

**DEPARTMENT OF COMPUTER SCIENCE**

**CS5803 Data Visualisation**

***Assignment Topic: Visualisation Design Task***

Student ID: 2375008

**Module Leader: Dr Timothy Cribbin**

April 2024

(2023-2024)

Contents

[1. Introduction 3](#_Toc164070243)

[1.1. Data 3](#_Toc164070244)

[1.2. Persona and questions 3](#_Toc164070245)

[1.3. Requirements 3](#_Toc164070246)

[2. Design 4](#_Toc164070247)

[2.1. Paper Landscape 4](#_Toc164070248)

[2.2. Final Implementation (Tableau) 5](#_Toc164070249)

[2.3. Discussion 5](#_Toc164070250)

[3. Implementation 6](#_Toc164070251)

[3.1. Tableau 6](#_Toc164070252)

[3.2. PowerBI 10](#_Toc164070253)

[4. Walkthrough 11](#_Toc164070254)

[5. Discussion 14](#_Toc164070255)

[5.1. Software comparisons 14](#_Toc164070256)

[5.2. Learning reflections 14](#_Toc164070257)

[6. Conclusions 15](#_Toc164070258)

[7. References 15](#_Toc164070259)

# Introduction

## Data

This is a dataset of hurricanes and severe storms occurring between 1975-2021 across North America. It was retrieved from the National Oceanic and Atmospheric Administration (NOAA) for the US and shared via Kaggle (NOAA, 2023). The dataset contains ~20k rows. For each row, 11 columns are provided as described below (a further 2 excluded as not used):

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Domain** |
| Name | Name given to storm | Nominal |
| Year | Year of measurement | Integer |
| Month | Month of measurement | Integer |
| Day | Day of measurement | Integer |
| Hour | Hour of measurement | Integer |
| Lat | Latitude for centre of storm | Float |
| Long | Longitude for centre of storm | Float |
| Status | Type of storm (e.g. hurricane or tropical depression) | Nominal |
| Category | Class of hurricane strength from 0 (min) to 5 (max) | Integer or NA |
| Wind | Storm's maximum sustained wind speed (in knots) | Integer |
| Hurricane\_force\_diameter | Diameter (NM) of the area experiencing tropical storm strength winds (>34 knots). Only available starting in 2004. | Integer |

## Persona and questions

The user is a meteorologist interested in understanding the patterns of major storms. This enables them to better forecast strengths and trajectories of hurricanes, estimate damaged caused to cities and design preventative or mitigation strategies. They have 4 research questions, 3 simple (Q1-Q3) and 1 complex (Q4):

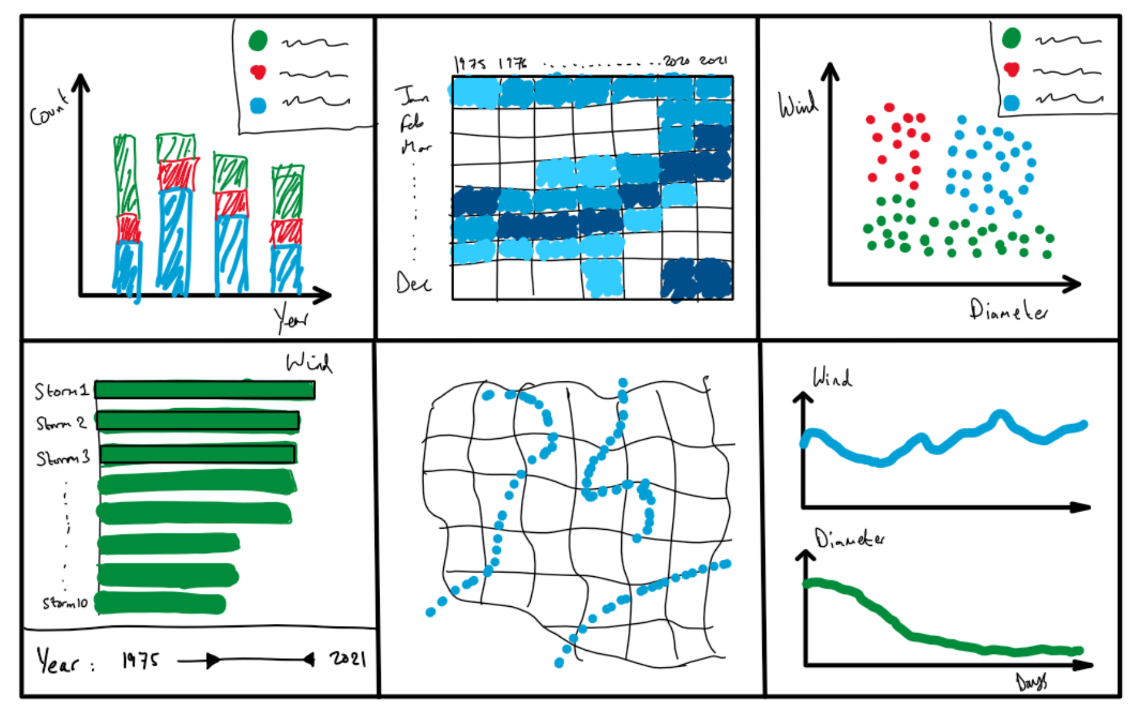
* + 1. How has the frequency of different storm types evolved over the years?
    2. How does the occurrence of storms vary by season? Has this changed over time?
    3. Considering diameter and wind speed, do hurricanes cluster into groups with similar damage potential?
    4. For the most severe hurricanes, how have the trajectories and intensities evolved over the course of their development?

