FIT5147 – Data Exploration and Visualisation

Data Visualisation Project

An Investigation into Data Science Remuneration Trends in 2023

Student Name: Tien Long Bui
Student ID: 33535779
Tutor: Megan Power, Bruno Mendivez
Applied session: 05

MONASH University

Faculty of Information Technology
Monash University
Australia
October 2023

Table of Contents

1. *Introduction* 3
2. *Design* 4
3. *Implementation* 6
4. *User guide* 7
5. *Conclusion* 7
6. *References* 8
7. *Appendix* 10

1. Introduction

The world of data science, with its emphasis on "science," has always attracted those who love working with data. Through my personal lens, people often say this profession is all about hard work and a dull routine of staring at screens for hours. To prevent falling into the trap of misconceptions in this field, my project proposal titled " An Investigation into Data Science Remuneration Trends in 2023" comes into play. Recent media and study reports indicate that Data Science job is more in demand than ever with employers and recruiters [1].

The project constructed Shiny dashboard designed to unveil the multifaceted aspects of remuneration trends in the Data Science sector in 2023. The designed narrative visualization aims to provide a coherent, insightful, and engaging overview of compensation, specialization, global distribution, and the evolution of remote work in the Data Science domain.

Key Messages:

Highlighting Job Titles and Specializations: The dashboard aims to cast light on various job titles and specializations within the Data Science domain, allowing users to discern areas of expertise and demand.

Showcasing Geographical Distribution: The visualization is designed to unveil the global proliferation of Data Science, highlighting regions where the domain has a substantial imprint.

Illustrating Remote Work Trends: Insight into the prevalence and acceptance of remote work within the Data Science industry is a pivotal focus of the dashboard.

Examining Salary Disparities: An exploration into the compensation trends that dominate the Data Science sector, providing valuable insights into earning potentials.

The project of "**An Investigation into Data Science Remuneration Trends in 2023**" has been implemented by the author will act as underlying fundamental for widespread of audience, including:

1. **Aspiring Data Scientists**: Student, particularly student who are concerning about their major preferences when applying to higher institution course.
2. **Current Data Scientists**: They can gauge the market trends in terms of salary expectations, frequently required skills or qualifications (from the word cloud), and the prominence of various job titles (from the donut chart). They can make informed decisions about what skills to acquire based on their prevalence in job postings. The tree map indicating online working ratios can be crucial.
3. **Educational Institutions**: These visualizations can offer insights into the most sought-after skills and qualifications in the data science domain, allowing institutions to design their curriculum accordingly.

4. **Researchers and Market Analysts**: Study the evolution and trends in the data science job market. Use the insights to forecast future trends or to fuel further research on related topics.
5. **Government and Policy Makers**: Understand the employment landscape and potential opportunities in the tech industry, especially if they are trying to promote technology growth in their region or country. Design policies or incentives that align with current market demands.

The visual narrative crafted was meticulously designed to shed light on the compensation trends prevalent in the Data Science domain as of 2023. The aim was to complex data into an accessible format that would resonate with an audience comprising data science professionals, recruiters, and industry analysts. The visual narrative was intended to not only inform but also engage stakeholders by presenting a lucid and compelling story of current trends, demands, and opportunities within the field.

2. Design

During the design phase, a scrutiny approach was adopted, heavily influenced by theoretical knowledge acquired throughout the unit. Each of the five design sheets, appended for reference, captures the iterative development of the visualization tools ranging from word clouds to Tree Map and donut charts.

A decision was made to employ a 'Paired' colour scheme from RColorBrewer to ensure a harmonious visual flow, thereby enhancing the user's ability to decode information effectively. This choice was underpinned by an understanding of the human visual system's response to colour and pattern. By using this colour scheme, the legend for job title will be interpreted understand by the audience.

Furthermore, the decision to depict data using specific visualization forms was informed by the narrative storytelling genre, ensuring that each graphical element contributed cohesively to the overarching narrative.

Five Design Sheet Methodology:

Design Sheet 1: Served as a canvas for brainstorming, which were organized to give shape to understand design pathways. 12 ideas with stacked bar chart, heat map, symbol map, histogram frequencies, word cloud, tree map, donut chart, line graph, multi-set bar chart, density plot, timeline and pie chart

The chart then filtered out line graph, timeline and pie chart as its inappropriate with the data. The sheet follows with the categorize of ideas, with the combination of histogram and density plot for counting the frequency, stacked bar chart with multi-set bar chart, heat map and symbol map, tree map and donut chart, and word cloud.

