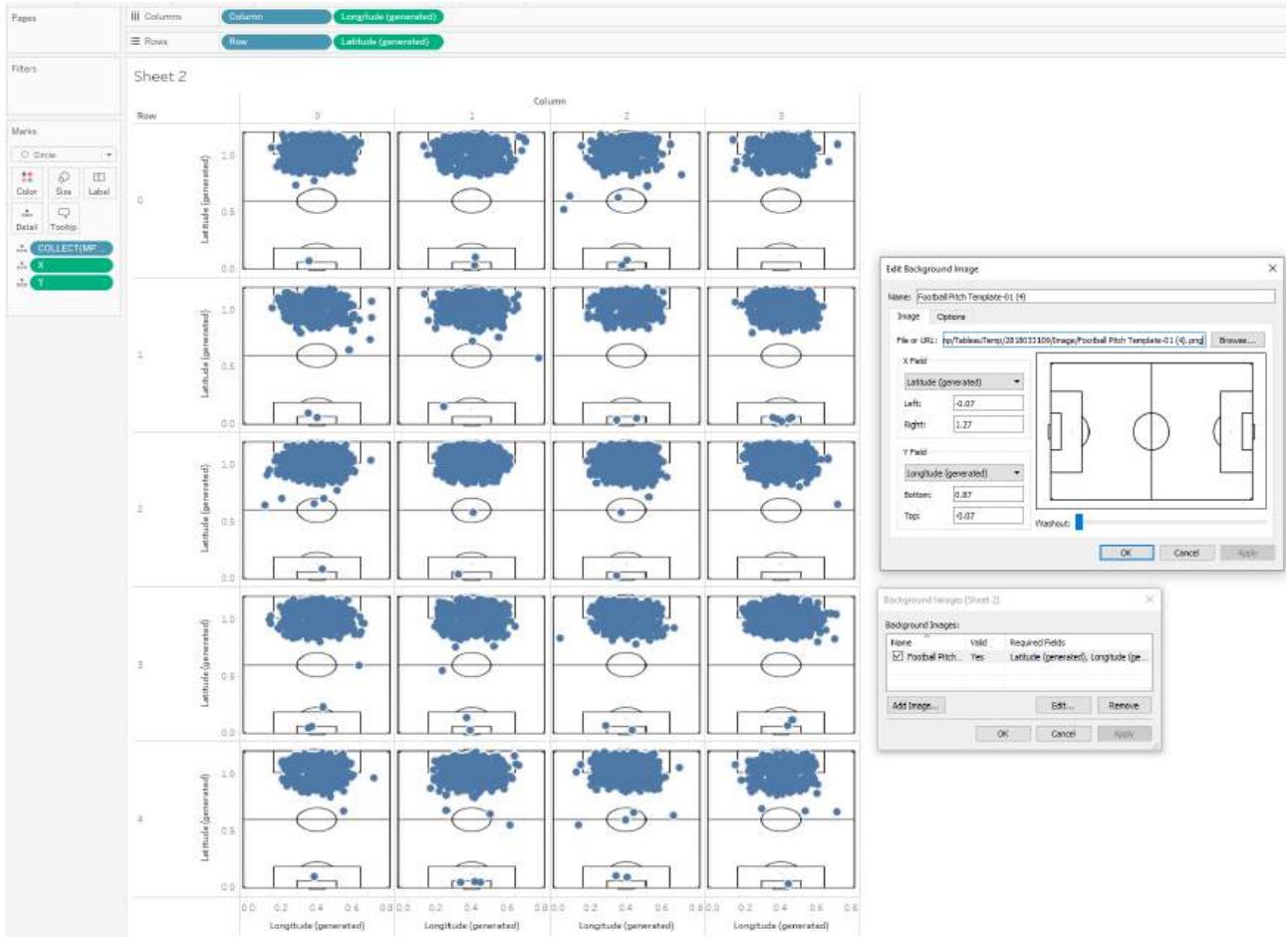
field, and export it to **player_shot_data_small_multiples.csv**. Again this file can be found in the repository if you want to compare the final output vs the original input. Now we have that small amount of transformation out the way we can go ahead and build our visualisation. **Tableau** Lets start our calculations with the standard X & Y scatter in make points. Reason being we can plot multiple layers this way, as well as add our background image. MP. Pitch MAKEPOINT(), Double click to bring it into the view.



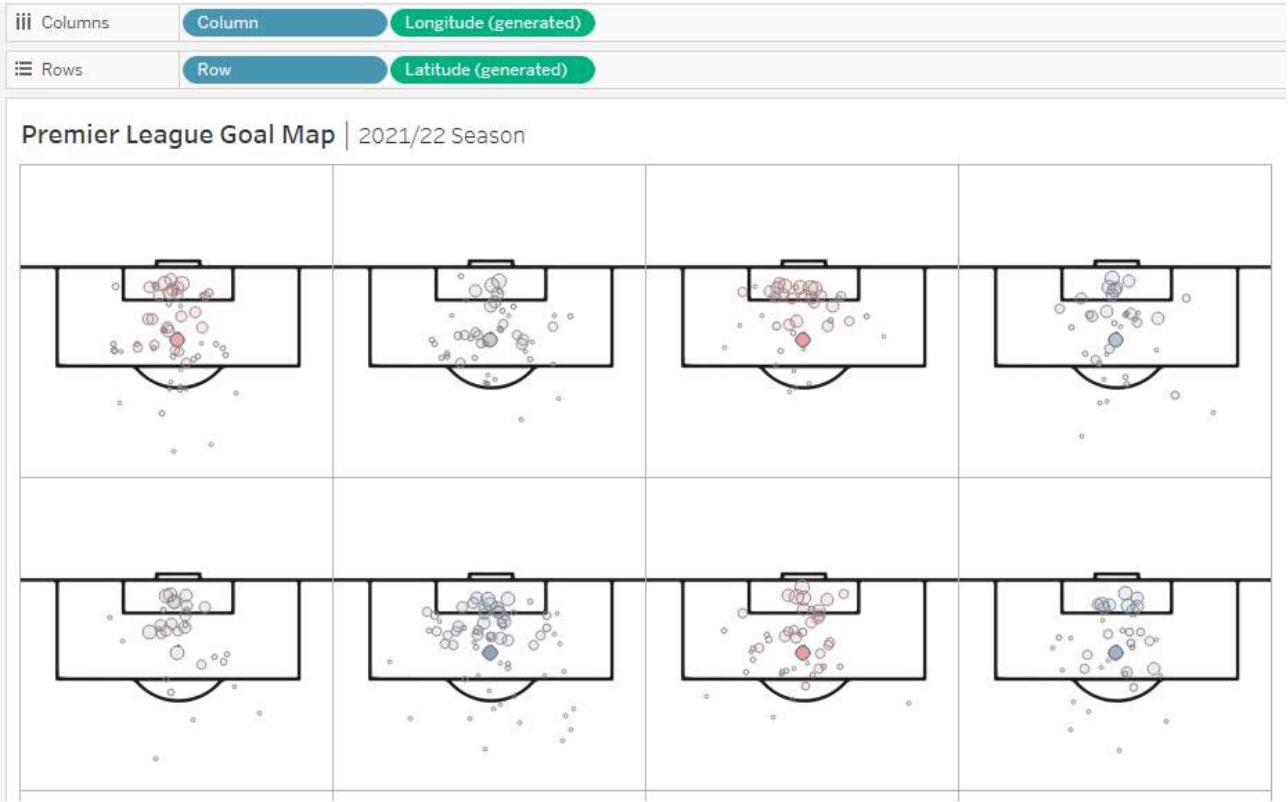
All our shots from the season! Looks messy without a pitch though doesn't it. Lets split our data by team. Drag Column onto Columns, and Row onto Rows. Make them both discrete dimensions. Make the marks a circle while adding X & Y onto the marks card also as dimensions.



Looks like blue paint splatter but we can soon fix this. Lets get the pitch in. A copy of the basic pitch can be found in the GitRepo.



Go to maps and turn off the back ground map. Then go to maps, background map. You can use the X & y Field co-ordinate mapping from my print screen if you have used my data flow, if not these will need to be configured based on the scale of your X and Y data. Tips to getting this correct. 1) Make sure you are using latitude and longitude in your x and y field. 2) Play around with the options whether you want to lock the aspect ratio and show the entire image. Because I only wanted a zoomed in version of the pitch I don't have these ticked myself. 3) On your dashboard, turn your map options off to stop the pitch zooming in. We can edit the axis of the latitude (i.e y-axis) to give some breathing space to the viz and rescale the pitch to amount of pitch we want to show. Same goes for the longitude (i.e x axis) Lets do some cosmetic changes, including tooltips, axis reshaping and colour.

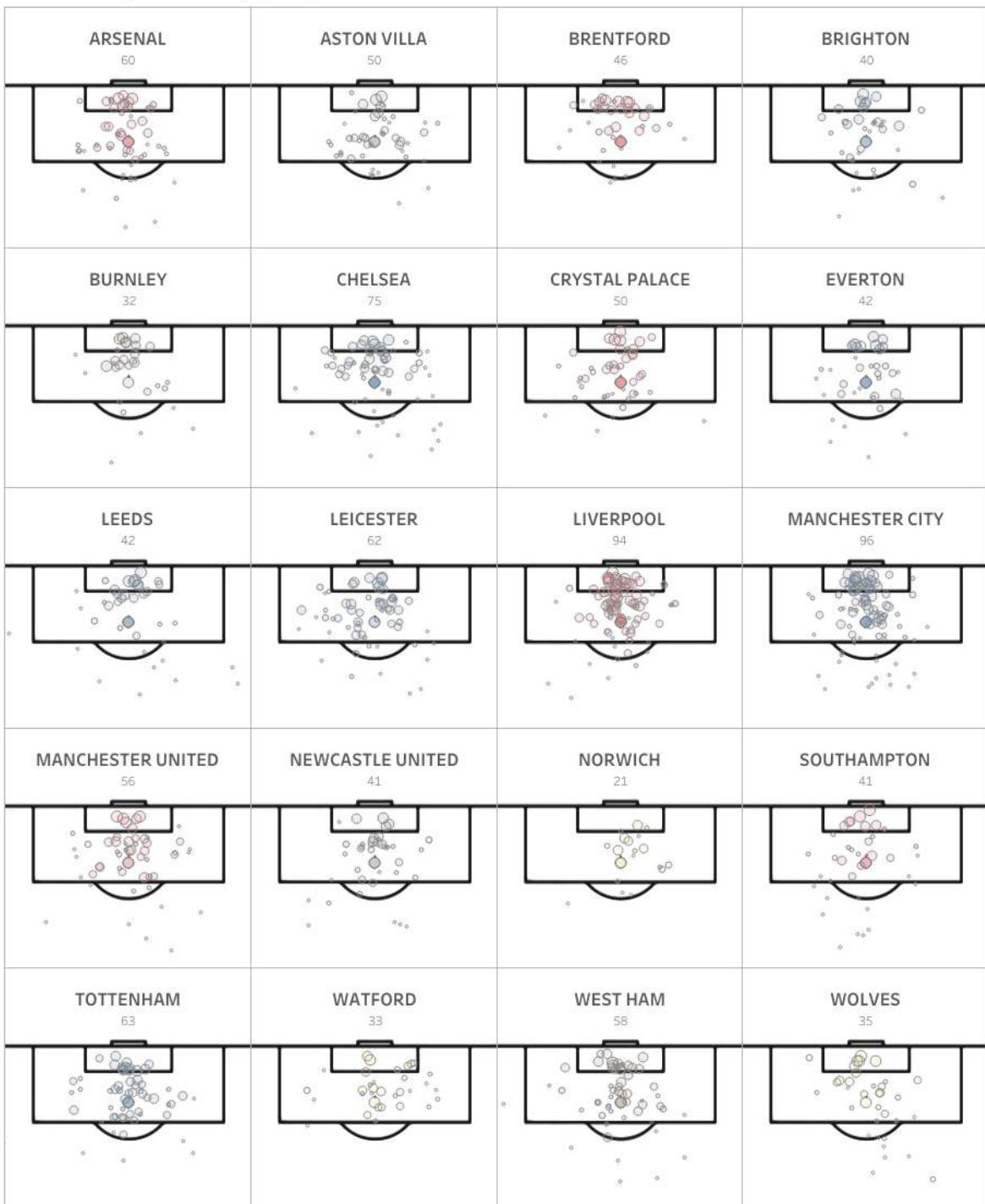


You will notice in the above print-screen I've made some space in my chart now for some Headers and where i'll add the number of goals. (result = Goal, on filter)



We create a placeholder with Y axis 1.3 in this case, and in the middle of the box with an X axis of 0.4. By dragging the team name onto the view and the fixed calculation of number of goals per team we can add in this final additional context to finalise the visual. Just remember to turn back on your map to drag the extra layer back on!

Premier League Goal Map | 2021/22 Season

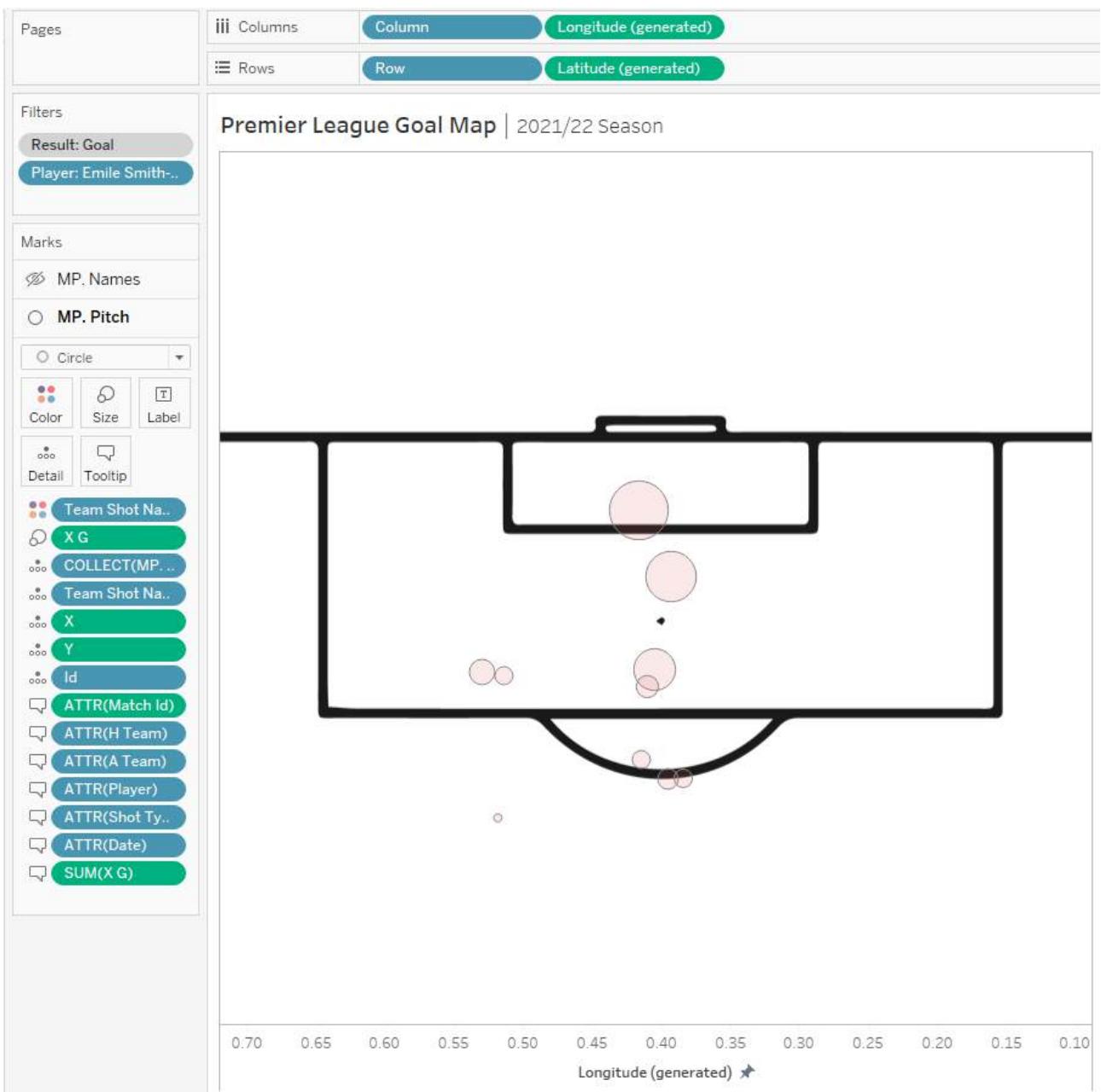


*Excludes Own Goals

All goes well, you should have your final visual. Truth be told we could still make this a lot prettier with a few more formatting concepts, but I think it's an adequate enough starting basis for people to run with the idea. **Some final things to consider** It's worth checking your data is correct with a cross reference. I ended up using [footystats](#). I find this important to see if I really have an accurate number of goals on labels. It instantly made me realise that I have excluded the own goals from the dataset. (I'm fine with that for now, but it would take extra work to amend whether you want this showing on the team it was for, or the team that conceded the own goals shot map) Hence why some discrepancies exist when comparing to the league table of Goals For. Another way to check is to actually use the understat dataset, For example I can take a player like Emile Smith-Rowe, This made me realise I had to actually mirror my x axis points.



As seen below.



Hopefully that has been useful as to how to build small multiple soccer shot maps. You can also see how I experimented previously with hardcoded calculations in this [hex shot map visual](#), and this [NBA visual](#). **LOGGING OFF, CJ**

TABLE HIGHLIGHTING IN TABLEAU

Hi all, Been a few weeks now since the last post. I've been really switching off on the weekends bar the odd #SportsVizSunday round-up which has been great for the soul, and the opportunity to get out for some good runs and gym sessions. I honestly probably feel the best I have in my whole life, physically at least. That being said... I can't believe we are reaching the end of Q1 already, so maybe time to get back to website content. This week we will look to create a few examples of different ways to highlight a table in Tableau! Yes you heard right, b2b tutorials on this website with content vaguely useful for work related activities and not just how to build whizzy chord charts. As always the resources can be found using the icon navigation under the title. Before we start, here are a few of my favourite table designs on Tableau to date. Autumn Battani - [Manufacturer Table](#) – A stunning way of ‘pinning’ a choice from the table to the top of the visual. phData – [KPI Dashboard for Executives](#) – Love the different chart types, the sort navigation as well as the see more detail ability. Also a nice way of having pagination at the bottom of the viz! Side note but how cool is the pop out navigations that reset the remaining views. Samuel Parsons – [State Sales](#) – Simple, clean and a great way of visualising different metrics. Had to include this one as its inspired at least 5 more visuals since then across the community! Luke Stanke – [Table with Map Layers](#) – Wanted to include this as an example of how you can really take your design to the next level introducing more layers. Anyway onto todays far less glamorous content. How to create a simple highlight table. Lets start simple

The screenshot shows a Tableau interface with a data source named 'Sheet 3'. The columns are 'Region' (Central, East, South, West) and 'Sub-Category' (Accessories, Appliances, Art, Binders, Bookcases, Chairs, Copiers, Envelopes, Fasteners, Furnishings, Labels, Machines, Paper, Phones, Storage, Supplies, Tables). A calculated field '00.Highlight' is being defined with the formula:

```
IF [Region] = 'East'
then 1 else 0
END
```

A tooltip for the 'ABS' function is open, stating: 'Returns the absolute value of the given number.' An example provided is ABS(-7) = 7.

We can highlight column by creating a calculation that takes into account the specific column, and using a highlight diverging colour between 0 and 1.

The screenshot shows a Tableau interface with a data source named 'Sheet 3'. The columns are 'Region' (Central, East, South, West) and 'Sub-Category' (Accessories, Appliances, Art, Binders, Bookcases, Chairs, Copiers, Envelopes, Fasteners, Furnishings, Labels, Machines, Paper, Phones, Storage, Supplies, Tables). A calculated field '01.Highlight' is being defined with the formula:

```
IF [Region] = 'East'
or [Sub-Category] = 'Binders'
then 1 else 0
END
```

A tooltip for the 'ABS' function is open, stating: 'Returns the absolute value of the given number.' An example provided is ABS(-7) = 7.

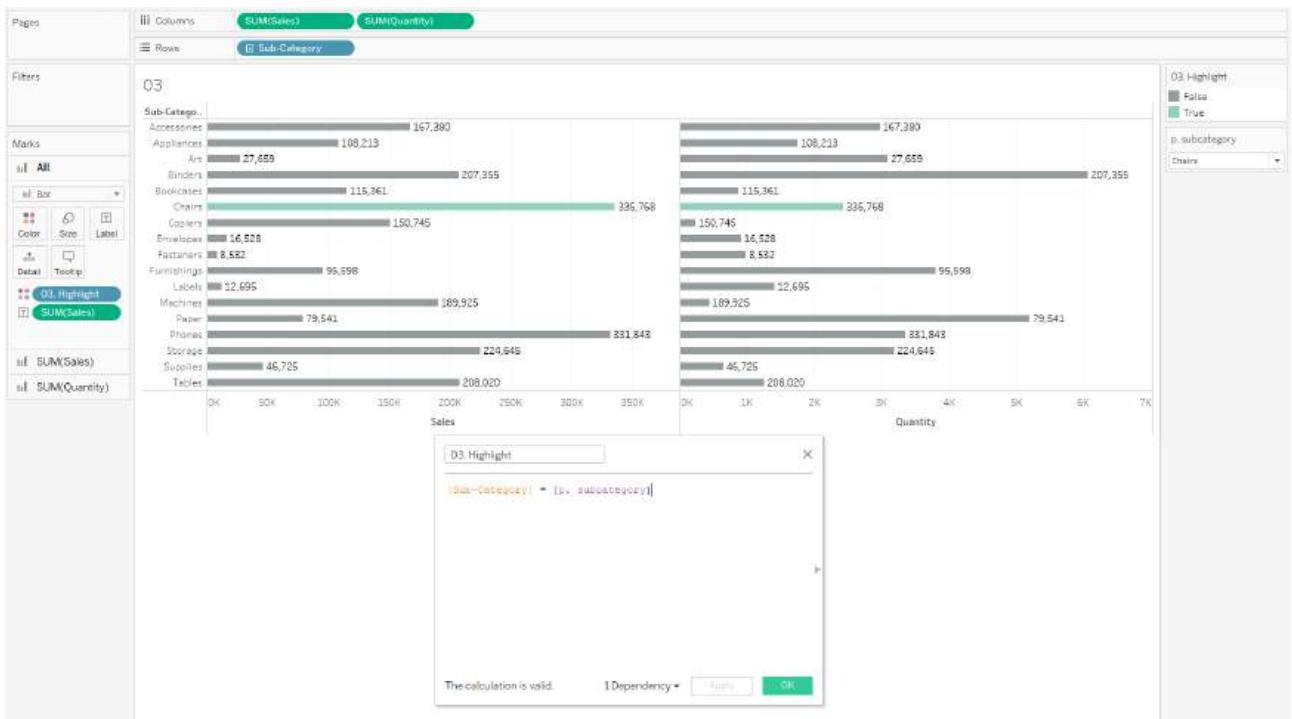
Of course, we can update this to also look at rows, like in the above example looking at both Binders or East.

The screenshot shows a Tableau interface with a data source named 'Sheet 3'. The columns are 'Region' (Central, East, South, West) and 'Sub-Category' (Accessories, Appliances, Art, Binders, Bookcases, Chairs, Copiers, Envelopes, Fasteners, Furnishings, Labels, Machines, Paper, Phones, Storage, Supplies, Tables). A calculated field '02.Highlight' is being defined with the formula:

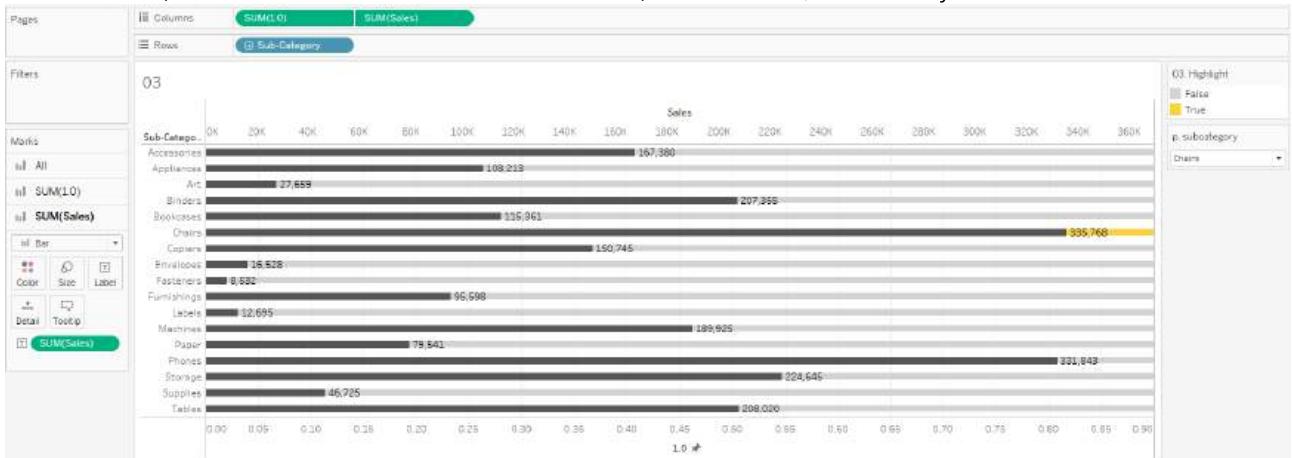
```
IF [Region] = p_region
or [Sub-Category] = p_subcategory
then 1 else 0
END
```

Parameters p_region and p_subcategory are set to 'Binders' and 'East' respectively. A tooltip for the 'ABS' function is open, stating: 'Returns the absolute value of the given number.' An example provided is ABS(-7) = 7.

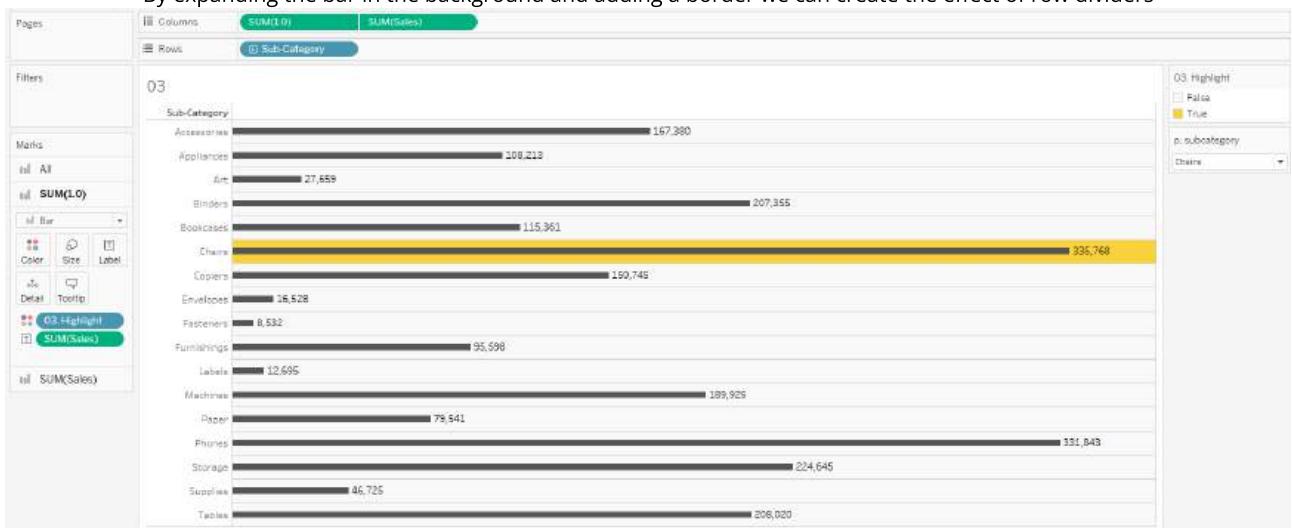
Then next logical solution is to have these colourings based on parameters – see how **p. subcategory** and **p. region** now drive the colour scale. So far nothing hugely ground breaking. But what happens when we want to look at something other than text tables? Well of course we will need to create ‘fake’ placeholders for each of the respective metrics. Let’s keep our subcategory on rows but now look at profit and sales as bars.



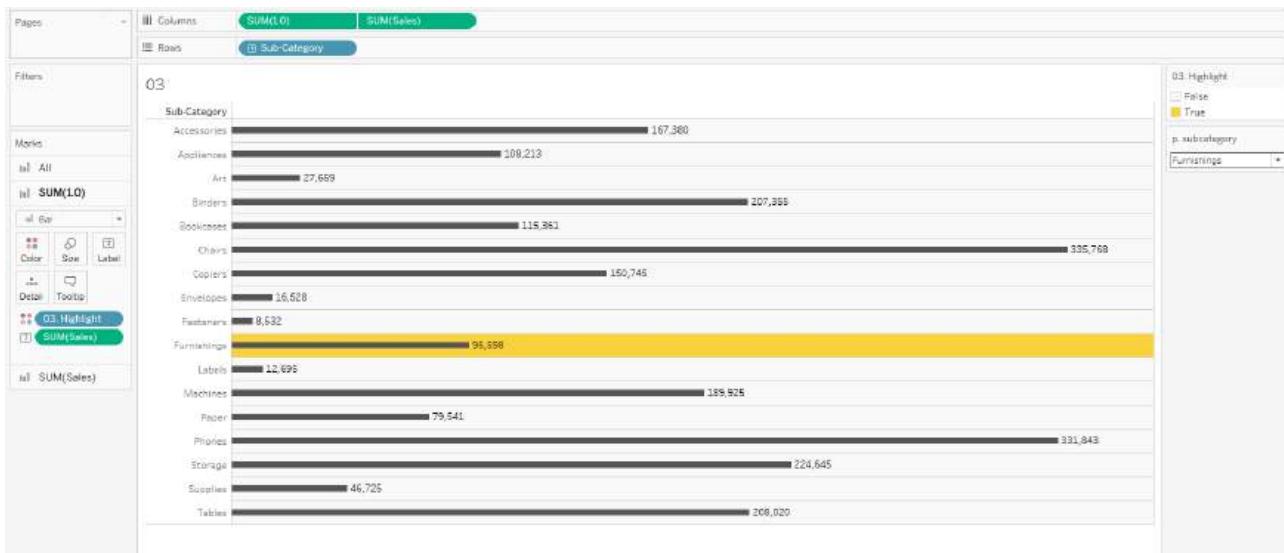
So adding the values to the columns and making the marks as bars now makes the bar the colour of our highlight instead of the row in the background. What can we do to bring back our functionality of highlighting the background of the row? Well lets just take sales as the example for now. Let's remove the colouring from the pill on our sales, and bring in a new field to columns of 1.0 (Fix the axis between 0 and a value less than 1.0) Dual axis these, no need to synchronise – of course.



By expanding the bar in the background and adding a border we can create the effect of row dividers



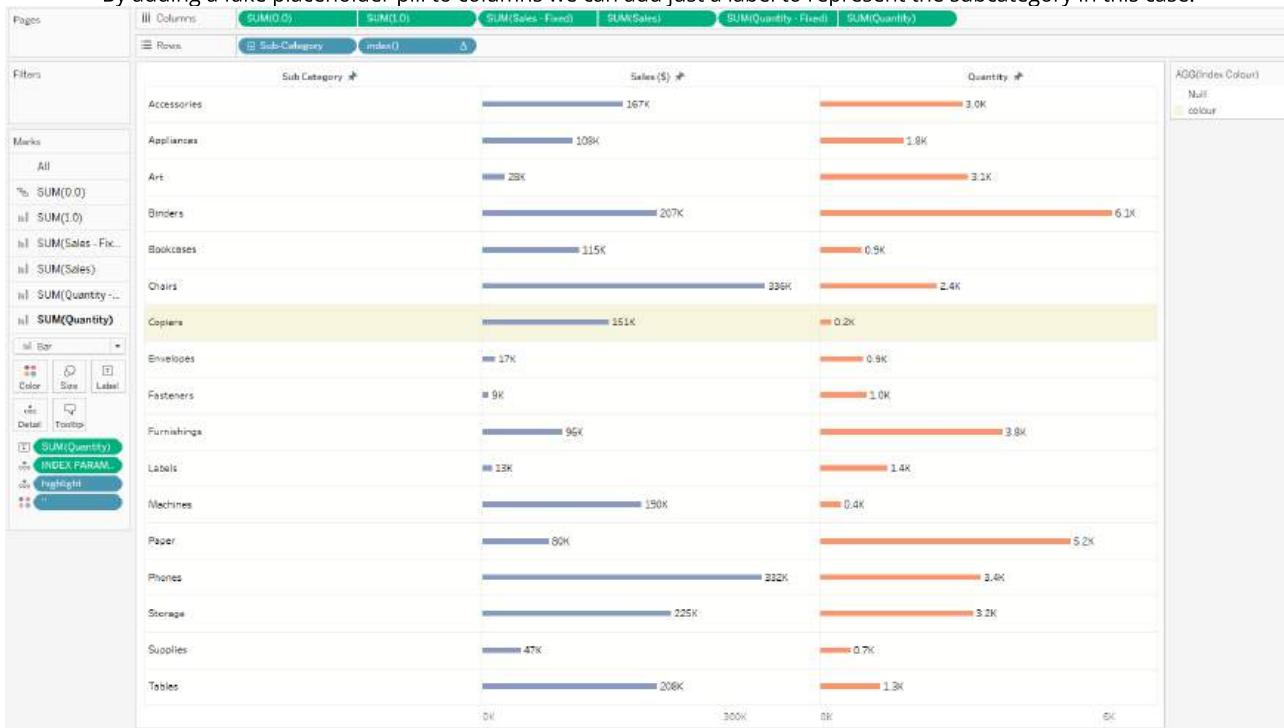
Changing the parameter now updates the chosen highlighted row



We can still take this further though, as it doesn't allow for the subcategory name to be highlighted at the moment.



By adding a fake placeholder pill to columns we can add just a label to represent the subcategory in this case.



The final things we can do are:

- Replace the 0.1 highlight bar actually with a fixed maximum bar to allow us to leave in an axis at the bottom if we wanted.

- We can edit our axis marks at the top to just show title headers
- We can add an action to our dashboard that changes our parameter based on the click to highlight the specific row.

Edit Parameter Action X

Name Insert ▾

Source Sheets
Highlight Dashboard Run action on
 Chart Hover
 Select Menu

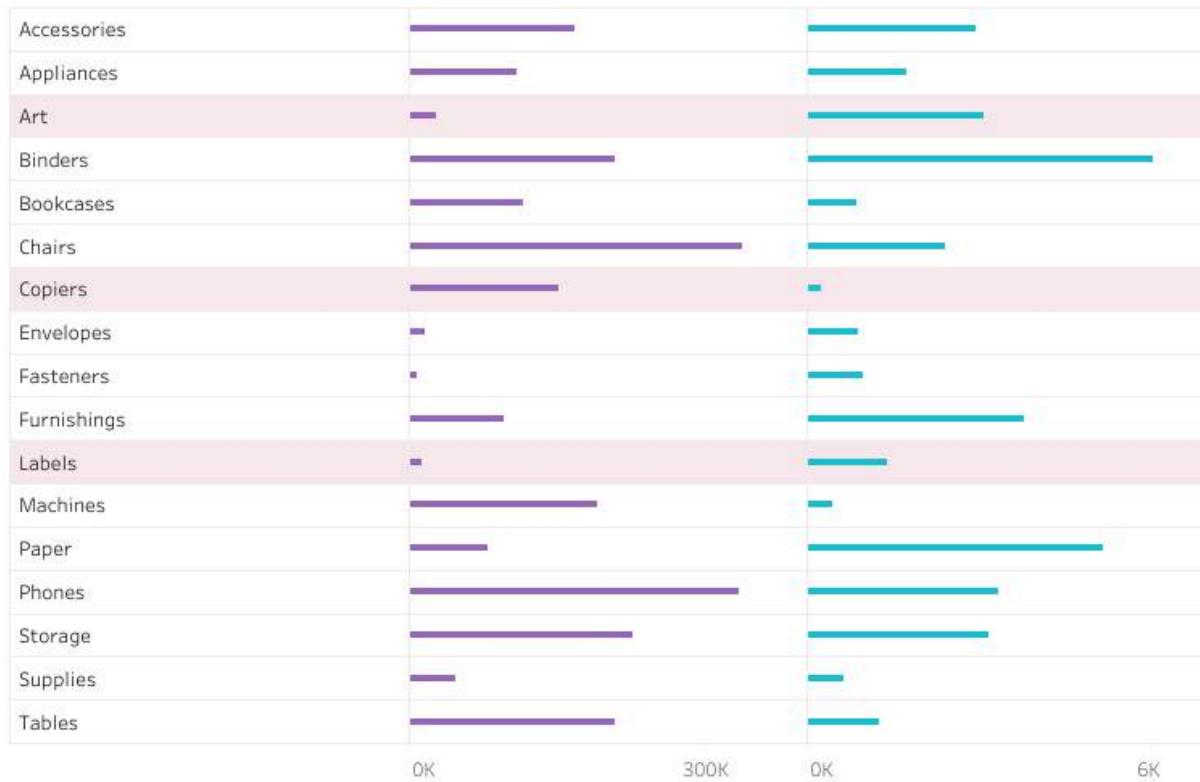
Target Parameter
INDEX PARAMETER Clearing the selection will
 Keep current value Set value to
 1

Source Field Aggregation
index() (Sa... ▾ Sum ▾

Cancel OK

Finally, we don't just have to use a parameter to colour or tables. We can control them using sets too. Check out my alternative method which allows the user to highlight multiple rows using a set action. You can find the [workbook here](#).

Sub Category Table Focus | Select To Highlight Row / Click Prompt To Remove



Going further:

- Create the select action to click once to highlight, click again to unhighlight as oppose to using a menu feature.
- Create a view with multiple views that can colour depending on a column or a row highlight
- Create a highlight row and column matrix that has the specific value that is both row and column as a darker shade than the rest of the row and column.

That's it for this week. Catch you all soon. **LOGGING OFF, CJ**

CALENDAR BUILDING WITH CHRIS WESTLAKE

Hi all, Building calendars in Tableau has been quite a popular trend as of late. You may have seen [Autumns content tracker](#), and Lindsay's [VOTD customised calendar](#). Both so elegant in design. Well with that in mind, I'm delighted to have Chris Westlake joins the blog today to talk through how to create a calendar in Tableau. Chris faced the challenge to upgrade one of our accounts calendars and made some major improvements too it, so I'm glad he is able to share a sanitised version with us today. All the resources (dataset and copy of the dashboard) can be found at the links at the top of the page under the title. I hope that you can easily implement this calendar style into your organisation. Chris, over to you.