## Requirements

1. To answer Q1, the user needs to visualise the change in count of storms over time. This time-series comparison can be done with an area chart. The x-axis would represent years (grouped for clarity) and the y axis a count of storms, segmented by storm type. This gives an overall trend with the top line, and growth of layers indicating which storm types are driving change. Each storm type would have different colours to easily distinguish them.
2. To answer Q2, we must compare both the year and month of storms. This lends well to a heatmap with years on the x-axis and months on the y-axis. Each cell indicates count of storms as both a label and colour gradient. This enables the user to see the seasonality via bright or dark patterns, which may also change over the years.
3. To answer Q3, the requirement is to understand the relationship between wind speed and storm diameter. A good view would be a scatterplot, with one axis representing “wind” and the other “hurricane\_force\_diameter”. Points grouped together may suggest the presence of clusters of similarly behaving hurricane types. Given a focus on hurricanes, “status” would need to be filtered for hurricanes only.
4. For the complex question Q4, the meteorologist will need to identify and explore each of the most severe storms in depth. This requires multiple interacting plots. Firstly, a bar chart might rank storms in the dataset (subject to filtering by year) based on maximum wind speed achieved. These can then be investigated by brushing (clicking a bar) and filtering 2 further charts for just the selected storm. A map view displaying bubbles for each latitude and longitude recorded can very effectively show the path of the storm. A second line chart can provide more granular time-series data on the wind speed and diameter by day to explore their relationship and evolution over time rather than space. By interactively selecting various storms, the meteorologist can identify behaviour patterns and understand regions most at risk of catastrophic damage.
5. Fit for scientific use, the dashboard must employ a consistent and detail-oriented approach to charts - clear labelling of axes, legends and titles, particularly stating measurement units for metrics such as wind speed to avoid confusion. Annotations where needed would point out nuanced features and trends to aid with quick comprehension.
6. Considering the user may present findings from the office or in the field, typical for weather researchers, the dashboard should be high contrast and easily visible with the use of bold colours. Additionally, charts encoding different metrics or segmentations should utilise distinct colour schemes to avoid confusion about inter-relationships.

# Design

## Paper Landscape



[Q1/R1] See trend in number of major storms each year, by “status”.

[Q2] Heatmap shows count of all storm types by month/year. Look for directional patterns [R2]

[Q3] Scatter plot of wind by diameter for hurricanes. Clearly demonstrating clustering with colour legend and labels [R3]

