Visualising Large Datasets - Exploration of Valuable Techniques

Raphael Fahey

Introduction

Visualising datasets such as the Fitforge allocation system is crucial in enhancing user understanding and decision-making. Highlighting the relationships between teams and projects is particularly important in this situation and thus a deeper insight on various visualisation techniques, which would provide beneficial for the system, will be challenged and sought after to improve and enhance the User Experience.

Effective data visualisation enables the user to digest complex information, identify patterns and make reasonable and informed decisions. For FitForge, the ideal visualisation method must balance simplicity and clarity/transparency between the system and user, this must be done whilst maintaining the ability to represent nuanced relationships between teams and projects. The section of the report will further investigate and reason potential methods to explore and justify its use in visualising project allocations

The Visualisation Problem - Why Graphs/Charts are Useful

Visually representing data is essential in ensuring clarity and transparency between the user and system. The complexity of team-project pairings can make the raw data difficult to interpret thus, when graphs and charts are effectively used the visualisation can simplify this complexity and allow users to understand a relationship, trend or pattern just with a glance. The basic purpose of visualisation is to essentially create representations of information that can exploit human's perceptual and cognitive capabilities of problem solving. (Khan, M. 2011).

Within the context of the FitForge system, our Industry Partner, David, stressed the importance of visualisation in decision-making. When analysing his method and procedure, visualising teams and projects had proven immensely beneficial, allowing him to create and justify said pairings with a lot more confidence whilst also being engaged. Consequently, choosing the right direction and method is critical for enhancing user understanding, providing flexibility for user interaction and aligning it with Davids ethics and the existing allocation algorithm

Factors to Consider When Designing a Visualisation

1. Purpose and Audience

The primary purpose of visualisation is to clearly define and convey insights and relationships between teams and projects. As the FitForge system will be used and modified by individuals

with varying techniques and expertise, the chosen system should be intuitive and easy for all users to understand. It is essential to balance sophistication with accessibility in order to avoid overwhelming the user.

2. Clarity & Simplicity

The visualisation method must adhere to avoid unnecessary complexity. As mentioned previously, overwhelming users with information that isn't relevant can hinder their ability to interpret the data effectively.A simple structured graph that uses labels, colours and shape meaningfully can help reduce the cognitive load, thus ensuring users can easily identify key relationships between the data points

3. Data Type and Structure:

It is important for users to understand the type of data that is being represented to them for an effective visualisation. Within FitForge, the key variables that must be considered would be the teams, projects and their fit scale. Given that large datasets are common in the system, the visualisation must adhere to handle this complexity whilst also ensuring the relationships between both team and projects remain interpretable.

4. Visual Appeal

Whilst functionality should be considered the primary goal of this project, the visualisation should remain engaging and aesthetically pleasing. Establishing and maintaining a consistency within the design across the entire system can help further reduce confusion and help users navigate the interface with much more ease.

5. Comparability

It is important to facilitate easy comparisons between all data points within this project thus helping users quickly identify which teams are the best for each project. By adhering to the previous points, comparing relationships at a glance will come at ease and will greatly enhance the users experience.

6. Interactivity

Incorporating and exploring interactive elements such as editing data points, filtering etc, can allow users to engage with the visualisation much more deeply. For instance, locking in certain team-project pairings or excluding repeated allocations will give users control over the data analysed thus resulting in improved decision making.

7. Accuracy and Integrity

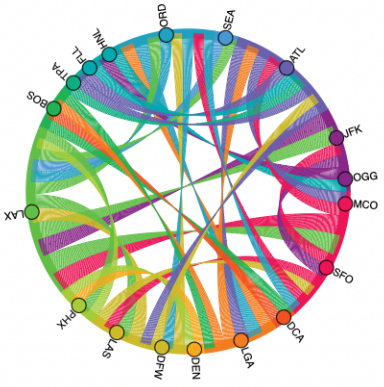
Lastly, it is extremely essential the visualisation accurately represents the data without misleading the user too much. The visualisation must strike a balance between both aesthetic appeal and a truthful representation of the dataset.