Dates I Should Remember

A reminder of [] birthdays and [] anniversaries in February 2023

Select Month
February 2023

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
			01	02	03	04
					Mark	
06	07	08	09	10	11	12
			Chris' Mum			
13	14	15	16	17	18	19
20	21	22	23	24	25	26
	Mairi's Mum					
27	28					

@Westlake_CJW

CW: I recently started a new job, and one of my first tasks was to improve the aesthetics of an events calendar. I was amazed by some of the techniques used in the workbook already, but also added some features of my own. My immediate thought looking at the workbook was to use map layers to achieve the desired result in one sheet, but sometimes business presents us with restrictions that we can't control. This tutorial will create the same calendar view, this time using data relating to birthdays and anniversaries in my family that I am simply terrible at remembering! **Step 1 - Data** You'll need two tables in your data source for this

1. Events

This is the events that you want to show on your calendar, with the dates. In this example we have 3 fields in here: Event Date, Event, Person

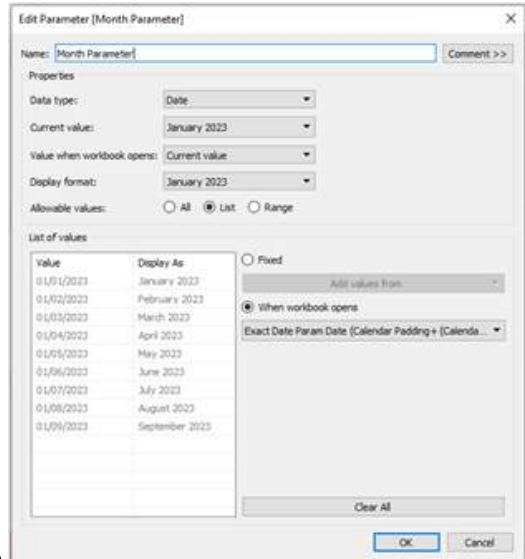
1. Calendar padding

This is simply a list of all the days in the year so that each day will feature in the calendar



These tables can be joined together

via a relationship on the date field **Step 2 - Basic calendar** We are going to use two worksheets in this tutorial: one for the background calendar layout, and one featuring text of events. The first step is to create a filter so that the view only displays one month at a time. To do this, we will create a calculated field using Event Date that returns the date of the first of the month of each event which ensures all events in the given month will still be in the view. *Exact Date Param Date*
`MAKEDATE(DATEPART('year'),DATEPART('month'),1)` Next, we create a date parameter which will populate with the values from our calculated field each time the workbook opens (I find it strange that no one in my family has a birthday or anniversary later in

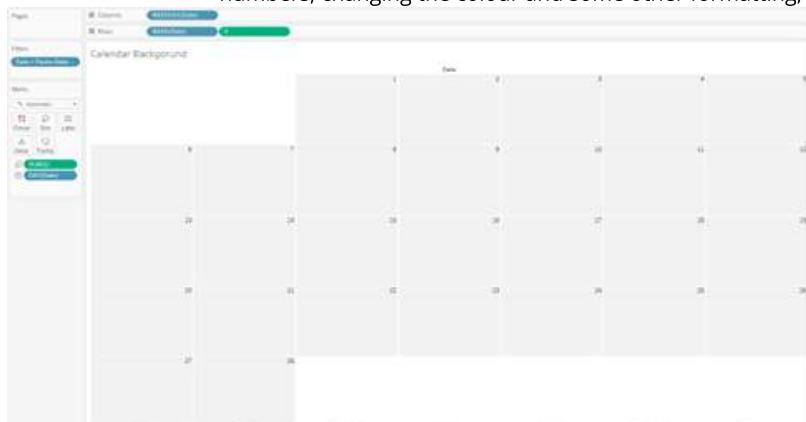


the year than September, but some things you can't control!

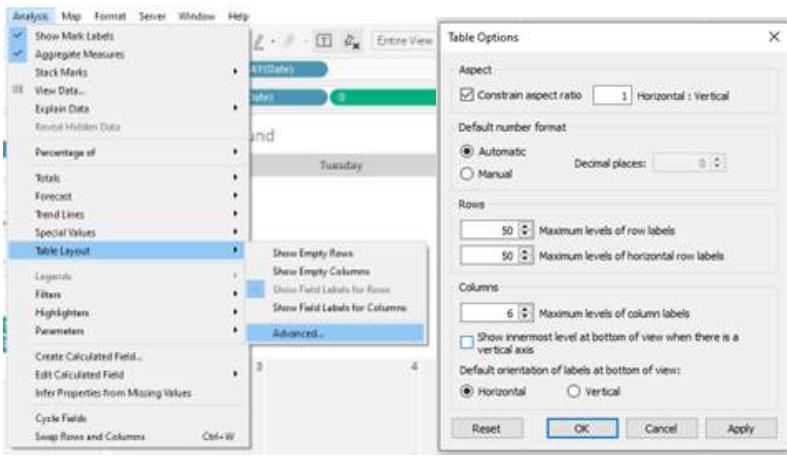
Finally, we can create the calculation to be used as a filter. We add this to Filters and set it to be TRUE $Date = Param Date$ $MAKEDATE(DATEPART('year'), DATEPART('month'), 1)$ = With that out of the way, we can start building the calendar view. Bring Date onto Columns at weekday level, and onto Rows at week number level. To give us maximum flexibility when building the background of the calendar, we are going to add a dummy field onto Rows. To do this, double click on the rows shelf and simply type a zero (0). Set this to be a dimension instead of a measure and you will see the marks card will have chosen to show this as a gantt chart automatically. Your sheet should look as follows:



To create the boxes, we need these straight lines to grow. Increasing the size with the size pill makes them wider, but not taller. To achieve this, double click below detail on the marks card and type a 1. The resulting sum(1) pill can be moved onto size and will give us some height. Manually change the axis of our dummy pill on rows to range from 0 to 1 and you will see something recognisable as a calendar. After adding date numbers, changing the colour and some other formatting, we now have the following:

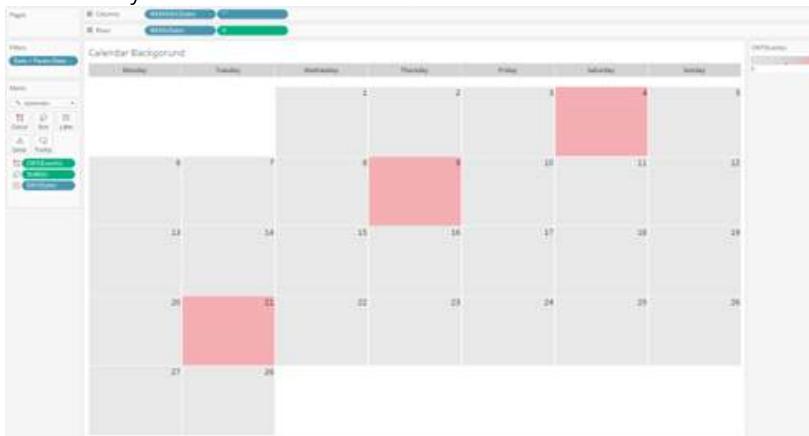


Obviously in a normal calendar, the days of the week appear at the top. To achieve this, navigate to Analysis > Table Layout > Advanced and deselect "Show innermost level at



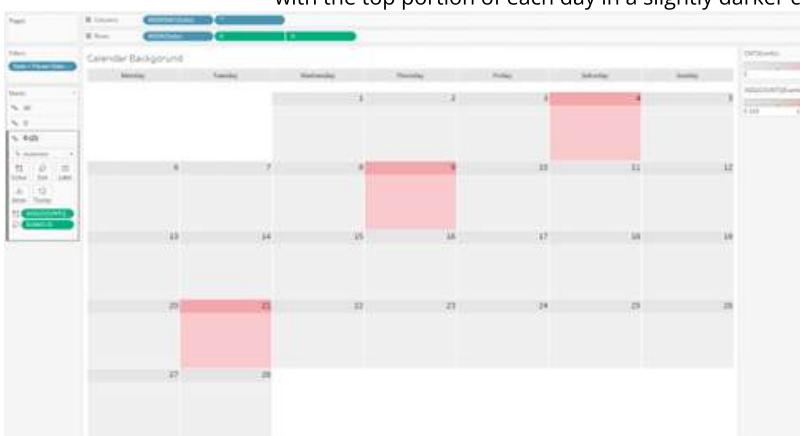
bottom of view when there is a vertical axis.

At this step I have also added another dummy pill to the columns shelf to give a little padding between the days of the week and the body of the calendar. We can also add the auto-generated Events (Count) field onto colour, formatting it to highlight days that have events on them. The sheet now looks like this and we have the basic outline of a calendar.



Step 3 – Banding on the calendar

We are now ready to add the darker band across the top of each day to help the rows stand out from one another. To do this, duplicate the dummy 0 pill and set to be a dual axis. On the marks card for the duplicated version, remove the text as we don't need this twice, and change the colours to be a shade darker so they stand out. When you change the colours you will notice it changes them for both axes. To avoid this, double click on the colour pill for the second axis and add 0.1 (so it reads $COUNT() + 0.1$). This creates a second colour legend that you will be able to change as distinct from the first. The second axis now covers the whole of the first because they have the same size element. Double click on the sum(1) pill and change it to sum(0.2). The 0.2 is right for the text size I have used – you may need to adjust it according to your own sizing. You can now see something close to what we want, but the band is at the bottom of each box instead of the top. This is because a gantt chart will always start at 0 and fill up to the size we have specified. To fix it, simply reverse the duplicated axis. Hiding the axis labels, we have a two-toned calendar with the top portion of each day in a slightly darker colour than the rest.



Step 4 – Text layover

Adding text to appear where we want it is tricky on this same page, especially where we have multiple events happening on the same day. For this reason, we are going to create another sheet with just the event details to float over the top of the first sheet in our dashboard.

There is a lot of crossover here with building the calendar layout so I will simply highlight some differences. **Difference 1:** Instead of Date on the marks card we are using Event Date so that the tooltip will only show for days with events on. **Difference 2:** The discrete `WEEKDAY(Date)` pill has become a continuous `ISOWEEKDAY(Date)` pill. It is continuous to make lining up the sheets easier when we come to building the dashboard, and `ISOWEEKDAY` so that the week starts counting on a Monday instead of a Sunday. **Difference 3:** In order to show the full month, we will need to show empty rows and columns. The filter has been added to context so that the rows are filtered before empty ones are shown in Tableau's order of operations. **Difference 4:** The index()

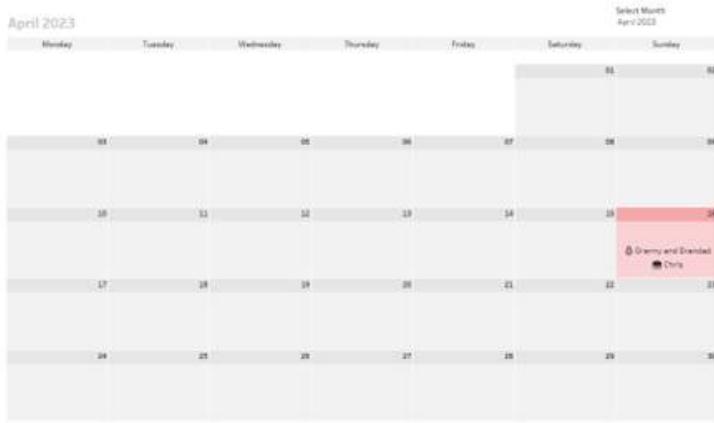
pill on rows has been added so that when there are multiple events on one day they appear in a column rather than on top of



one another.

now have our sheets built and are ready to combine them into a dashboard. Add a title and don't forget to add the date parameter so that you can change which month is displayed!

Dates I Should Remember



Step 5 – Dashboard creation

We

together. It feels like a very robust solution and love the ability to highlight specific dates in the calendar and the nifty trick of reversing the axis for the red banner across the top. Do tag Chris in your own creations if you end up using his calendar view.

LOGGING OFF, CJ

GOAL SEQUENCES

Hi all, Kind of incredible we are already into March. These months really are rolling past. I've got a few days off this week, I'm excited to catch up with family and get a few workouts in before doing the Paris half mara this weekend.

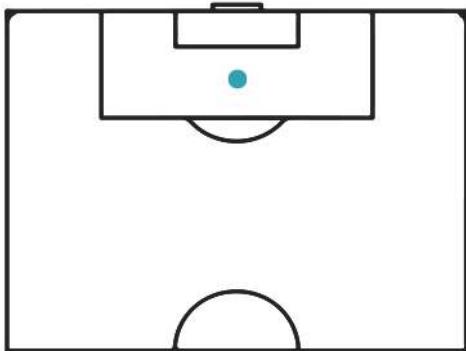
Goal Sequence Events

USA vs Netherlands

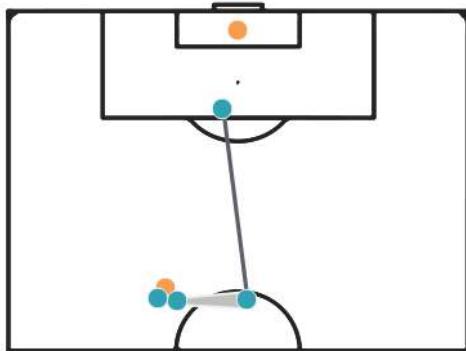
World Cup Final (2-0)

7th July 2019

Goal 1: Penalty Kick (61st Minute)



Goal 2: Long Carry (69th Minute)



Dutch defender Stefanie van der Graft fouled Alex Morgan while attempting to control the ball in the penalty area, which was left uncalled until a VAR review awarded a penalty to the United States. The penalty was scored in the 61st minute by U.S. captain Megan Rapinoe.

Eight minutes later, Rose Lavelle scored the second goal of the final for the U.S. on a solo run through the Dutch defence that ended with a left-footed strike from 17 yards.

Thought piece on CJ-Mayes.com

https://www.youtube.com/watch?v=_6VeZAZdffQ @ 1:17

Todays blog is a bit of a thought blog around goal sequencing using soccer data. I'll share some code, some data transformations in Tableau as well some overall thoughts in general when it comes to creating a visual of sequenced events. Here is an example from [Opta's Twitter page](#).



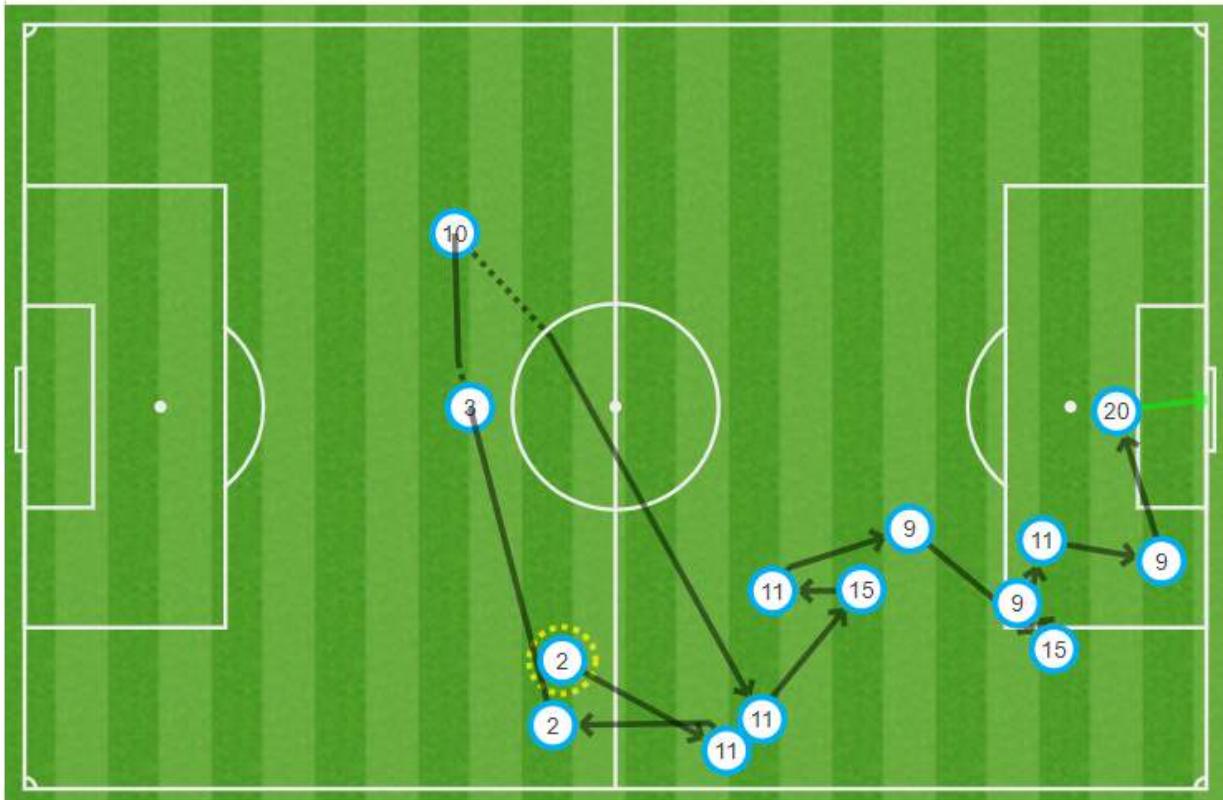
2

 Spanish La Liga
Real Madrid v. Alavés
 Santiago Bernabéu


0

79'20 Vinícius Júnior - Goal

Alavés events



So

what are goal sequences? Sequences are passages of play. They can be anything from defensive actions, duals, passes or shots. Sequences end when there is a stoppage in play. When a team has possession they will join these sequences up together. We can then take these events and try join them up, plotting them on the pitch. [Jonny Whitmore](#) wrote a great article on sequenced events at the Analyst last year. You can read it [here](#). Jonny goes into further detail around how a series of uninterrupted passes leading to a shot would be counted as a sequence. The shot ends this sequence but suppose the keeper saves it and pushes it out for a corner. Although the sequence ends, the possession continues until the opposition gains control of the ball. The original sequence and the sequence initiated from the corner would both be included under the same possession for the team taking the shot. Now I was pretty excited to go away and start looking at this in action, and realised instantly it was a little tough than first imagined.

```

import json
from mplsoccer import Pitch
from statsbombpy import sb
import pandas as pd
import numpy as np

from kloppy import statsbomb

match_events = statsbomb.load_open_data(
    match_id= 69321,
    # Optional arguments
    coordinates="statsbomb"
)
events = match_events.to_pandas()
events.to_csv('events.csv', index=False)

```

Above is a code snippet I used to access some simplified StatsBomb data, using the Womens 2019 World Cup Final as the example. So I gave it a go, you can check the visualisation and the code from the top of the page under the title header using the icons.

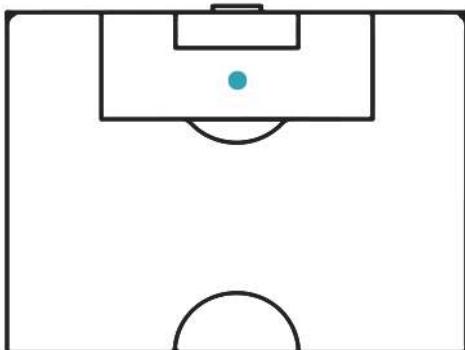
Goal Sequence Events

USA vs Netherlands

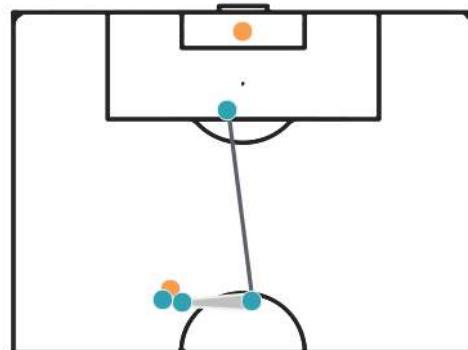
World Cup Final (2-0)

7th July 2019

Goal 1: Penalty Kick (61st Minute)



Goal 2: Long Carry (69th Minute)



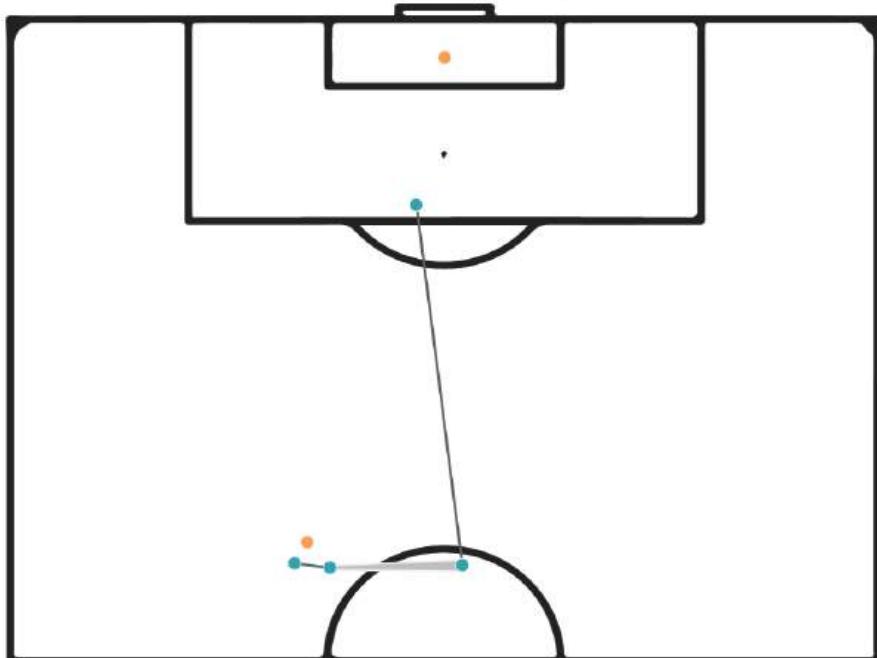
Dutch defender Stefanie van der Graaf fouled Alex Morgan while attempting to control the ball in the penalty area, which was left uncalled until a VAR review awarded a penalty to the United States. The penalty was scored in the 61st minute by U.S. captain Megan Rapinoe.

Eight minutes later, Rose Lavelle scored the second goal of the final for the U.S. on a solo run through the Dutch defence that ended with a left-footed strike from 17 yards.

What did trying to plot this data teach me? Football can be a scrappy game. Unless you have proper context behind the sequence of events, it's actually quite hard to understand sometimes from data on its own. In my example below you can see how Rose Lavelle scored the second goal of the final for the U.S. on a solo run through the Dutch defence that ended with a left-

footed strike from 17 yards.

Goal 2: Long Carry (69th Minute)



I've attached the summary video of the game showing the different actions on the dashboard, this actually helped me understand the data more having seen the game.



USA v Netherlands | FIFA Women's World Cup France 2019 FINAL | Full Match Highlights



It's really hard to show a sequence of play when possession changes rapidly, during duels. For example, I had looked previously at some games that scored from a corner that had multiple duels in it. Really tough to visualise! The second thing is, some events need to be excluded from the sequence. You'll see in my data there will be additional rows of position of players.

Whilst we can string the sequence of events together, some carries and passes don't necessarily 'come alive' when visualising. I think this could be for a combination of reasons, like the number of moving parts. 22 players are on the pitch but if we only map those that are on the ball at a moment in time, we effectively lose a lot of context. On the flip side, how do we show continuous

pressure/challenges from players over a duration of multiple sequences? See how it becomes a little complex. Finally and I guess this is a general rule of thumb with data visualisation, does it showcase and truly reflect the story in the best means possible? **So when do goal sequences work well?** Well this example of Mo Salah taking 11 seconds from box to box was a nice example.

Goal Sequence.

Mohamed Salah (36')

Liverpool vs. Everton
Premier League 2022/23

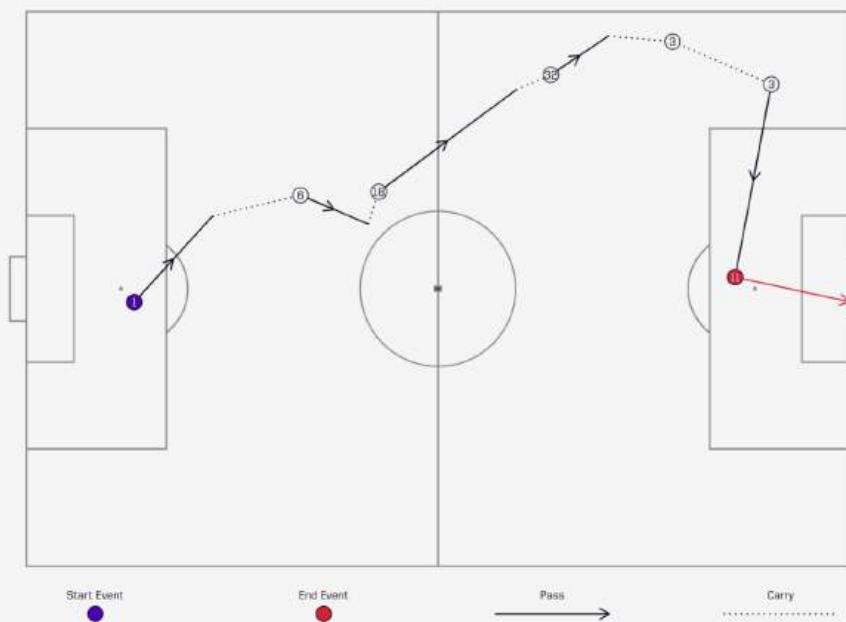


StatsBomb
Data Champions.

What this visual and the one below from the analysts blog have in common is the way of showing passes vs carries using dashed and block lines. (dashed lines are a little hard to do in Tableau, less so in Python/R etc i'm sure.)

Martin Ødegaard goal (43:50)
Arsenal 2 - 1 Tottenham Hotspur | English Premier League 2020-21

 The Analyst



Simple in design, beautifully presented and is able to provide context to the different carries, passes and part of the pitch used in building up to the goal. But sequences can also be used at the aggregated level too. Whilst these goal sequences provide pretty pitch charts, we can also start to analyse the impact of players through sequences. A few examples could be analysing turnover, build up of attack, number of successful passes in a row, direction of attack. You will see in a lot of games on telly some more simple metrics around possession, metres run and potentially duels too. Think of this as just that next step on. So, what tips do I have for those wanting to build the visual in Tableau?

- For lines you will want to union your dataset to create a start and an end node! If you read any of my other soccer blogs you will have seen anything with lines gets unioned, whereas player position or shot data can stay as it is, as you only need one mark to represent that player/action.
- In Tableau if you map the pitch onto the background you can benefit from creating multiple 'layers'.

- Check your data to make sure that you don't need to reflect the data depending if your co-ordinates are plotted on a full or a half pitch.
- Cross reference your data with summary video snippets
- Label passes and carries with different indicators as these are different actions. Dashed vs block works well, if not try a different colour or thickness of line. I attempted to use a comet style for the pass to indicate direction.
- Decide how much detail you want to put into your chart. Mine were very basic, but you could add action numbers, player details and use colour more effectively.
- Narrow your dataset down to start and end at appropriate events. For example a throw in or free kick would be a great starting point. A goal seems like a plausible end point. Perhaps you also may want to showcase possession up the the point of a goal scoring free kick. You'll see in my case, there was a penalty. Wouldn't it be nice to know what events happened for that to occur?
- Generally good to rank your events in your data.

Anyway, thanks for listening to my rambling. Excited to see where you take it. **LOGGING OFF, CJ**

PASS NETWORK IN TABLEAU USING STATSBOMB & ALTERYX

Hi all, Time is really flying this year, it has been a bit of a head down month or two and just finally coming up for fresh air. I'm particularly excited to go away with a few best friends and then go over to Paris to compete in my first half marathon. I hope everyone has had a somewhat settling few months and are making time for things that they enjoy. This week we return to some more soccer content. This time looking at how we can create Pass Networks in Tableau. I am a bit of a broken record when it comes to talking about different data tools, but I am on a journey of wanting to highlight to those in the sports community that dabble mainly in R and Python, that they can also create visuals in Tableau (and vice versa of course!) So today we will use python to retrieve the data (don't worry I will provide this for you in the GitRepo under the header) We will then transform the data using Alteryx, and finally we will export or data and showcase how it can be used to create a pass network in Tableau. Small disclaimer, there may be a few tweaks along the way that you want to add to your data process, but for now I've taken a few examples from the community and tried to merge them into a structured learning process. Okay? lets begin!

```
import json
from mplsoccer import Pitch
from statsbombpy import sb
import pandas as pd
import numpy as np

# Get competitions - using statsbomb open data
comp = sb.competitions()
comp.to_csv('competitions.csv', index=False)

# UEFA Women's Euro taken from competition csv
df = sb.matches(competition_id=53, season_id=106)
df.to_csv('matches.csv', index=False)

# Find a match_id required - England Germany 2-1 in the final, blog example
match = 3847567
match_events = sb.events(match_id=match)

# split locations into x and y components
match_events[['location_x', 'location_y']] =
    match_events['location'].apply(pd.Series)
match_events[['pass_end_location_x', 'pass_end_location_y']] =
    match_events['pass_end_location'].apply(pd.Series)

# split the shot_end_locations into x,y and z - some z missing
match_events['shot_end_location_x'], match_events['shot_end_location_y'], \
    match_events['shot_end_location_z'] = np.nan, np.nan, np.nan

end_locations = np.vstack(match_events.loc[match_events.type ==
    'Shot'].shot_end_location.apply(
        lambda x: x if len(x) == 3 else x + [np.nan]).values)

match_events.loc[match_events.type == 'Shot', 'shot_end_location_x'] =
    end_locations[:, 0]
match_events.loc[match_events.type == 'Shot', 'shot_end_location_y'] =
    end_locations[:, 1]
match_events.loc[match_events.type == 'Shot', 'shot_end_location_z'] =
    end_locations[:, 2]

# clean up and choose the columns wanted
events_df = match_events[[
    'index', 'id', 'duration', 'match_id', 'minute', 'second', 'pass_angle',
    'pass_height', 'pass_length', 'pass_outcome',
    'pass_recipient', 'pass_shot_assist', 'period', 'play_pattern', 'player',
    'player_id', 'position', 'possession',
    'possession_team', 'possession_team_id', 'shot_type', 'shot_outcome', 'team',
    'timestamp', 'type', 'location_x',
    'location_y', 'pass_end_location_x', 'pass_end_location_y',
    'shot_end_location_x', 'shot_end_location_y',
    'shot_end_location_z'
]]
events_df.to_csv('match_events.csv', index=False)

formation = match_events.loc[match_events['type'] == 'Starting
XI'].set_index('id', 'player')
df = formation[['tactics']]
data = df['tactics'].tolist()

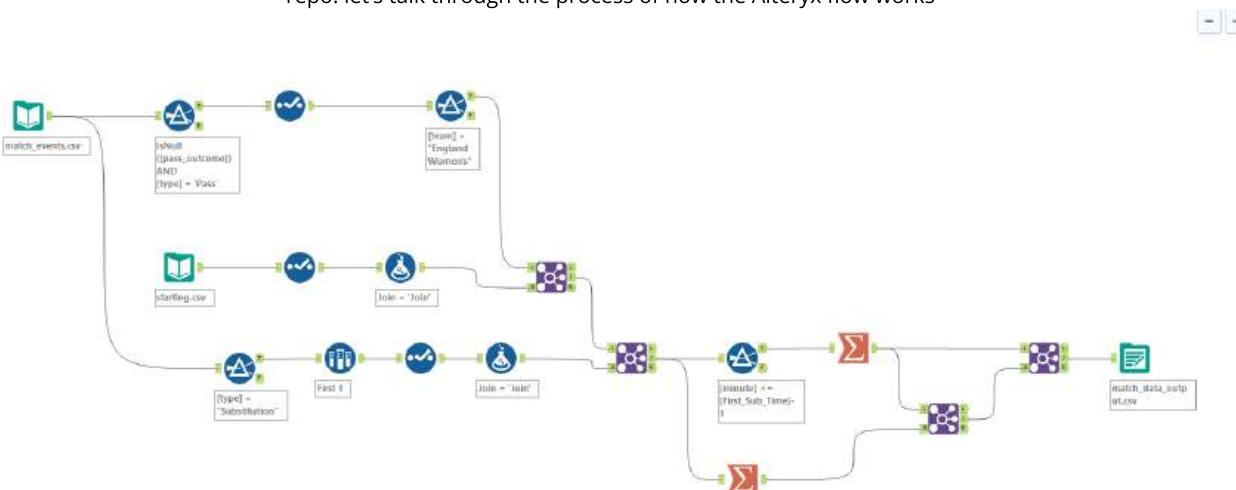
n = 0
starting = pd.DataFrame(columns=['Names'])

#print both teams lineup and export that to a CSV
for item in data:
```

All the csv files names = [player['player'][‘name’] for player in data[i][‘lineups’]] just look at the output that's df_a = pd.DataFrame(names, columns=[‘Names’])g? Well,

- starting = starting.append(df_a, ignore_index=True) to add bomb free data
- We search for ID to see what competitions are available, and then within that the different matches. You'll see I went for the starting.to_csv(‘starting.csv’, index=False)
- We do a few transformations here just to tidy up some location X & Y values, as well as choose what columns we want in the data frame Of course you can do this in Alteryx if you wanted but I had the code from a previous [Intro to Statsbomb tutorial](#).
- Line 50 onwards is really the only new part to the code where we are searching through a json to find the starting XI for each pitch = pitch = Pitch(pitch_type='statsbomb', pitch_color='#ffffff') these 11 players.
- line_color = '#555555' looks to create a plain football pitch background to save down for when we use it in Tableau.
- fig, ax = pitch.draw() Once you've run the code you will end up with the following outputs:
- fig.show()
- fig.savefig(‘fig.png’) Competitions.csv – All competitions in Statsbomb repo
- matches.csv – The associated matches from the chosen competition ID
- match_events.csv – All match events associated to the chosen match ID (Where we will get all our pass information from)
- starting.csv – The starting line up for each team
- fig.png – the background football pitch

So with little transformation I thought best to do the rest of the prep in Alteryx. You can find the Alteryx workflow attached in the repo. let's talk through the process of how the Alteryx flow works



Results - Filter (17) - Out - True

9 of 9 Fields | Cell Viewer | 120 records displayed | ↑ ↓ |

Record	minute	pass_recipient	player	team	location_x	location_y	pass_end_location_x	pass_end_location_y	First_Sub_Time
1	0	Millie Bright	Bethany Mead	England Women's	61	40.1	35.7	45.3	45
2	18	Lucy Bronze	Bethany Mead	England Women's	120	0.1	105.9	58.6	45
3	18	Lucy Bronze	Bethany Mead	England Women's	120	0.1	111.2	48.5	45
4	19	Lucy Bronze	Bethany Mead	England Women's	120	0.1	109.1	52.3	45
5	27	Lauren Hemp	Bethany Mead	England Women's	84.5	61.3	94.8	63.9	45
6	35	Lucy Bronze	Bethany Mead	England Women's	73.5	73.8	69.5	73.3	45
7	36	Lucy Bronze	Bethany Mead	England Women's	64.4	67.9	66.3	69.3	45
8	37	Ellen White	Bethany Mead	England Women's	116.3	66.3	105.6	46.4	45
9	40	Rachel Daly	Ellen White	England Women's	11.2	10	nn.2	44.4	45

The next part of the flow is all about creating average locations for each of the players. So we take their position when on the ball and take an average of those marks.

4 of 4 Fields | Cell Viewer | 11 records displayed | ↑ ↓ |

Record	Player X	Player Y	player	team
1	94.9625	38.7125	Bethany Mead	England Women's
2	92.733333	51.516667	Ellen White	England Women's
3	72.55	42.75	Francesca Kirby	England Women's
4	54.6	48.9125	Georgia Stanway	England Women's
5	49.13125	34.4875	Keira Walsh	England Women's
6	86.444444	28.433333	Lauren Hemp	England Women's
7	42.428571	24.609524	Leah Williamson	England Women's
8	58.033333	73.253333	Lucy Bronze	England Women's
9	12.000000	47.722222	Mary Alexandra Earps	England Women's

Finally we glue back in data to do with the recipient in order to know which player is passing to who! We also count how many times each player combination happened.