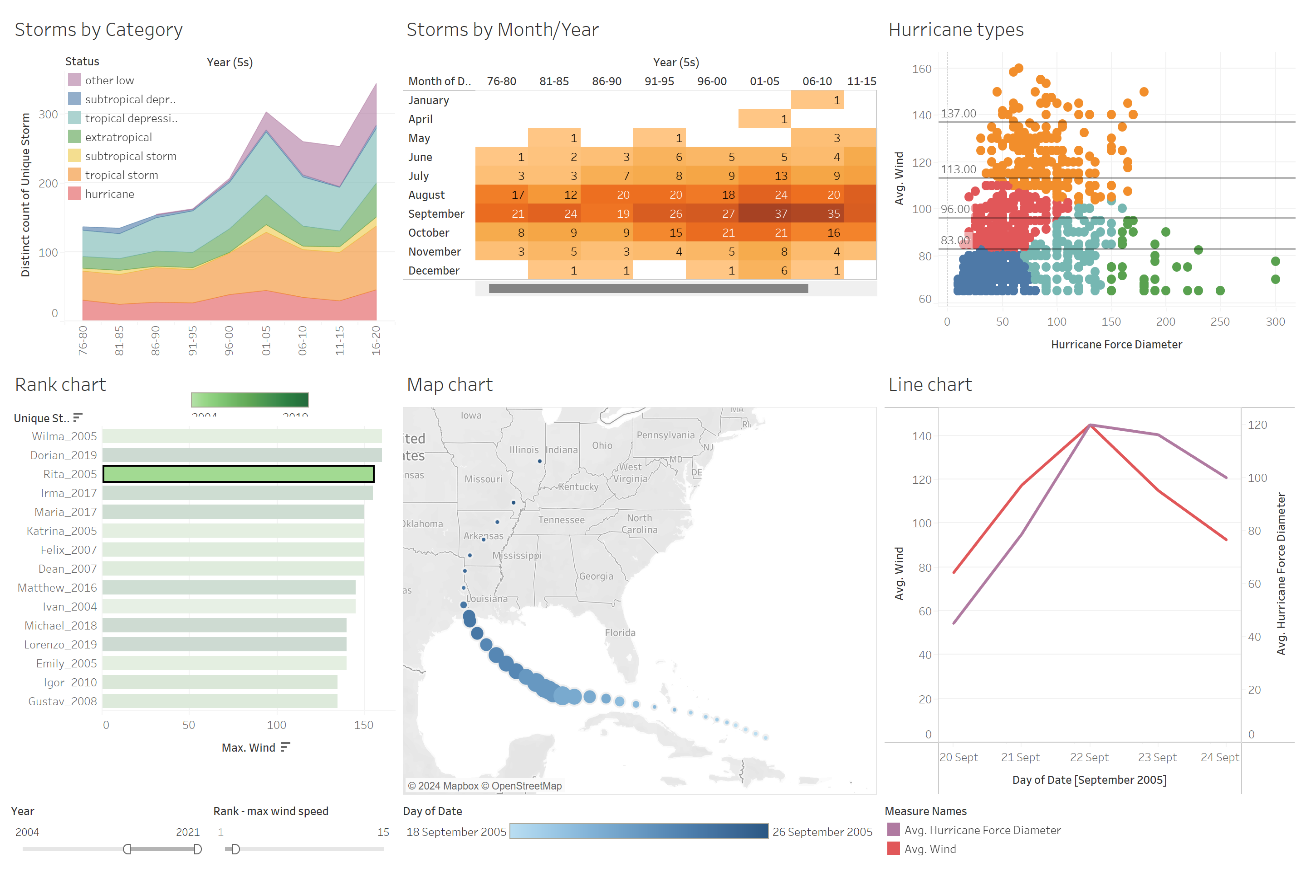
[R5] Clear axes labels with units will be used to aid technical comprehension

[R6] Use of bold colours to aid visibility

[Q4] Bottom 3 charts used in combination to understand top storm behaviours. Filter by year range and brush for selected top storms of interest on left. Then view trajectories (middle) and evolution of wind/diameter (right) to understand relationships between location and intensity for various storms [R4]

|  |
| --- |
| **Historical Hurricane Patterns** |
| **Goal: To analyse the historical patterns, seasonal variation, clustering based on damage potential, and development progression of severe hurricanes – in order to address Q1-Q4.** |
| Filters for category / status allows for drilling down and comparison, synchronised between all views |
| **Result**: The researcher gains a strong understanding of how major storms have varied over time. They can brush the Q4 charts together to investigate storm behaviours, enabling insights for better storm categorisation and prediction of potential damage to people and infrastructure. |

## Final Implementation (Tableau)



[Q1] Chart type changed to area chart to better visualise growth of segments. Years grouped into 5-year intervals to smooth out annual variations

[Q2] Bright orange colour scheme used to aid contrast and visibility. 5Y groupings for clarity

[Q3] Scatter plot and color-coding used to clearly differentiate similar groups of storms. Legend removed. Horizontal lines added for category bounds

[Q4] Click-able rank chart of top storms used to identify storms of interest for brushing. Year filter and additional rank filter applied to control number and age of storms shown (with age indicated by colour gradient for clearer visibility)

[Q4] Chloropleth map used to show the trajectory of selected storm. Bubble size and colour encoding introduced to show category and storm age (in days) at a glance.