Potential Graphing Techniques

1. Chord Diagrams

Represents flow between team and projects allocations, The thickness of the arcs may represent the volume or best fit of the flow

Pros:



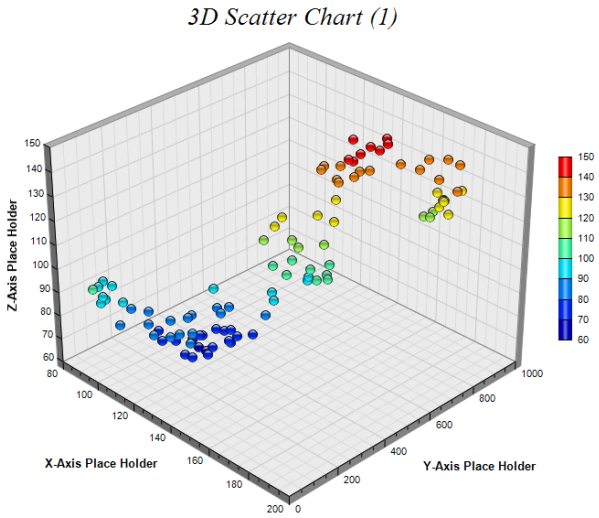
- **Visualises complex relationships**: Excellent for showing the interconnections between different entities, such as teams and projects
- **Aesthetic and engaging**: Chord diagrams look extremely engaging and capture a users attention, especially in complex datasets
- **Summarises flow or interaction**: Great for summarising the flow of resources or interactions between different groups.

Cons:

- **Complexity**: Difficult to interpret, especially for non-expert audiences or when the dataset is large or overly complex.
- **Requires clean data**: Chord diagrams need very well-organised, specific data to be meaningful; messy or unclear datasets can result in a confusing visualisation. Sometimes therefore hindering the data
- **Not scalable**: As the number of nodes or relationships increases, the visualization becomes harder to interpret, losing clarity - could be difficult as projects and teams increase.

2. 3D Scatter Plots

Can involve three dimensions (e.g., team, project, and another variable like priority or completion rate) allowing for a more complex understanding and visualisation in comparison to other techniques. Users could rotate and zoom the plot to explore the relationships between variables.



Pros:

- **Displays multiple dimensions**: 3D scatter plots allow for three variables to be visualized at once, which can show patterns that may not be obvious

in 2D - would mean a third variable is needed to be considered
- **Interactivity**: Interactive 3D plots allow users to rotate and explore the data from different perspectives, which can help in identifying patterns.
- **Great for visualising clusters**: Particularly useful for datasets with multiple clusters or groupings, making it easier to see outliers or patterns such as having multiple teams in a single project

Cons:

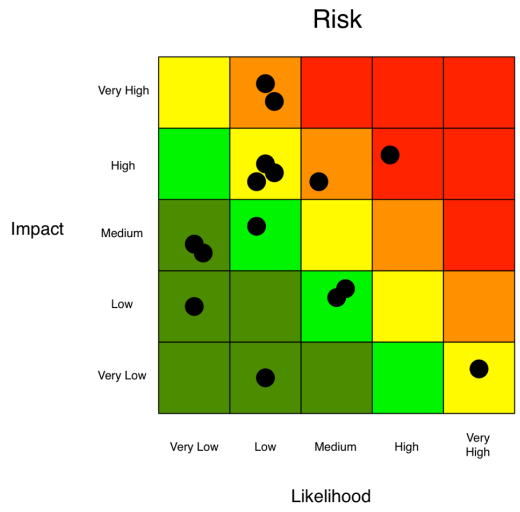
- **Hard to interpret**: Non-interactive 3D scatter plots can be difficult to interpret as data points may overlap, and perspective can distort the view.
- **Overcomplicated for some data**: In many cases, 2D representations provide clearer insights without the complexity of adding a third dimension - Unsure of what the third variable could be at the moment
- **Potentially difficult to code**: Due to the 3D nature and it's immense use of User interactivity, it could be extremely difficult for a user to code representing the data efficiently

3. Heatmap chart

Each cell in the matrix could represent a team's fit for a specific project, with varying colours which intensifies indicating how strong a match is

Pros

- **Transparent**: Easy to understand and provides a immediate overview of team-project relationships
- **Compact visualisation**: They allow you to display large amounts of data