Design Sheet 2: Focused on conceptualizing primary visual elements like the word cloud and Tree Map, nurturing the visualization's foundational aspects.

Design Sheet 3: Enhanced interactivity was introduced, conceptualizing a more engaging user experience through interactive maps and dynamic charts.

Design Sheet 4: Refinement of visual elements was undertaken, fine-tuning the design to enhance coherence and visual appeal.

Design Sheet 5: The culmination of the design from all the previous. step, integrating key elements from previous sheets to craft a comprehensive and engaging visualization narrative.

Dashboard design on theoretical unit in the **Table 1** below

Table 1: Visualisation genre and style in project's Dashboard

Visualisation ideas	Genre	Style	References class
Wordcloud	Magazine Style	Author	Week 5
Treemap Visualisation	Annotated Chart	Reader	Week 3
Donut Chart	Magazine Style	Hybrid	Week 2
Mapbox Visualisation	Magazine Style	Hybrid	Week 4
Stacked Bar Chart	Annotated Chart	Hybrid	Week 2

(Source: Author)

The dashboard predominantly employs the Magazine Style and Annotated Chart genres, with a mix of Author, Reader, and Hybrid styles. This blend ensures that the dashboard is both informative and interactive, catering to a wide audience range. To be more specific,

The **word cloud** aligning with the "Magazine Style" genre. The author style is evident as the visualization seems to be curated without explicit user interaction for exploration.

Tree Map visualization with radio buttons for user interaction, fitting the "Annotated Chart" genre. The reader style is emphasized through interactivity, allowing users to explore and draw conclusions.

The **donut chart** in the code corresponds with a "Magazine Style" genre, and the hybrid style is seen as it provides a curated view with interactive elements for exploration.

The **map visualization** aligns with the "Magazine Style", enhanced with cues and annotations. The hybrid style is portrayed through user-interactive elements coupled with curated content.

The **stacked bar chart** fits the "Annotated Chart" genre with annotations providing clarity. The hybrid style is evident through its combination of curated visualization and user-driven exploration.

The table follow the conceptual framework of the visualizations. It categorizes each visualization into specific genres and styles, reflecting the design intentions. The word cloud and Tree Map are intricately designed to facilitate a guided yet explorative user experience, aligning with the respective genres of "Magazine Style" and "Annotated Chart".

Simultaneously, the donut chart, map visualization, and stacked bar chart harmonize curated content with interactive elements, embodying a "Hybrid" style that enriches the user experience through a blend of author-driven design and reader-driven exploration. This thoughtful categorization in the table mirrors the implementation in the code, illustrating a meticulous design approach grounded in enhancing user engagement and interpretative ease.

3. Implementation

The implementation was executed using the R programming environment, coupled with the Shiny package to bring interactivity to the fore. A wide range of libraries was employed to bring the narrative to life, including wordcloud2 for textual visualizations, plotly for dynamic charts, and leaflet for geospatial mapping. The code's skeleton remained true to the **Five Sheet Method** as has been presented. **Table 2.** below will summarized all libraries that have been used in the project.

Table 2: Libraries used in the source code.

Library name	**Method explains**
wordcloud2	Word cloud visualisation
leaflet	Map visualisation
shiny	Visualize interactive dashboard
shinydashboard	Shiny dashboard visualisation
RColorBrewer	Enhancing visualisation with colour package
plotly	Donut chart and Tree map visualisation
tidyverse	Cleaning data with tidy verse library
countrycode	Retrieving the country name and country code in iso2c format

(Source: Author)

The challenge lay in the intricate data wrangling required to shape the dataset for visualization functionality within the Shiny framework to achieve the desired level of interactivity and user engagement. Some of the challenging can be mentioned below:

- Retrieving the coordinate data outsource
- Library ggplot2 and plotly were unable to make a scrollable horizontally, and highchart library offer it, time is limited [6]
- Wordcloud2 does not always render the most frequency term [7]
- Implementing a stacked bar chart can be time-consuming when dealing with a large number of variables.