[Q4] Wind speed and diameter lines combined with dual axes for better comparison and use of space.

## Discussion

For Q1, ordering of storm segments by severity (average wind speed) was added. This improved visual clarity and adhered with best practices for ordering categorical data with numerical metrics to aid pattern discovery (Smart, et al., 2019). In addition, using a rainbow colour gradient was further aligned with HCI principles advocating the use of intuitive colour schemes to categorise data, reducing cognitive strain and improving overall usability (Tory & Moller, 2004).

Tackling Q3, horizontal lines were added to the scatterplot to delineate standard hurricane category boundaries, addressing verbal feedback around improving the interpretability of clusters. This was supported by the principle of adding reference points which enhance user readability (Lidwell, et al., 2010), allowing more informed assessments of storm intensity and size.

Introducing k-means clustering produced insights beyond basic windspeed categorisations. Combining windspeed and diameter yielded a better depiction of potential damage, ultimately enriching the narrative and revealing outliers which challenge typical classifications via a granularity that windspeed alone fails to capture.

With Q4, the addition of a rank count filter in the rank chart boosted the meteorologist’s ability to control the number and age of storms displayed. The use of bars to denote rank here aligns well with visualisation theory that suggests bar charts are the most effective way to represent comparisons (Few & Principal, 2004) (Spence, 2001). This choice also stemmed from the UCD concept that direct manipulation interfaces promote deeper engagement and empower by providing a sense of control (Dirin, et al., 2023).

Using feedback received, the map visualisation was refined in favour of a single storm path with bubble sizes and a colour gradient to illustrate the magnitude and evolution of storm intensity at a glance. Circular dimensions representing intensity is rooted in the idea of size as a pre-attentive attribute that quickly conveys quantitative differences to the viewer, also a key part of Bertin’s image theory as well as others in the field (Bertin, 2000) (Ebert, et al., 2000) (Green, 1998).

# Implementation

## Tableau

A screenshot of a graph

Description automatically generated**Storms by Category (sheet 1)**

First, 2 additional fields had to be created as there was no native way to categorise unique storms or group years into 5-year intervals:

1. **Unique Storm** – created by concatenating ‘name’ and ‘year’ fields in a calculated field using the formula: *[Name]+"\_"+STR([Year])*
2. **Year (5s)** – created by using the create > group option on ‘year’ and assigning groups

Filters for ‘status’ were added to exclude the very small disturbance and tropical wave categories (for visual clarity) and for Year (5s) to exclude the outer years 1975 and 2021. Status was also sorted by field (wind) > ascending > average aggregation to order roughly by severity.

A screenshot of a calendar

Description automatically generated**Storms by month/year (sheet 2)**

Before making this heatmap, year, month and day fields had to be combined into a date field using a new calculated field *‘Date’: MAKEDATE([Year],[Month],[Day])*.

A simple table was then created with marks as above. Year (5s) also had to be sorted manually to ensure the correct ordering of year groups.

**Hurricane types (sheet 3)**

A screenshot of a graph

Description automatically generated

A scatter plot was filtered for non-zero diameters to filter for only hurricanes. Both unique storm and day were dragged onto the detail marks card to represent every daily record. Finally the ‘cluster’ option was dragged from the analytics tab onto the chart. Both avg(wind) and hurricane force diameter were selected as clustering fields, and 5 target clusters selected. Additional constant lines were added to signify hurricane category thresholds for windspeed from the dataset metadata.

A screenshot of a graph

Description automatically generated**Rank chart (sheet 4)**

First a calculated field had to be created to rank unique storms by maximum windspeed using: RANK\_UNIQUE(MAX([Wind])). This was then added as an at-most filter on the right-hand side, alongside a filter for year. The rank chart was created using Max(wind) as columns and unique storm as rows (sorted descending by maximum wind), along with a year gradient.