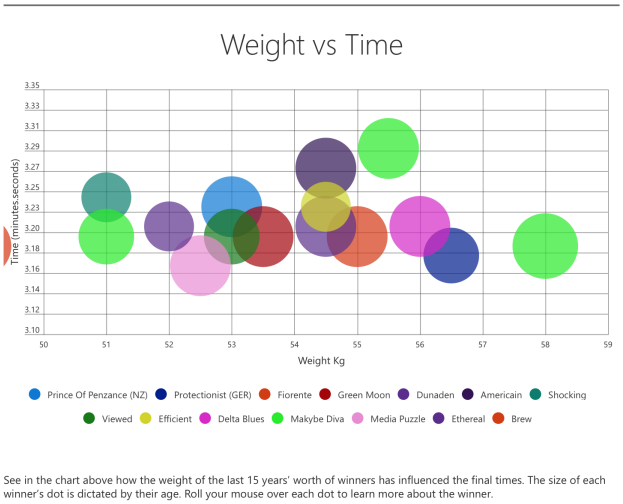
Cons:

- **Visuals:** Colour perception may be an issue as it may be difficult for users to distinguish
- **Lack of information:** Doesn't often show detailed numerical data unless hovered over

4. Bubble Charts

For each team-project pairing, the size of the bubble could represent the strength of the fit. This works well to emphasise how strongly or weakly a team fits a particular project, while keeping the overall chart clean and intuitive.

Pros:

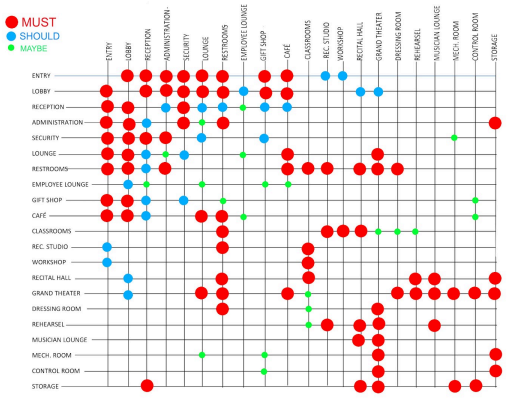


- **Great Visualisation**: Represents the strength of team-project fits in intuitive way
- **Easy Translation:** Can allow for comparisons across multiple dimensions in one chart
- **Aesthetic**: Appealing and engaging look that stands out

Cons:

- **Clutter:** Can get extremely messy and busy with more team and projects reducing clarity
- **Innacurate:** Can potentially be hard to read exact values

5. Matrix Diagrams (Adjacency Matrices)



These provide a clean and structured way to represent relationships between two entities. Each cell in the matrix can represent the fit or connection between a team and project alongside colours that indicate the strength of the fit

Pros:

- **Clear and Compact:** Matrix diagrams are a simple and easy to interpret and allow for quick side-by-side comparisons between all team-project pairings
- **Scalable:** Even with extremely large numbers of teams and projects, a matrix remains legible and structured, thus avoiding extreme clutter like other charts

Cons:

- **Limited Interactivity:** Matrix diagrams lack greatly in regard to dynamic features and may not be visually engaging to an individual

Exploring Bubble Charts in Depth

Through researching bubble charts are well-suited for representing multidimensional data, which is extremely important in this project setting apart and distributing a mass group of teams and projects, thus it's highly recommended to look further into this. Each axis of the chart can represent teams and the other for projects. The size of the bubble would point to/show the fit quality between the team and the project, providing a direct visual comparison. This allows users to test/evaluate the relative suitability of teams for projects at a quick look, which is extremely important when making quick decisions based on setting apart and distributing (wasting very little while working or producing something).

Unlike other visualisation methods like heat maps or bar charts, bubble charts offer an intelligent way to display more than two dimensions of data (Kirk, 2016). By incorporating the fit strength through the size of each bubble, the chart can simplify what would otherwise be a complex and dense dataset into a visually digestible format. This is one of the first reasons bubble charts are

chosen, as they allow for fast pattern recognition and decision-making based on visual hints/signals.

Clarity and Scalability

One of the key considerations for implementing a bubble chart is ensuring that it can represent data clearly, without overcrowding the visual space. Since the project allocation system involves a limited number of teams and projects, the bubble chart can remain legible even with moderate data volumes (Few, 2012). For larger datasets, interactive elements such as filtering or zooming can be implemented to prevent bubbles from overlapping or becoming indistinct (Munzner, 2014). This would assist the user in navigating and making decisions with the data represented