8 of 8 Fields | Cell Viewer | 69 records displayed | ↑ ↓ |

Record	Player X	Player Y	player	team	Recipient X	Recipient Y	Recipient	Count
1	94.9625	38.7125	Bethany Mead	England Women's	92.733333	51.516667	Ellen White	2
2	94.9625	38.7125	Bethany Mead	England Women's	54.6	48.9125	Georgia Stanway	1
3	94.9625	38.7125	Bethany Mead	England Women's	86.444444	28.433333	Lauren Hemp	2
4	94.9625	38.7125	Bethany Mead	England Women's	58.033333	73.253333	Lucy Bronze	3
5	94.9625	38.7125	Bethany Mead	England Women's	16.883333	42.733333	Mary Alexandra Earps	1
6	94.9625	38.7125	Bethany Mead	England Women's	38.025	58.325	Millie Bright	2
7	92.733333	51.516667	Ellen White	England Women's	94.9625	38.7125	Bethany Mead	1
8	92.733333	51.516667	Ellen White	England Women's	72.55	42.75	Francesca Kirby	1
9	92.733333	51.516667	Ellen White	England Women's	86.444444	28.433333	Lauren Hemp	2
10	92.733333	51.516667	Ellen White	England Women's	58.033333	73.253333	Lucy Bronze	1
11	92.733333	51.516667	Ellen White	England Women's	16.883333	42.733333	Mary Alexandra Earps	1
12	92.733333	51.516667	Ellen White	England Women's	54.233333	11.78	Rachel Daly	3
13	72.55	42.75	Francesca Kirby	England Women's	49.13125	34.4875	Keira Walsh	2
14	72.55	42.75	Francesca Kirby	England Women's	86.444444	28.433333	Lauren Hemp	1
15	72.55	42.75	Francesca Kirby	England Women's	54.233333	11.78	Rachel Daly	4
16	54.6	48.9125	Georgia Stanway	England Women's	92.733333	51.516667	Ellen White	1
17	54.6	48.9125	Georgia Stanway	England Women's	49.13125	34.4875	Keira Walsh	6
18	54.6	48.9125	Georgia Stanway	England Women's	86.444444	28.433333	Lauren Hemp	3
19	54.6	48.9125	Georgia Stanway	England Women's	42.428571	24.609524	Leah Williamson	2
20	54.6	48.9125	Georgia Stanway	England Women's	58.033333	73.253333	Lucy Bronze	7
21	54.6	48.9125	Georgia Stanway	England Women's	16.883333	42.733333	Mary Alexandra Earps	1
22	49.13125	34.4875	Keira Walsh	England Women's	54.6	48.9125	Georgia Stanway	4

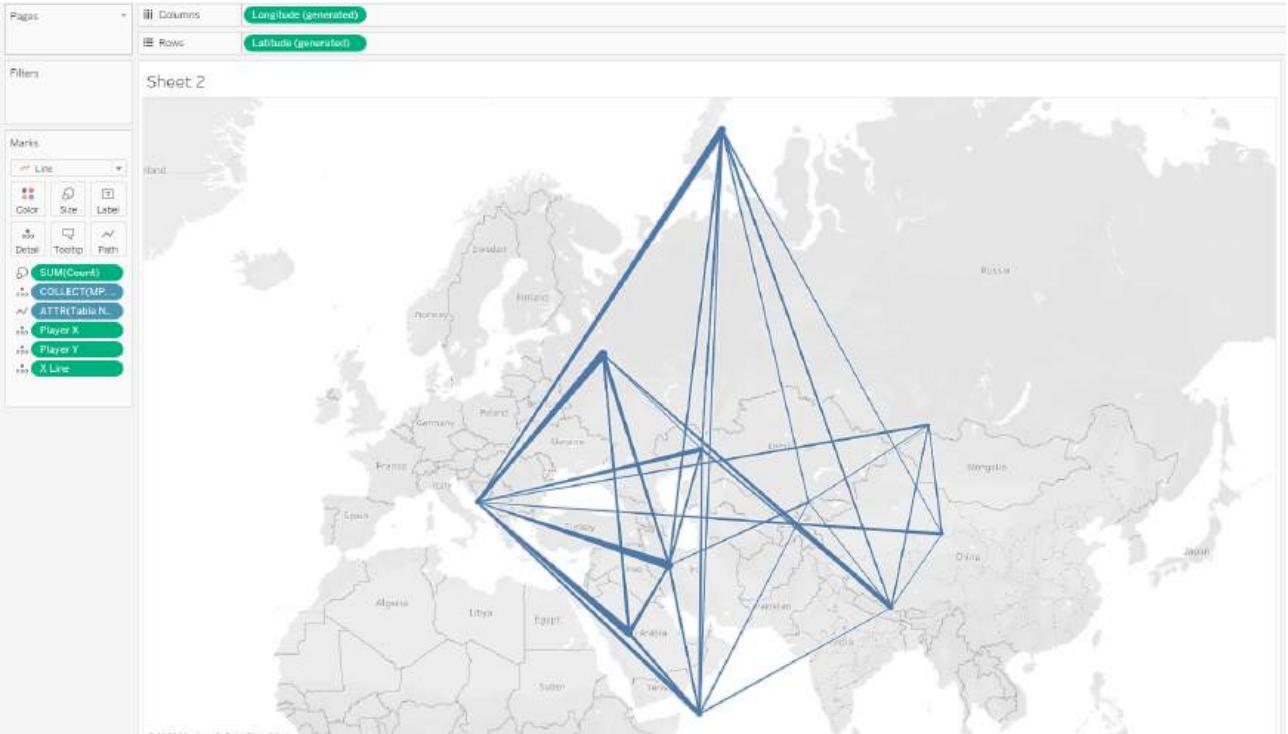
All goes well, you will end up with an output such as seen in the **match_data_output.csv**. From here we are ready to build in Tableau. This is the easy bit! Connect to the **match_data_output** and union it on itself by opening it up and dragging it on top of one another. Reason we do that is because to create our lines between the players.

match_data_output.csv is made of 2 tables. ⓘ

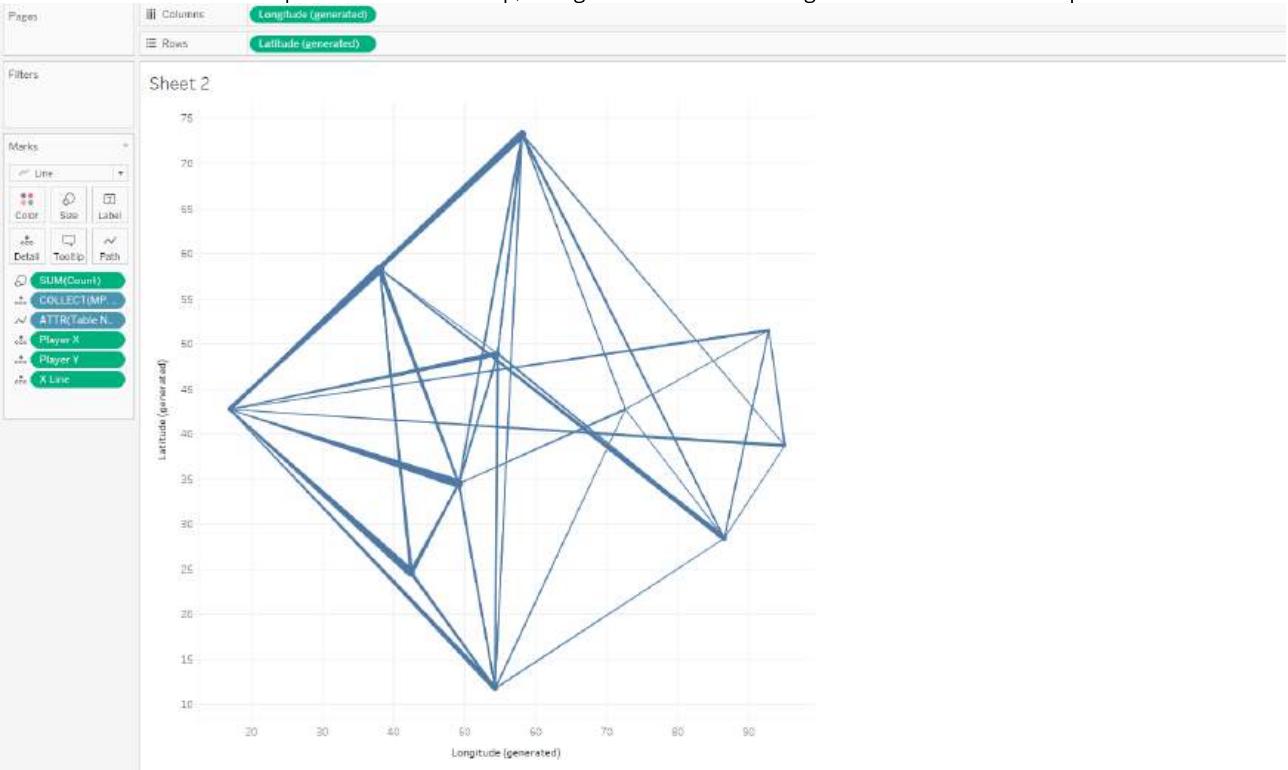
match_data_output.csv+

You will need the following calculations X

Line IF = 'match_data_output.csv' then else END Y Line IF = 'match_data_output.csv' then else END MP. Line MAKEPOINT(,) MP. Player MAKEPOINT(,) From here we are ready to create our chart. Double click MP Lines to bring it into the view, then add the required dimensions to the marks card. Make sure ATTR(Table Name) is on path. This is because where we have unioned the data we want to connect the points to create the lines!



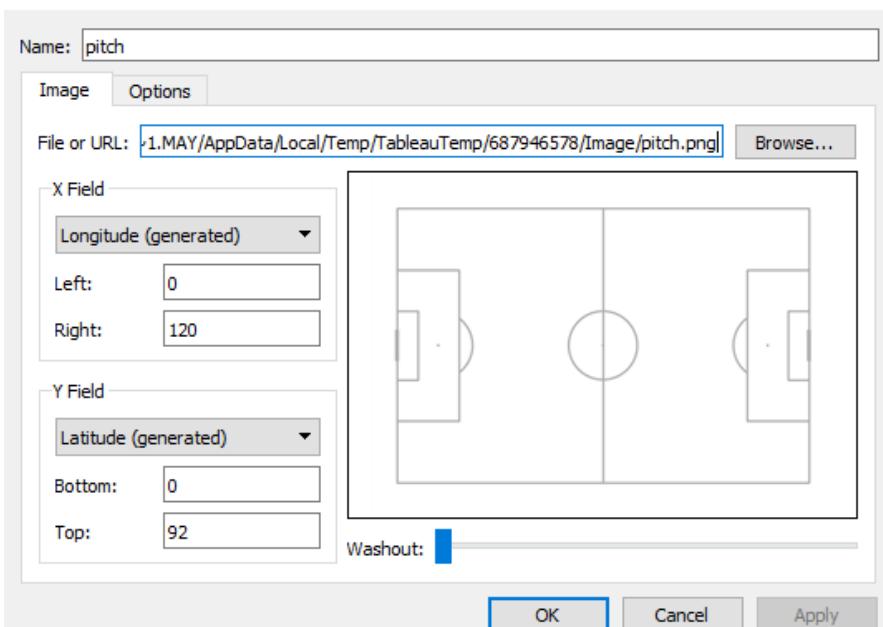
The next step is to turn off the map, background. We will reconfigure this with our football pitch.



Go to map, background images and configure the pitch. Make sure you get your longitude and latitude correct for your pass network. You can easily check you've got this correct because your goal keeper shouldn't be in the position of your right winger,

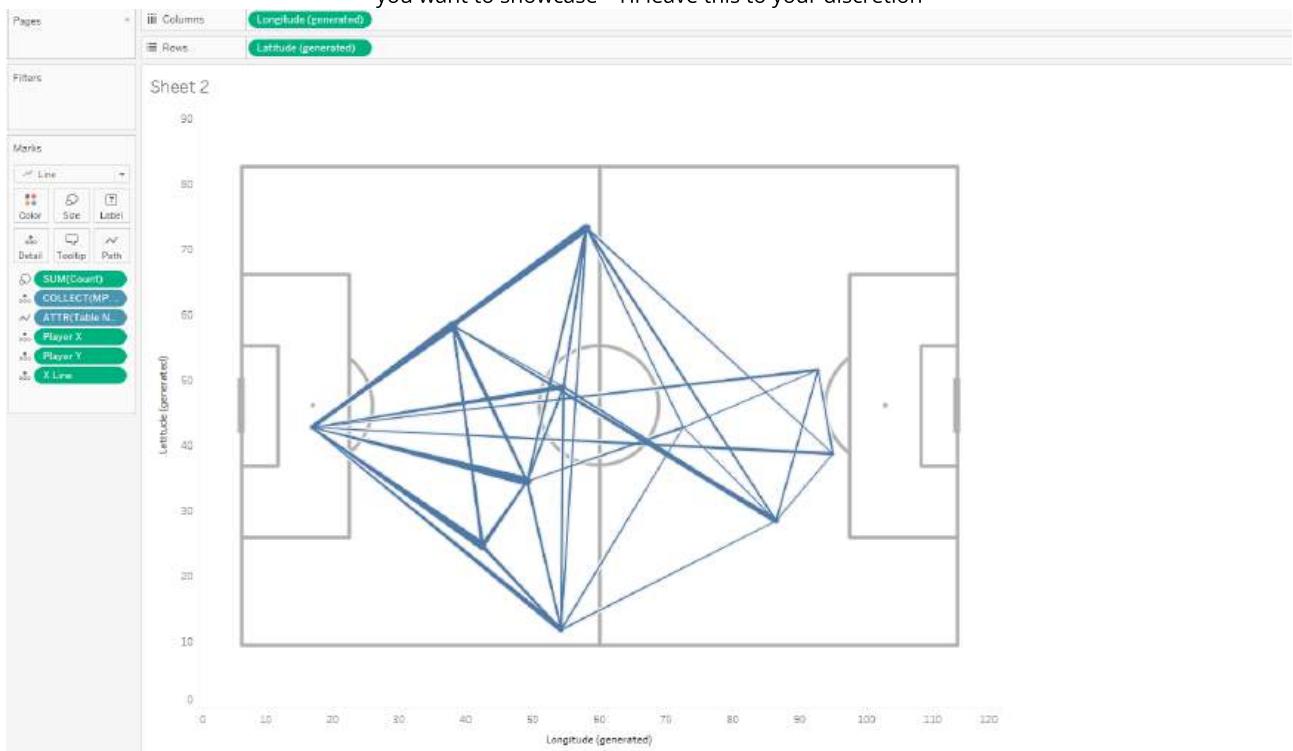
Edit Background Image

X

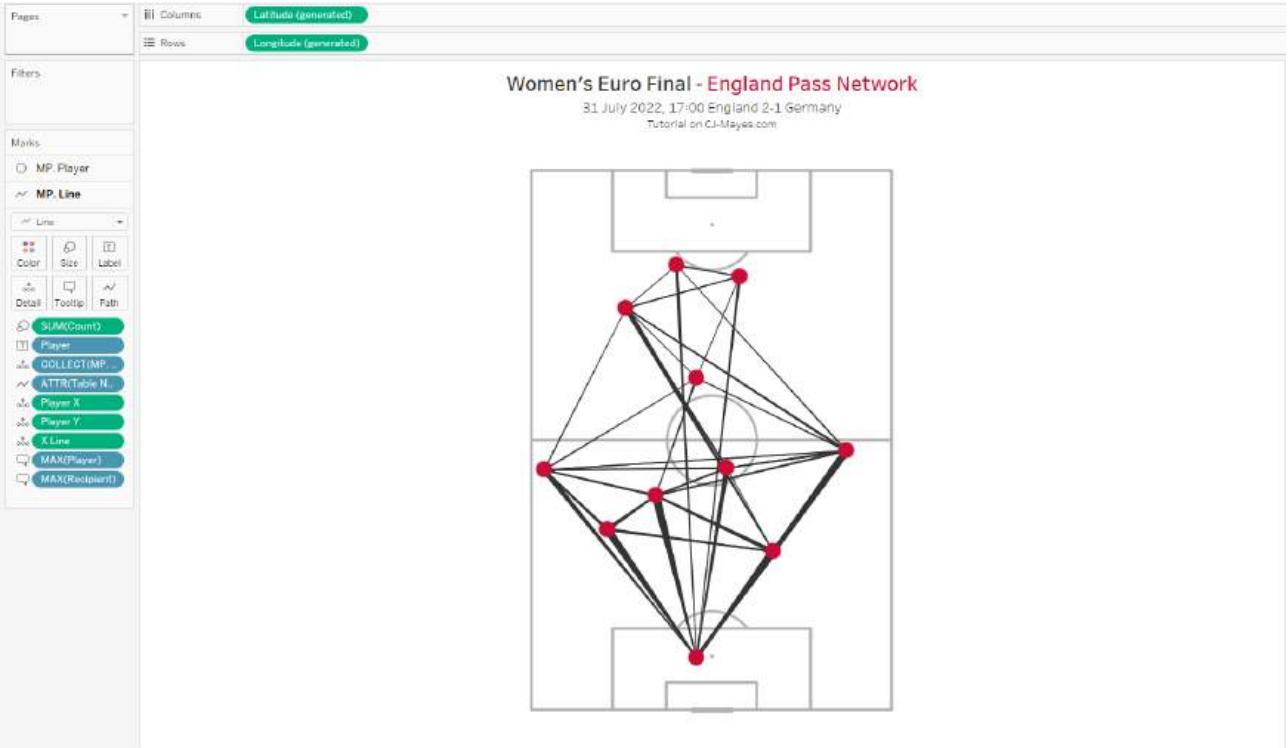


ha.

either add more layers, such as where I've introduced the overlay of player circles, perhaps even a few more design elements you want to showcase – I'll leave this to your discretion



By the end of it, you can have something simple like below.



As always all the resources are found in the GitRepo as well as the dashboard is downloadable from my Tableau Public profile.
Why not go further?

- Can you create a small multiple of all games leading up to the final?
- Can you add in jersey numbers into the code and Alteryx flow?
- Can you change the design of the pitch to be different colours?

Hope this has been a fun one to follow along to. Dabbling in python for retrieving data, Alteryx for transforming the data, and Tableau for showcasing the data. Catch you in the next one. **LOGGING OFF, CJ**

SPIEL DES JAHRES WITH BRITTANY ROSENAU

Hi all,

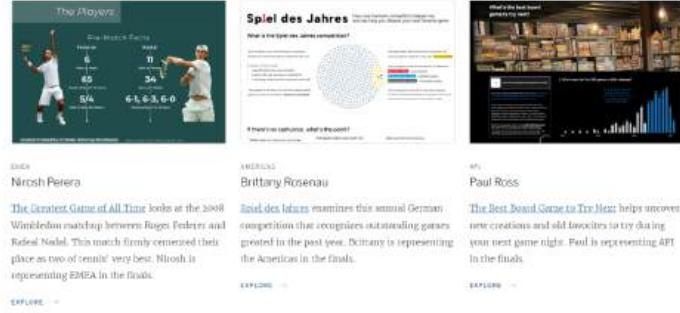
In December last year we got to see the results of the IronViz Feeder for 2023 and the Top 15 visuals from around the world. These included finalists Nirosh Perera from EMEA, Paul Ross from APAC and Brittany Rosenau from AMER. I'm so excited to see them battle it out on stage! Wishing them all the best of luck, and I would urge you to take a look at each of the winning entries.

Check the full 15 out [here](#).

Game over: 2023 Iron Viz Qualifier Round

Congratulations to our [viz-tastic qualifier participants](#). We are excited to announce the elite three who move on to the finals at Tableau Conference – May 9–11, 2023.

Want to learn more about the rest of our top 15? [Watch a replay](#) of our results show or [read the recap blog](#).



Whilst the three finalists are probably putting together their workflows, filming their introductions and settling their nerves prior to presenting live to thousands of people, I want to wind the clock back to look at what it took to get them in the first place. Now, I have the pleasure of working with Brittany over at JLL and we have a bit of a running joke about the JLL curse and always the bridesmaid never the bride when it comes to winning IronViz. So who knows we will see..... but for now I've had the chance to catch up with her to dive a little deeper into her feeder viz. I've really seen Brittany come into her own this year, frequently sharing new tips and ideas on the forums, promoting content for upvotes to help the product teams as well provide her own tutorials both internally and in the wider community. If you'd like to check out Brittany's viz you can find it [here](#). If you're not already following her, she can be reached on Twitter [here](#), and Tableau public [here](#).



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

Brittany Rosenau

Sr Analyst at JLL | Tavares, Florida, United States

Design Nerd, Aquarium Enthusiast, Percussionist, #VOTD

[Follow](#)

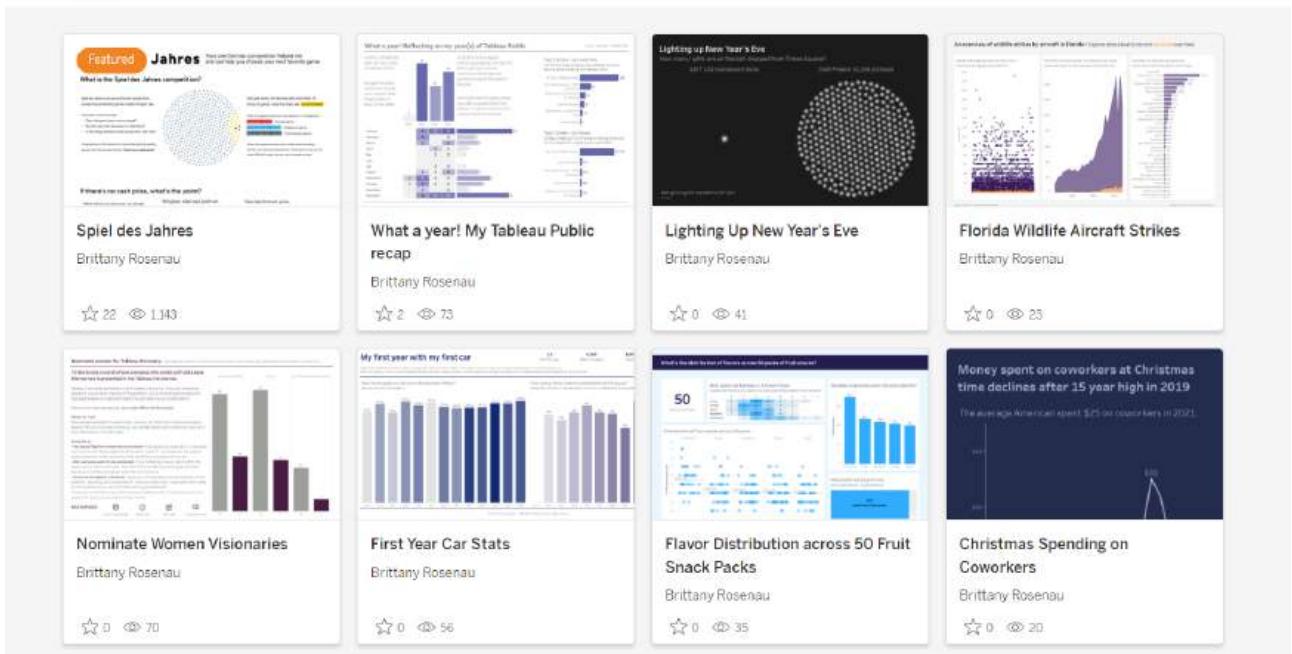
[Hire Me](#)

Vizzes 75

Favorites 140

Following 109

Followers 130



CJ: Hi Brittany, Congratulations, Let's start with the visual at a high level. Why did you land on visualizing Spiel des Jahres? Was this your first idea when the topic was released?

B: It was not my first idea – at first I wanted to visualize data from just one board game, but I struggled picking a game or coming up with a sort of story. So I took a step back and thought about why I had bought the games I owned in the first place. In 2020 I was living at home for a bit, and needed new games to play. Board games can be expensive, so I wanted to make sure I was getting a good value game that would be fun for everyone in my family. I noticed a number of highly rated games advertised that they won Spiel des Jahres, so I wanted to dig into not only why those games won, but also why they might appeal to a broad audience.

CJ: Data collection can always be tricky for personal projects from scratch. What did this look like for you in terms of data collection as well as data availability?

B: At one point during the build, I laughed to myself and thought one day I'll actually enter a competition where I'm not sourcing the data myself. Data collection for this was a lot of copy and pasting into Excel (I'm sure I could have learned something like Python to make this easier, but I had limited time).

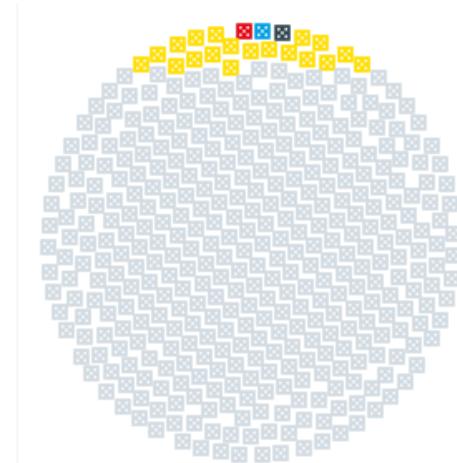
Once I had the game list, I need to enrich it with more data. I got most of this from the website Board Game Geek which has loads of stats. The tricky part there was to be selective- I was tempted to grab all the details but had to remind myself to keep focused and not fall into the trap of just showing a bunch of data without much meaning.

One piece of data I wish was more readily available was board game sales, whether that's in dollars or units sold. I was able to find some data on a handful of games from press releases, but I couldn't find solid sales data overall. I think that would have strengthened the story showing the impact of what winning the award does for a game.

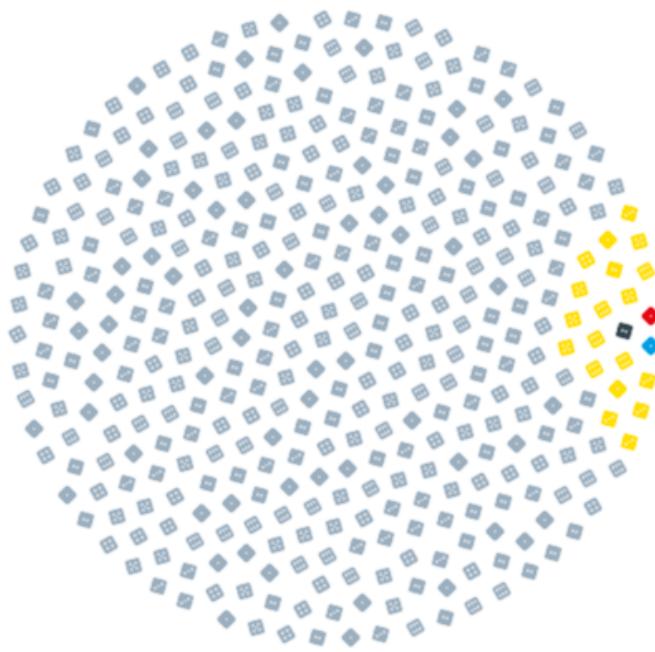
CJ: You have some really nice design effects that are so subtle but help elevate the visual. Elements such as the title having a chess piece, and your pudding chart including dice of varied rotation and numbers. Are these things that you will consider early on in the design of your work, or after multiple iterations?

B: The chess piece in the title was pretty early on. The Spiel des Jahres website has their title with a red letter I, and I wanted to put my own spin on it by incorporating a game piece.

The pudding chart went through several iterations. At first I had only one dice type, with all the same rotation and it just didn't look that great to me.



While it took more time to apply different shapes and rotations, I was really happy with how much more polished the final result looked. Big shoutout to [Lindsay Betzendahl's blog post on pudding charts](#), and her [collaboration viz](#) with Kevin Flerlage – they have a plum pudding chart with cheetah spots rotated at different angles. I was able to crack open how they did the rotated shapes and learn how to apply it to my own viz.



CJ: You ended up using just the primary colors, in addition to white black and gray.
Were these colors chosen with specific intent?

B: They were! From the start, I knew I wanted to work with a limited set of colors. A previous year I needed 8 colors across categories, and it was difficult to come up with that many that looked good together and were visually distinct. This time, I started by going to the Spiel des Jahres website. They used the red, blue, and dark grey for their awards and website, so I grabbed the hex codes and called it a day. Knowing I only had a few colors to work with helped me focus on keeping it as simple as possible.

CJ: From a storytelling perspective it was great that you introduced what Spiel des Jahres really was for those unsure. How important is it to give context before diving into the details? It's interesting to see how you transition from context to impact through to how it impacted you and the reader getting to select a game. Why did you choose this route?

B: A lot of the structure came from the sort of classic funnel method for writing papers in school – you start with something general, narrow in on specifics, and then reverse the funnel to zoom back out into something more general.

A big piece of feedback that I got out of the **Iron Viz Feedback Initiative** was I needed to actually get my explanations into the viz. I'd go into zoom calls and find myself talking about a lot of context that wasn't on paper yet – but since I wouldn't get the chance to talk to the judges, I needed to make sure whatever I thought people needed to know made it on the page.

The middle gets into more specifics – including details about games I purchased and why. I like vizzes where people show a personal connection to the data, so I rolled the dice and hoped people would enjoy anecdotes from my own experience.

Finally, the last part zooms back out to the general pool of games, with a sort of “so what” ending that lets people interact with and see themselves in the data. Maybe someone wants a game they can play on their own, or maybe a parent wants a game for their kids that doesn’t take forever to play. Bouncing off ideas with **Zach Bowders** led me to using another plum pudding chart to bring it back full circle (and thanks to **Nicholas Pillsbury** for helping me understand some functionality logic).

CJ: Your hand drawn images are very complimentary to the visual without being overpowering. Can you tell us a little more about the considerations taken in terms of placement, balancing against the charts and the style of these drawings. Also, How did you create them!?

B: A big part of why games win Spiel des Jahres is the build quality of the game itself, and I really wanted to show what some of these games looked like because that influenced my own purchases. But, I have to admit I was terrified of accidentally disqualifying myself by using copyrighted images that I didn’t have the rights to.



What I ended up doing was taking some product shots I felt were good representations of the game, traced the outline, and added my own embellishments. I actually used a whiteboarding tool, **InVision Freehand** for the drawings. I'd used the tool before in a work setting to make visuals for presentations, and it was quicker for me to use that than learn how to draw properly in another tool. What I really like is as you draw, it automatically smooths out the lines a bit. You retain the hand drawn look but it's a little more polished than what I could achieve with just a mouse.

CJ: I would say your pudding chart, stacked dice and even perhaps the gantt style bar aren't the standard types of charts. How come you landed on these? Were there other ways you considered showing the same metrics?

B: I'm not sure why, but I'm really drawn to unit charts. Some of it might come from work situations where data is aggregated, but stakeholders inevitably want to look at record level detail. The gantt bars, pudding chart, and stacked dice not only let me see the larger shape of the data, but also make it easy to inspect an individual element without using a table. I was lucky my data sets were small enough for that to work effectively. If I was working with thousands of records that would go out the window pretty quickly.

The real origin of the pudding chart was I hadn't ever done one, and I wanted to learn how to do it. Iron Viz – with the theme of win or learn, you can't lose, seemed like the perfect time to try it out, and it seemed to work well enough for what I wanted to show. I

hoped that by using it at both the beginning and end of the viz, the story would come around full circle (*ba dum tss haha*) and provide a sense of completeness or closure.

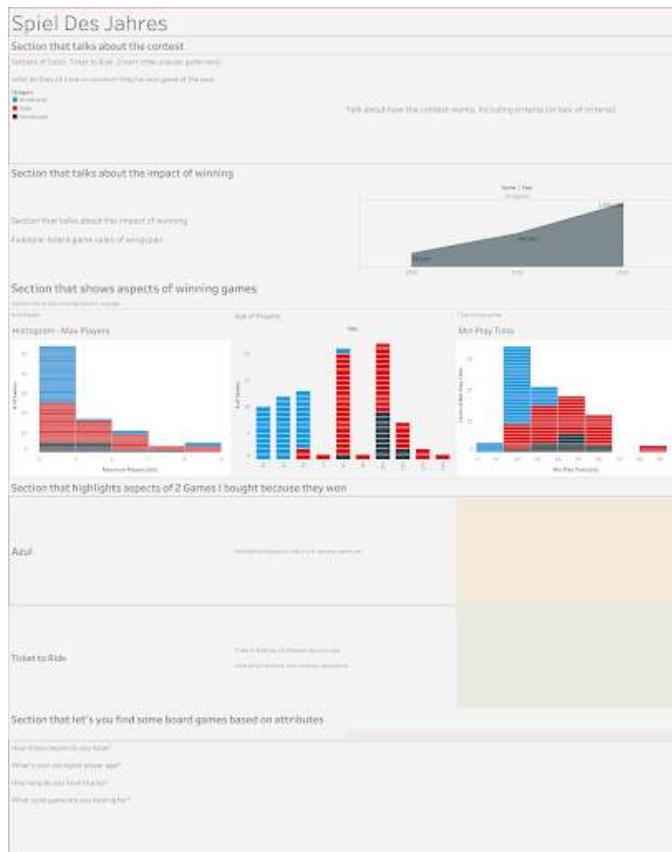
CJ: Do you have any previous iterations of the visual that you can share? I'd love to see the different considerations taken along the design process journey if you have any.

B: I sure do 😊 Last year I saved an image each time I hit a new iteration of my entry, and this year I tried to do the same. I find it helpful to have different versions saved in case I need to revert back, but also it's really satisfying seeing the progress throughout. My first iteration of the viz, I thought I was going to do a horizontal style viz. I blocked out the dimensions of the Ticket to Ride board game, and wanted to move the story along through the tiles of a board game. I abandoned that pretty early for a long form vertical viz.

From this....



through to this..



Playing with alignment, text and images,

Spiel des Jahres

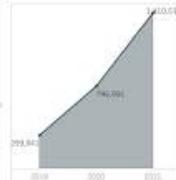
What is Spiel des Jahres?

Sometimes referred to as the Oscars of Board Games, Spiel des Jahres awards recognition for board games. Games published in 2022-23 that year are submitted, and a jury members that evaluate them. The jury consists of people that are living game experts (they play board games, card games, video games, etc.).

Nearly 400 games were submitted in 2022-23 (from 22 countries recommended).

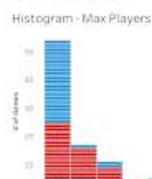


If there isn't a prize, why win?



What can I expect from a game of the year?

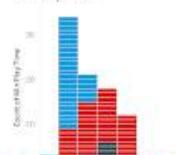
Event Description here



b. age (2)



Min Play Time



Highlight: 2 Games I've bought



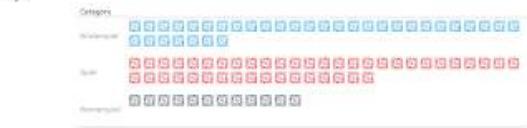
azul



Ticket to Ride

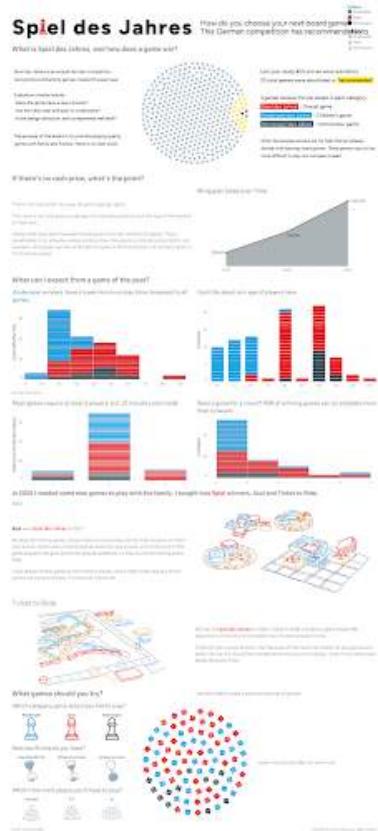
What games should you try?

Min Play Time:
All values.
Max Play Time:
All values.
Max/Min Played:
All values.
Minimum Played:
All values.

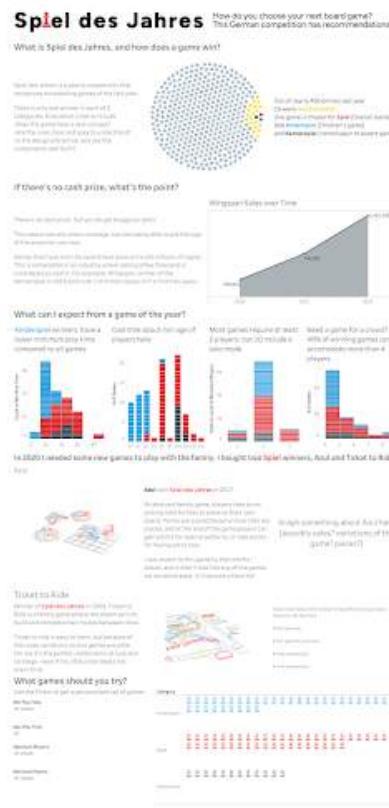


©2024 Artfully Board | ArtfullyBoard.com

Adjusting height and framing, introducing softer imagery,



And finally looking at flow and whitespace,



After this you'll see the final touches to reach the end visual on my profile. It's a bit of an iterative process.

CJ: Long form visuals can be a tough process in terms of laying them out on the page. Do you have any tips for formatting your viz and giving the visual alignment in charts, text and appropriate spacing?

B: Containers and blanks! For me, containers help keep everything organized. For anyone new to containers or intimidated by them, I'd recommend watching Curtis Harris's video "**things I know about containers**" as it completely changed how I develop.

Putting in blank placeholders as well lets me swap in sheets or other elements fairly easily without disrupting the structure of the viz.

I'd also recommend adding padding fairly early on – Tableau's default of 4 pixels make things look cluttered fast. I'd start with adding 20-30 pixels of padding to the outside of the dashboard and adjust from there. It's easier to adjust the size of things down the line if you've given yourself some breathing room in the beginning.

CJ: Having received feedback from the judges system, combined with your own reflections, is there anything about your visual you'd actually re-visit?

B: Reading the scores and feedback was really interesting – obviously I knew I must have scored well, but I was curious to see what they thought worked or didn't work. I don't think I'd change anything about my actual charts – other than perhaps adding some more ways to interact with the data. I really wish more sales related data had been available as well.

From a less technical standpoint, I created almost all the text in PowerPoint and loaded it in as images...that bit was kind of painful. I probably won't repeat that effort for a while unless I'm really committed to using a custom font again. It's also quite a large viz – I'd maybe see if I could make it a little more friendly for smaller screens.

CJ: Were there any other visuals this year you'd like to call out in appreciation of?

B: There were so many great entries this year, I'll inevitably leave someone out. But a few personal favorites include Ann Pregler's **Going for Broke**, the **The Friendship Ruining Power of Monopoly**, Michelle Frayman's **Games Prestons Play** and Ali Tehrani's **Steph Curry**.

What really interested me was even though this feeder came quicker than folks expected, there was still a great turnout. Of the 209 entries, 111 of them were submitted by folks who had 10 or less published vizzes, and 25 folks published their Iron Viz as their first Tableau Public viz. I think that's incredible, and I hope everyone who looks at the gallery in the future is inspired to throw their hat in the ring.

CJ Round-up:

Thank you so much for joining Brittany. It's been a long time coming to try get you on the blog. I hope others reading this feel inspired by your work, can resonate with your build process and the iterations in design.

I would love to close out this blog with the same sentiment as you, I whole heartedly agree that I hope this piece motivates more individuals to get involved in coming years.

Welcoming someone new can be the start to something quite magical.
Wishing you all the best come May, alongside Nirosh & Paul. Cherish the moment, you deserve it.

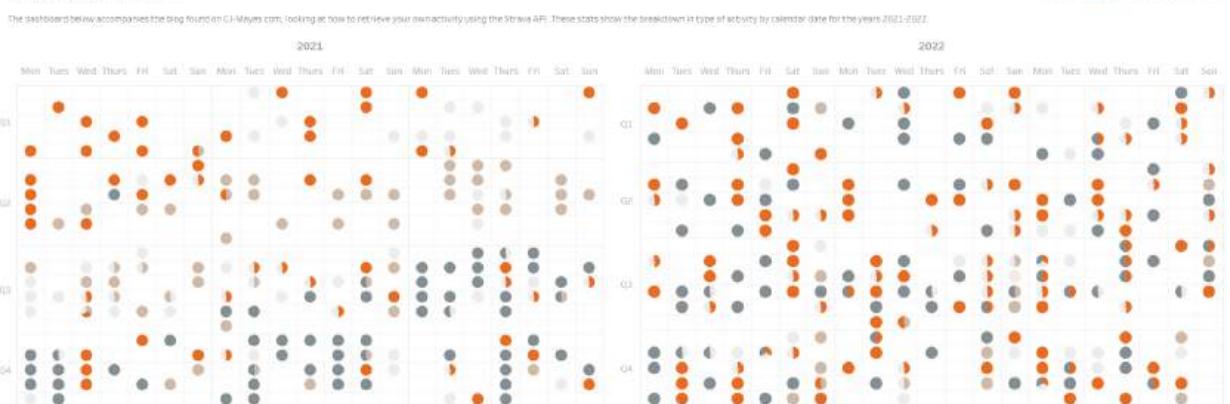
LOGGING OFF,

CJ

USING THE STRAVA API

Hi all,

2 Years of Strava Stats



API's are scary huh? I feel like us data people just use fancy words sometimes for the sake of it. That, and we write documentation that makes zero sense, no shade intended. Well, welcome to my dummies guide on how to access your Strava data using python and postman. What I will go through today is how to successfully set up your Strava API connection and retrieve ALL of your Strava stats from your account. A few *really useful* resources before we start:

- [The Strava Documentation](#)
- **Pre-Requisites.**
 - A Strava Account!
 - Ability to run (not necessarily understand) python code, as well as access to postman (We will cover access)

Okay let's crack on. You can login to Strava, [here](#). Navigate to your Settings [API Page](#).