A screenshot of a map

Description automatically generated**Map chart (sheet 5)**

A choropleth chart was created using longitude and latitude, with date (colour), category (size) and unique storm (detail) as marks. Note the regular day field would be incorrect, stepping down to 1 when crossing into a new month, unlike the datetime field. The action filter was applied later via the dashboard to link charts.

A screenshot of a graph

Description automatically generated**Line chart (sheet 6)**

A simple line chart was created here. Dual axis was selected for the second one to place them both on the same chart.

**Dashboard**

A screenshot of a computer

Description automatically generated

A dashboard was created by dragging in the 6 selected sheets and arranging the associated filters and legends as shown. One encountered issue was to only allow interaction between the bottom 3 charts for Q4 - a filter was added under dashboard menu > actions. The source sheet was selected as rank chart and target sheet selected as the relevant Q4 sheets. Keep filtered values only was activated on single selection.

## PowerBI

A porting of this design was also successfully created in PowerBI, a popular alternative to Tableau, with some minor differences.

[Q3] Scatter plot of wind speed against hurricane force diameter using PowerBI clustering functionality and constant lines, exactly as per Tableau version

[Q2] Heatmap of unique storms by month/year, identical in function to Tableau

[Q1] Area chart of unique storm by year, very similar to Tableau except for segment ordering (below).

A screenshot of a data presentation

Description automatically generated

[Q4] Interactive rank-chart for top storms with colour gradient for date. Note year/rank filters are applied but situated on the right-hand filters pane instead on PowerBI.

[Q4] Wind speed and diameter dual-axis line chart exactly as per Tableau implementation

[Q4] Linked chrolopleth map of the selected storm, almost identical to Tableau (except for visual design).

The PowerBI implementation differed as follows:

1. The calculated fields required to appropriately aggregate and transform the given variables required much more complex formulae, e.g. Unique Storm = data[name] & "\_" & FORMAT(data[year], "General Number"). This is due to PowerBI’s design more catered towards power users and complex query functionality.
2. Sorting was more complex in PowerBI, requiring a new custom field to manually order a field, meaning that status sorting in Q1 was not applied in this case.
3. Unlike Tableau, PowerBI doesn’t plot small circles in a chloropleth map for null or zero-valued size marks. Hence during initial data input and transformation, any null values were replaced by a float value of 0.2 to achieve the same visual effect.

# Walkthrough

A graph of different colored lines with Crust in the background

Description automatically generated with medium confidenceTo answer Q1, looking at the area chart - the below stacked area chart reveals a clear upward trend in the total number of storms across the last 5 decades, driven by a rise in tropical storms, hurricanes and other low speed storms

For Q2, observing patterns in the below heatmap shows a distinct seasonality with peaks of storm occurrence in early Autumn (Aug-Oct) and almost none in winter and spring. The gradual upward trend overall is observed again here, but seasonality has remained consistent.

A screenshot of a computer

Description automatically generated

For Q3, the scatter plot categorises storms based on windspeed and diameter. Category 4/5 storms present a distinct high-speed profile, whereas lower-category storms form three groups of differing diameters. This nuanced view on storm classification helps our understanding on the varied impact potential of storms.

A screen shot of a graph

Description automatically generated

Next for Q4 considering 3 charts – rank, map and line, we can obtain a comprehensive analysis of the most severe storms. For example, selecting ‘Wilma\_2005’ from the rank chart, the corresponding storm trajectory and wind speed over time (several days) are highlighted. This allows us to explore any storm for its characteristics and affected regions.

A screenshot of a computer

Description automatically generated

The bar chart is used first to identify the most severe storms by maximum wind speed, each of which can be selected. It also offers insights into intensity changes across different storms over time.

A screenshot of a graph

Description automatically generated

With a selected storm e.g. Wilma, the plotted points on the map illustrate the storm's trajectory and intensity change over time. Darker to lighter shades show progression, with bubble size indicating landfall impacts, helping to model behaviour.

A map of the united states

Description automatically generated

Lastly, the line chart for the selected storm shows interactions between wind and diameter evolution, helping to understand and anticipate how total damage potential varies and when it peaks.

A graph on a screen

Description automatically generated

# Discussion

## Software comparisons

The final dashboard was an overall success, allowing the user to obtain both high-level and deep insights through easy interactions. However, considering the profile of a typical user (scientist/meteorologist), there is a slight risk of overcomplication if the dashboard is presented to a non-technical audience. This was largely mitigated through well thought out design in-line with established theories/methodologies such as the use of high-contrast colours and minimalist graphs to prioritise the most relevant information for the viewer.