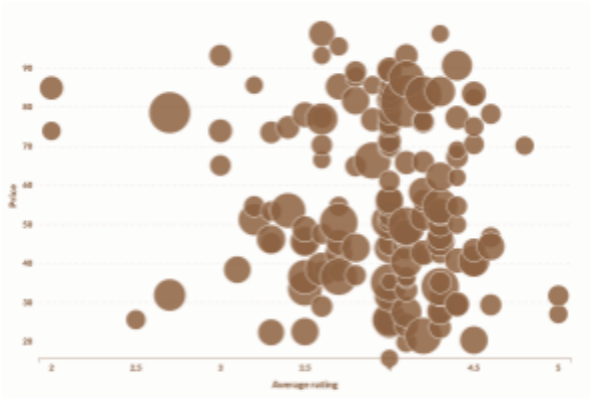


Image 1: Example of poorly executed bubble chart under this context

Another aspect is ensuring that each bubble's size accurately reflects the strength of the fit. A standard approach would involve using a logarithmic scale for bubble sizes to avoid bubbles becoming disproportionately large or small, which could skew the perception (Tufte, 2001). Additionally, it is important to ensure that the bubbles are distinct in terms of colour and size to enhance differentiation, particularly when the fit strengths are close so organising a bubble chart requires mass clarity and understanding to be performed well.

Interactive Features

The interactivity of bubble charts provides added value, allowing users to click on specific bubbles to view more detailed information, such as team strengths or project requirements. For instance, this will offer more granular details on the exact fit score and for the user to make a stronger decision behind locking in a team's placement. This interactivity also helps to manage any limitations caused by overlapping bubbles or data density issues by allowing users to explore subsets of the data dynamically. Another interactivity feature would also be to lock in and filter team placements as mentioned, this will help clear the data as more allocations are made

Managing Complexity and Clutter

One challenge with bubble charts is the potential for clutter when dealing with large datasets. If there are too many team-project combinations, bubbles may overlap, making it difficult to discern individual matches (Few, 2012). However, this can be mitigated by allowing users to filter the data based on

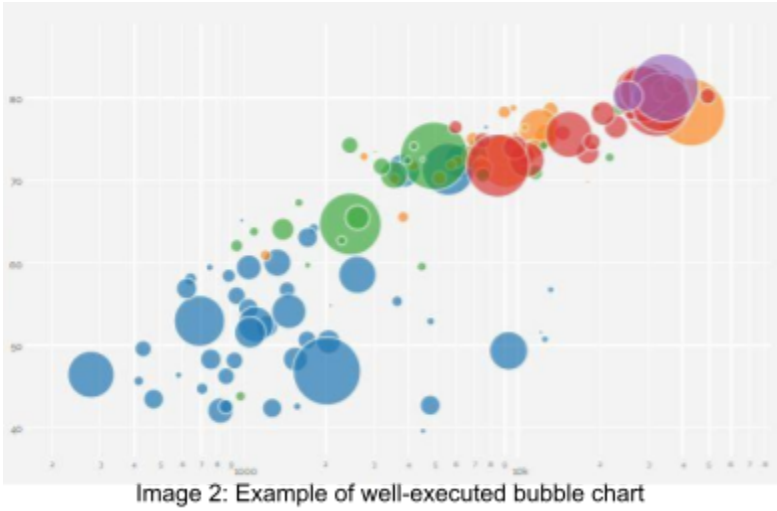


Image 2: Example of well-executed bubble chart

certain criteria as mentioned above, such as only showing the top project matches for each team. Another option is to implement a zoom feature that enables users to focus on specific regions of the chart where more detailed analysis is required.

Recommendations: Justifying Design Elements

When designing visual representations for complex data in general, as mentioned above there are several core principles that must be considered to ensure usefulness to a user. The main goal of the visualisation is to make the data accessible without overwhelming the user while maintaining a balance between aesthetics and practicality. Below are design considerations drawn from leading authorities in data visualisation

1. Focus on Simplicity and Clarity

The data visualisation pioneer, Edward Tufte advocates, “good design is clear thinking made visible, bad design is stupidity made visible” (Tufte, 2001). To simplify complex data is crucial to the project, without of course losing important details. Ensuring key points such as team-project fit visually stand out and are well-distributed across the axes is crucial. Data points should stand out and minor elements should be minimised. This approach allows individuals to grasp patterns and make decisions quickly without a cognitive overload.