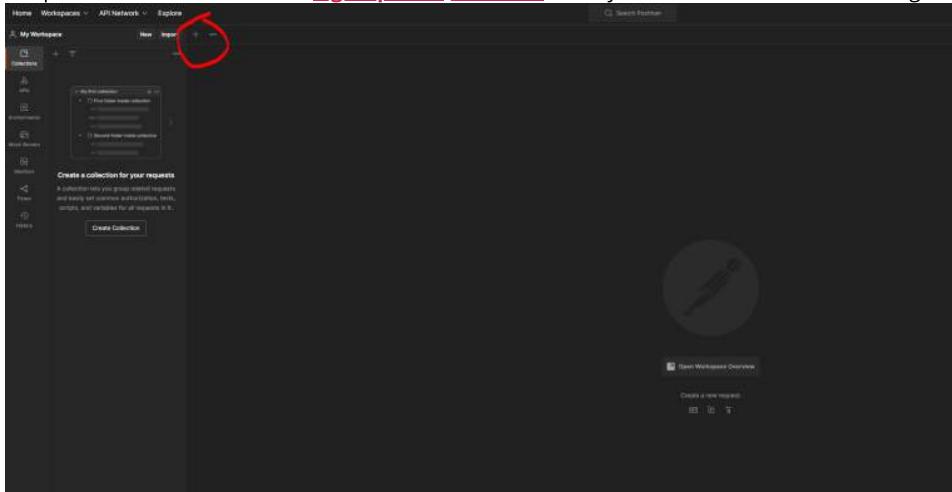
This will store all your starting information. You may need to create the initial API account application.

1. After you are logged in, go to <https://www.strava.com/settings/api> and create an app.
2. You should see the "My API Application" page now. Here is what everything means:

- Category: The category you chose for your application (doesn't matter)
- Club: not necessary
- Client ID: Your application ID (ID for your API)
- Client Secret: Your client secret (confidential!)
- Authorization token: Your authorization token which will change every six hours (confidential)

- o Your Refresh token: The token you will use to get a new authorization token (confidential)
- o Rate limits: Your current rate limit (lets not spam it)
- o Authorization Callback Domain: set to 'localhost'

Once we've done that, We can then refer back to the Strava documentation where it is asking to make a curl request. What does that mean?? Well its a command line tool that helps make http requests, but we will use Postman to do this, so lets set up our postman account. You can [sign up here, for free.](#) Once you've made an account, click get started with something new.



Postman makes requests alot easier to build. What it'll help us do is run our commands based on the access token we feed it. So the documentation tells us we need to then make a GET request based on this link. <https://www.strava.com/api/v3/athlete> For it to work we need to add in our Bearer token. Go to the **header** and add 'authorization' in the key. Add into the value 'Bearer YOURACCESSTOKEN'

 A screenshot of a POST request in Postman. The URL is 'https://www.strava.com/api/v3/athlete'. The method is 'GET'. In the 'Headers' tab, there is a single entry: 'Authorization' with a value of 'Bearer YOURACCESSTOKEN'. The 'Body' tab shows a JSON response with various fields like id, username, resource_state, etc. The 'Pretty' option is selected in the body view.


```

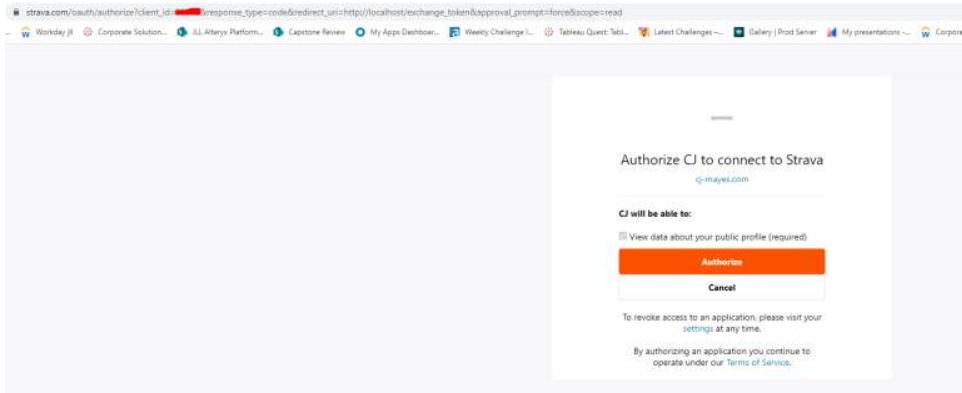
1 {
2   "id": 123456,
3   "username": "cj_mays",
4   "resource_state": 2,
5   "firstname": "Cj",
6   "lastname": "Mays",
7   "bio": null,
8   "city": null,
9   "state": null,
10  "country": null,
11  "sex": "M",
12  "premium": false,
13  "summit": false,
14  "created_at": "2020-01-07T12:18:48Z",
15  "updated_at": "2023-02-04T10:07:25Z",
16  "badge_type_id": 0,
17  "weight": 0.0,
18  "profile_medium": "https://dgalwyrs063hv.cloudfront.net/pictures/athletes/49726306/22262922/1/medium.jpg",
19  "profile": "https://dgalwyrs063hv.cloudfront.net/pictures/athletes/49726306/22262922/1/large.jpg",
20  "friend": null,
21  "follower": null
22 }
```

Amazing, if it pops up with your account details you've done it correctly. We now know who we are.... If we go back to the [documentation](#) now we now notice in part D we have to authorise our account. This is because our account is set to 'read' and not 'read_all' mode. This will allow us to retrieve all our data. First we need to get our authorisation code from the authorisation page.

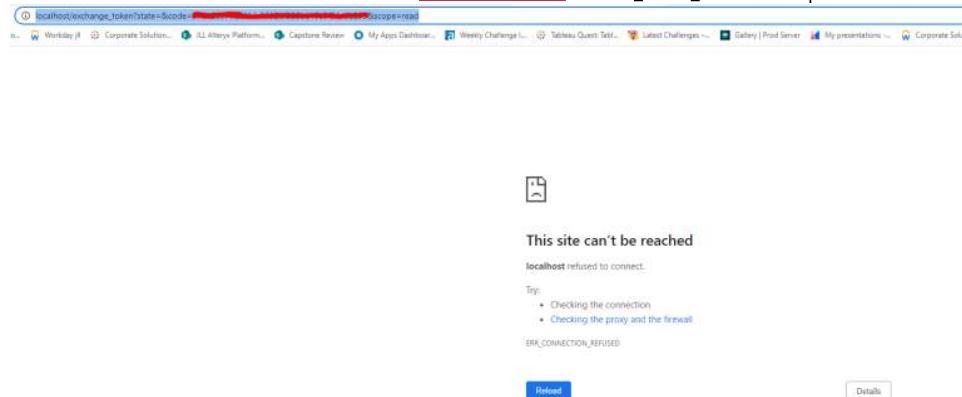
1. Go to <https://www.strava.com/settings/api> and copy your Client ID
2. Paste your Client ID into this

https://www.strava.com/oauth/authorize?client_id=your client id&redirect_uri=http://localhost&response_type=code&scope=activity:read_all

** Note, it is really important to have this say read_all, this will allow us to be able to find all our data.



Click Authorise and save down the new url, it stores the code that we will need later. http://localhost/exchange_token?state=&code=STORE_THIS_CODE&scope=read_all



Don't worry if the page errors out. This gives us a new code for our access token and refresh token for a read_all. This will give access to the Strava App. Go back into to Postman. We want to generate a new token with the scope that we have now requested.

https://www.strava.com/oauth/token?client_id=CLIENT_ID&client_secret=CLIENT_SECRET&code=CODE_SAVED_FROM_PREVIOUS&grant_type=authorization_code

KEY	VALUE
client_id	CLIENT_ID
client_secret	CLIENT_SECRET
code	CODE_SAVED_FROM_PREVIOUS
grant_type	authorization_code

Copy and past the url above into postman. Make sure the setting is on POST. I personally would start a new tab for this. Replace your client_id, client_secret from your API Application Page. Add the new code that we just retrieved into the value of code_saved_from_previous.

The screenshot shows a Postman request to https://www.strava.com/oauth/token. The method is POST, and the URL is https://www.strava.com/oauth/token?client_id=[REDACTED]&client_secret=[REDACTED]&code=[REDACTED]&grant_type=authorization_code. The 'Params' tab is selected, showing four parameters: client_id, client_secret, code, and grant_type, all with their values redacted. The 'Body' tab is selected, showing a JSON response with fields like token_type, expires_at, expires_in, refresh_token, access_token, and athlete details (id, username, first name, last name, bio, city, state, country, sex, premium, sub, created_at, updated_at, badge_type_id, weight). The 'Pretty' tab is selected, displaying the JSON in a readable format.

```
1 "token_type": "Bearer",
2 "expires_at": 1672840449,
3 "expires_in": 18443,
4 "refresh_token": "[REDACTED]",
5 "access_token": "[REDACTED]",
6 "athlete": {
7     "id": 4972636,
8     "username": "cjhayes",
9     "resource_state": 2,
10    "firstname": "CJ",
11    "lastname": "Hayes",
12    "bio": null,
13    "city": null,
14    "state": null,
15    "country": null,
16    "sex": "M",
17    "premium": false,
18    "sub": false,
19    "created_at": "2019-01-07T12:18:46Z",
20    "updated_at": "2023-01-04T18:07:26Z",
21    "badge_type_id": 8,
22    "weight": 10.8
```

Sorry lots of red pen, but if it works when you hit the send button this should generate your **refresh token and access tokens!** ** Do note your new refresh token and access token **will be different** to previous** We can copy these for our code. Finally we are ready to run some code. Time to open up Pycharm. Copy and paste the main.py script from the repo at the top of the page. The code is written in a way that it'll loop through all your activities until there are no more to find. Some of the code has been pulled together from various other articles. The payload will include the client_id, client_secret from your main Strava page. The refresh token is the token we just generated in Postman. This means we can continue to request data.

```
import requests
import urllib3
urllib3.disable_warnings(urllib3.exceptions.InsecureRequestWarning)

auth_url = "https://www.strava.com/oauth/token"
activities_url = "https://www.strava.com/api/v3/athlete/activities"

# replace variables with values for your account
# Leave grant_type as 'refresh_token'
payload = {
    'client_id': "xxxxx",
    'client_secret': 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx',
    'refresh_token': 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx',
    'grant_type': "refresh_token",
    'f': 'json'
}

print("Requesting Token...\n")
res = requests.post(auth_url, data=payload, verify=False)
access_token = res.json()['access_token']

print("Access Token = {}\n".format(access_token))
header = {'Authorization': 'Bearer ' + access_token}
```

Once it authorises, it will print out the access token that you will have just seen in our latest postman POST request.

```
Requesting Token...
```

```
Access Token = 00000000-0000-0000-0000-000000000000

all_activities is needing to be populated
all_activities populated
all_activities populated
all_activities populated
No more activities
```

So what does the remainder of the code do? Well according to the Strava documentation, there is pagination that we need to account for. I.e if we didnt set our parameters we would only get a small view of the true number of activities we've done. By being able to flick through the pages we can now store all the activities.



```
# First loop starts on page 1.
# It will then loop through any further pages required.
request_page_num = 1
all_activities = []

while True:
    param = {'per_page': 200, 'page': request_page_num}
    # initial request, where we request the first page of activities - 200 per
    # page
    my_dataset = requests.get(activites_url, headers=header, params=param).json()

    # If the response is empty we will leave the loop, otherwise we will keep
    # collecting data
    if len(my_dataset) == 0:
        print("No more activities")
        break

    # If the all_activities populated, continue
    if all_activities:
        print("all_activities populated")
        all_activities.extend(my_dataset)

    # Create all activities from the dataset
    else:
        print("all_activities is needing to be populated")
        all_activities = my_dataset

    request_page_num += 1

import pandas as pd

df = pd.DataFrame(all_activities)

# Check dataframe has been created
print(df)
# Create Dataset
df.to_csv(r'Strava_data.csv', index=False, header=True)
```

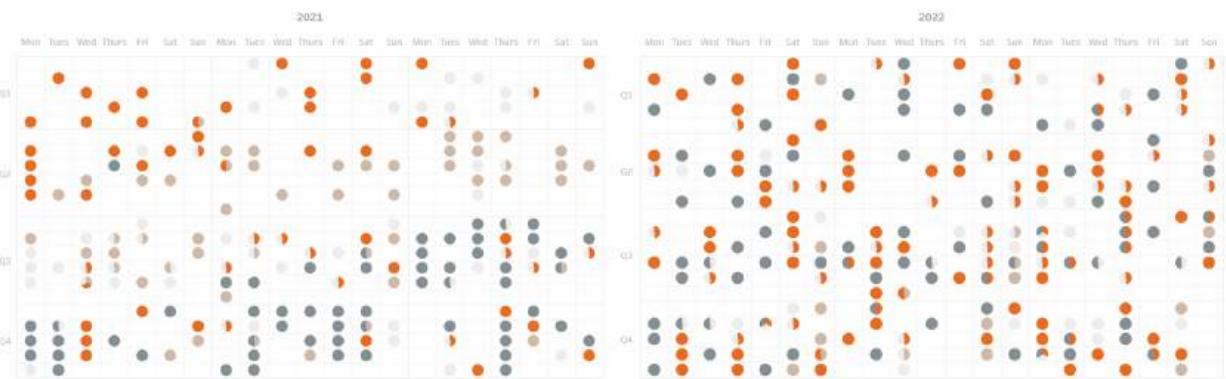
I'll put some more sanitised example data from my results on the GitHub repo so that you can take a look at what the fields may look like. Let me know how you get on with setting up and connecting your own API. It can be a little fiddly, especially with the need to set your tokens for read_all access. A few common errors if you get stuck during the process:

- If you're using postman for the first time, do not get params and headers mixed up.
- GET & POST are two very different things. Double check you're trying to do the correct thing.
- When you paste in your codes / tokens make sure not to leave a leading or trailing space!
- Make sure that each part of the postman requests are working before you try running the code, otherwise it will fail if you try jump straight into the code.
- For the code, make sure you have all the packages pip installed that are required.

In future weeks we will look to see how we can visualise some of these elements. For now, here is a quick calendar I put together in Tableau, nothing too fancy for now.

2 Years of Strava Stats

The dashboard below accompanies the blog found on CJ-Mayers.com looking at how to retrieve your own activity using the Strava API. These stats show the breakdown in type of activity by calendar date for the years 2021-2022.



LOGGING OFF, CJ

PLOTTING XG IN TABLEAU

Hi all, Today we are going to be revisiting some of the content I wrote last year based on expected goals in football, or soccer depending where you are in the world. In this [blog](#), I specifically look at a python package that helped me retrieve all the shots taken in a league for a specific year. To build a catalog of data for the whole season really was a dream. What I've noticed over the years are a lot of analysts will tend to make all their shot charts and summary dashboards for football using python or R. Something I have no problem with. However I want to give users the ability to chart as many of the charts also in Tableau. One of these charts is showcasing the cumulative xG (Expected Goals) over the duration of a match. In short, a line graph! (lol) But it may not be as simple as you first suspect. This tutorial will talk through collecting the stats needed for a game and then how best to emulate the visuals seen on the Understat website. Where to begin? Well I had the pain of watching my team lose 3-1 to Arsenal. We can take a look at how that game played out [here](#).



We can see over the course of the game West Ham had an early chance that they capitalised on. Since that goal, they really didn't have many chances to score more. Arsenal had a few chances in the first half, but started having more dominant goal scoring opportunities in the second half. So how can we get this data? Well, the code is attached at the top of the page. Usually I would use the python package, but I've amended what was originally Mckay Johns code to pull out the xG components we need using BeautifulSoup.

```
#pip install lxml - you will need a parser
# Install all the packages below
#Thanks to McKay Johns for the original code

import requests
from bs4 import BeautifulSoup
import json
import pandas as pd

#scrape a single game.shots. We will allow users to enter a base url
base_url = 'https://understat.com/match/'
match = '18362'
url = base_url+match

#Use requests to get the webpage and BeautifulSoup to parse the page
res = requests.get(url)
soup = BeautifulSoup(res.content, 'lxml')
scripts = soup.find_all('script')

#get only the shotsData from the script
strings = scripts[1].string

# strip unnecessary symbols and get only JSON data
ind_start = strings.index("')+2
ind_end = strings.index("')")
json_data = strings[ind_start:ind_end]
json_data = json_data.encode('utf8').decode('unicode_escape')

#convert string to json format
data = json.loads(json_data)
print(data)

x = []
y = []
xG = []
result = []
team = []
data_away = data['a']
data_home = data['h']
minute = []
shotType = []
player = []
match_id = []
situation = []
player_assisted = []
lastAction = []

for index in range(len(data_home)):
    for key in data_home[index]:
        if key == 'X':
            x.append(data_home[index][key])
        if key == 'Y':
            y.append(data_home[index][key])
        if key == 'h_team':
            team.append(data_home[index][key])
        if key == 'xG':
            xG.append(data_home[index][key])
        if key == 'result':
            result.append(data_home[index][key])
        if key == 'minute':
            minute.append(data_home[index][key])
        if key == 'shotType':
            shotType.append(data_home[index][key])
        if key == 'player':
            player.append(data_home[index][key])
        if key == 'match_id':
            match_id.append(data_home[index][key])
        if key == 'situation':
            situation.append(data_home[index][key])
        if key == 'player_assisted':
            player_assisted.append(data_home[index][key])
        if key == 'lastAction':
            lastAction.append(data_home[index][key])

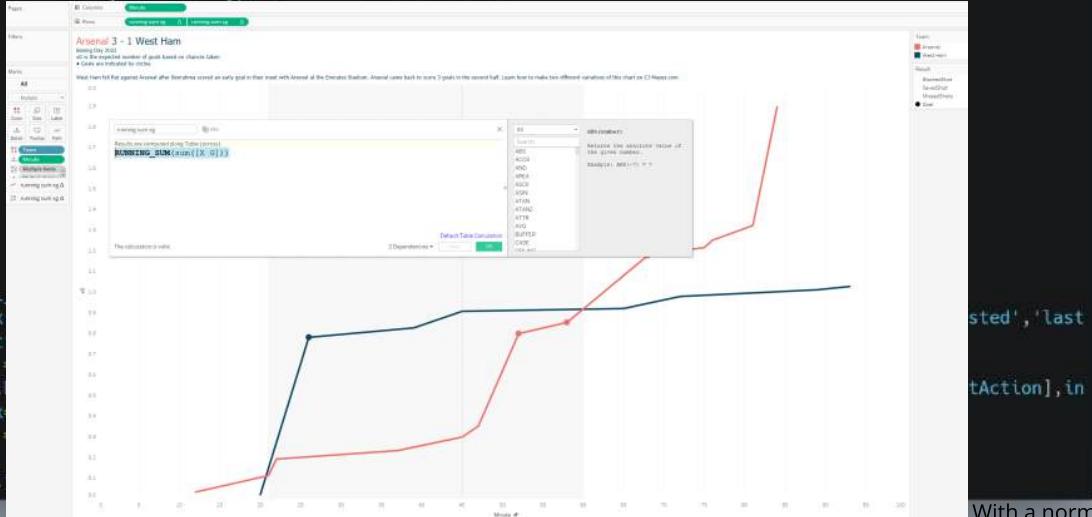
for index in range(len(data_away)):
    for key in data_away[index]:
        if key == 'X':
```

```

Once again, I can't credit or McKay's code but would recommend it for those wanting to access some of the key
components from the site. You'll notice I added a few extra columns to the original code from here. A few things I'd say is
• If key == 'Y': install all the packages you need. A common error is not installing the lxml parser.
• If you add more attributes, make sure you clarify if its home or away and what list you want to append the value to.
• You can amend the match event to whatever match you are after, why not even try to loop through a teams for the whole
  season?
• Each result.append(data[match][index][key]) one we are looking at is https://understat.com/match/18362
  if key == 'minute':
    minute.append(data[match][index][key])

```

A copy of the Arsenal vs West Ham game can be found in the repository. On to the visual. So, you will see that the xG is cumulative across `key == 'shotType'` pretty our dataset contains the xG of each event of the game, it will also have the minute of



that event.

With a normal line

graph it will connect all the points. We can add in a second axis with our marks and drag result onto shape to highlight the goals.

When we calculate along the axis we have to make sure we don't have the team box ticked. But is this a reflection of what we



see on Understat?

Not in it's

entirely no. This is because we are joining the points up of the marks we do have. On understat, the increase in xG is based on specific minute events. This means we will need to create a minute per minute analysis. One method we could do is join in a new dataset of every single minute of the game. That, or we can create bins! Right click on minute and create a bin. Make the bin size 1 (i.e a bin for every minute of the game) We can now drag our discrete minute bin onto columns instead of our minutes field.

This way, we only see spikes in the sum(xG) where the event happened.

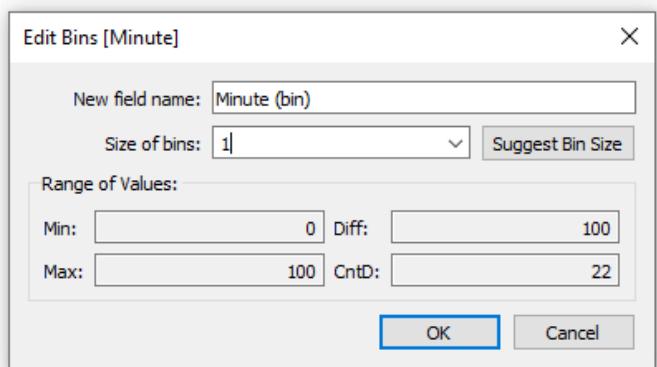


Much better! One final thing I've

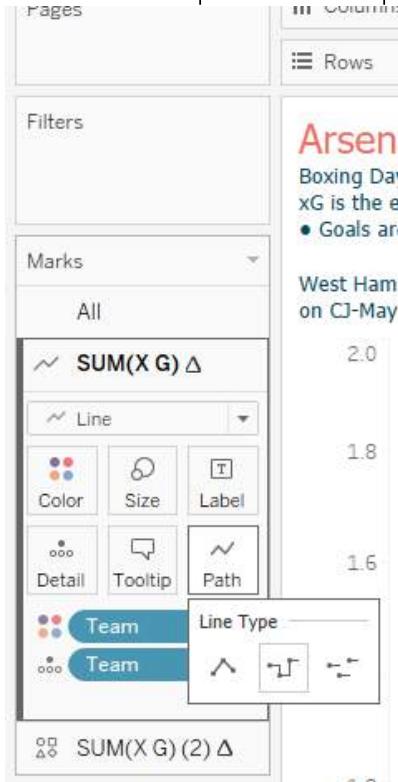
noticed is that when you go to create bins, it will do it based on your data. So in this case, it creates the bins from a minimum of 12, as that was when the first minute was. Obviously, though the game started earlier than this, so to be a true representation of the data we want the 1 minute mark in the data, (and perhaps a 90 minute, or longer if added time) One way to get around this is to add in two extra rows into the dataset.

A	B	C	D	E	F	G	H	I	J	K	L	M
x	y	xG	result	team	minute	shotType	player	match_id	situation	player_as:	lastAction	
0.782	0.401	0.031748	SavedShot	Arsenal	12	RightFoot	Granit Xhaka	18362	OpenPlay	Martin Od	TakeOn	
0.815	0.518	0.07922	BlockedShot	Arsenal	21	RightFoot	Eddie Nketiah	18362	OpenPlay	Granit Xhaka	Pass	
0.953	0.377	0.080291	MissedShot	Arsenal	22	RightFoot	Martin Od	18362	OpenPlay	Bukayo Saka	HeadPass	
0.844	0.686	0.040984	BlockedShot	Arsenal	37	LeftFoot	Eddie Nketiah	18362	OpenPlay	Kieran Tierney	Pass	
0.89	0.542	0.065564	MissedShot	Arsenal	45	Head	Eddie Nketiah	18362	SetPiece	Martin Od	Cross	
0.793	0.606	0.029734	BlockedShot	Arsenal	47	LeftFoot	Martin Od	18362	FromCorner	Gabriel Martinelli	Pass	
0.766	0.418	0.023141	MissedShot	Arsenal	47	RightFoot	Thomas Partey	18362	OpenPlay	Martin Od	Pass	
0.698	0.419	0.014648	MissedShot	Arsenal	52	LeftFoot	Martin Od	18362	OpenPlay	Ben White	Pass	
0.906	0.472	0.432865	Goal	Arsenal	52	LeftFoot	Bukayo Saka	18362	OpenPlay	Martin Od	Rebound	
0.969	0.696	0.054474	Goal	Arsenal	58	LeftFoot	Gabriel Martinelli	18362	OpenPlay	Granit Xhaka	Pass	
0.912	0.361	0.324268	Goal	Arsenal	68	RightFoot	Eddie Nketiah	18362	OpenPlay	Martin Od	Pass	
0.7	0.488	0.037983	BlockedShot	Arsenal	75	LeftFoot	Granit Xhaka	18362	DirectFreekick		Standard	
0.805	0.35	0.035636	MissedShot	Arsenal	76	LeftFoot	Martin Od	18362	OpenPlay	Bukayo Saka	Pass	
0.823	0.421	0.070935	MissedShot	Arsenal	81	LeftFoot	Martin Od	18362	OpenPlay		BallRecovery	
0.9	0.365	0.106037	BlockedShot	Arsenal	84	LeftFoot	Martin Od	18362	OpenPlay	Bukayo Saka	Pass	
0.966	0.464	0.467093	SavedShot	Arsenal	84	Head	William Saliba	18362	FromCorner	Bukayo Saka	Aerial	
0.775	0.608	0.018991	SavedShot	West Ham	20	RightFoot	Said Benrahma	18362	FromCorner		None	
0.885	0.5	0.761169	Goal	West Ham	26	RightFoot	Said Benrahma	18362	Penalty		Standard	
0.841	0.655	0.045767	MissedShot	West Ham	39	LeftFoot	Michail Antonio	18362	SetPiece	Lucas Paqueta	Chipped	
0.909	0.322	0.080357	SavedShot	West Ham	45	RightFoot	Michail Antonio	18362	OpenPlay	Tomas Souza	Pass	
0.692	0.575	0.013417	MissedShot	West Ham	65	RightFoot	Thilo Kehrer	18362	OpenPlay	Declan Rice	Pass	
0.871	0.676	0.058455	MissedShot	West Ham	72	LeftFoot	Said Benrahma	18362	OpenPlay	Lucas Paqueta	Pass	
0.782	0.468	0.032316	MissedShot	West Ham	89	LeftFoot	Lucas Paqueta	18362	FromCorner	Declan Rice	Pass	
0.752	0.64	0.016043	SavedShot	West Ham	93	RightFoot	Pablo Mari	18362	OpenPlay	Said Benrahma	Pass	
					0							
					100							

Now when you go to create your bins, the minimum and maximum value will change.



This allows you to be able to plot the full game. BUT, there is a way to avoid having to do this if **we use a continuous minute bin!** After asking around a little, Val helped me with a solution that doesn't need the additional rows of data. Below is the better option. You can keep the minute bin as continuous but change



the path to the second option on the path list.

1.2 What you will notice in the difference of having the minutes bin continuous is that it allows for more than one event to happen in the same minute. It also allows us the flexibility in being able to set our axis to zero without the need for amending the data.

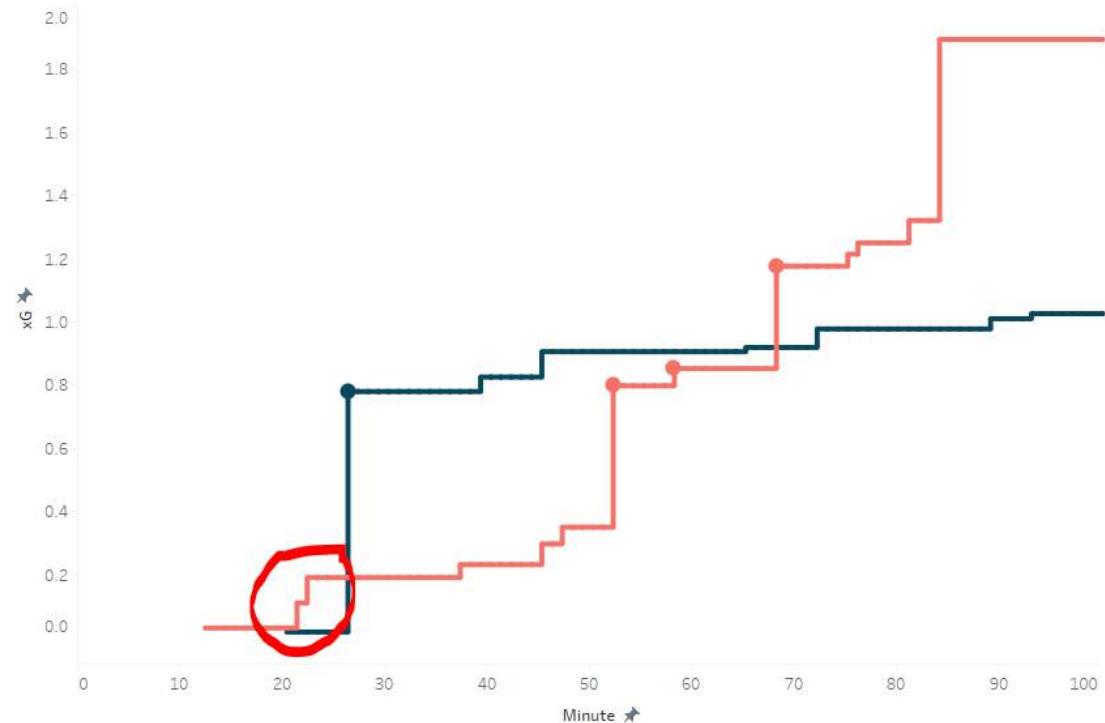
Arsenal 3 - 1 West Ham

Boxing Day 2022

xG is the expected number of goals based on chances taken

- Goals are indicated by circles.

West Ham fell flat against Arsenal after Benrahma scored an early goal in their meet with Arsenal at the Emirates Stadium. Arsenal came back to score 3 goals in the second half. Learn how to make two different variations of this chart on CJ-Mayes.com



I've left both sheets in to compare! Feel free to download the workbook off Tableau Public from the top of the page if you get stuck. Thanks for reading. **LOGGING OFF, CJ**

TWITTER CIRCLES (SELENIUM) WITH JESSICA MOON

Hi all, I hope everyone has settled into the year well, as we come up to being 1/12 of the way through 2023! Today it is an absolute pleasure to have Jessica Moon join the site to teach us some tricks with web scraping using Selenium and how then to visual display the output of Twitter Circles in Tableau. Like with all tutorials, the resources can be found in the GitHub repository under the title – and the Tableau Public link points to Jessica's dashboard. You can follow Jessica for more updates on Twitter [here](#). Over to Jessica for the run through. When TwitterCircle.com hit the scene, Twitter was ablaze with Circle photos and I enjoyed looking at the #datafam circle posts and would instinctively inspect to see if there was a circle with a dark background and a woman with dark hair among those featured. I ran my own to see who I was interacting the most with. But then I wondered—was the engagement mutual? The website would allow you to search yourself or any other user without a login. This was the opportunity to write another Python program!

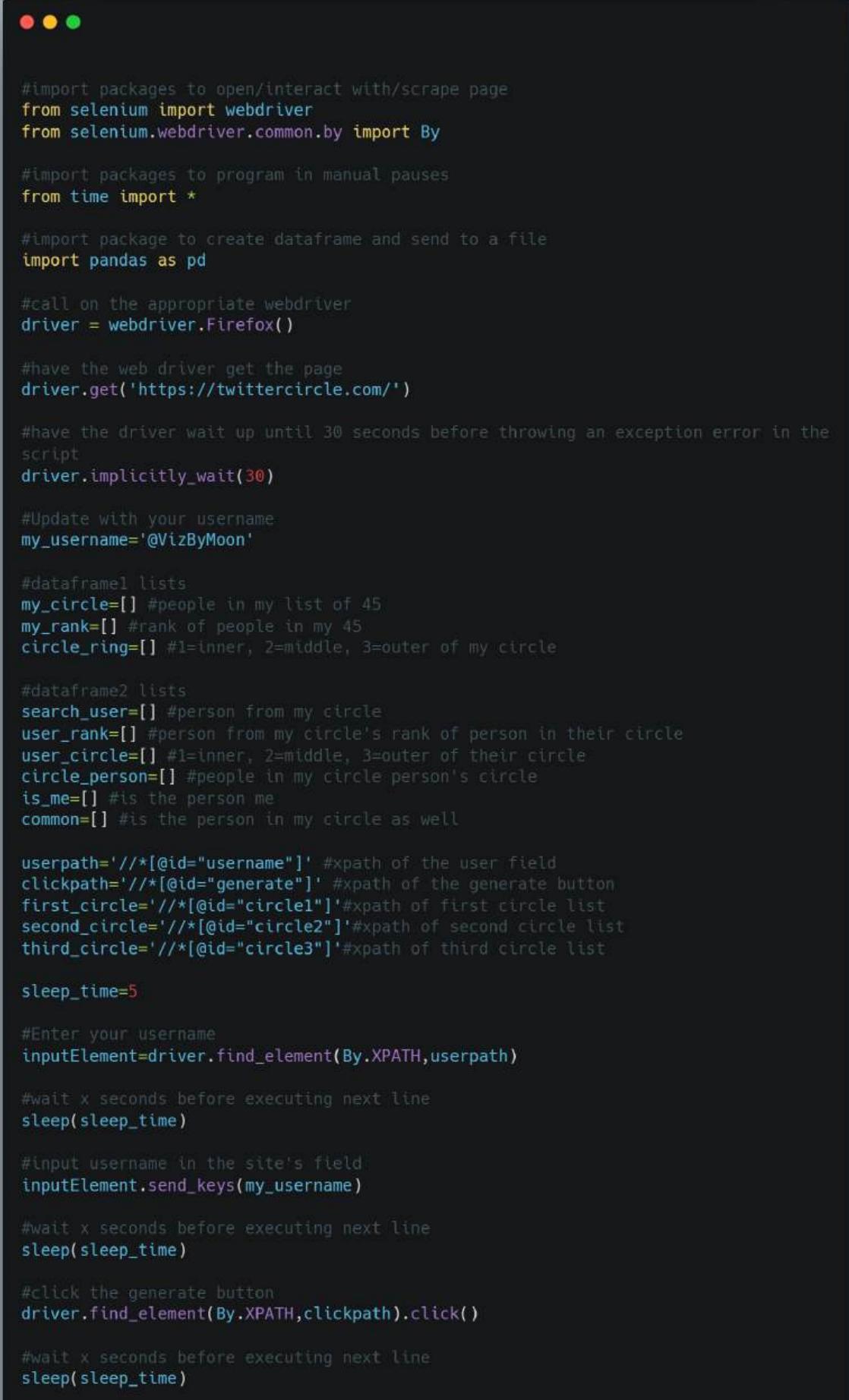


While I prefer to use

Beautiful Soup for web scraping, it has limitations like it can't click elements of a webpage and if the page uses a lot of JavaScript—the soup isn't going to return much of use. These are situations where I use [Selenium](#). Selenium can not only scrape the rendered html of a page, but it can also enter text, click elements, etc like a user could interacting with a site. Before we discuss specifics, web scraping can be a real grey area. Make sure to check out sites' terms of use and /robots.txt files. In the program I wrote, I made sure that between waits and sleeps, I would not be bogging down the site as each user ran (45 in total) would take approximately 35 seconds before the next user was ran. If you're new to Selenium, I recommend checking out these 3 videos ([video1](#), [video2](#), [video3](#)) to learn how to install the package + driver, and understand what all Selenium can do in regards to

web scraping. I honestly found these in preparation for this blog, and I discovered more elegant ways to slow down the program so elements could be rendered than I had coded. But, y'all, bad code can work too. If you're just writing a program to get the data you need and not to put food on the table—don't be afraid of some bad, ugly, not optimally efficient code. It can get the job done. You might notice in my code I use a FireFox driver whereas these videos use Chrome. I'm pretty sure I found out about Selenium on Stack Overflow and the example code I started from used FireFox as the driver, so that's what I've been rolling with. The program I wrote does the following:

- Opens TwitterCircle.com
- Enters my username
- Clicks generate
- Scrapes the ranks and usernames of circle 1+circle 2+circle 3
- Then it does the same cycle for the list of usernames to get their circles
- Looks to see if the circle username is my username or in my circle list.
- Two dataframes are created: one that combines the ranks, usernames, and circle numbers of my Twitter circle (45 rows) and one that gets similar data for each person in my circle for their respective circles (2025 rows assuming everyone is engaging with 45+).
- The dataframes are merged (like a SQL join) on My Circle and user, so the resulting set is 2025 rows.



```
#import packages to open/interact with/scrape page
from selenium import webdriver
from selenium.webdriver.common.by import By

#import packages to program in manual pauses
from time import *

#import package to create dataframe and send to a file
import pandas as pd

#call on the appropriate webdriver
driver = webdriver.Firefox()

#have the web driver get the page
driver.get('https://twittercircle.com/')

#have the driver wait up until 30 seconds before throwing an exception error in the
script
driver.implicitly_wait(30)

#Update with your username
my_username='@VizByMoon'

#dataframe1 lists
my_circle=[] #people in my list of 45
my_rank=[] #rank of people in my 45
circle_ring=[] #1=inner, 2=middle, 3=outer of my circle

#dataframe2 lists
search_user=[] #person from my circle
user_rank=[] #person from my circle's rank of person in their circle
user_circle=[] #1=inner, 2=middle, 3=outer of their circle
circle_person=[] #people in my circle person's circle
is_me=[] #is the person me
common=[] #is the person in my circle as well

userpath='//*[@id="username"]' #xpath of the user field
clickpath='//*[@id="generate"]' #xpath of the generate button
first_circle='//*[@id="circle1"]' #xpath of first circle list
second_circle='//*[@id="circle2"]' #xpath of second circle list
third_circle='//*[@id="circle3"]' #xpath of third circle list

sleep_time=5

#Enter your username
inputElement=driver.find_element(By.XPATH,userpath)

#wait x seconds before executing next line
sleep(sleep_time)

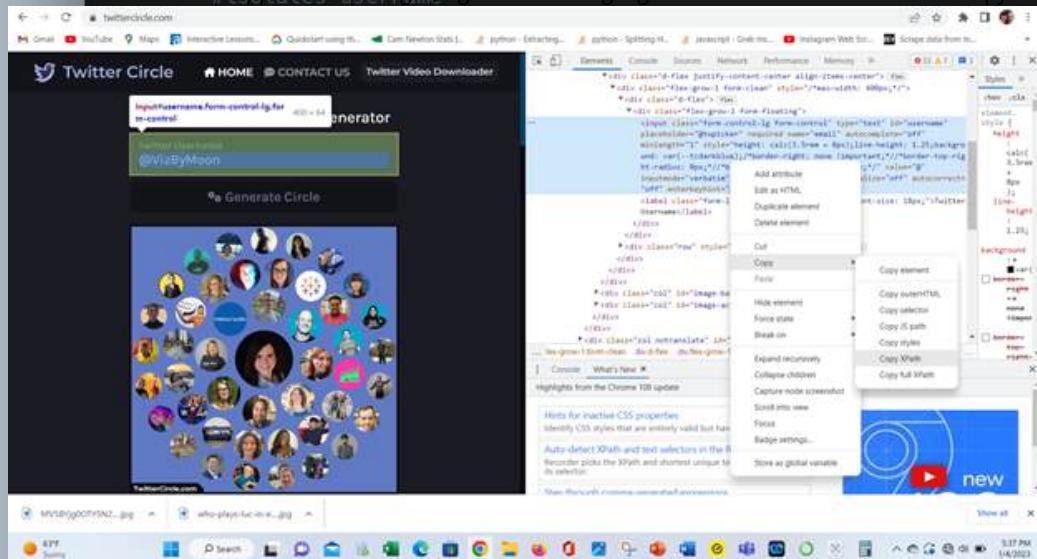
#input username in the site's field
inputElement.send_keys(my_username)

#wait x seconds before executing next line
sleep(sleep_time)

#click the generate button
driver.find_element(By.XPATH,clickpath).click()

#wait x seconds before executing next line
sleep(sleep_time)
```

I chose to find elements by xpath for each of the `your_clickhere` button because they are very specific to elements whereas doing something like `my_circle_peeps(iterator_list, num_ring)`: the element you intended. How do you find this? Open the page of interest, `for twinge in iterator_list:` you want your program to interact with. Right click and "Inspect". In the code that is output to the right click the highlighted html for the element and Copy->Xpath



The code can be found

[on GitHub.](#) (Circle Scrape) Things to be aware of:

- You can `append` this to your preferred IDE after you've installed Selenium + the driver.
- Make the `print` statements like `update_my_username` and possibly increase `sleep_time` if your internet lags.
- The csv file will be available in the same folder as your IDE, like mine is `C:\Users\jessi`.
- The code may take about half hour to fully complete given the number of times it needs to loop through each of the different individuals.