4. User guide

Description for audience: When we're talking about Data Science, what do you know for sure?

1. **Word Cloud: Data Science - Big Picture:**
    - Using Word Cloud presents a visual representation of the most frequent terms associated with Data Science. Words displayed larger are more common terms, allowing to quickly identify key topics and trends in the field.
2. **TreeMap Visualisation: Remote Ratio in Data Science World**
    - Explore the proportions of remote work opportunities in Data Science. Utilize the radio buttons to toggle between different views and filters, allowing for a customized exploration based on your interests.
3. **Donut Chart: Data Science Specialization**
    - This section provides a circular statistical graphic that is divided into slices to illustrate numerical proportions of various Data Science specializations. Engage with the chart to discover the distribution of different specializations in the Data Science field.
4. **Mapbox Visualisation: Understand the Widespread of Data Science**
    - An interactive map is presented, showcasing the global popularity of Data Science. Use the map's interactive features to zoom and pan, allowing to explore the geographical distribution of Data Science roles and trends.
5. **Stacked Bar Chart: Data Science Remuneration**
    - This visualization allows for the exploration of Data Science remuneration trends. Bars are color-coded, and a legend is provided for easy interpretation. Hovering over each segment of the bar provides.

Navigating the narrative visualization is an intuitive process designed to facilitate user exploration. Interactivity is at the heart of the experience from my Data Visualisation Project. Users are advised to engage with the filter options to uncover different layers of the data story. Attention should be given to the TreeMap's remote work filters and the interactive map's display of Data Science's global spread. The stacked bar chart, a key element of the visualization, warrants a closer examination for its detailed salary breakdown by job title, which may otherwise be overlooked.

5. Conclusion

In conclusion, the project "**An Investigation into Data Science Remuneration Trends in 2023**" serves as an in-depth exploration of the economic aspects of the data science profession, which contrary to popular belief, is not just about hard work and monotonous routines, but also a domain of diverse opportunities and financial rewards.

The narrative visualization stands as a testament to the insightful analysis of Data Science compensation trends. It has successfully mapped out the industry's remunerative landscape, shed a light on the spectrum of specializations within the field, and captured the essence of Data Science.

Reflecting upon the project, it becomes evident that there are opportunities for refining the alignment between the visualization design and the underlying data. The study sheds light on the most and least popular job titles within the field and examines how these vary across companies of different sizes, complemented by an analysis of experience levels in conjunction with company locations. It also delves into how qualifications affect salary and benefits relative to job titles.

The narrative visualization, supported by 2 data sources combined into one file not only maps the current compensation landscape but also taps into the vitality of the field, as emphasized by Davenport in the Harvard Business Review. As the project reflects on the potential for enhancement in the visualization-data alignment, it also expresses an eagerness to integrate predictive analytics for forecasting future trends in Data Science roles and remuneration, thereby weaving a story that resonates with academics and professionals alike, across a spectrum that includes vocational passion and financial aspirations.

6. References

[1] Davenport, Thomas H. (2022), "Is Data Scientist Still the 'Sexiest Job of the 21st Century'?" Harvard Business Review, July 2022. URL: https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century. [Accessed on: [07.08.2023]]

[2] MapboxAPI. "Mapbox API R Documentation." CRAN Repository. URL: https://cran.r-project.org/web/packages/mapboxapi/mapboxapi.pdf [Accessed on: [10.09.2023]].

[3] Walker, K. "Mapbox API Geocoding Reference." Walker Data. URL: https://walker-data.com/mapboxapi/reference/mb_geocode.html [Accessed on: [10.09.2023]]

[4] "Country Codes," IBAN.com, 2023. URL: https://www.iban.com/country-codes. [Accessed on: [07.09.2023]].

[5] "Geocoding with Google API," Duke University Library, 2023. URL: https://guides.library.duke.edu/r-geospatial/geocode. [Accessed on: [10.09.2023]].

[6] "How to make highcharts scrollable horizontally when having big range in x-axis," https://stackoverflow.com/, [Online]. Available: https://stackoverflow.com/questions/16706068/how-to-make-highcharts-scrollable-horizontally-when-having-big-range-in-x-axis.

[7] "R - WordCloud2 does not always render the most frequent words," stackoverflow.com, [Online]. Available: https://stackoverflow.com/questions/41654007/r-wordcloud2-does-not-always-render-the-most-frequent-words.