2. Consistent Use of Color

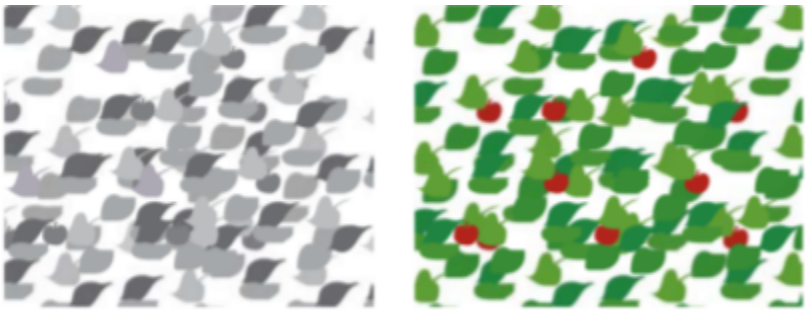


Image 3: Excerpt from *‘Information Visualization: Perception for Design’* where cherries are much easier to visualise with colour

Color plays a significant role in making visualisations intuitive and guiding the users attention, thus it must be used purposefully rather than sporadically. Colin Ware, highlights that colour vision has a critical function as it helps us break camouflage as things differ visually from their surroundings by colour. (Colin, W. 2012). This is important to know as important elements especially within visualisation methods can appear almost incoherent if colours are not expertly used and considered. Rather than using multiple colours that can distract an individual, a consistent colour scheme, such as gradients, ensures a user can intuitively distinguish strong and weak matches in data.

3. Interactive Elements

Modern visualisation techniques should greatly encourage users to explore data via interactive elements. Computer Scientist Ben Shneiderman’s Visual Information Seeking Mantra “Overview first, zoom and filter, then details on demand” supports this idea of interactivity to allow for a greater exploration of data. (Shneiderman, B. 1996). By incorporating features such as a hover

effect to show more detail (such as team details, project importance) or dynamic filters where a user can reduce clutter and gain essential information, it can help enhance the users engagement and ensure a chart can remain concise and clear to the audience.

4. Minimal Use of Axes and Gridlines

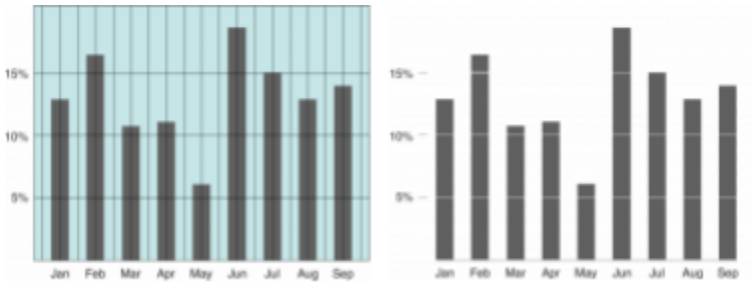


Image 4: Example of high data-ink ratio to low data-ink ratio

It is important to consider axes and gridlines within a chart as too many can overwhelm a user and detract from the core message. The concept of “Data-Ink Ratio”, introduced by Edward Tufte reveals that the non-essential visual elements, data-ink, should be minimised to focus on the important relationships of a chart. This method can be done by minising gridlines, lightening them or placing labels which will ensure the design remains clear and transparent for the audience to interpret the data.

5. Size and Scaling Considerations

Sizing is key to a powerful visualisation method as it is the requirement to represent the data’s magnitude, however it should be used cautiously. Tufte warns against the idea of “chartjunk” which are visual elements that distort the underlying data (Tufte, 2001). Implementing logarithmic or proportional scaling will help balance the design of the visualisation tenfold, this will ensure even subtle differences are noticeable without distorting the representation.

6. Aesthetics vs. Functionality

Few’s work emphasises the importance of functionality over flashy aesthetics (*Few, 2011*). While some charts can include visually appealing design elements, they should almost never compromise the ability to understand the data as that is the main point that is being aimed for. For example, animations and a submenu can be helpful in a visualisation method but it should be used sparingly as to not overwhelm the user or distract them from the data relationships being conveyed. The goal is always to convey the data story as clearly as possible and to achieve the main task at which the project is trying to serve rather than making the user sidetracked. If the chart takes too much time to understand, it won’t be effective for the user.

Incorporating these design principles from various experts ensures the representations that could be produced for the allocation project will remain not only visually appealing but also easy to interpret and functional for the users.

Recommendations: Exploring Advanced Techniques

Innovating and pushing the FitForge system beyond conventional methods is important to create a wider, more expansive view on how the data can be visualised. While the charts and

design features offer a great starting point, there are numerous techniques that could enhance the system's functionality and provide users and developers of the FitForge system a richer experience. The following are advanced techniques that could be integrated with various visualisation formats.