An Example of Jessi's Output can be found on GitHub. On the Tableau side, I decided to take a self-centered approach in my WZ, attempting to mimic the spacing in the circle image generated. I'll run through the main calcs that you can use with the data generated from the program. With Radials you need to:

1. #loop through the `first` list out the angle interval that evenly divides a circle (or semi-circle)
2. `my_circle_peeps(follower_list, 1)` Figure out the angle that each item will be placed at.
3. Calculate the X coordinate using the angle and radius
4. #get the `circle 2` users Calculate the Y coordinate using the angle and radius

If you're `second_followers=driver.find_element(By.XPATH, second_circle)` viz. That's usually what I'm looking at each time. `follower_list2=second_followers.find_elements_by_tag_name("li")` controls the space in the middle for your program. `sleep(sleep_time)` Coordinate zhuzh (Float, default 1.1) –this parameter boosts or minimizes the radius and controls how tightly the pies surround the inner circle. Calcs you'll need: Origin //center point for you to be in the middle

MAKEPOINT(`my_circle_peeps(follower_list2, 1)`) –code or use a single parameter, but the three values kind of mimics TwitterCircle's placement and helps it look a little uneven which I like here */ CASE MIN() WHEN 1 THEN -50 WHEN 2 THEN -100 ELSE -180 END Adjusted Rank /*take the actual rank and subtract the minimum rank for the ring (1, 2, or 3) so lowest ranked user

`third_followers=driver.find_element(By.XPATH, third_circle))N)-MIN({ FIXED :MIN()}) Angle Part 1`

`//determine Angle (deg) spaces points in each circle 360/MIN(({FIXED :COUNTD()})) Angle Part 2 /*Adjust the angle to start at follower_list3=third_followers.find_elements_by_tag_name("li") observed*/ ()-((0)* X /*Determine x`

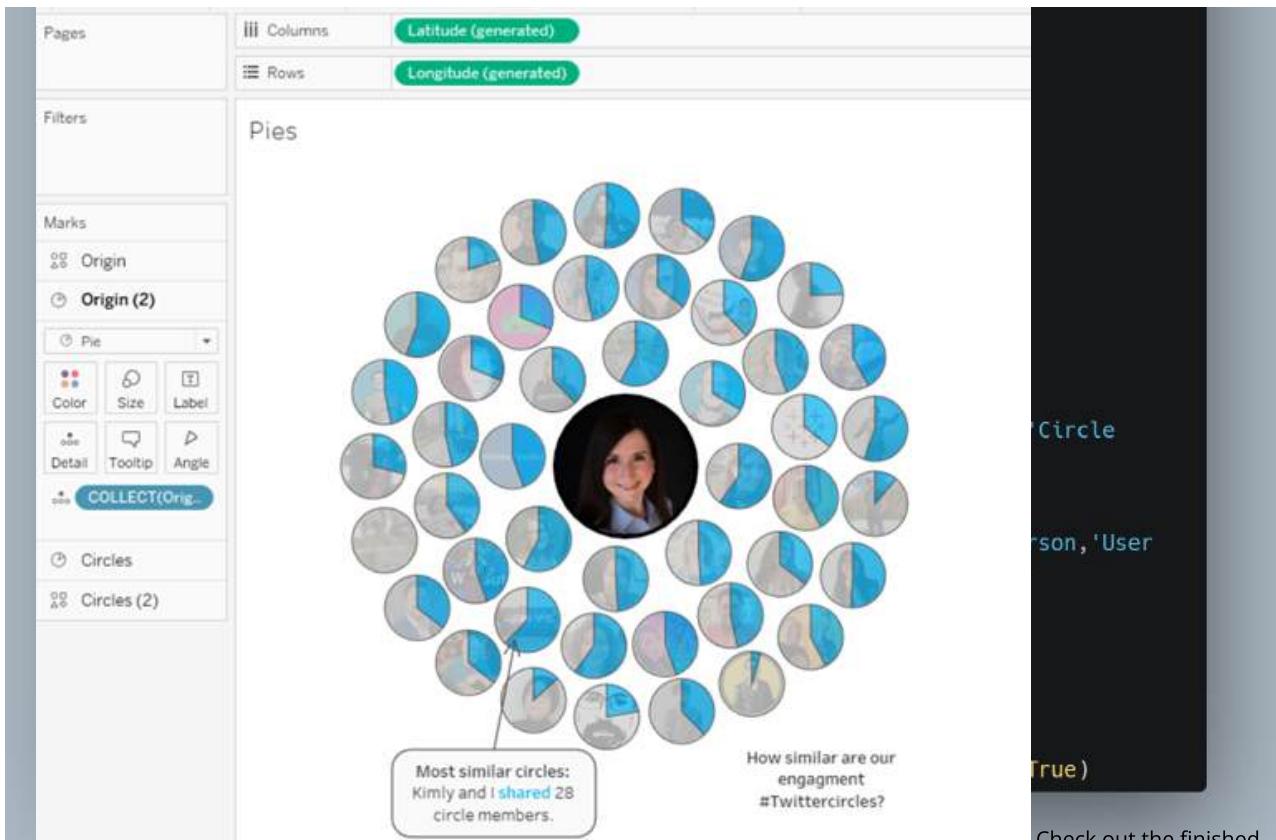
coordinate using Angle Part 2, adjusting for the inner circle and coordinate zhuzh./* COS(RADIANS())*(+MIN(*)) Y /*Determine y coordinate `my_circle_peeps(follower_list3, 1)` inner circle and coordinate zhuzh./* SIN(RADIANS())*(+MIN(*)) Circles

MAKEPOINT() Here's the build in action: So remember:

- #loop through the users in your circle to find who is in their circle
- for follower in my_circle: Top Circles mark is Pie
- driver.get('https://twittercircle.com/')
- driver.implicitly_wait(30) COUNTD(Circle Person) goes on Angle.
- inputElement=driver.find_element(By.XPATH,userpath)
- sleep(sleep_time)
- inputElement.send_keys(follower)
- sleep(sleep_time)
- driver.find_element(By.XPATH,clickpath).click()
- sleep(sleep_time)
- first_followers=driver.find_element(By.XPATH,first_circle)
- sleep(sleep_time)
- follower_list=first_followers.find_elements_by_tag_name("li")

- #loop through the follower 1 list to extract the rank and user
- circle_circles(follower, follower_list, 1)

- second_followers=driver.find_element(By.XPATH,second_circle)



product [here](#)! If you have any questions, send me a message on [Twitter](#)! Round-Up: I can totally resonate with how Jessica so eloquently put “Bad code can work too. If you’re just writing a program to get the data you need and not to put food on the table” If it runs, it runs. Unless you’re bringing down the server and costing your business a fortune, ha. Anyway, I want to thank Jessica once again for putting together such a fantastic piece of code, blog and accompanying visual. It’s not often we get to see a full E2E, and I hope this may prompt others to visualise their own Twitter Circles down the line or apply the logic of the code to their own web scraping abilities. **LOGGING OFF, CJ**

HEATMAPS USING ALTERYX WITH FUAD AHMED

Hi all, I am so excited today to have Fuad join the site today to showcase an alternative method of creating heatmaps, this time using Alteryx and Tableau. Fuad can be found on twitter, [here](#).



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

Fuad Ahmed

Senior Data Strategy Consultant at Swift Insights | London, United Kingdom

[Follow](#)

[Hire Me](#)

Vizzes 22

Favourites 59

Following 88

Followers 761

The grid displays eight Tableau visualizations:

- The Cost of Living** by Fuad Ahmed: A dark-themed viz showing various metrics related to living costs.
- Tableau Engagment (PostgreSQL) Dashboard** by Fuad Ahmed: A dashboard showing engagement metrics for PostgreSQL.
- Golden Boot** by Fuad Ahmed: A viz titled "The Race for the Golden Boot" showing data for the 2018 World Cup.
- Finance Business Dashboard** by Fuad Ahmed: A dashboard showing financial performance metrics.
- Brian Charles Lara** by Fuad Ahmed: A viz showing data related to Brian Charles Lara.
- Economic Freedom** by Fuad Ahmed: A dashboard showing economic freedom metrics.
- KPI SUPERSTORE DASHBOARD** by Fuad Ahmed: A dashboard showing KPIs for a supermarket.
- THE TABLEAU BOOK OF CALCS** by Fuad Ahmed: A viz showing calculations from the Tableau Book of Calculations.

Many of us by now have the logic to create them probably using either python or r, and whilst that may do the job justice, I am a huge fan of being able to emulate similar outputs in other software. I think with different software come different methodologies (whilst perhaps a similar approach), but also allow flexibility in output and certainly have allowed Fuad to flex his design skills. Speaking of output, for me, Fuad Ahmed created one of the most beautiful soccer vizzes I have seen all year. I would urge you to check out his [full visualisation](#).

THE TOP CONTENDERS

WHO ARE THE TOP CONTENDERS FOR THE GOLDEN BOOT & HOW WELL ARE THEY TAKING THEIR CHANCES?
HOW ARE THE TOP CONTENDERS CONTRIBUTING IN EACH GAME THROUGHOUT THE SEASON?

Move the mouse to read
Storytelling with Data

Erling Haaland leads the way this season as the number one contender for the Golden Boot. Tottenham's Harry Kane, and Fulham's Aleksandar Mitrovic follow behind in second and third place. Using shot coordinates we can plot on a football pitch where the top contenders are taking their chances and scoring their goals.

Haaland has converted 36% of shots taken into goals. 15 percentage points higher than his closest counter part, Harry Kane. You can see how tightly grouped Haaland's shooting patterns are in comparison. 89% of his shots were taken inside the box. This is far less spartan than the others highlighting how good his positioning has been in front of goal.



CJ: Fuad, great to have you join. Can you tell us a little more about your passion for football and the types of metrics you focused on for your IronViz entry? Hey CJ! Thanks for having me, it's a pleasure to feature in one of your blogs. As you already know I'm a huge fan of your work and I look forward to seeing what you have in store for us in 2023. My earliest memories of football actually started with La Liga. I remember watching Ronaldinho at Barcelona before the GOAT (Messi) came along. The football Barcelona played and the flair Ronaldinho had really made me fall in love with the sport. For my IronViz I wanted to highlight the significance of Haaland's goal-scoring contribution and compare it to others in the Premier League. My focus was on 3 key areas:

1. Where was he scoring/taking chances from?
2. How did he perform against the big 6?
3. What impact was his goal-scoring having on Man City's league position?

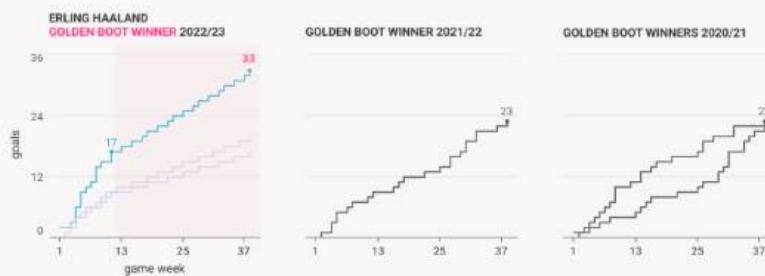
These were all key aspects to telling my IronViz story but I think the 3rd one is often overlooked in football analytics at least from what I have seen. There's always a lot of attention given to the most goals or most assists but as you know looking at just a select few data points will not tell the full story. People tend to use this as a stand-alone argument over who is a better player. But I think it's equally important to highlight whether the impact a player has led to better team performance and ultimately a team win. This was the most significant factor for me and I spent a lot of time thinking about what other data points I can use to present this. In the end, it made sense to me to bring league points into the equation and analyse the direct contribution of goal scoring by the top goal scorers was having on the number of league points gained.

THE RACE FOR THE Golden Boot

Erling Haaland to win the Golden Boot 22/23 & finish with 33 goals.

Manchester City's striker has been unstoppable this season, dominating his competition in every aspect. If my predictions are correct and Haaland ends with more than 30 goals he will become 1 of only 10 players in the history of the sport to score 30 goals in a Premier League season. Interestingly, only 3 out of those 9 players have gone on to win the title during their 30 goal season.

Please read with care as this is a **prediction** for the end of season taking into account all of the data we have used so far. See methodology for how I projected 33 goals.



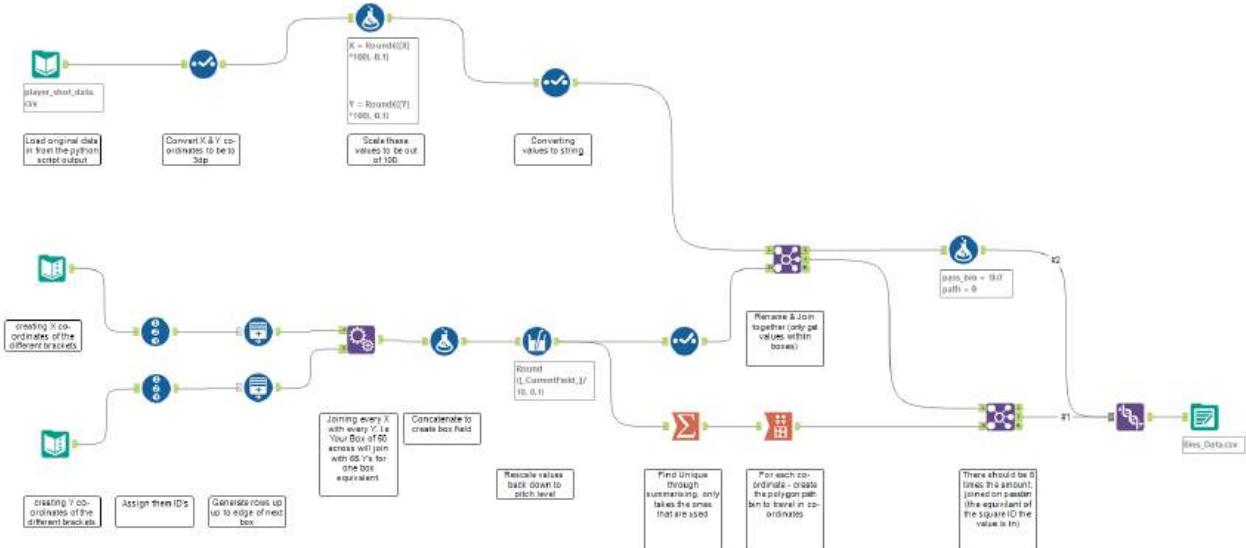
I have no doubt that Haaland will win the race for the Golden Boot this year as long as he maintains fitness. Will he also help Manchester City win their own race for the Premier League title?

[Data & Methodology](#)
IronViz 2023



CJ: What was the data collection process like? Can you explain a little about the Alteryx flow for those that are new to Alteryx? (The Alteryx workflow can be found on the GitHub repository at the top of the page, alongside some of the original datasets) Firstly, I want to thank [Peter Silvester](#). Peter really helped me create the workflow and data prep, without him being able to replicate the polygon binning would have been incomplete. The latter parts of the flow especially were particularly tough to

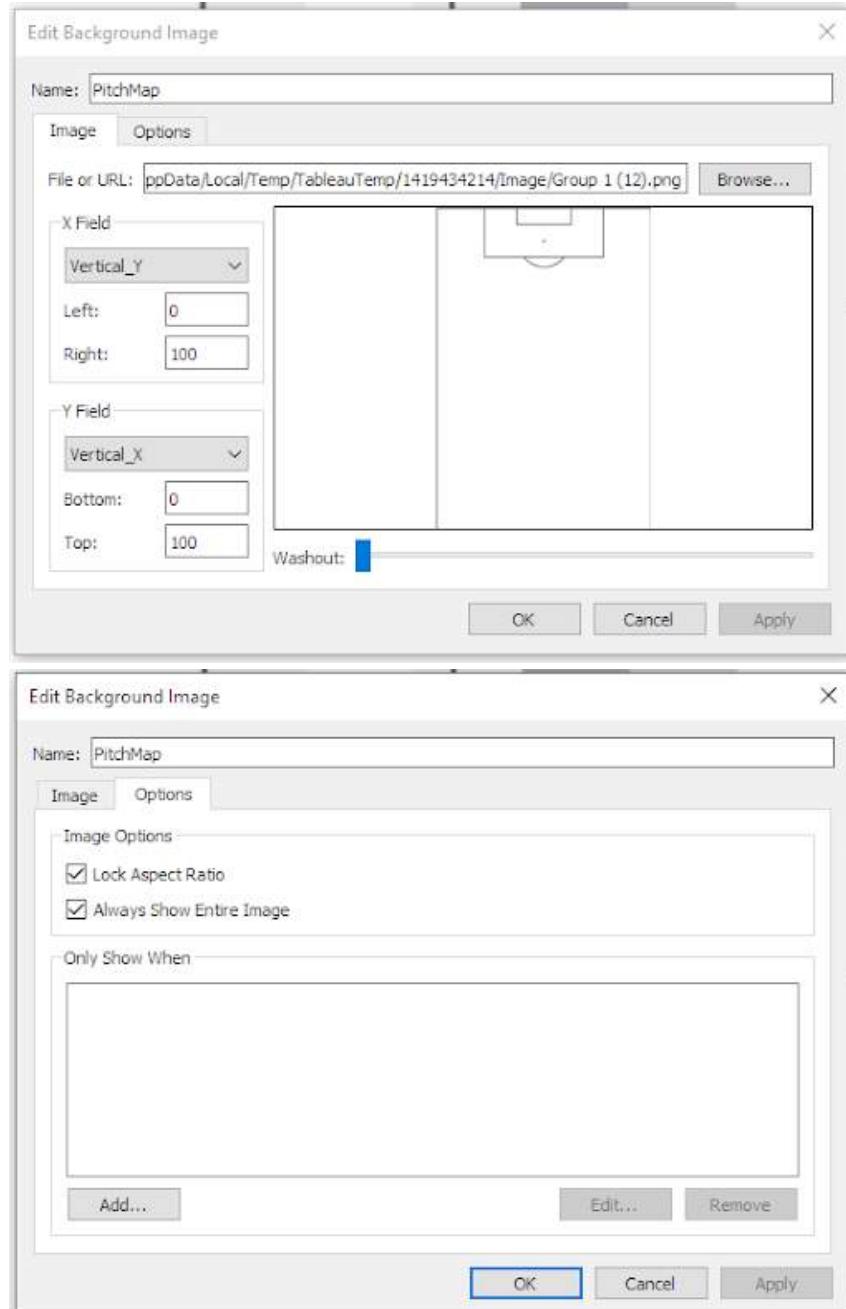
piece together!



1. The top half of the flow takes the original data from the python script output from a blog post you've previously written, actually. I then convert all the values to be eventually 1 decimal place. We have to remember that these X and Y coordinates are important when building the background football pitch.
2. The bottom half of the flow creates the different X and Y boxes and assigns them each an ID.
3. The generate rows help to add in all the different decimal places between each 1×1 square.
4. Next, the find and replace tool gives all the different possibilities of shot boxes. This explodes the data up to a really big size.
5. By concatenating values we create essentially a box id, assigning a shot to a specific box.
6. After rescaling the values of our grid to be in line with our original dataset, we can glue them together to only take the box values of the data that is actually needed. Remembering that there is no point to create the boxes for the full pitch.
7. For polygons to work, the final thing we need to do is create a pathID, so our dataset then takes the edges of the box and creates 5IDs for it to pass through. Essentially building our polygon square.
8. The final data export is saved back down as Bins Data. It's worth noting that our final dataset will be 5x the size of our original data.

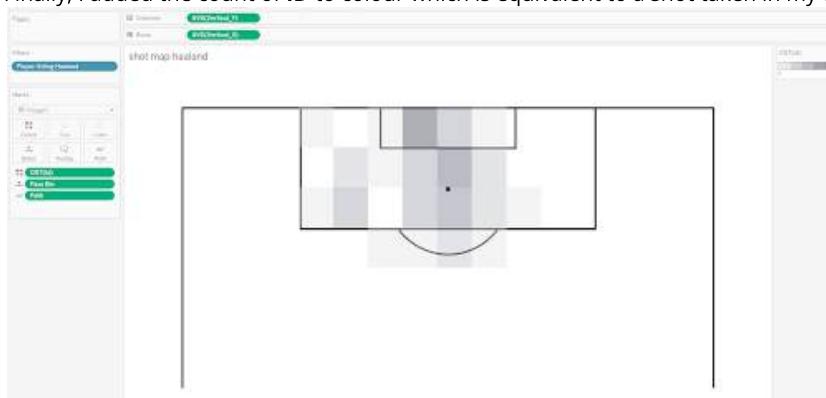
CJ: Now you have the output, can you showcase in Tableau how to create a shot heatmap using polygons like on your visualisation second page? Once you have the output there are a few more steps required to get to the end result.

1. I imported a background image by going into the Map tab in Tableau and added an image of a football pitch outline. You can play around with the options depending on what type of image you used but here is how I set mine up. Again, the image can be found in the repository at the top of the page to download.



- After I set up the background image correctly I used the bins that I created in Alteryx called **X bin** and **Y bin** and dragged them into the columns and rows and changed them to be aggregated as an average because the data is duplicated.
- I then changed the mark type to a polygon and added the fields **Pass Bin** to detail and **Path** to Path.

Finally, I added the count of **ID** to colour which is equivalent to a shot taken in my data. This will give you the heatmap-style look.



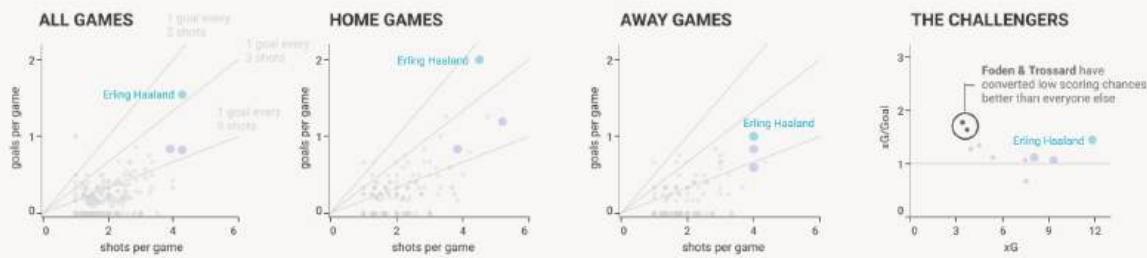
CJ: Besides the technical expertise and the beauty in creating such a delicate shot map, what did you want people to take away from reading these types of charts when looking at Kane, Haaland and Mitrovic?

COMPETITION & GOAL IMPACT

HOW DOES HAALAND COMPARE AGAINST THE REST & IS HE GETTING MORE VALUE PER SHOT COMPARED TO HIS CHALLENGERS?
WHAT IMPACT ARE THE GOALS SCORED BY THE TOP CONTENDERS HAVING ON THE TITLE RACE?

Want to know more? Check
Soccerology and this

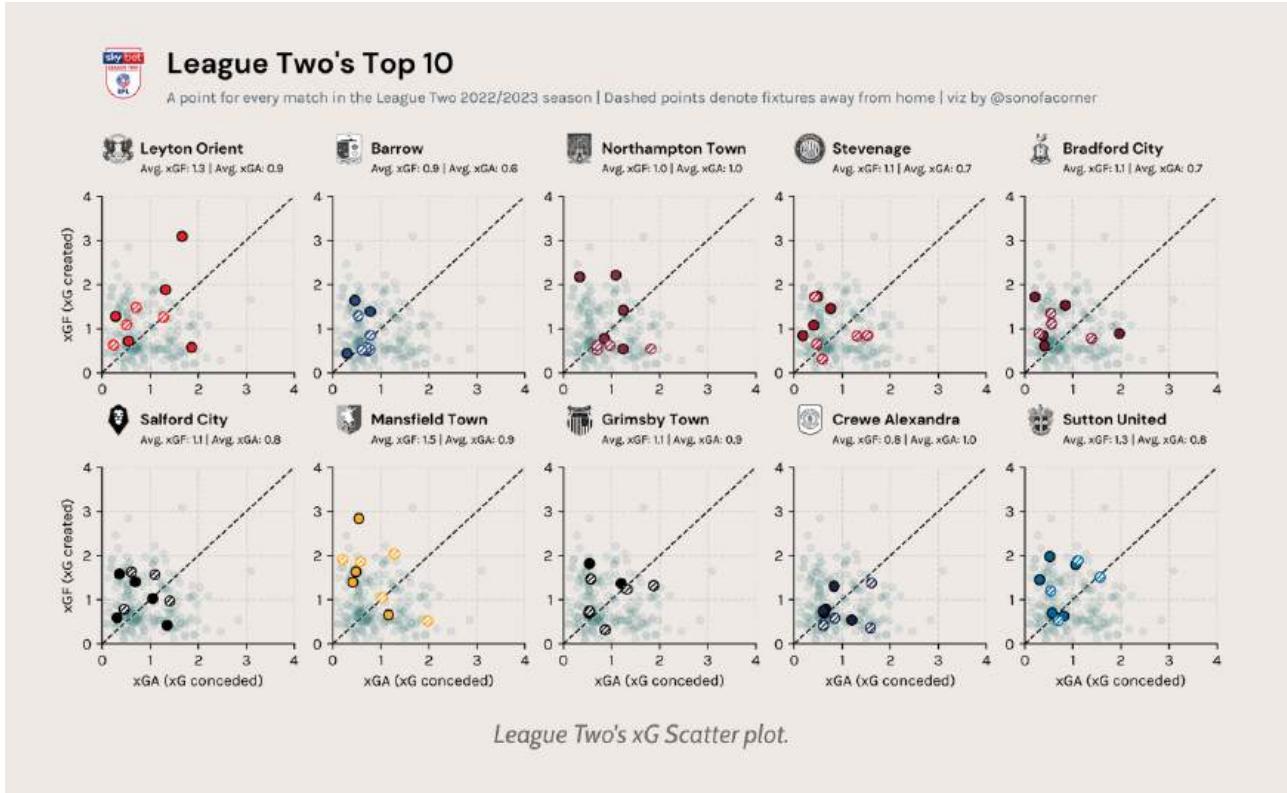
- Manchester City have not dropped a single point at home so far with Haaland scoring 12 goals throughout those games, averaging almost 1 goal every 2 shots. The only aspect Haaland has competition in is the xG/Goal ratio. Both Foden and Trossard are getting high goal value from their chances. However, Haaland's slightly lower ratio is likely to be a result of his positioning to receive better chances as opposed to him not converting as well as Foden and Trossard.



- Goal scoring is an important factor to any team winning the league title. However, goals can be scored in vain if they are not converted into wins which is ultimately the point of the game. I have therefore grouped goals scored in a match by each player to identify how many points on average the contenders' teams are getting when they score or don't score.



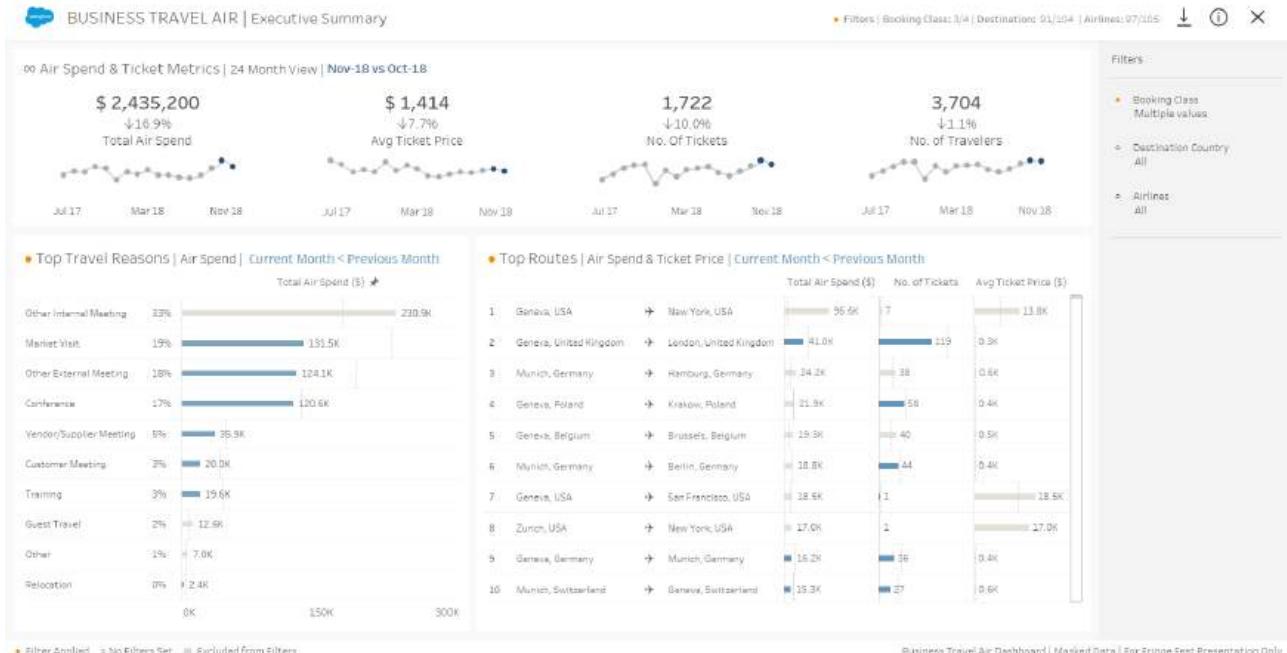
I wanted to show where exactly each striker was taking their chances from in front of the goal. Did they favour a certain area or what percentage of their chances were taken inside the box? We know that if you're taking chances in front of the goal in key areas there's a much higher likelihood you will score than if you're taking chances outside the box for example. I think the shot map shows this clearly and the key takeaway for me is that Haaland's spread of chances was tight and right in front of the goal as opposed to the others who were far more sporadic and took more chances outside the box. Haaland's positioning in front of the goal this season has been incredible and in my opinion a big reason for his success so far. Cj: You mentioned your work was inspired by a few other pieces/people in the community. Are there any specifics you'd like to call out, especially from a design perspective? Yes, definitely I'm always on the lookout for inspiration. First and foremost, I like to look outside the data viz bubble for design inspiration. I have a particular interest in cinematography and web/app design and I think there are a lot of lessons that can be learned from these industries when it comes to good storytelling and design principles. For this particular project, I did spend a lot of time on Twitter, specifically the accounts [@jonollington](#) [@sonofacorner](#) [@petermckeever](#). They all have great content so if you're interested in football they are a must-follow. [@sonofacorner](#) for me especially stood out. I love the aesthetic look of his work. The minimalist style of his designs and the way he is able to convey his analysis directly and clearly is something I tried to incorporate in my IronViz.



CJ Round-Up: Awesome thanks Fuad. Hopefully a few takeaways from today. With football analytics being such a hugely competitive market it's interesting to see what catches the eye of the community online. I think there is a great balance to reporting meaningful metrics, against beautiful design. If you have both you are on for a winner. As I mentioned, Fuad's visual this year was one of my favourites to read through both of the subject matter being soccer and also because of the elegance in design. Really looking forward to what is next from him. **LOGGING OFF, CJ**

FILTERS FOR EXECUTIVES

Hi all, Welcome to another blog. At the end of last year I published a dashboard for Tableau Fringe Festival. You can view it on Tableau Public using the link under the title above.



What I wanted to achieve with this dashboard is by simply applying a few visual design elements to existing metrics (i.e re arrange the page a little bit, amend typography and colours and balance, but not make huge chart changes or changes to metrics looked at), the user can better understand the visual. If you'd like to view the original, a print-screen can be found in the version on Tableau Public, else you can find it on the Tableau [Accelerators page](#).

Executive Summary

Current Month: May | Current Year: 2019 | Airlines: All | Destination Country: All

Current Period vs Reference Period: - denotes color coding | Current Month vs Previous Month: +

Nov-18 compared to Oct-18

What is our volume of Air Spend?



How and when do travelers book their tickets?



To what extent do travelers accept the lowest fare?



How many travelers do we have?

**\$ 1,414**
Avg Ticket Price**1,722**
Nb of Tickets**16.1 d**
ABW - Advance Booking Window**0.5%**
Lowest Fare Accepted Rate**\$ 2,351**
Air Spend per Traveler

What are the top Reason Travels?



What are the top Routes?



Salesforce | Business Travel Air Dashboard | Confidential Information | For internal use only.



Now, this blog won't look at whether you believe I have achieved that or not, but I do want to explain how and why I chose to use a colour indicator for filters. The first thing to realise is that this is an executive summary. Now we all know how much people sadly want to just take a print-screen of a dashboard and send it on. So it seems reasonable enough that they want the filters on the page. For me this takes up quite a lot of screen real estate where the title is. So is there anything we can do about this? Well yes – we could put all the filters in a show hide me container but then the executive that wants a print-screen of the page won't know what filters are applied! So I chose to add in the filter bar as a show hide panel. (Admittedly still taking up real estate on the page) The difference in this case the user now has the ability to hide them.

BUSINESS TRAVEL AIR | Executive Summary

Filters: Booking Class: 3/4 | Destination: 91/104 | Airlines: 97/105 | (1)

Air Spend & Ticket Metrics | 24 Month View | Nov-18 vs Oct-18:



Top Travel Reasons | Air Spend | Current Month < Previous Month



Top Routes | Air Spend & Ticket Price | Current Month < Previous Month

Filter Applied: (1) | No Filters Set: (1) | Excluded From Filters: (1)

Business Travel Air Dashboard | Masked Data | For Fringe Test Presentation Only

So a good half way house. We now can claw back some of that screen real estate by having our filters pop out in a show hide me on the page. Allowing the user the flexibility. One other aspect I love about a pop out filter bar is that it doesn't cover your charts up, meaning from an interactivity perspective it's less clicks. That's all well and good but then we come back to our Exec print screening problem. What is a visual method I can prompt the user if filters are applied? I ended up with a colour ascii solution. You'll see that in the footer bar, I prompt the user that an orange circle means that filters are applied. How can we create this logic?

ALL SELECTED TRUE

X

Totals summarize values from Table (across).
if TOTAL(COUNTD([Airlines])) = SUM({ FIXED : COUNTD([Airlines]) })
and TOTAL(COUNTD([Booking Class])) = SUM({ FIXED : COUNTD([Booking Class]) })
and TOTAL(COUNTD([Destination Country])) = SUM({ FIXED : COUNTD([Destination Country]) })
then
'o'
else ''
END

Default Table Calculation

7 Dependencies ▾

Apply

OK

The calculation is valid.

If a filter has been applied our total number of records will be less than our current value. So I create two calculations. One saying if they are equal, then no filters have been set, so assign it an empty ascii. I create the reverse logic, if they are not equal then we want a filled ascii circle.