[8] "Pie Charts in R," plotly.com, [Online]. Available: https://plotly.com/r/pie-charts/.

[9]"countrycode: Convert Country Names and Country Codes," rdrr.io, [Online]. Available: https://rdrr.io/cran/countrycode/.

[10] "Treemap Charts in R," plotly.com, [Online]. Available: https://plotly.com/r/treemaps/

[11]"Geocode an address or place description using the Mapbox Geocoding API," Walker-data.com, [Online]. Available: https://walker-data.com/mapboxapi/reference/mb_geocode.html.

[12] "Geocoding API," doc.mapbox.com, [Online]. Available:

https://docs.mapbox.com/api/search/geocoding/?size=n_10_n.

[13]"How do I change the formatting of numbers on an axis with ggplot? [duplicate]," stackoverflow.com, [Online]. Available: https://stackoverflow.com/questions/11610377/how-do-i-change-the-formatting-of-numbers-on-an-axis-with-ggplot.

[14] "Event handler," rstudio.github.io, [Online]. Available: https://rstudio.github.io/shiny/reference/observeEvent.html.

[15] "Build My Medium Dashboard with R Shiny," medium.com, [Online]. Available: https://medium.com/analytics-vidhya/build-the-medium-dashboard-with-r-shiny-618263243393.

[16] "shinydashboard," studio.github.io, [Online]. Available: https://rstudio.github.io/shinydashboard/.

[17] "Combine ggplotly and ggplot with patchwork?," stackoverflow.com, [Online]. Available: https://stackoverflow.com/questions/61574401/combine-ggplotly-and-ggplot-with-patchwork.

[18] "R Fluidrow/Column Font Size," Stackoverflow.com, [Online]. Available: https://stackoverflow.com/questions/52820202/r-fluidrow-column-font-size.

[19] "HTML - The Head Element," W3Schools, [Online]. Available: https://www.w3schools.com/html/html_head.asp.

7. Appendix

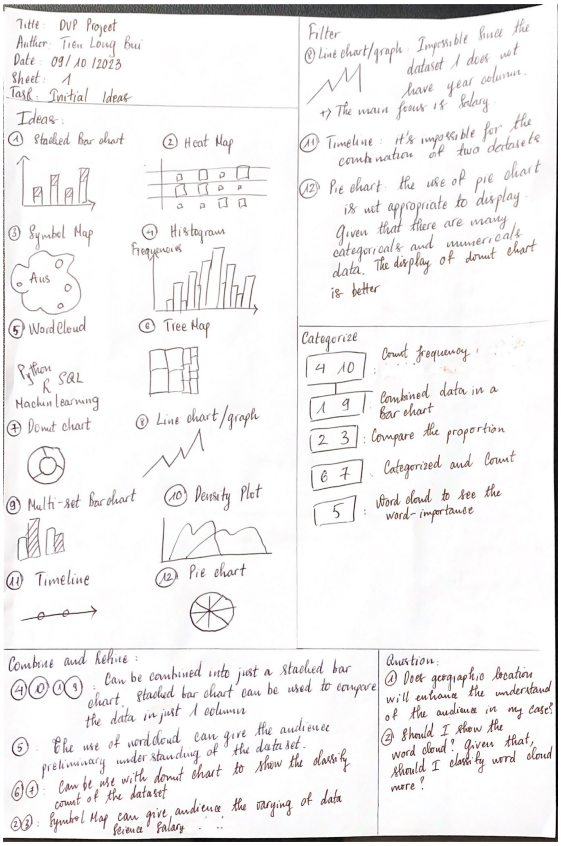

Figure 1: Five sheet method (Sheet 1)