Furthermore, the successful porting of the dashboard to another popular software, PowerBI, highlights its adaptability and readiness to accommodate the users needs. Whilst Tableau offered less complex formulae and a simpler GUI in addition to more logical data-sorting techniques, PowerBI excelled at creating single page dashboards with universal filtering and a streamlined design process using minimal clicks between pages and elements. Overall, it became clear that Tableau was positioned towards systematic, foolproof and straightforward visualisations, whilst PowerBI was more suited to power users needing advanced functionality and rapid prototyping capabilities.

## Learning reflections

This project acted as an ideal medium for concentrated learning and application of data analysis principles/techniques – ultimately providing a valuable experience. Examples of such learnings were finding an effective data-to-ink ratio that improves visual clarity in a complex dashboard, in addition to using bold, distinct colours to avoid confusion and allow for easy data segmentation. Furthermore, the discovery of brushing as a powerful tool for detailed analysis showed how effective Tableau stories were in communicating findings in both a structured and presentable way. Observing similarities with other packages such as PowerBI provided a stronger foundational understanding of core data analysis techniques, yet there was still a healthy learning curve and appreciation for advanced methods when exploring the feature sets of this software in comparison to Tableau. One such example is integrating k-means clustering directly within visualisations to provide a new dimension of interpretation.

Looking ahead, I aim to use more statistical and machine learning techniques to achieve stronger and more effective insights, showcased better by learning advanced charting or design principles from literature. These can then be published online for users to benefit and provide feedback upon, continuing the cycle of improvement from this project.

# Conclusions

**Objectives**: The core aim was to explore data visualisation techniques and their application using popular software packages and design principles from literature, to assist a weather researcher in effectively making use of historical hurricane data. Research outcomes were an understanding of hurricane patterns, discovering seasonal trends and the damage potential/development of a hurricane over its lifespan.

**Approach**: Initial data selection and a user persona led the formation of key research questions to be answered by the subsequent data analysis. From there, brainstorming and planning using design principles such as paper landscapes allowed for rapid prototyping and feedback for various dashboard designs. After a preliminary data cleanup/import the chosen dashboard was then built using two industry standard data visualisation packages. This was then coupled with a story-like presentation which communicated the overall research and findings for a typical user and an audience of similar peers.

**Main Outcomes**: In the process of successfully answering the research questions via two well-functioning dashboards and a Tableau story, there was significant learning and application of foundational data visualisation techniques alongside building proficiency with industry standard tools. This was enhanced with considerable efforts to implement advanced methods such as cluster analysis – ultimately defining a clear learning plan for improving the effectiveness of data analysis in any future projects.

# References

Bertin, J., 2000. Matrix theory ofgraphics. *Information Design Journal,* 10(1), pp. 5-19.

Dirin, A., Nieminen, M. & Laine, T. H., 2023. Feelings of being for mobile user experience design. *International Journal of Human–Computer Interaction,* 39(20), pp. 4059-4079.

Ebert, D. S. et al., 2000. Procedural shape generation for multi-dimensional data visualization. *Computers & Graphics - Elsevier,* 24(3), pp. 375-384.

Few, S. & Principal, P. E., 2004. *Designing Tables and Graphs to Enlighten.* 1st ed. California: Springer.

Green, M., 1998. Toward a perceptual science of multidimensional data visualization: Bertin and beyond. *ERGO/GERO Human Factors Science,* Volume 8, pp. 2-27.

Lidwell, W., Holden, K. & Butler, J., 2010. *Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design.* 1st ed. Houston: Rockport Pub.

NOAA, 2023. *NOAA Atlantic Hurricane dataset - Kaggle.* [Online]   
Available at: https://www.kaggle.com/datasets/utkarshx27/noaa-atlantic-hurricane-database  
[Accessed 15 Apr 2023].

Smart, S., Wu, K. & Szafir, D. A., 2019. Color crafting: Automating the construction of designer quality color ramps. *IEEE transactions on visualization and computer graphics,* 26(1), pp. 1215-1225.

Spence, R., 2001. *Information visualization.* 1st ed. Switzerland: Springer.

Tory, M. & Moller, T., 2004. Human factors in visualization research. *IEEE transactions on visualization and computer graphics,* 10(1), pp. 72-84.