ALL SELECTED FALSE

X

Totals summarize values from Table (across).
if TOTAL(COUNTD([Airlines])) = SUM({ FIXED : COUNTD([Airlines]) })
and TOTAL(COUNTD([Booking Class])) = SUM({ FIXED : COUNTD([Booking Class]) })
and TOTAL(COUNTD([Destination Country])) = SUM({ FIXED : COUNTD([Destination Country]) })
then
''
else 'o'
END

Default Table Calculation

7 Dependencies ▾

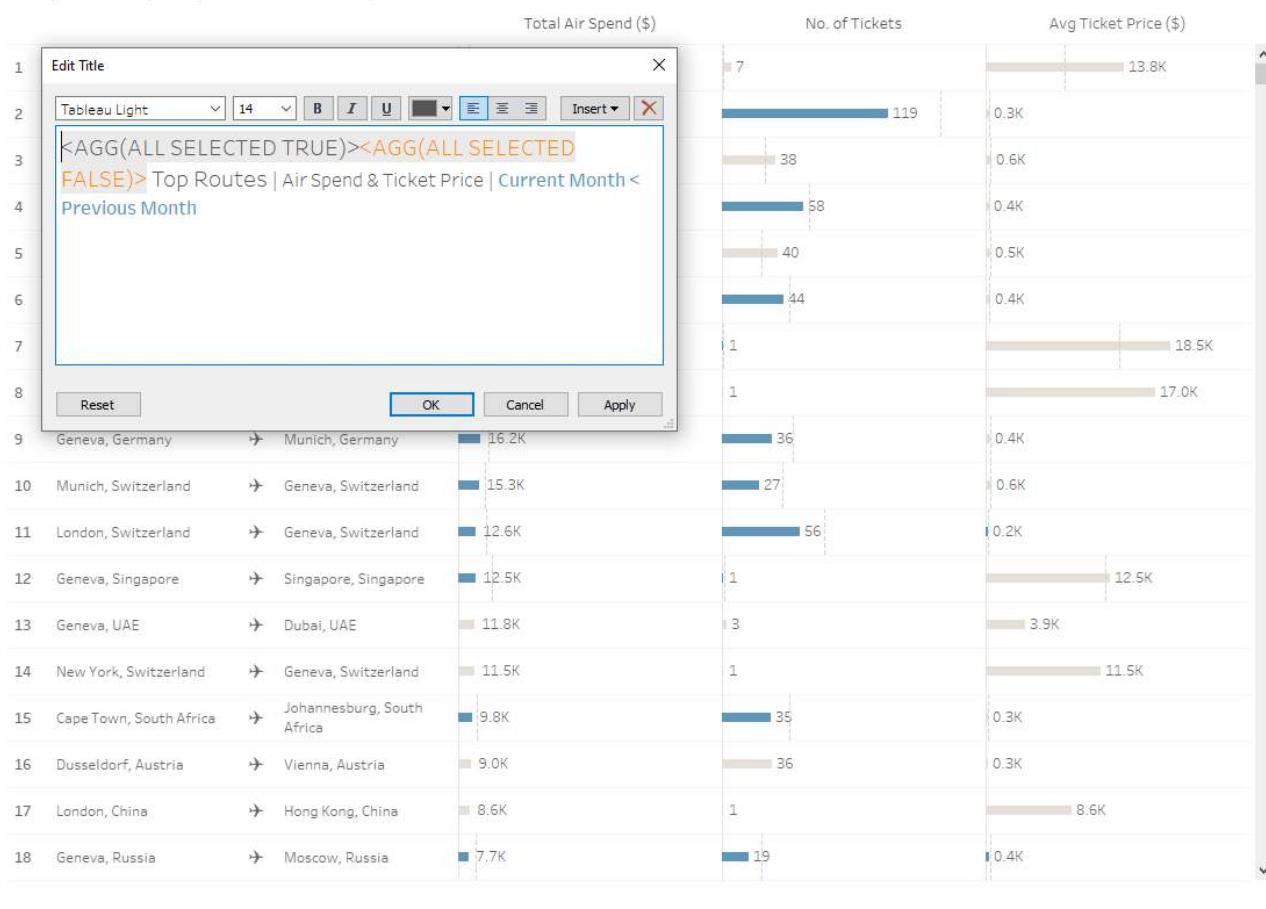
Apply

OK

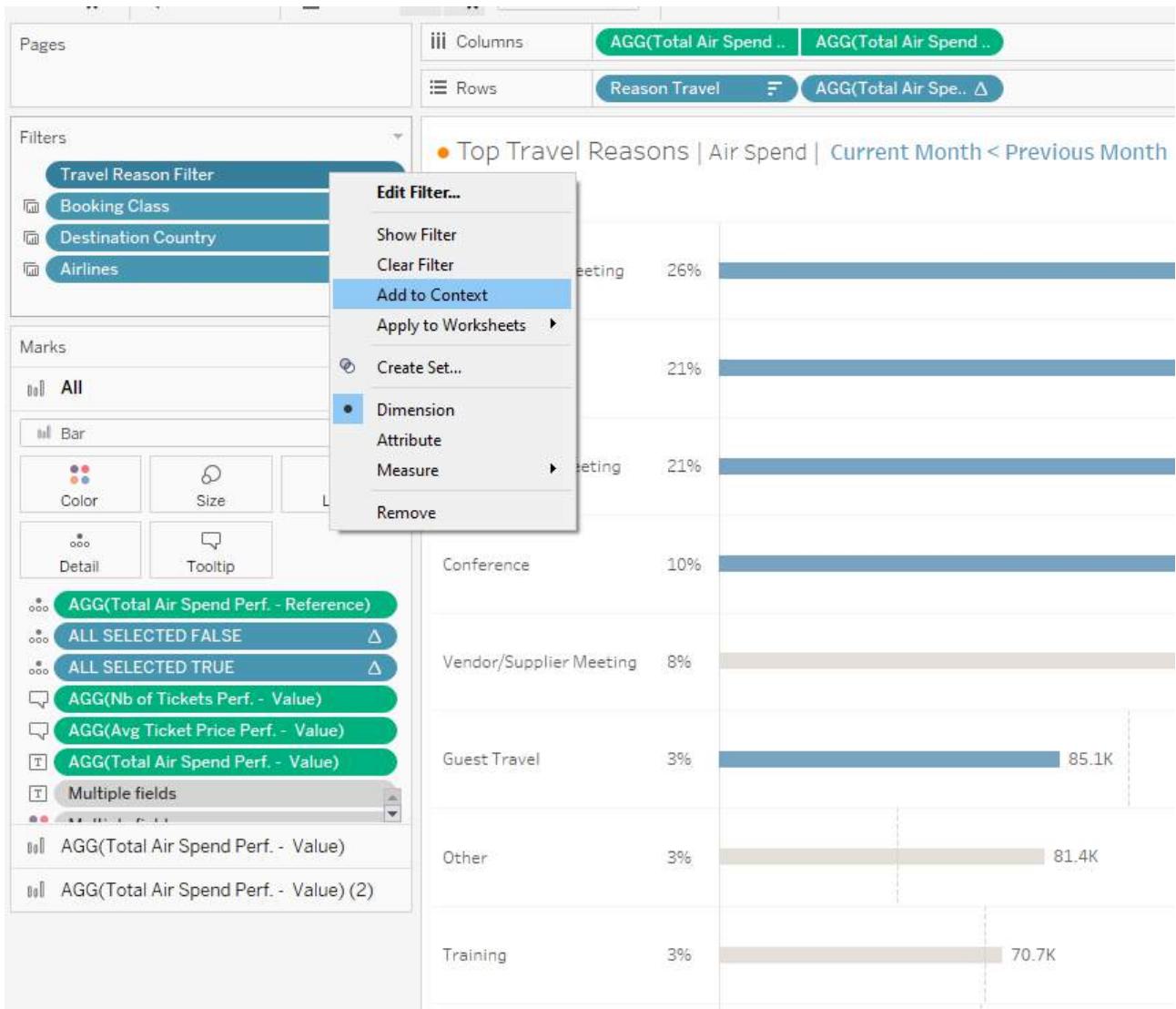
The calculation is valid.

Of course, you can create this as one calculation if you like – but you'll notice I wanted to also colour the labels separately. Therefore I actually put these two calculations side by side, and colour the false one orange. By bringing both these calculations onto the details pane of our chart we can now add them into the title as calculations.

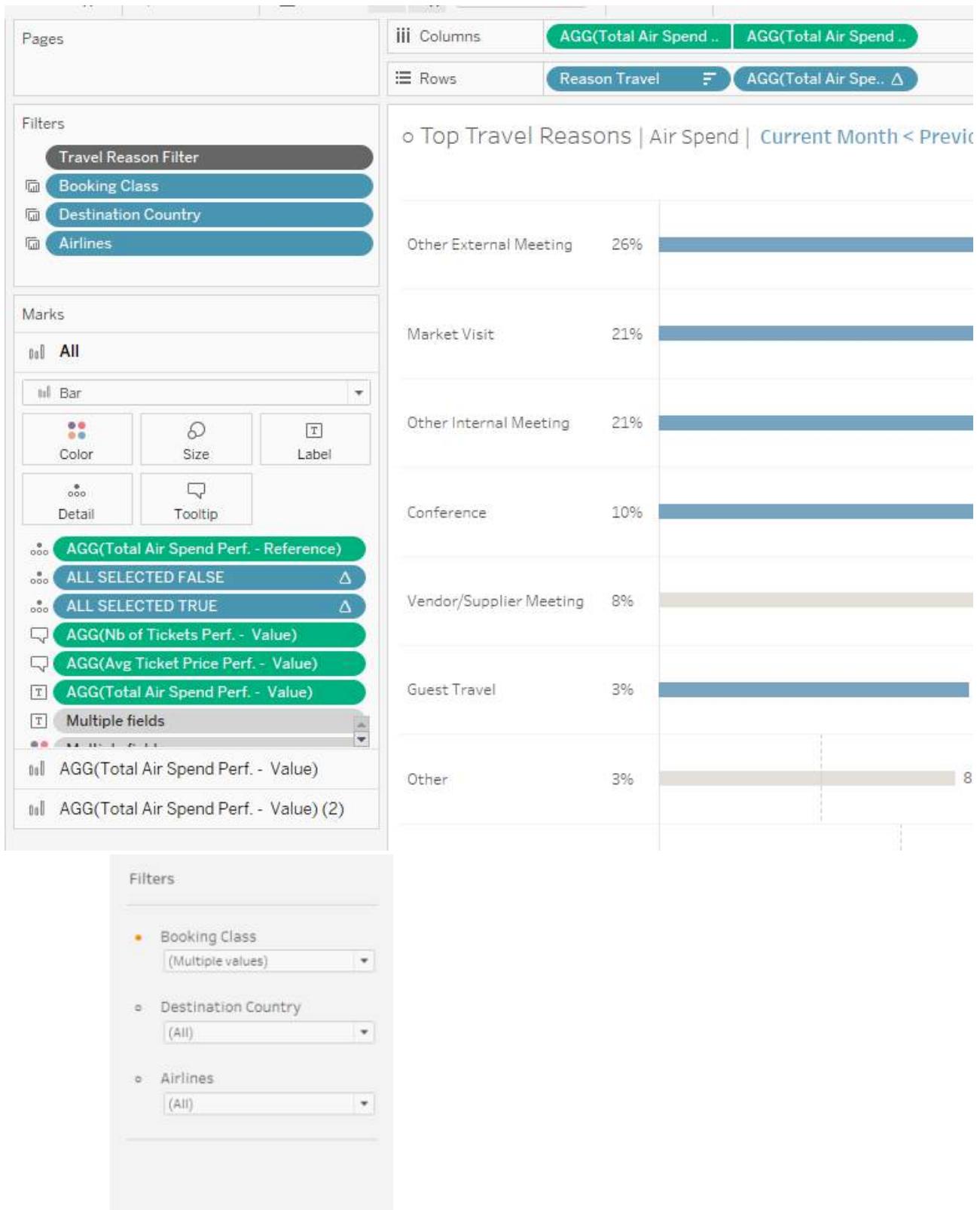
• Top Routes | Air Spend & Ticket Price | Current Month < Previous Month



This way, when our filter is applied we get an orange circle ascii, and when it isn't applied we get the empty circle. One thing to be cautious of with this method is that if you have some sheets that contain filters that you haven't accounted for, of course your total volumes won't be equal! This means you would get an orange circle when theoretically.... you have the filters you're looking at set to all.



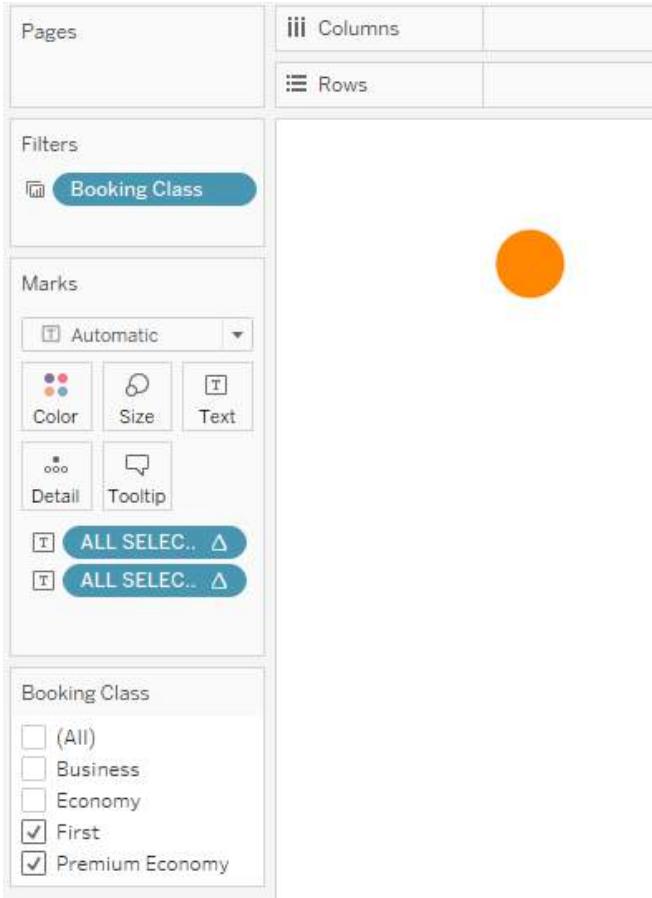
For example, take the above where we set our chart to show only the Top 10. We have to add this to context so that it does our fixed calcs after the filter has been applied.



That's better.

The final thing I wanted to do, was be able to quickly show the user per filter

whether the filter was applied or not. We look to create a new sheet. We create a sheet for EACH of the filters to place next to



the filter. The important thing here is that we make sure to only apply to this sheet the filter of *the filter*. I.E if we are creating our dot for our booking class filter, we will only want to add the booking class filter onto this sheet. Reason being, we only want the dot to change if this filter count changes, irrespective if we are clicking the other filter counts. (We have to remember our original true, false has all three calculations in) I don't necessarily think people will apply this to business dashboards moving forwards as its quite the faff, but I think the use of colour to draw your eye to the filter really helps quickly identify which filters are being used. The final touch I added, was a small filter count at

the top of the page! I completely ripped this off Ken Flerlage from about 5 years ago. You can see how the caption sheet summary works in this Tableau Public thread [here](#). Unfortunately, my filters are in the 10's or even 100's so if you go down this method and have more than 3 you end up getting something like 'Countries: United Kingdom, United States, India +96 more' which makes sense because they'd not want to print over 100 countries. So I explored a different option of just finding what the distinct count of each filter possibility is, and then a count based on the filters that are applied. I think this gives the user additional context to how much of there dataset has been filtered out.



That's it for todays thoughts, Let me know if you think it's overkill or not. I do enjoy playing around with different design ideas. I can be reached on LinkedIn or Twitter. **LOGGING OFF, CJ**

ATP TENNIS STANDINGS (RAPID API)

Hi all, The season is upon us! Yes I'm referring to the fact the Australian Open has begun. Who's excited? I certainly am. It is my plan to go to all 4 Grand Slams at some point in my lifetime. I'd assume that's quite an easy bucket list task, but funny enough living in London the only one I've done so far is the French Open. Tell you what I was excited to see the news that they will slowly be making Tennis data more available. Unfortunately not to the general public, so Nick Kyrgios if you are reading this, I am

Simon Briggs @simonrbriggs

...

Tennis “Moneyball” alert. The ATP plan to make ball-tracking data - which previously was an expensive nightmare to source from HawkEye (at £150 per match) - freely available to their players from the spring



telegraph.co.uk

The tennis data revolution that can level the playing field

Exclusive: Historically the sport has kept quality data out of the reach of the players, who could benefit the most from their analysis

happy to come consult on your team.

Today we are going to look at accessing data through an API. What do we really mean when we talk about API's? In typical fashion we overcomplicate the data world, but really it is just an intermediary that sends a message of a request, i.e give me this data/ information, and then brings back a response. Kind of like writing a letter, posting it, and then someone responding to the

letter with whatever the answer to the question you asked. The reason I use the postman analogy is because someone conveniently created an application called postman where we can see that request happen. I'll be writing a follow up blog at some point about how to use Postman when accessing Strava data, but for now I'll simply use it to showcase the request in action.

Welcome to the RapidAPI Hub

Discover and connect to thousands of APIs

Public APIs Private APIs

Categories

- Data
- Sports
- Finance
- Travel
- Entertainment
- Location
- Science
- Food

Discover More APIs

Browse through our collections to learn about new use cases to implement in your app



There are plenty of different API connectors you can retrieve information from. Today we will be using a website called RapidAPI.

Feel free in your own time to look at some of the other API's available. (You will need to make an account, I am assuming if you want to learn about API's I don't need to explain how to click sign up to a website!) The one we will demo today is the [TennisAPI1](#)

The screenshot shows the RapidAPI platform interface. At the top, there is a navigation bar with links for "API Hub", "My Apps", "My APIs", "Docs", a notification bell with a red badge, and a user profile icon. Below the navigation bar, the "TennisApi1" API is displayed. It has a "FREEMIUM" status and was updated a month ago by "REcodeX". The API has a popularity rating of 9.7 / 10, latency of 420ms, and a service level of 100%. Below the API card, there are tabs for "Endpoints", "About", "Tutorials", "Discussions", and "Pricing", with "Subscribed" status indicated. The main content area is titled "TennisApi1 API Documentation". It includes a brief description of the API's purpose, mentioning tennis scores for over 500 tournaments, ATP and WTA rankings, and live results. It also notes that high API calls rate plans can be found at <https://rapidapi.com/fluis.lacasse/api/allsportsapi2>. A section titled "What are some things we should probably know?" lists bullet points about latency and API strikes. To the right of this text is a sidebar titled "Search endpoints" containing a list of endpoints: GET Flags, GET Search, > Events, > Players, < Rankings, GET ATP Rankings, GET WTA Rankings, > Tournaments. Each endpoint has a small info icon next to it.

- Latency is pretty much how quick or slow the API is. (I.e the delay time of your post, luckily there are no Christmas strikes with API's)
- End points refer to the url of where we want to get the information from. Just think of this as when you send a letter you have to write your recipients address on the envelope. (Yes I am going to continue this fairly shoddy analogy)
- It's worth noting that the different end points retrieve different bits of data, we can see the URL update when we click through to the ATP Rankings end point.
- Here is the url we want to get information from <https://tennisapi1.p.rapidapi.com/api/tennis/rankings/atp>

```

const axios = require("axios");

const options = {
  method: "GET",
  url: "https://tennisapi1.p.rapidapi.com/api/tennis/rankings/atp",
  headers: {
    "X-RapidAPI-Key": "na0dffsb75730",
    "X-RapidAPI-Host": "tennisapi1.p.rapidapi.com"
  }
};

axios.request(options).then(function (response) {
  console.log(response.data);
}).catch(function (error) {
  console.error(error);
});

```

- Next thing to consider is the API Key. Imagine you are sending your friend a personal letter, that you only want them to read. The key is your way of protecting the information so that only you can request it. In this case, we use it to retrieve the data required.

So how do I know what information we can get? Okay well we can test the endpoints in an application called Postman. **Postman** is an API interface development tool, which will help you build test and modify API's. I.e Am I sending my requests to the correct place? Is it telling me the correct information? What do I need to amend to ensure save retrieval of the information? Like I mentioned, the set up of postman will come in a future blog, but for now. How do we check our end point is correct, Well we add in our url from the site. We also need to add in our key. You'll see the key name is X-RapidAPI-Key, so I will need to add this to the headers section. I'll need to add the super long value for this key.

Once I've done this, I can click send.

The data it retrieves tends to be in JSON format. Fine for sending and retrieving data, but if I actually want to do anything substantial in a visualisation tool I will need to see this data probably in a dataframe/table/csv format. This is where I load up python and recreate the steps there.

```
import requests
import json
import pandas as pd

# set the endpoint URL and headers
endpoint = "https://tennisapi1.p.rapidapi.com/api/tennis/rankings/atp"

headers = {
    "X-RapidAPI-Key": "REDACTED",
    "X-RapidAPI-Host": "tennisapi1.p.rapidapi.com"
}

# send a GET request to the endpoint
response = requests.get(endpoint, headers=headers)
print(response)

# convert the json response to a python dictionary
data = json.loads(response.text)
print(data)

rank_list = []
name_list = []
points_list = []

a = 0

# loop through the data and write the values to the csv file
for item in data['rankings']:
    # access the first item in the rankings list
    ranking = data['rankings'][a]
    # access the name
    name = ranking['rowName']
    # access the ranking
    rank = ranking['ranking']
    # access the points
    points = ranking['points']
    # print the details
    print("Name:", name)
    print("Ranking:", rank)
    print("Points:", points)
    rank_list.append(rank)
    name_list.append(name)
    points_list.append(points)
    a = a + 1

# Calling DataFrame constructor after zipping
# both lists, with columns specified
df = pd.DataFrame(list(zip(rank_list, name_list, points_list)),
                  columns=['Rank', 'Name', 'Points'])
df.to_csv(r'ATP_rankings.csv', index=False)

print("ATP Rankings successfully written to ATP_rankings.csv")
```

What does the code do?

- First we want specific packages. One to help us make requests, one to convert json format, and another to be able to build our data into a dataframe.
- The endpoint like I mentioned is the custom url we go to retrieve our information.
- The headers are taken from the API page.
- We end up retrieving the data and it is in the json format, like when we tested it in Postman.
- All we then do is take the attributes we want from this file and glue them together in a dataframe.

Rank	Name	Points
1	Carlos Alcaraz	6820
2	Rafael Nadal	5770
3	Casper Ruud	5720
4	Stefanos Tsitsipas	5715
5	Novak Djokovic	5070
6	Andrey Rublev	3930
7	Felix Auger-Aliassime	3895
8	Daniil Medvedev	3860
9	Taylor Fritz	3545
10	Holger Rune	2876

There we have it our nice shiny list of the Top 500 Mens Players in the ATP. Hopefully this was somewhat useful for an introduction into API's. As always the code can be found on the Github repo under the title. Going Further:

- Why not try use a different endpoint and retrieve the information from it?
- Why not try use a different API in its entirety?
- Why not try automate a flow that makes multiple requests a day / over a number of days?

LOGGING OFF, CJ

GRADIENT CIRCLES

Hi all, Ahhhh gradients, what a blast from the past in the design world. A couple years ago you may have seen Ivett publish



about colour gradient circles. I actually think they look amazing.

Ivett's example uses bins to

create each of the different circle gradients, and has applied some beautiful palettes to each of them shown above. You can see the creation [here](#). If you are more a bins person, why not also check out Toan's blog on how to create gradients, [here](#).

For me, I'm not huge on bins, they're simple to grasp but I have my way of making circles now and I like to stick to it. Not very accommodating of me, I know but joining data in just feels so much easier. So today is a mini tutorial on how you can emulate gradient circles (or portions of, but using a simple dataset. What does it look like? You'll see I have one column called dot. 1 through 360, coincidentally the number of degrees in a circle. You can download the workbook and dataset from the top of the page under the Github and TP links under the title. Create a calculation ranking these

001. Rank

Next we want to find an equal constant spacing between them so find the angle the dots will be separated by. i.e 360 degrees divided by the number of dots.

002. Angle

(360/{fixed : MAX()}) We then times these values together, to space them out

003. Rank Angle

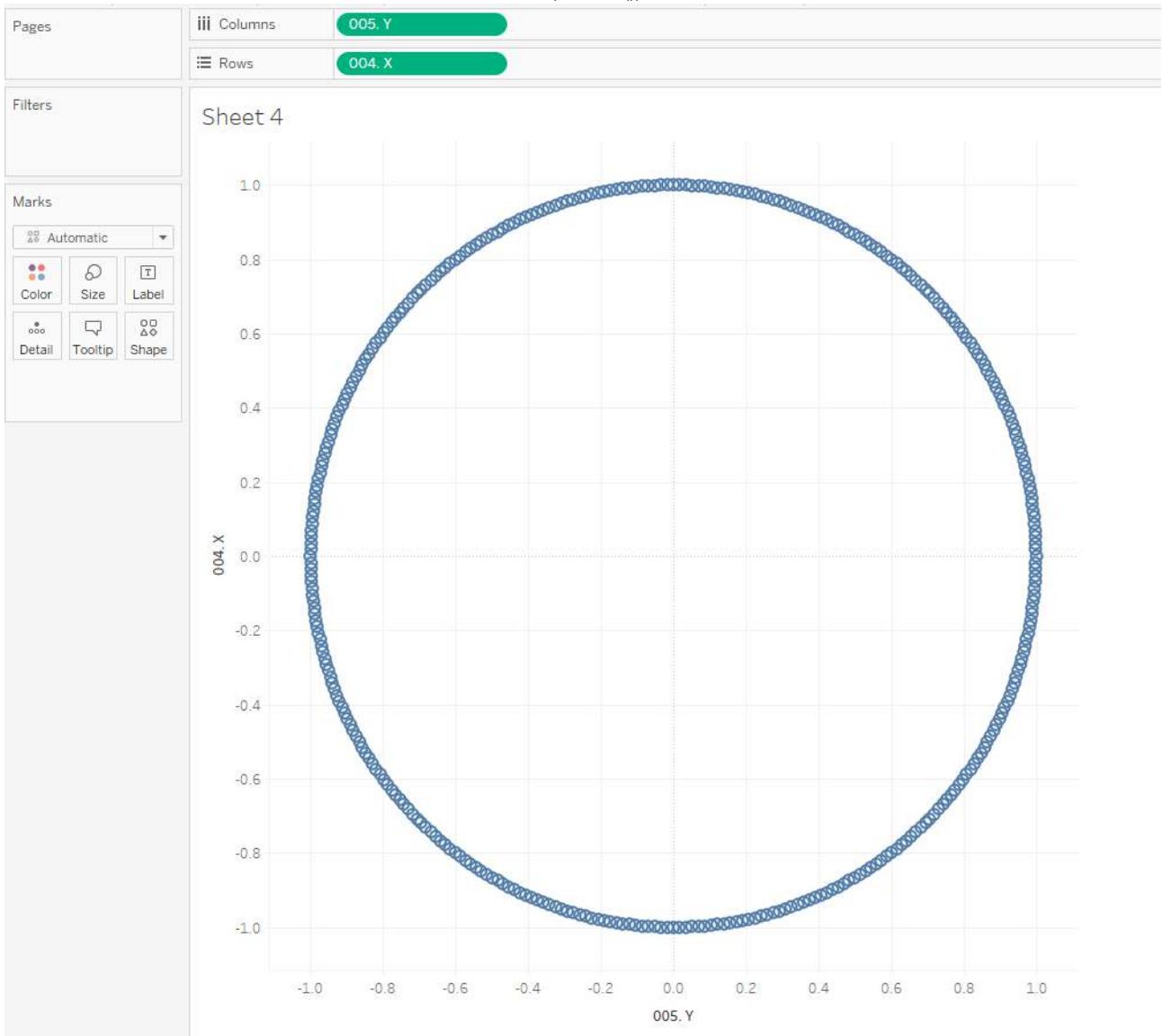
* Finally a bit of trig, and we have our circle when we plot them against each other as dimensions, easy.

004. X

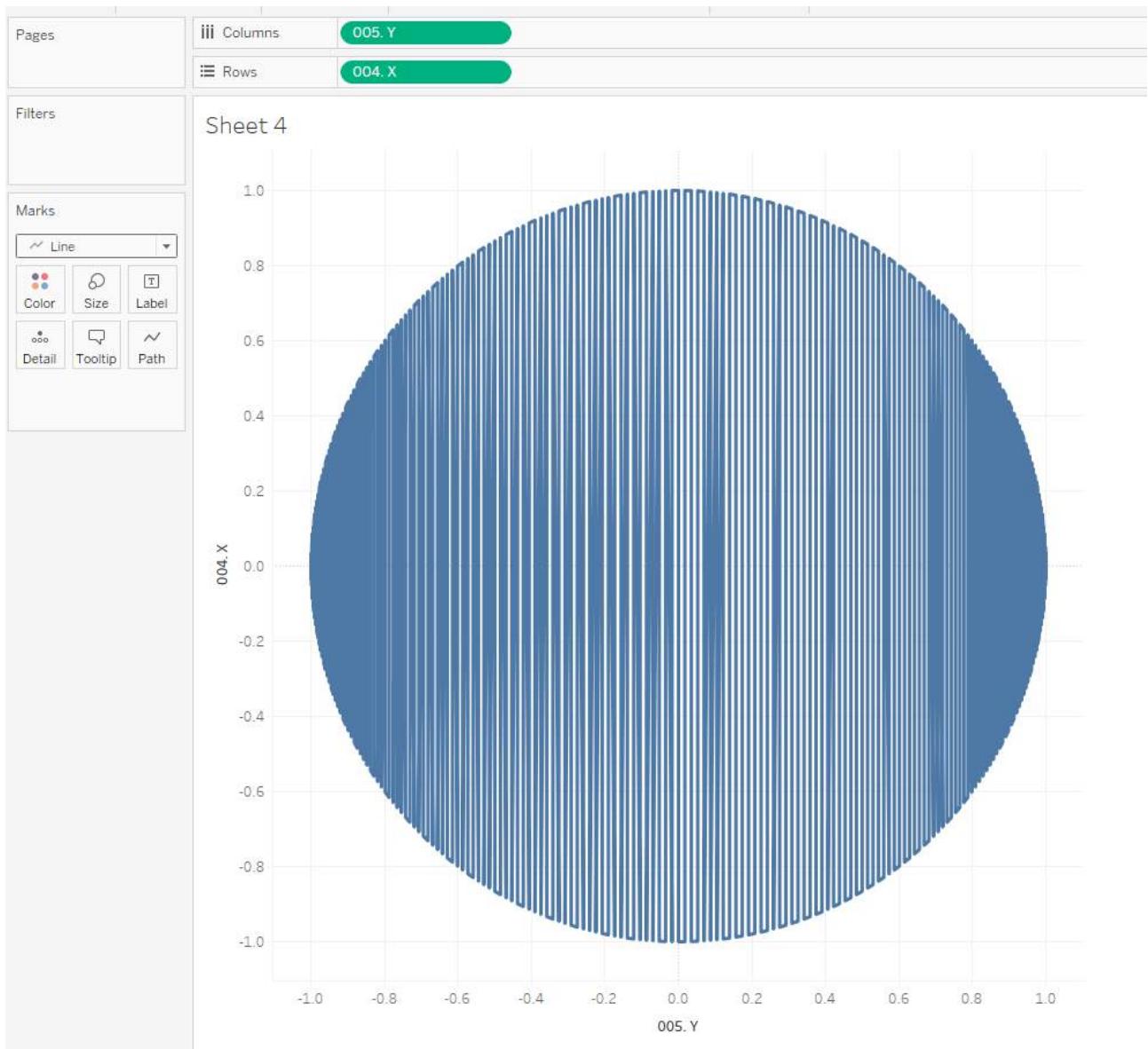
$\cos(\text{radians}())$

005. Y

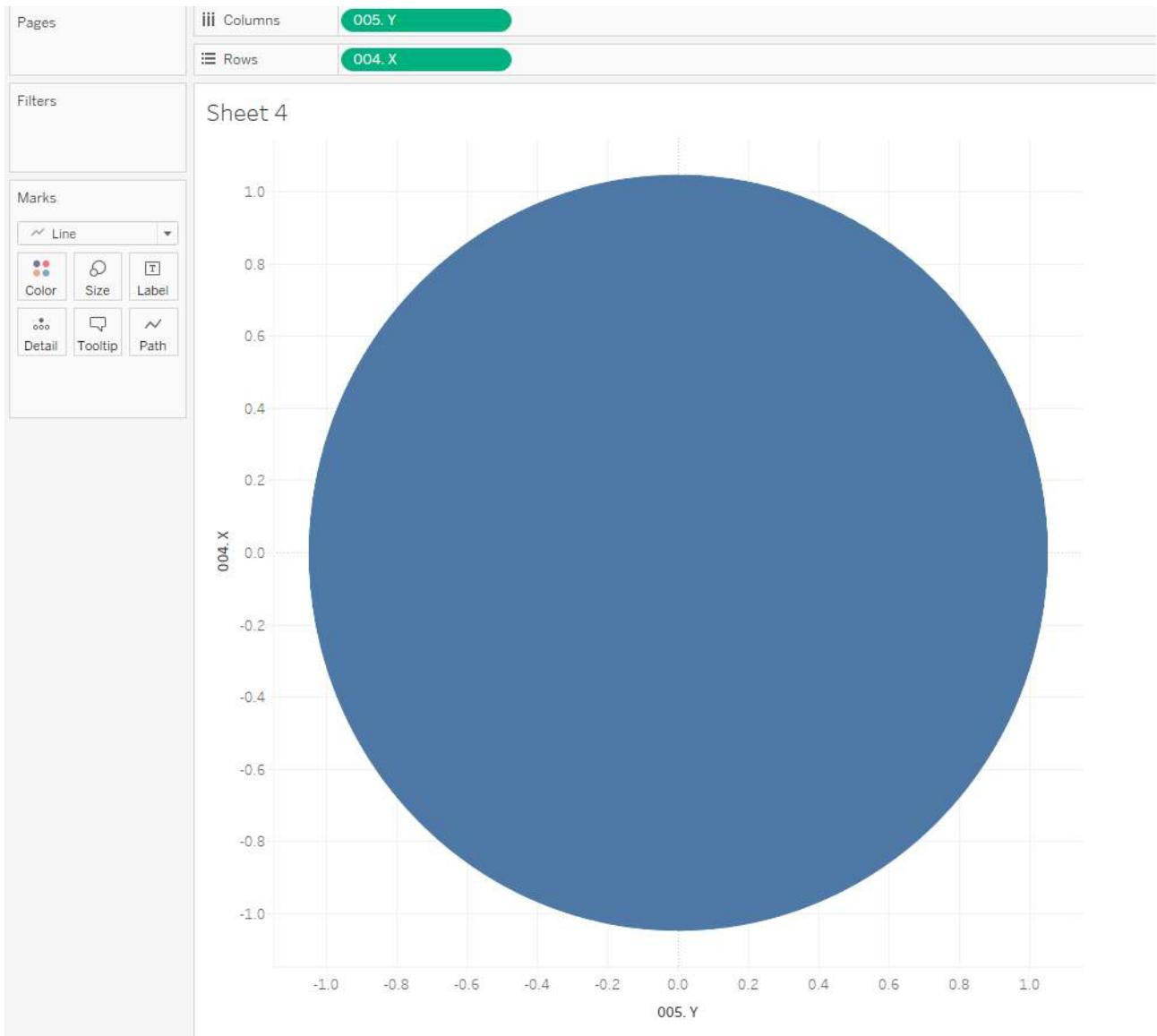
$\sin(\text{radians}())$



Next thing we can do is change path to line, Now it will automatically pick the path of the line, as we have nothing yet on the path.



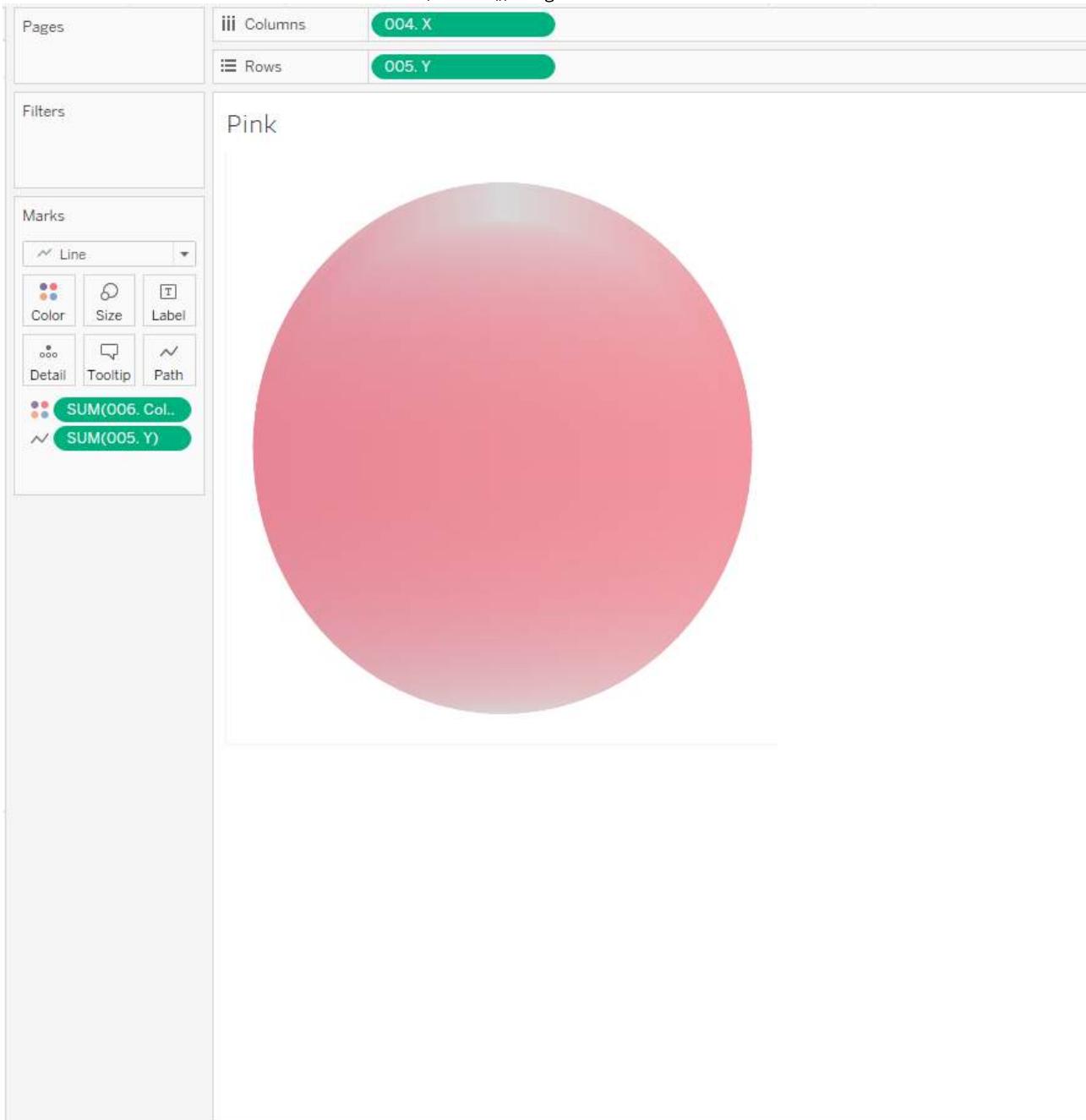
If we increase the size of these marks, then it fills the circle in nicely.



Now we just need to add a bit of colour. Lets drag Y onto path, this means we will have our gradient working horizontally, the thing to recognise with this is, the way we draw the path is the way our gradient will end up. Create a colour calculations

006. Colour

$\cos(\text{radians}())$ Drag that onto colour



Of course we can create a parameter to help the angle of our gradient.