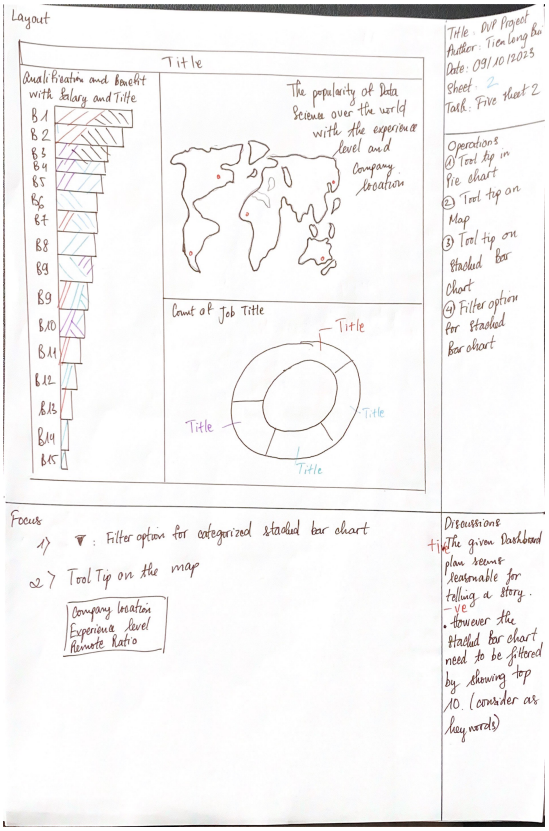

Figure 2: Five sheet method (Sheet 2)

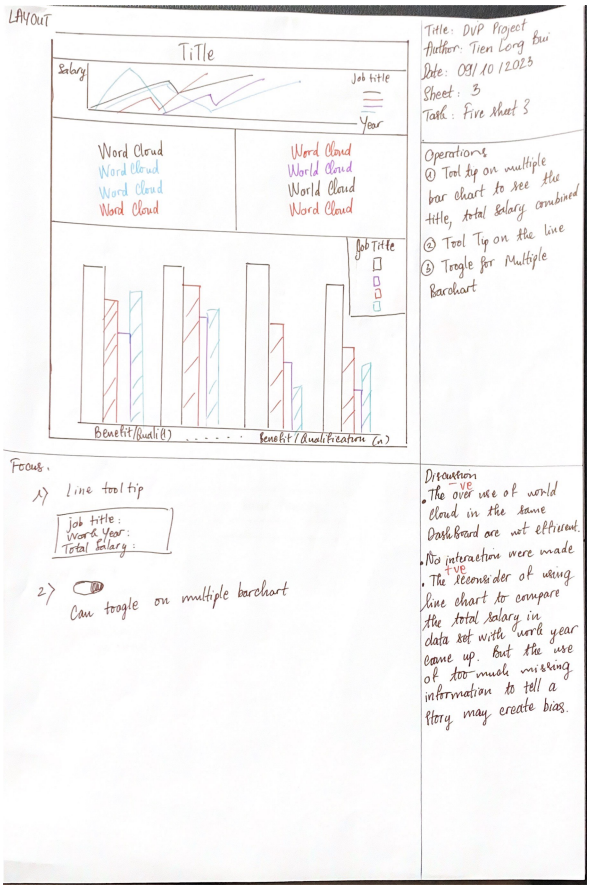

Figure 3: Five sheet method (Sheet 3)

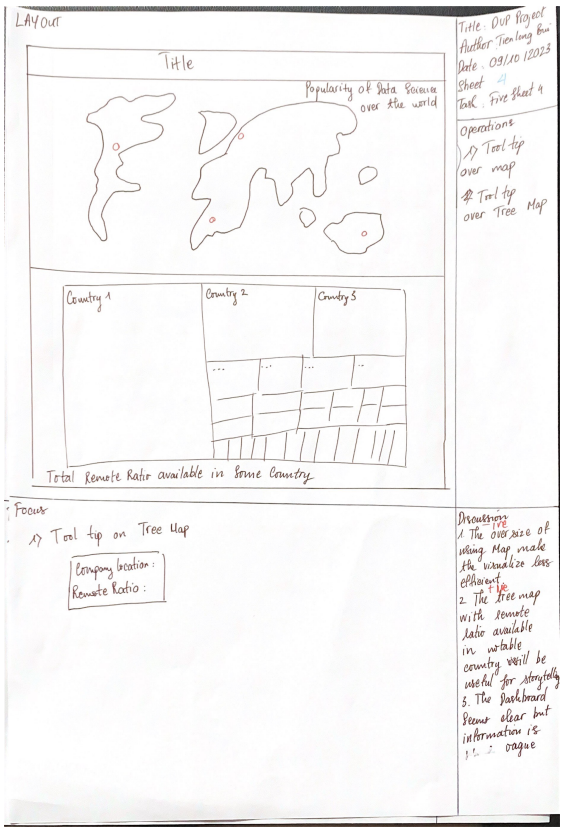

Figure 4: Five sheet method (Sheet 4)