Hybrid Visualisation Models

As is regular for a system to be used by many individuals, providing hybrid visualisations can enhance the system greatly. Combining multiple techniques into a single view can increase a users ability to make more connections across the data that would be difficult for a single visualisation method to do. For example, in resource-heavy/collaborative projects, showing certain dependencies can assist in identifying bottlenecks and optimise team placements across all projects. This can especially be helpful when users require a different view to organise their thoughts when allocating projects to teams.

Multi-Layered Visualisation Dashboards

Similar to the previous method, a multi-layered dashboard allows users to toggle between different visualisation methods. For example:

- A bubble chart can provide a quick overview of team-project fits
- A chord diagram could show team allocations or project overlap
- A radar chart can show team or project detail (fit score, importance, experience, etc)

Using several overlays effectively can provide great flexibility whilst maintaining the benefits of each technique. This method can cater greatly to different user preferences and experience. With more context behind the decision making process it can create more exploration and comparison rather than being locked into one visualisation style.

Conclusion

By integrating the above innovative visualisation methods and techniques, individuals using the FitForge system will be empowered to explore the data more comprehensively as it provides many views and directions for allocation. Moving beyond methods such as bubble charts, techniques such as hybrid visualisation models, multi-layered dashboards and _____ can offer a wider field of knowledge and engagement from the user with the theme-project allocation data. These approaches would not only improve decision-making but also inspire future developers of the FitForge system to interact and develop new dynamic ways, thus improving the system over the years.

References

Angori, L., Didimo, W., Montecchiani, F., Pagliuca, D., & Tappini, A. (2019, January 1).

ChordLink: A new hybrid visualization model. Springer International Publishing.

https://link.springer.com/chapter/10.1007/978-3-030-35802-0_22

Aparicio, M. (2015). *Data visualization* (C. J. Costa, Ed.). ACM Digital Library. https://dl.acm.org/doi/abs/10.1145/2721882.2721883

Das, R., & Soylu, M. (2023). A key review on graph data science: The power of graphs in scientific studies. *Chemometrics and Intelligent Laboratory Systems*, *240*, 104896. https://doi.org/10.1016/j.chemolab.2023.104896

Elmqvist, N. (2008). *Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation* (P. Dragicevic & J.-D. Fekete, Eds.). IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/4658123

Finnegan, A., Sao, S. S., & Huchko, M. J. (2019). Using a chord diagram to visualize dynamics in contraceptive use: Bringing data into practice. *Global Health: Science and Practice*, *7*(4), 598–605. https://doi.org/10.9745/GHSP-D-19-00205

Fitzgerald, B. (2016, February 26). *Schneiderman's mantra*. Data Visualization. https://hampdatavisualization.wordpress.com/2016/02/26/schneidermans-mantra/

Gu, Z. (2022). *Complex heatmap visualization*. iMeta. https://onlinelibrary.wiley.com/doi/full/10.1002/imt2.43

Islam, M. (2019). *An overview of data visualization* (S. Jin, Ed.). IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9012031

ML+. (2012). *MATRIX AND BUBBLE DIAGRAM*. 2012. https://arch3611fall12ml.blogspot.com/2012/09/matrix-and-bubble-diagram.html

Piringer, H. (2004). *Interactive focus+context visualization with linked 2D/3D scatterplots* (R. Kosara & H. Hauser, Eds.). IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/1319526

Sadiku, M. N. O. (2016). *DATA VISUALIZATION*. https://www.researchgate.net/profile/Adebowale-Shadare/publication/311597028_DATA_VISUALIZATION/links/5851945608aef7d0309f20a7/DATA-VISUALIZATION.pdf

Shi, J. (2024). *High-Precision data recovery for bubble charts*. IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10503739

Ware, C. (2020). *Information visualization*. Morgan Kaufmann. https://books.google.com.au/books?id=3-HFDwAAQBAJ&pg=PA95&source=gbs_toc_r&cad=2#v=onepage&q&f=false

Wu, M., Wang, Y., Yan, W., Gou, J., Liu, B., Shi, Y., Zhao, H., Gide, E., Cui, X., Wang, P., Wang, Y., Fan, J., & Li, Q. (2021). *Bubble-Wall plot: A new tool for data visualization*. Figshare. https://acquire.cqu.edu.au/articles/conference_contribution/Bubble-Wall_Plot_A_New_Tool_for_Data_Visualization/24531658?file=43091356