Edit Parameter [P6. Change Gradient Angle]

X

Name: Comment >>

Properties

Data type:

Current value:

Value when workbook opens:

Display format:

Allowable values: All List Range

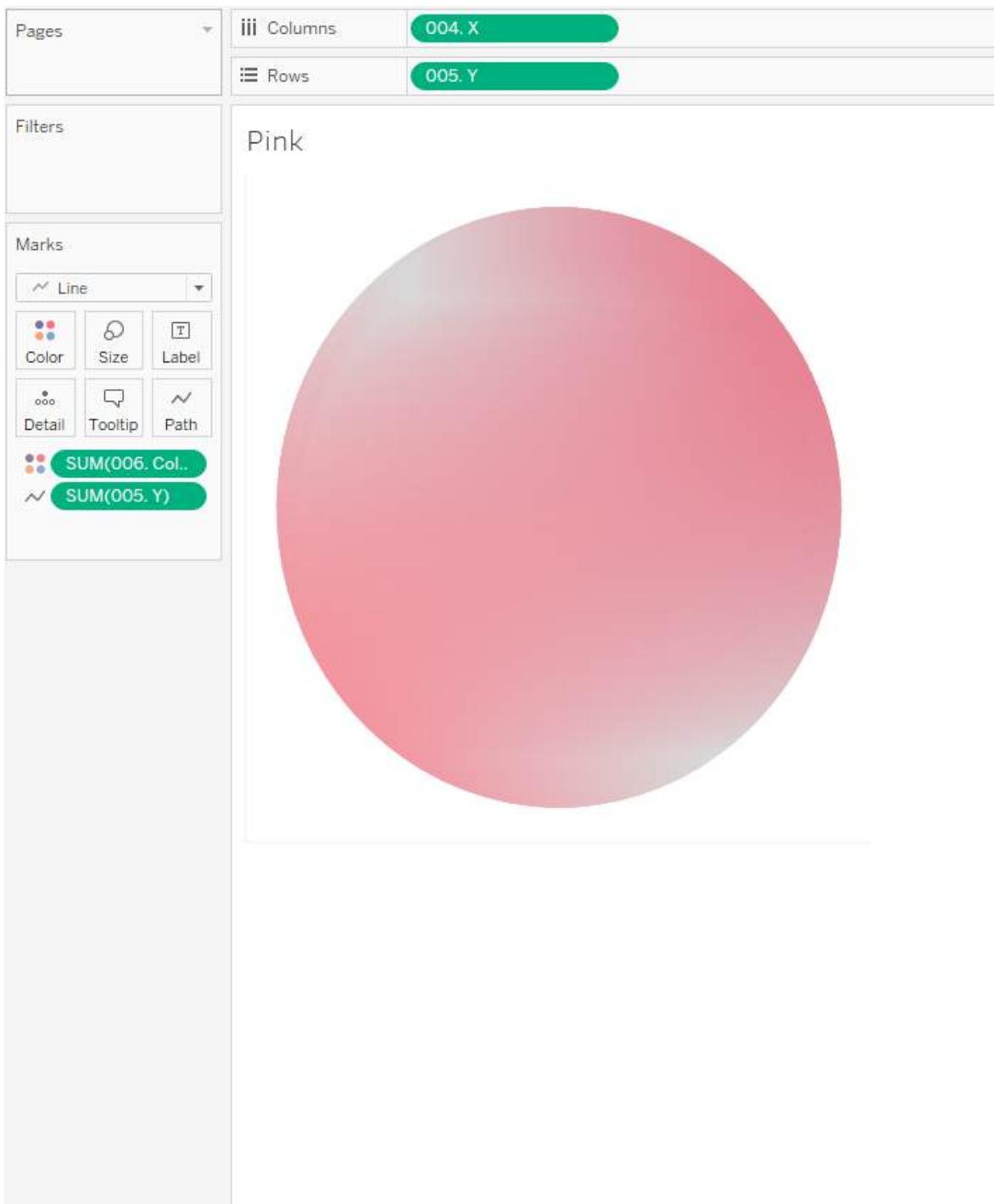
Range of values

Minimum: Fixed

Maximum:

Step size: When workbook opens

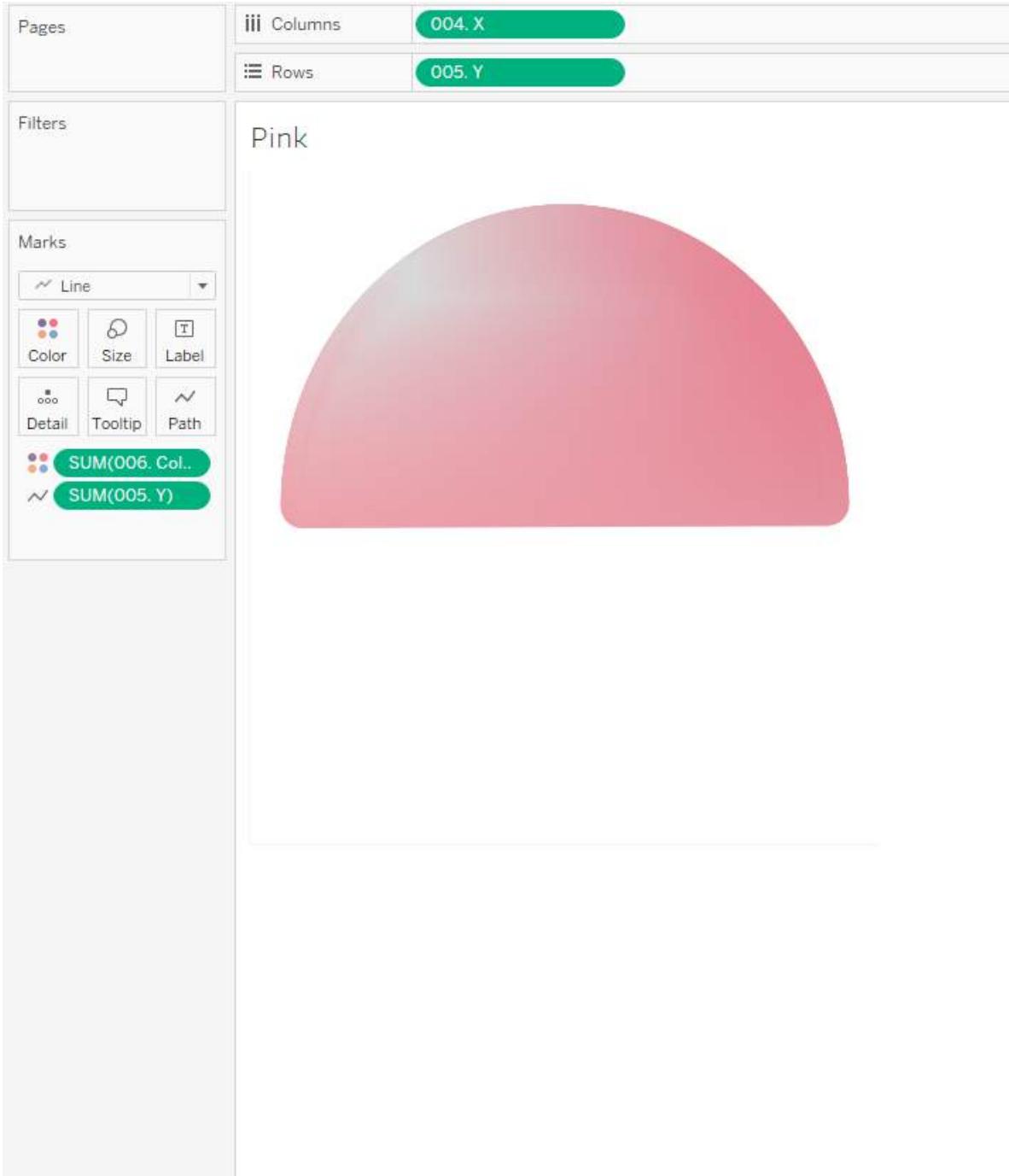
Finally we can go back and amend our colour calculation $\cos(\text{radians}(-))$. Now when we change the value of the parameter the gradient changes rotation.



I think the beauty of building gradient circles this way is that you can also start to look to bring in more layers!



One final thing that we will have learnt in previous blogs is we don't even have to make our circles full circles. Perhaps we want to create some geometric gradient patterns. You can play around with the path lines to be able to create custom shapes.

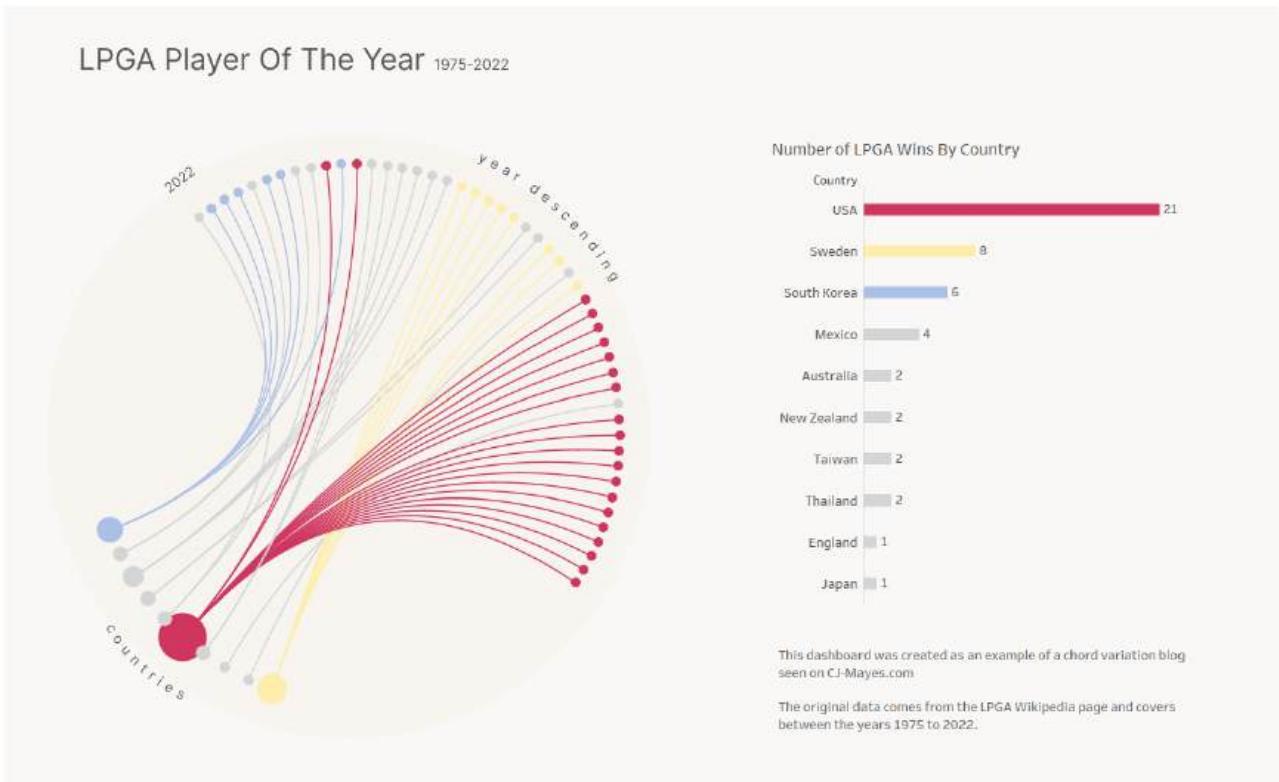


Bit of a

fun one today. Probably worth thinking about how it may explode your data to be 100's of times the size, but pretty no doubt.
LOGGING OFF, CJ

MORE CHORD VARIATIONS

Hi all, At the end of last year I revised one of my Tennis visualisations for a third time that included a chord diagram. The [accompanying blog](#) seems to be gaining a fair amount of traction in terms of views. I was also pretty glad to see one of my favourite designers in the community, Lindsay Betzendahl use the chart as inspiration for her Iron Viz feeder for her [Wordle games entry](#). Do Check it out, it was certainly one of my favourites from this years entries. So here we are, doing more on chords.



For those that follow my blogs, I create all my chord charts in the same structure as follows:

- Rank your points
- Find the angle needed to separate them
- Spread them equally using the rank and angle
- Make them circular using trigonometry

That logic goes for any circle really. Today we will take some of that logic I've previously written about but create a chart that has:

- The start and end points as flexible points
- An ability to increase and decrease the spacing
- Re-introduce middling points for the chords (Bezier curves)

As always, the dataset can be found on the Github page at the top of the page. The visualisation can be found on the following link on Tableau Public. I think by this stage I don't need to do all the disclaimer malarkey of telling you where and when is appropriate to use a chart like this, so lets just crack on. This will be a run-through as oppose to template for that very reason.

The Data We have a list of the LPGA Number 1's of each year split by Player and Country. Each of these in our final chart **will need their own mark**, so I've given them a rank ascending from the most recent year down to the earliest record. This will form

A	B	C	D	E
Year	Player	Outer Rank	Country	
2022	Lydia Ko		1	New Zealand
2021	Ko Jin-young		2	South Korea
2020	Kim Sei-young		3	South Korea
2019	Ko Jin-young		4	South Korea
2018	Ariya Jutanugarn		5	Thailand
2017	Sung Hyun Park		6	South Korea
2017	So Yeon Ryu		7	South Korea
2016	Ariya Jutanugarn		8	Thailand
2015	Lydia Ko		9	New Zealand
2014	Stacy Lewis		10	USA
2013	Inbee Park		11	South Korea
2012	Stacy Lewis		12	USA
2011	Yani Tseng		13	Taiwan
2010	Yani Tseng		14	Taiwan
2009	Lorena Ochoa		15	Mexico
2008	Lorena Ochoa		16	Mexico
2007	Lorena Ochoa		17	Mexico
2006	Lorena Ochoa		18	Mexico
2005	Annika Sörenstam		19	Sweden
2004	Annika Sörenstam		20	Sweden
2003	Annika Sörenstam		21	Sweden
2002	Annika Sörenstam		22	Sweden
2001	Annika Sörenstam		23	Sweden
2000	Karrie Webb		24	Australia
1999	Karrie Webb		25	Australia
1998	Annika Sörenstam		26	Sweden
1997	Annika Sörenstam		27	Sweden
1996	Laura Davies		28	England
1995	Annika Sörenstam		29	Sweden

the out spread of our chord.

We will want our starting co-ordinates to come from the country, then fan out to the players. Therefore they all need their own rank value. In this case, I

A	B
Country	Inner Rank
New Zealand	1
South Korea	2
Thailand	3
USA	4
Taiwan	5
Mexico	6
Sweden	7
Australia	8
England	9

simply find the distinct in a new sheet and number them and rejoin that data in. The final tab on the excel is the value T. This is 100 points (don't change this) We are just going to use them to plot the chord in between our start and end points. Here's what the join should look like in Tableau. Glue in your inner tab on the inner rank names, i.e country. and then create a custom join of 1=1 to blow the dataset up in terms of 100 new records for each individual row in our original dataset.

⌚ Data+ (LPGA)

Data is made of 3 tables. ⓘ

The screenshot shows the Power BI Data+ interface. At the top, there's a diagram illustrating a join between two tables: 'Data' (represented by a grey bar) and 'T' (represented by a grey bar). The join type is set to 'Inner'. Below this is a 'Join' configuration window with four options: Inner (selected), Left, Right, and Full Outer. It also includes fields for 'Data Source' (set to 'T') and 'Condition' (set to '1 = 1'). A button 'Add new join clause' is available. Below the join configuration is a data table with the following columns and rows:

#	Year	Player	#	Outer Rank	#	Country	#	Inner Country (Inner)	#	Inner Rank	#	T
	1995	Annika Sörenstam		29		Sweden		Sweden		7		1
	1996	Laura Davies		28		England		England		9		1
	1997	Annika Sörenstam		27		Sweden		Sweden		7		1
	1998	Annika Sörenstam		26		Sweden		Sweden		7		1
	1999	Karrie Webb		25		Australia		Australia		8		1
	2000	Karrie Webb		24		Australia		Australia		8		1
	2001	Annika Sörenstam		23		Sweden		Sweden		7		1
	2002	Annika Sörenstam		22		Sweden		Sweden		7		1
	2003	Annika Sörenstam		21		Sweden		Sweden		7		1
	2004	Annika Sörenstam		20		Sweden		Sweden		7		1
	2005	Annika Sörenstam		19		Sweden		Sweden		7		1
	2006	Lorena Ochoa		18		Mexico		Mexico		6		1
	2007	Lorena Ochoa		17		Mexico		Mexico		6		1
	2008	Lorena Ochoa		16		Mexico		Mexico		6		1
	2009	Lorena Ochoa		15		Mexico		Mexico		6		1
	2010	Yuko T		14		Japan		Japan		5		1

Let's create the original circle

00. Rank

01. Angle

360/{fixed: COUNTD()}

02. Rank Angle

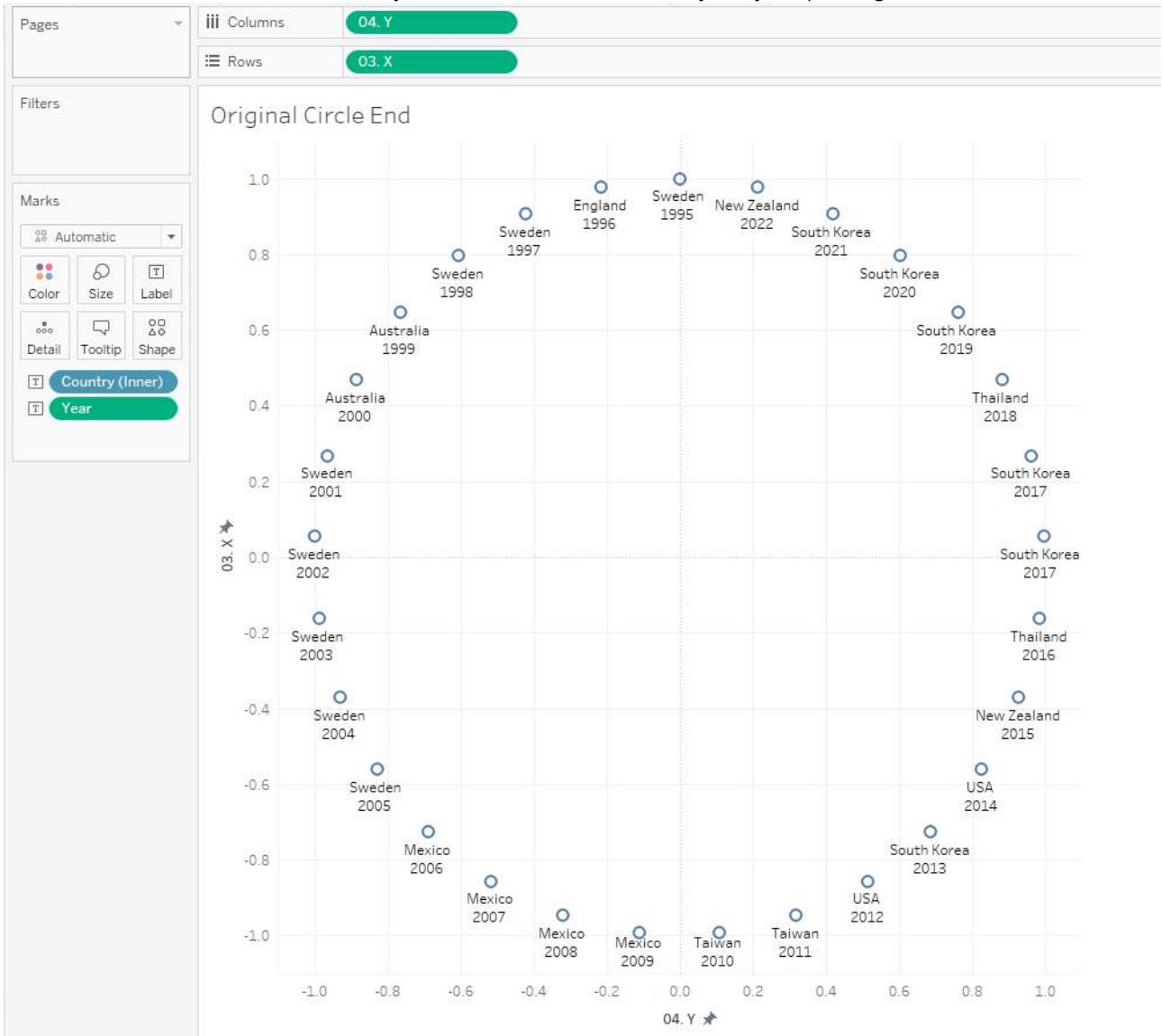
*

03. End X

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

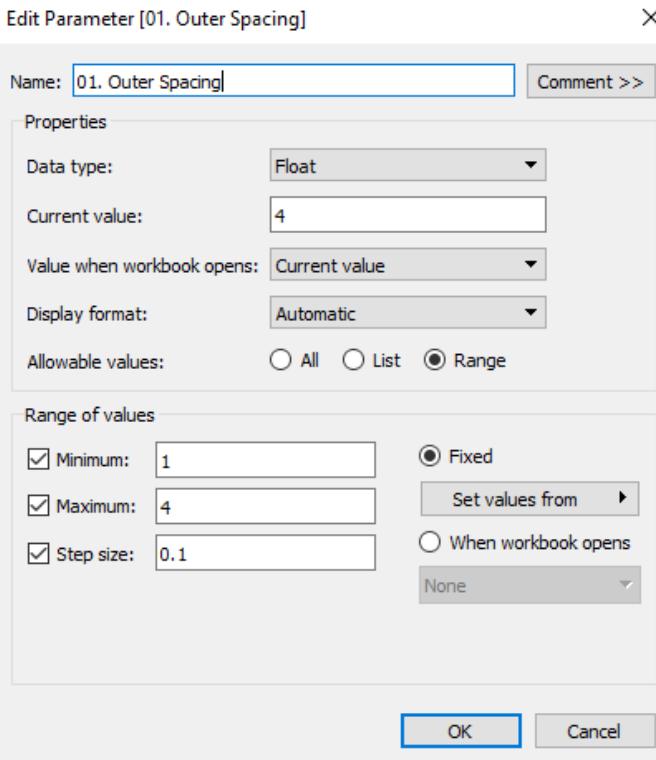
04. End Y

$\sin(\text{RADIANS}())$ Told you I make all circles the same way. Okay lets plot X against Y.



So at the moment we have a circle, but really we will want to squash these points together and move them around. Let's go back

and amend the Angle calculation. First we will need to make a parameter, this will be a float between 1 and 4 with increments of

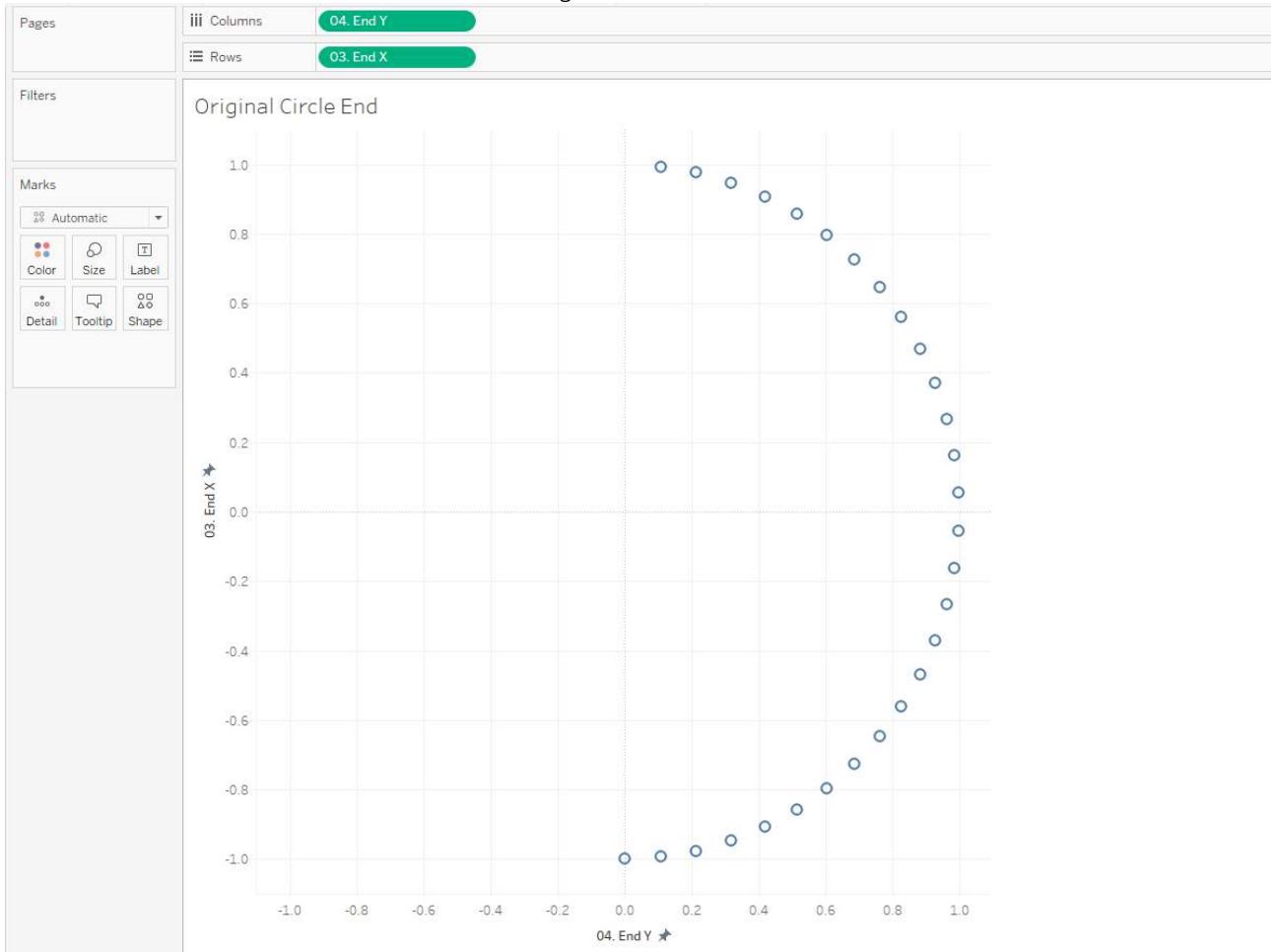


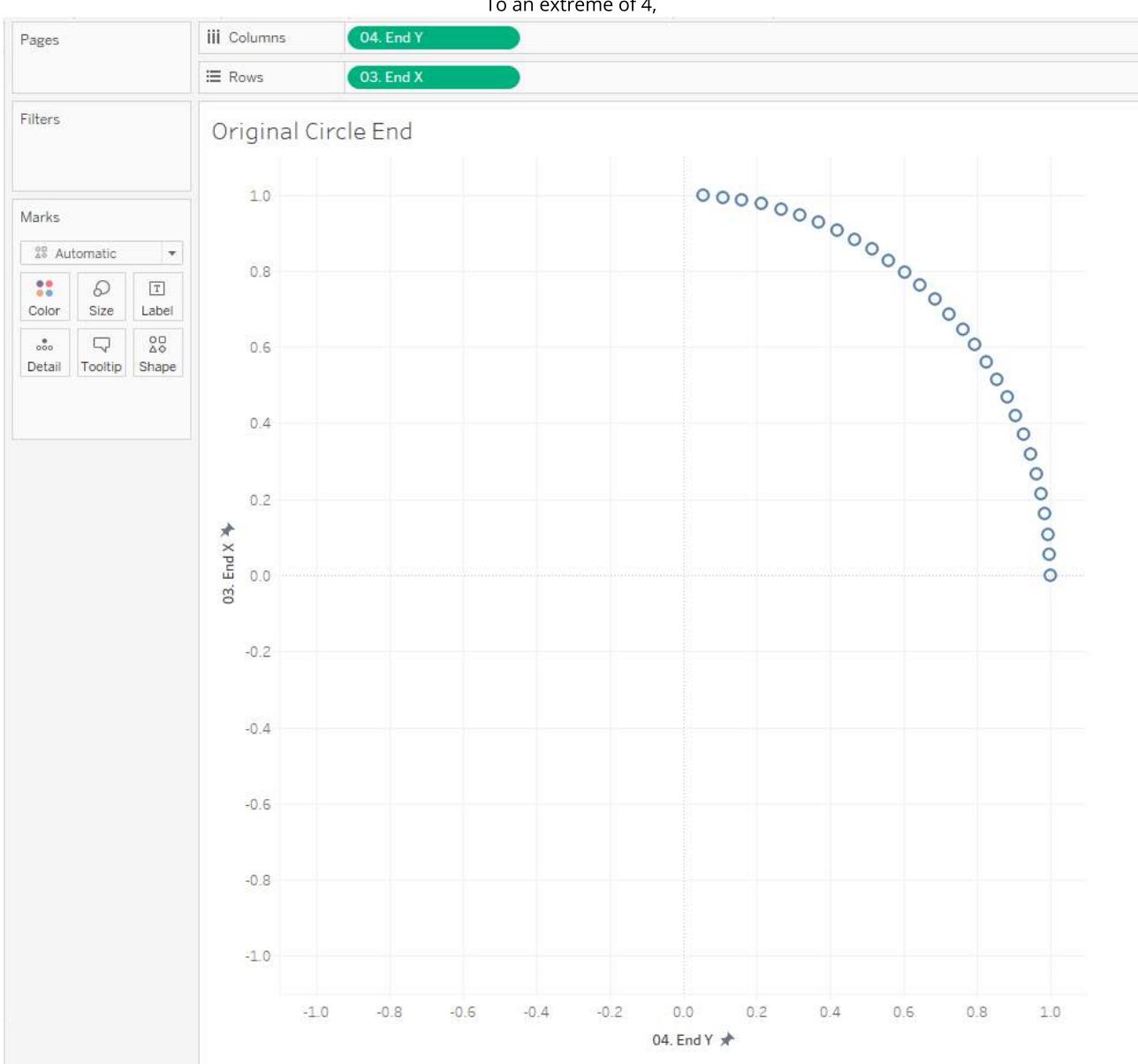
0.1.

Amend your angle calculation now to

01. Angle

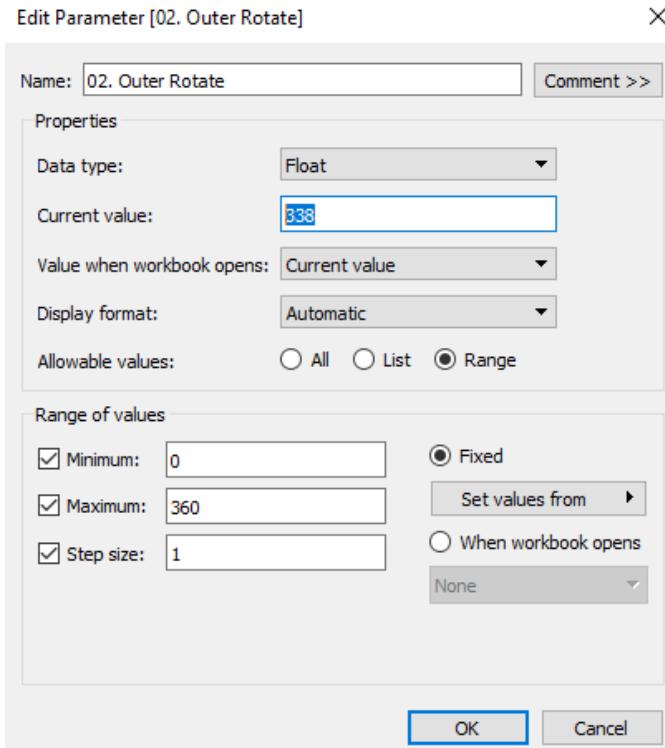
$360/\{\text{fixed: COUNTD()}\}$ / Have a play around with the parameter value to see how it impacts the circle, here are the two extremes. 1, which is our original circle. 2 Which is a half circle





Technical term, see how they get squished together. Okay, But now we may want to actually rotate these values. So lets add in some new calculations that allows us to rotate these points about our origin 0,0. First lets create a parameter

02. Outer Rotate



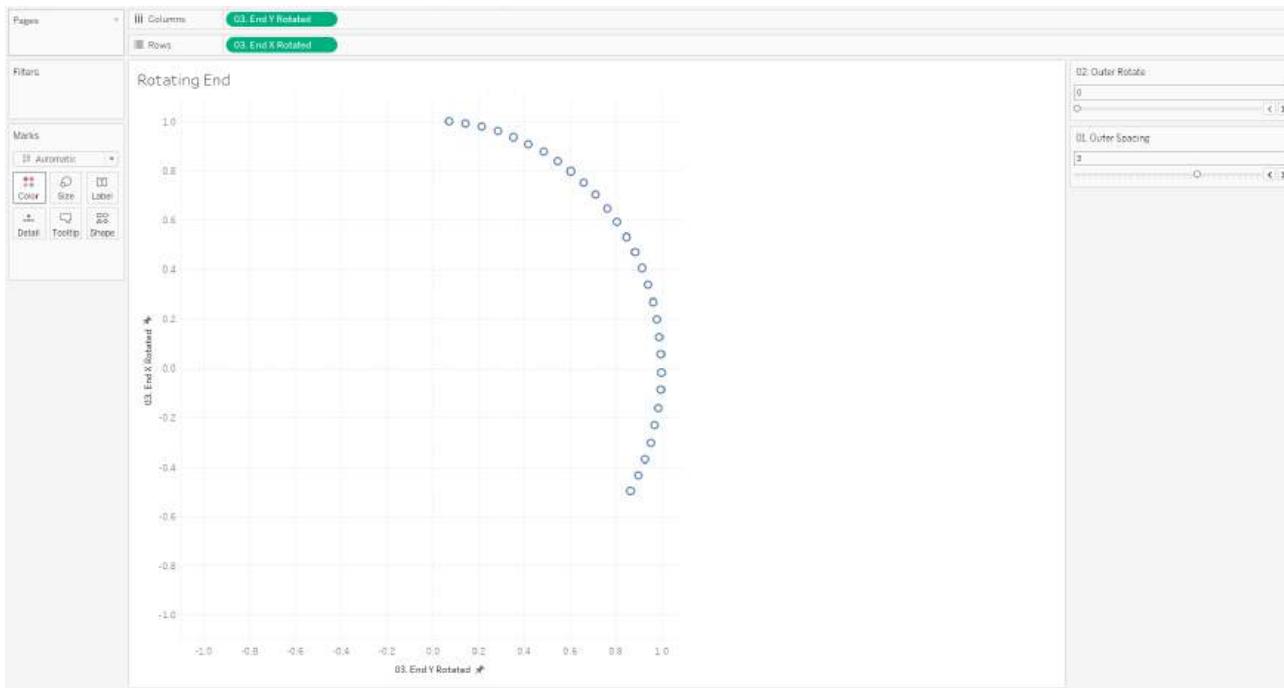
You'll see that we make the parameter between values 0 and 360, so we can spin our points to any position on the circle. Next create our two new calculations

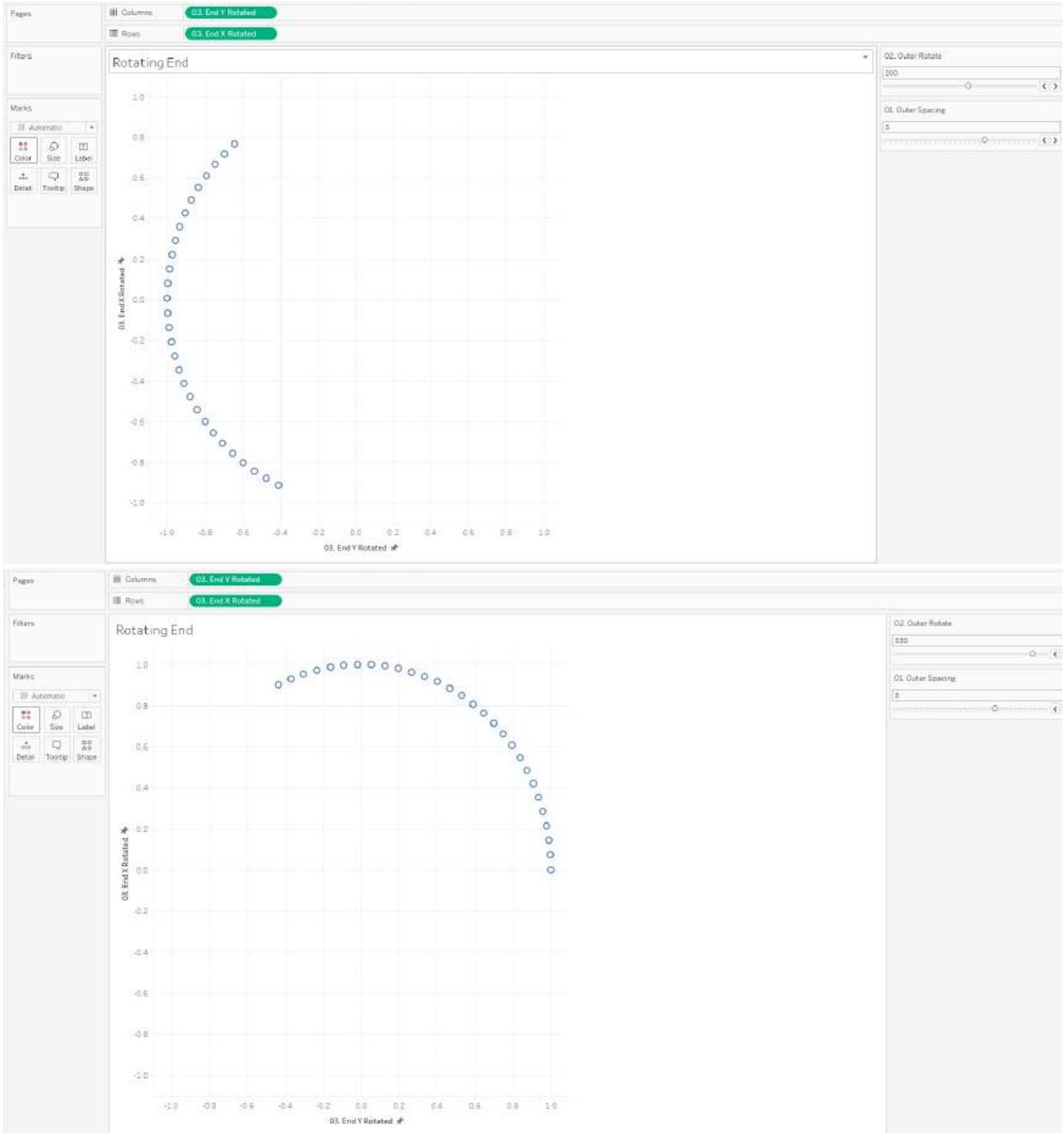
03. End X Rotated

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

03. End Y Rotated

$(\sin(\text{radians}()) + (\cos(\text{radians}()) - 0))$ If you'd like to learn more about what these calculations do, you can read more from this [blog here](#). Or even better read my blog on [Understanding Polygons 1](#). All goes well you can now plot your end points and rotate them. Here are a few variations





Cool, now we have some end points finally. Time to repeat the process for our start points.

10. Rank

Create a new parameter 11. Inner Spacing, to be the same set up as the outer spacing.

11. Angle

$$360/\{\text{fixed: COUNTD()}\} /$$

12. Rank Angle

*

13. Start X

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

14. Start Y

$\sin(\text{RADIANS}())$ All goes well you should be able to plot your start X and Y against one another to create a circle



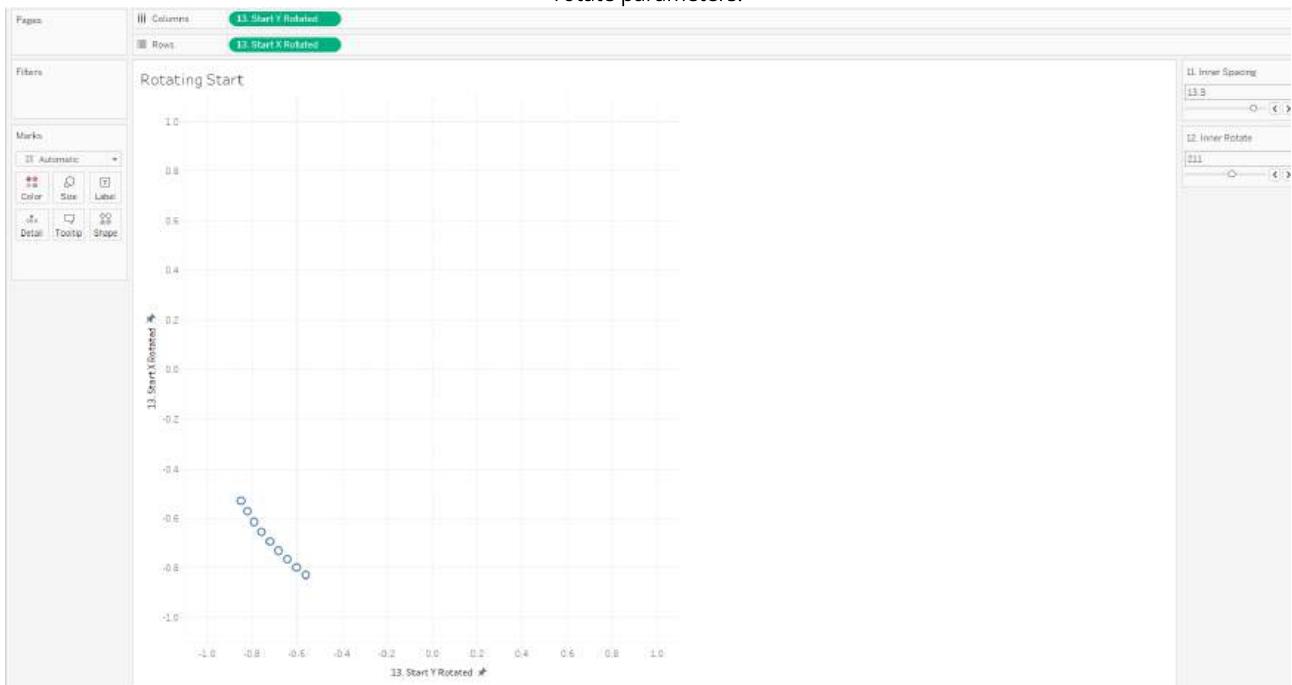
Of course we can amend our Inner Spacing parameter to squeeze our points together. Finally we want to rotate these around our origin 0,0. Lets duplicate our Outer Rotate parameter details but name it **Inner Rotate** Lets apply our final rotation calculations to our start points.