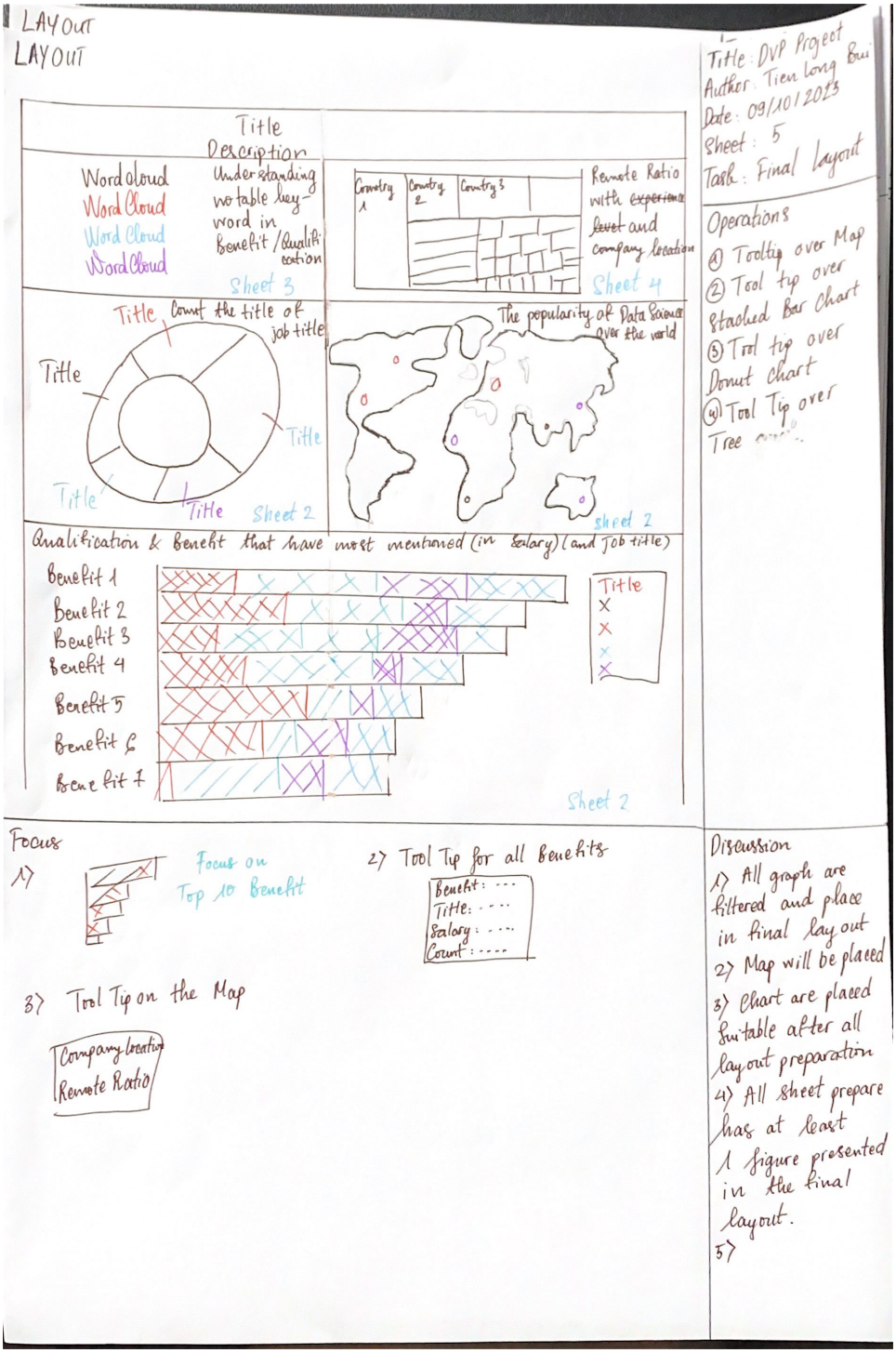

Figure 5: Five sheet method (Final Layout)