13. Start X Rotated

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

13. Start Y Rotated

$(* \sin(\text{radians}()) + (* \cos(\text{radians}())))$ All goes well, Now you should have your start points radially plotted, as well as your end points. These are both dynamic so we will be able to amend these moving forwards using our inner and outer spacing and rotate parameters.



The final part we need is the chords! Lets create two new calculations for some mid points.

21. Mid X

0

21. Mid Y

0 as well as a calculation for making sure our additional 100 points are to scale.

22. TT

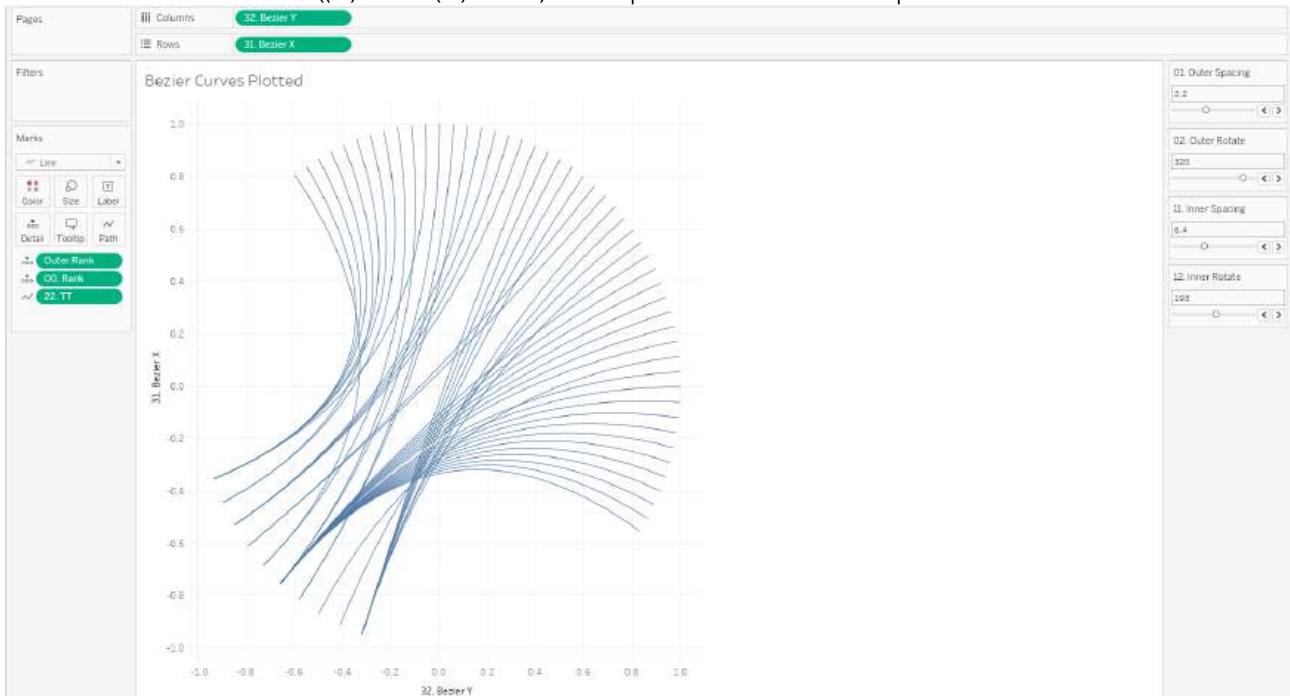
(T-1)/{MAX()-1} Finally we can build the chords,

31. Bezier X

$$((1-)^2 + 2*(1-)^{**} + ^2*)$$

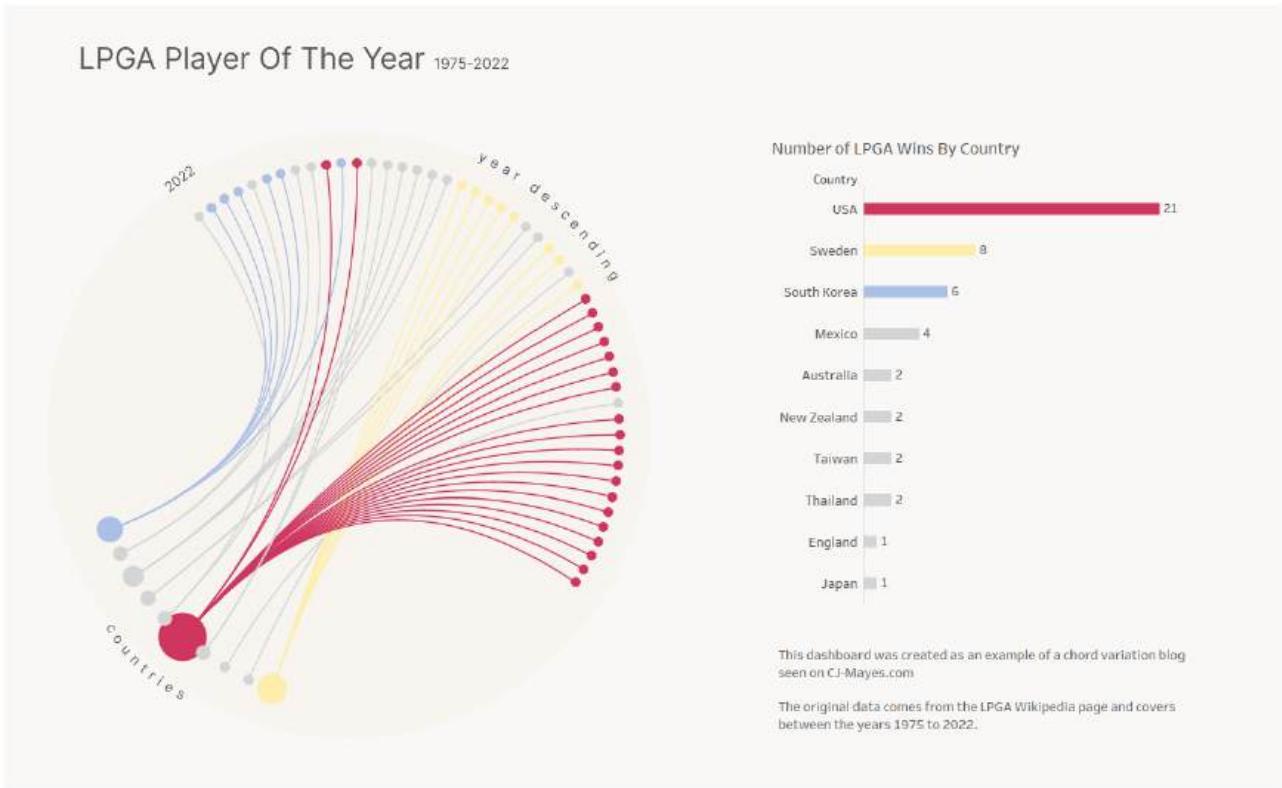
32. Bezier Y

$((1-)^2 + 2*(1-)^{**} + ^2*)$ We can plot these as our final components



By playing around with our spacing parameters, Rotation parameters as well as our midpoints we can now customise our chord to our desired effect. To add a few more details, we can use map layers, to bring in the start and end dots, with different tooltip information. Check out my quick visualisation showcasing the LPGA Player Of The Year data. I made a few minor adjustments to close out the shape of the visual.

- I re-arranged the ranking of my countries as I wanted the chords to bend in specific ways, left and right.
- I did consider moving my mid points but ended up reverting them back to zero.



To recap, we cover off how to make circles, we then understand how to squeeze these points together and move them radially around our origin of 0,0. Finally we build a bezier curve between the start and end points. Some pretty deep calculations putting this one together, let me know how you get on with it. **LOGGING OFF, CJ**

IMAGE ROLES NBA PLAYOFFS

Hi all, Welcome back. It's been about a year since I did a write up on NBA data. You can refresh your memory on how to create NBA shot maps from a previous blog I wrote called '[Python: From Benchwarmer to MVP](#)' Safe to say my code still works. Talk about building robust code huh! Anyway, for todays blog we will take some of that data and look at different ways we can perhaps chart some of the teams using the new image role functionality in Tableau. Truth be told, you don't actually need any of the NBA python data, but two birds one stone with these things. As a small recap: You can access the code for yourself at the top of the blog following the GitHub link. I've updated it to look at the 2021-22 Playoffs.

```
import requests
from nba_api.stats.endpoints import leaguegamefinder
from nba_api.stats.endpoints import playbyplayv2
from nba_api.stats.endpoints import shotchartdetail
import json
import pandas as pd
import time

"""Get the games so we can choose what one we want to look at"""
try:
    result = leaguegamefinder.LeagueGameFinder(
        season_nullable='2021-22',
        season_type_nullable="Playoffs"
    )
    time.sleep(3)
    all_games = result.get_data_frames()[0]
    all_games.to_csv('Games.csv', index=False)
    game_id = all_games.iloc[:, 4]
    game_id_list = game_id.values.tolist()
    game_id_list = list(set(game_id_list))
except requests.exceptions.ReadTimeout:
    print('Timeout error for games')
    time.sleep(3)

"""Print the First Game with header"""
pbp = playbyplayv2.PlayByPlayV2(game_id_list[0])
pbp = pbp.get_data_frames()[0]
pbp.to_csv('PBP.csv', mode='w', index=False, header=True)

game_counter = 1

"""Print the remainder games without header"""
for i in game_id_list[1::]:
    try:
        pbp = playbyplayv2.PlayByPlayV2(game_id_list[game_counter])
        pbp = pbp.get_data_frames()[0]
        pbp.to_csv('PBP.csv', mode='a', index=False, header=False)
        time.sleep(3)
        print(f"{game_counter} has been exported for plays")
        game_counter = game_counter + 1
    except requests.exceptions.ReadTimeout:
        print('Timeout error for plays')
        time.sleep(3)

"""Print the first with a header"""
response = shotchartdetail.ShotChartDetail(
    team_id=0,
    player_id=0,
    game_id_nullable=game_id_list[0],
    context_measure_simple='FGA',
    season_type_all_star='Playoffs'
)

content = json.loads(response.get_json())
results = content['resultSets'][0]
headers = results['headers']
rows = results['rowSet']
df = pd.DataFrame(rows)
df.to_csv('ShotData.csv', mode='w', index=False, header=True)

"""Find the shot details for the chosen game"""
game_counter = 1
```

If you'd like to cross reference the data a little bit, I'd recommend reading about the playoff bracket [here](#). Today's blog will be more about [Image Roles](#) [Draft] – but always nice to tie it back to sports. You can read a little more on [Image try: Roles](#) [here](#). Worth noting, you will need to be on 2022.4 – so get your upgrade on.

Clipboard_20230105T191306 4 rows 2 fields

Show Fields Download

	Clipboard_20230105T191306	Abc Clipboard_20230105T1913...
Image		Team
https://content.sportslogos.net/logos/6/213/full/boston_celtics_logo_primary_19977628.png	Boston Celtics	
https://en.wikipedia.org/wiki/Dallas_Mavericks#/media/File:Dallas_Mavericks_logo.svg	Dallas Mavericks	
https://en.wikipedia.org/wiki/Golden_State_Warriors#/media/File:Golden_State_Warriors_logo.svg	Golden State Warriors	
https://upload.wikimedia.org/wikipedia/en/f/fb/Miami_Heat_logo.svg	Miami Heat	

Above is my dataset. It has been exported for shots. game_counter has been exported for shots. game_counter = game_counter + 1. They are all svgs from wikipedia, the other is a png from a sports logo website.

Pages Columns

Filters Rows Team Name Image

Marks Shape

Colour Size Label

Detail Tooltip Shape

Team Name

Testing Out Image Roles - January 2023

Team Name	Image
Boston Celtics	
Dallas Mavericks	
Golden State Warriors	
Miami Heat	

Okay looks like SVG's don't work yet, that's the first thing to recognise. This also will mean no to GIF's! No problem, we can work with jpgs and pngs for now. Probably time to dig out the documentation to see what we are working with.. Check it out [here](#).



CONTENTS

- Tableau Cloud Help
- › Get Started
- › Administer a Site
- › Manage Data
- › Create and Interact with Views on the Web
 - Web Authoring and Tableau Desktop Comparison
 - Set a Site's Web Authoring Access
 - Create and Edit Private Content in Personal Space
 - › Create a Tableau Data Story (English Only)
 - Add Web Images Dynamically to Worksheets
 - Connect to Published Data Sources

Prepare your data source

Image Role can be assigned to discrete dimension fields that contain URLs that point to web images. To prepare your data, be sure that your image fields meet the requirements set by Tableau to be assigned an image role:

- Make sure your URLs navigate to image files with .png, .jpeg or jpg extensions.
- Verify that each URL begins with http or https. If a transport protocol isn't included, Tableau assumes https.
- Optimise the number of images used in your data set. Usually, you can load up to 500 images per field.
- Ensure that each image file is smaller than 128 KB.

Note: Depending on the complexity of your viz, Tableau may default to a server-side render, which limits the number of images to 30 per field. To

The other thing that sticks out to me here is the image file needs to be smaller than 128KB. Okay that instantly wipes out a bunch of resources online. But the Boston logo comes in at about 53 so fits the requirements. What you will notice if the images don't meet the requirements is the grey small box appear. Fortunately Tableau has listed a whole bunch of guides as to why an image may not render as expected. It includes everything from nothing being rendered,

None of the images are being displayed in my viz

- ⌚ There are too many images in the viz
- ⌚ Tableau couldn't access the images

to some not rendering,

Some of the images aren't displaying in my viz

- ④ The image file is too large
 - ④ The image URL must begin with http or https
 - ④ The image file must be a URL
 - ④ The image file extension is not supported
 - ④ The image file contains bad characters
-

to specific user cases

The images aren't being displayed outside of my worksheet

- ④ The images aren't displaying in Viz in Tooltip
 - ④ The images aren't being displayed in View Thumbnail
 - ④ The images aren't being displayed when I export my workbook
 - ④ The images aren't being displayed on a mobile device
-

So we can either... find images online that fit this criteria. Or... force a way for images to fit it. Well luckily I have some code that will help us resize and format any image saved from the internet.

```
from PIL import Image

foo = Image.open('Original/Boston_Celtics.png') # Large
Image
foo.size # (815, 905)

# downsize the image with an ANTIALIAS filter (gives the
highest quality)
foo = foo.resize((815, 905), Image.ANTIALIAS)

foo.save('Scaled/Boston_Celtics.png', quality=95) # Reduced
Quality
```

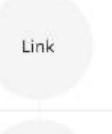
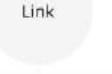
This code will take your file, for example our Boston Celtics logo, and resize it and reduce the quality of the image in order to compress it. Fantastic. But then we are stuck with a saved down file. How are we going to be able to access it for image roles? Well, luckily I have a public GitHub repo. Once I rescale my images and store all my 4 images in the public repo, they each will have their own links. The code has helped resize the images and reduce the quality of the images in order to bring them down to roughly 100kb, give or take. Let's find the URL's on the repo for the four images. Here are the new links

https://raw.githubusercontent.com/CJ-Mayes/Website-Tutorials/main/2023%20Blogs/2023-01-06%20Image%20Roles%20NBA%20PlayOffs/Scaled/Boston_Celtics.png https://raw.githubusercontent.com/CJ-Mayes/Website-Tutorials/main/2023%20Blogs/2023-01-06%20Image%20Roles%20NBA%20PlayOffs/Scaled/Dallas_Mavericks.png https://raw.githubusercontent.com/CJ-Mayes/Website-Tutorials/main/2023%20Blogs/2023-01-06%20Image%20Roles%20NBA%20PlayOffs/Scaled/Golden_State_Warriors.png https://raw.githubusercontent.com/CJ-Mayes/Website-Tutorials/main/2023%20Blogs/2023-01-06%20Image%20Roles%20NBA%20PlayOffs/Scaled/Miami_Heat.png

They Meet the requirements on size, link style and image format. Interesting enough when working in desktop the image url doesn't render amazingly, but the shape file looks okay.

	Boston Celtics		
	Dallas Mavericks		
	Golden State Warriors		
	Miami Heat		

But once published to server, the image quality is far better, and the shape file loses some of its quality.

	Boston Celtics		
	Dallas Mavericks		
	Golden State Warriors		
	Miami Heat		

So to summarise:

- I'm pretty happy with the quality of the rendering.
- The size limit may cause some problems.
- I think it opens doors when storing images in public repositories and drives that are accessible to all!
- Once published image roles have a much higher resolution than shape files.
- I'd like to see more flexibility in scaling the images to proportions and sizes. I think demo's have tended to have the perfect circle, maybe even introducing a masking effect or circular overlay may help keep images tidy? I've seen Tim do some nice work in this space [using Cloudinary](#).
- At the moment shape files have the advantage of being able to positioned anywhere on a chart, I can't seem to find a way to do the same with an image role, other than in a table.

Let me know how you get on with the Image Role functionality, I hope this blog has helped you, access NBA data, use python to amend image sizes and understand the pros and cons of Image roles. **LOGGING OFF, CJ**

INFOSYS TENNIS SHOT MAPPING

Hi all,

Welcome back to a new year, I hope the festive period has been one filled with joy for you.

Those of you who know me well, knows tennis is one of my favourite sports, despite not playing it often. Unfortunately, tennis data isn't as detailed as I really wish it was in terms of being publicly accessible – but it doesn't stop me from trying to make some creative tennis visuals.

For years, I have wanted better access to shot location data, and its seemingly been hard to find. I did however stumble across the infosys stand when I visited Roland Garros last year and so thought to take a bit of a deeper look at the website.

This blog will look to showcase how you can recreate one of the visuals from the site in Tableau and some of the logic behind it.

If you'd like to follow along, you can find the links to all the resources at the top of the page. The dataset and court image will be stored in the repo, and the visual on my tableau public page.

Let's take a look at the final from 2022, [here](#).

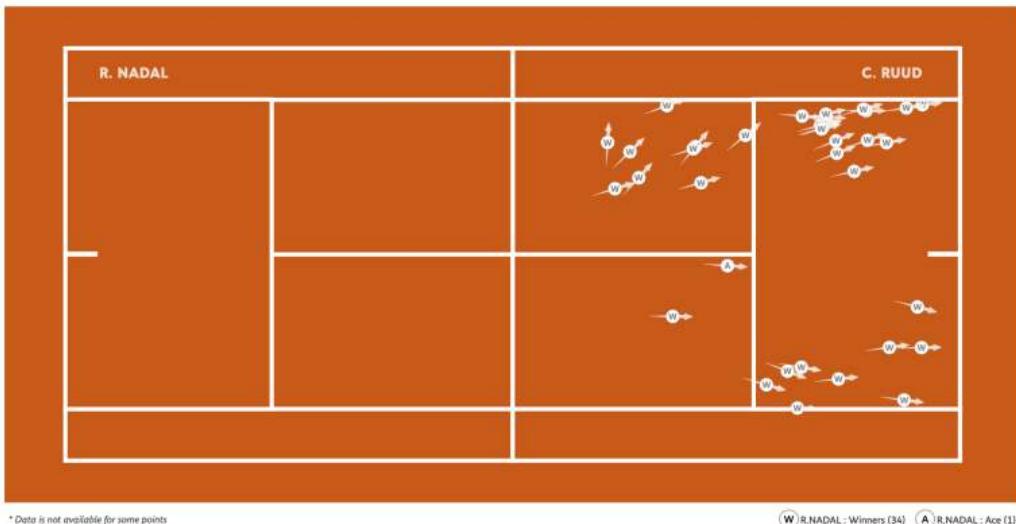


Rafa ended up winning 3 sets to love in the final of the French Open.

The exciting part of that is we can navigate to the court vision page and see what some of these shots look like mapped.



Now right away I'm thinking, yeah this is cool, but it's going to be pretty hard to sketch out in Tableau, so i switch to the 2D mode.



* Data is not available for some points

W R.NADAL: Winners (34) A R.NADAL: Ace (1)

Much better.

Some really smart people are probably able to automate this next bit, but really I wanted just to test if the mapping co-ordinates would work.

```

<div class="rfk-player-name-row"></div>
<div class="rfk-court">
  <div class="rfk-court-svg-wrapper">
    <div class="rfk-p2-court-svg" device-type="desktop" player-name="1" width="709" height="346.4879585831493" preserveAspectRatio="minMid meet" viewBox="-354.5 -173.24397925157464 789 346.4879585831493"> ...
      <def id="</def>">
        <g id="CourtMapper-1">
          <g id="OuterGround"></g>
          <g id="CourtServiceBox"></g>
          <g id="CourtDoublesAlley">
            <use xlink:href="#doublesAlley"/>
            <use xlink:href="#doublesAlley" transform="scale(-1, 1)" />
            <use xlink:href="#doublesAlley" transform="scale(-1, -1)" />
            <use xlink:href="#doublesAlley" transform="scale(1, -1)" />
          </g>
          <g id="CourtINMansLand"></g>
          <g id="CourtCenterMark"></g>
          <g id="CourtPlayersName"></g>
          <g id="plottedBallNotSelected" style="cursor: default;">
            <g class="rfk-court-ball-1_1_5_1">
              <use x="-288.98895361985922" y="-86.34616367824823" xlink:href="#winnersArrow" transform="rotate(193.783972198881)" />
              <use x="-212.98895361985922" y="-86.34616367824823" xlink:href="#winner" />
            </g>
            <g class="rfk-court-ball-1_4_2_1">
              <use x="-225.6434878721749" y="99.58573256798815" xlink:href="#aceArrow" />
            </g>
          </g>
        </g>
      </div>
    </div>
  </div>
</div>

```

div.rfk-court 708.75x349.74

R. Nadal C.Ruud Match Bluff count

Winners (15)

R. NADAL C. RUUD

* Data is not available for some points

W C.RUUD: Winners (15) A C.RUUD: Ace (0)

First I inspect the page and see the ratio between the height and width of the court.

I use a chrome extension in order to save a copy of the court.

You can find the extension [here](#).



SVG Export

[svgexport.io](#) Featured

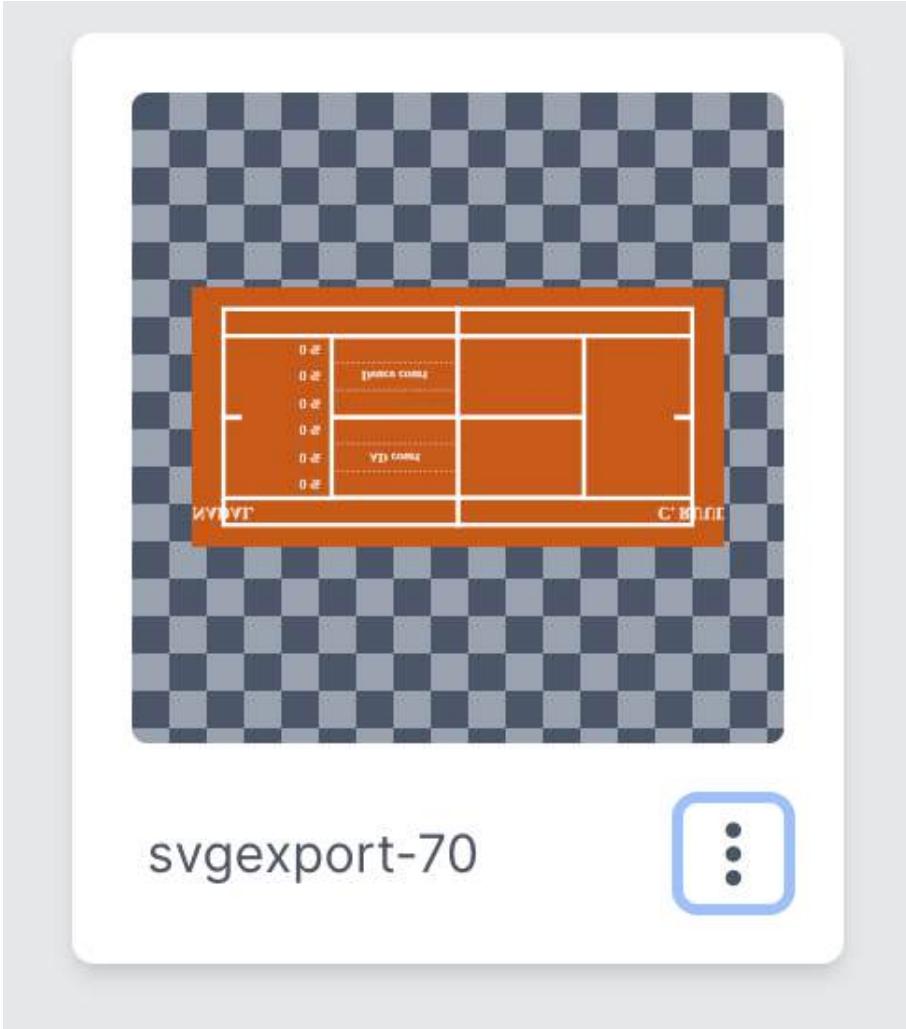
★★★★★ 109 | Productivity | 100,000+ users

I actually really love this extension because you can also start to use it in terms of helping build out other branding aspects, simply by pasting in the website link. The best part is, because they are SVG's you can amend the different layers to each of the items.

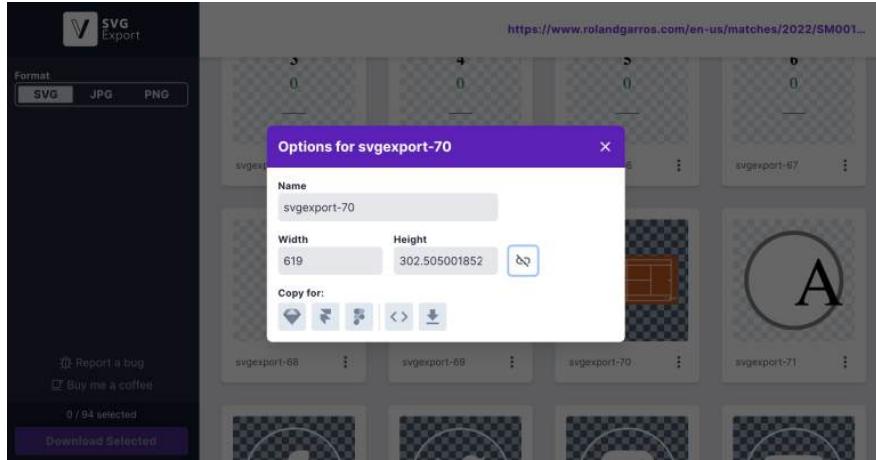


Here are a few examples from the extension in terms of what the type of output files you can get are.

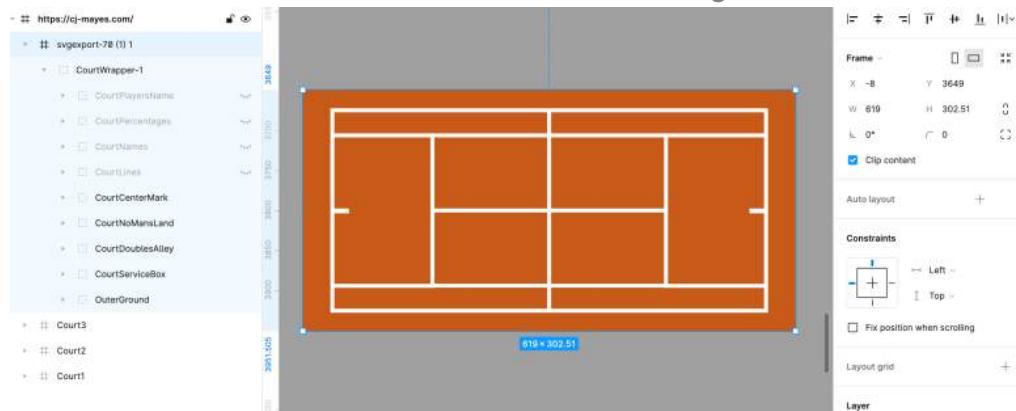
Anyway I digress, the one we are after is the SVG of the court.



Next let's download the SVG, constraining the proportions to make sure we keep the right scale in the court.



We can then move this SVG file into Figma to edit.



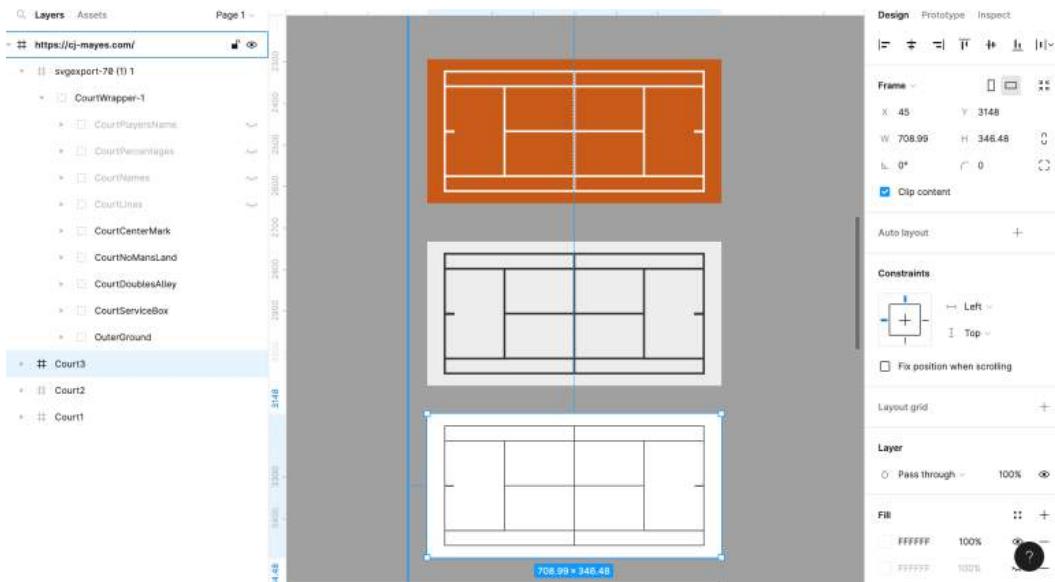
What you will notice is that we exported it at constrained width and height, but we actually want to make these equal to whatever is on the web page.

This is because we want everything to be proportional.

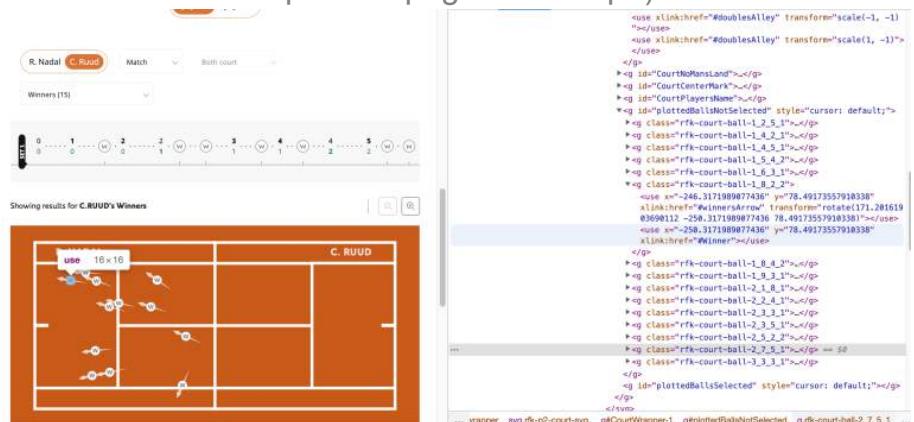
You'll see in the output files I scale this up to 708 * 346, but these values will need to be dependent on what you have set your web browser court to.

Reason being, the x and y values on the court also adjust based on the screen. You can test what I mean by changing the size of your web browser and court.

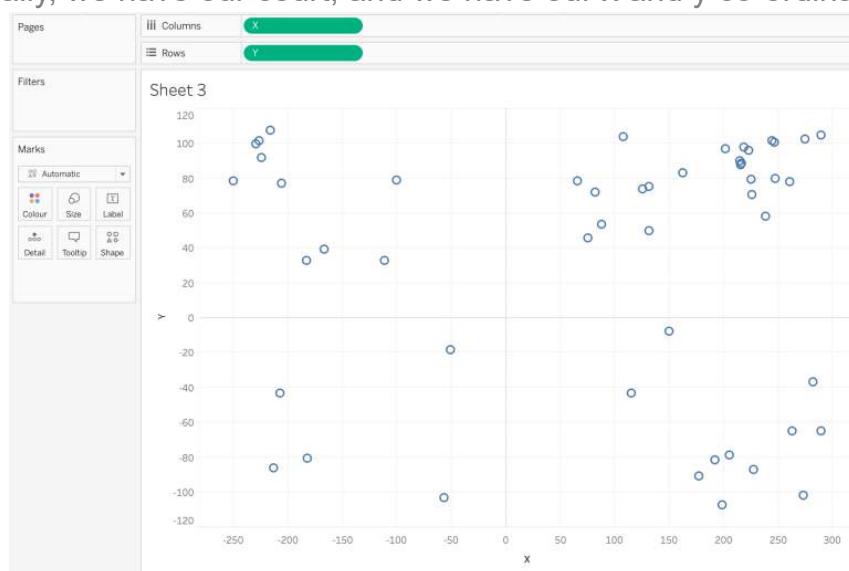
Now in Figma we have rescaled our court. But I also want to play around a little with how the court is designed. I remove the orange background, opting for a light grey and also thin the stroke of the white out to be a thin black court outline. These are simply aesthetic choices, but don't impact the proportions of the court.



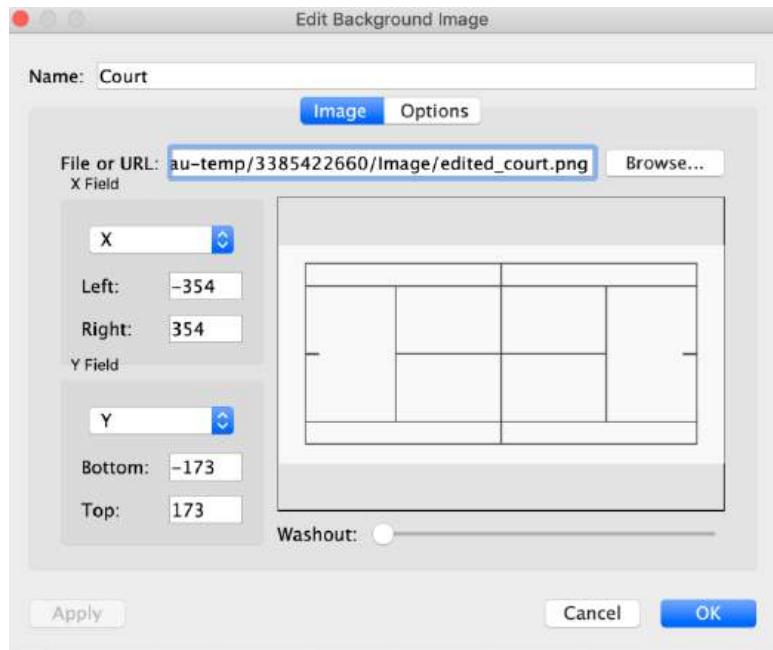
Next we need our data points. Remembering to keep our court the same size, we can then look through the rfk-court-balls within the html elements to find all the x and y co-ordinates of the shots. In my case I took winner shots. (This dataset is attached at the top of the page in the repo)



Finally, we have our court, and we have our x and y co-ordinates.

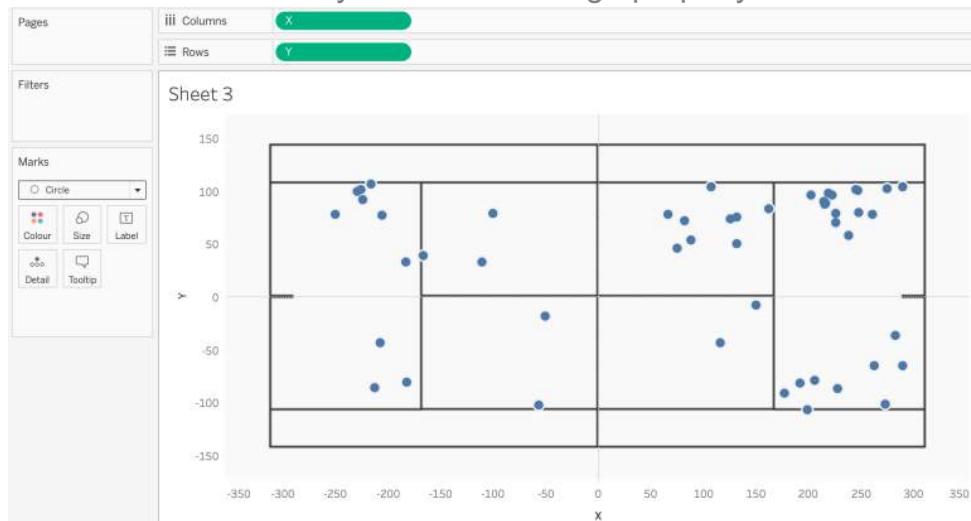


Plot your X on columns and Y onto Rows.
Go to Map > Background Images



Add in your court. Remember at the start when I mentioned my court size was based on $708 * 346$.

I halve these values so that my co-ordinates align properly around the centre 0,0.



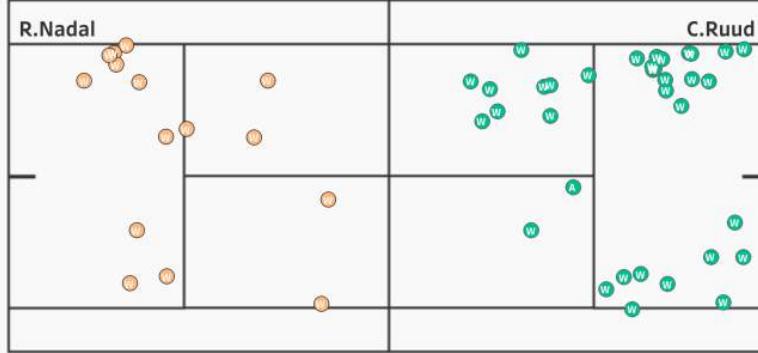
And there we have it, our tennis court with each winner hit mapped in Tableau. Final thing to do is add some colour, spruce it up a bit.

INFOSYS MATCH CENTRE

Roland-Garros | Paris 2022

R.Nadal (5) vs C.Ruud (8)

		Set 1	Set 2	Set 3
R. Nadal	✓	6	5	5
C. Ruud		3	3	0



Rafael Nadal hit **35 winners** (including one ace) in the final, compared to Rudd's 15 winners
 // Data taken from Roland Garros InfoSys Match Centre - Final 2022.
 (W) shows winners, (A) shows Aces.

Going Further:

- Why not see if you can streamline the data collection process?
- Add colour to the chart to highlight each player
- Try mapping a different game from Rolland Garros
- Try mapping a different shot type metric

LOGGING OFF,

CJ