Monash University: Assessment Cover Sheet

Student name	Tah	Wen Zhong	
School/Campus	Monash University Malaysia	**Student's I.D. number**	29940672
Unit name	FIT3179 Data visualisation - S2 2021		
Lecturer's name		**Tutor's name**	Ting Chai Wen
Assignment name	Data Visualisation II Report	**Group Assignment: No** **Note, each student must attach a coversheet**	
Lab/Tute Class: Thursday	**Lab/Tute Time:** 12pm - 2pm	**Word Count:** 874	
Due date: 18-10-2021	**Submit Date:** 17/10/2021	**Extension granted** ☐	

If an extension of work is granted, specify date and provide the signature of the lecturer/tutor. Alternatively, attach an email printout or handwritten and signed notice from your lecturer/tutor verifying an extension has been granted.

Extension granted until (date): ...../...../..... Signature of lecturer/tutor: .....

Late submissions policy	**Days late**	**Penalty applied**
Penalties apply to late submissions and may vary between faculties. Please refer to your faculty's late assessment policy for details.		

Patient/client confidentiality: Where a patient/client case study is undertaken a signed Consent Form must be obtained.

Intentional plagiarism or collusion amounts to cheating under Part 7 of the Monash University (Council) Regulations

Plagiarism: Plagiarism means to take and use another person's ideas and or manner of expressing them and to pass these off as one's own by failing to give appropriate acknowledgement. This includes material from any source, staff, students or the Internet - published and unpublished works.

Collusion: Collusion means unauthorised collaboration on assessable written, oral or practical work with another person. Where there are reasonable grounds for believing that intentional plagiarism or collusion has occurred, this will be reported to the Associate Dean (Education) or nominee, who may disallow the work concerned by prohibiting assessment or refer the matter to the Faculty Discipline Panel for a hearing.

Student Statement:

- I have read the university's Student Academic Integrity Policy and Procedures
- I understand the consequences of engaging in plagiarism and collusion as described in Part 7 of the Monash University (Council) Regulations (academic misconduct).
- I have taken proper care to safeguard this work and made all reasonable efforts to ensure it could not be copied.
- No part of this assignment has been previously submitted as part of another unit/course.
- I acknowledge and agree that the assessor of this assignment may, for the purposes of assessment, reproduce the assignment and:

  i. provide it to another member of faculty and any external marker; and/or

  ii. submit to a text matching/originality checking software; and/or

  iii. submit it to a text matching/originality checking software which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking.

- I certify that I have not plagiarised the work of others or participated in unauthorised collaboration or otherwise breached the academic integrity requirements in the Student Academic Integrity Policy.

Date: 17/10/2021 Signature:  *

Privacy Statement:

For information about how the University deals with your personal information go to http://privacy.monash.edu.au/guidelines/collection-personal-information.html#enrol

Name: Tah Wen Zhong
StudentID: 29940672

FIT 3179

Assignment 2: Data Visualization 2

Name: Tah Wen Zhong

StudentID: 29940672

Word count: 874

Visualization URL:
https://wtah0001.github.io/FIT3179_Assignment_2/

Repository URL:
https://github.com/wtah0001/FIT3179_Assignment_2

Name: Tah Wen Zhong
StudentID: 29940672

Assignment 2: Report

Domain, why & who

The domain I selected is on university rankings, mainly the ranking systems behind the global university rankings. In general, students will grow an interest towards the world's top universities as these are the ideal choice for students to secure a future career. As such, my target audience are students as their interest towards top universities may lead to an interest on understanding the ranking system for universities. With that said, my aim for this visualization is to analyze the relation between score ranking and metric, as well as understanding the total score's progression rate between different ranks and scores. Additionally, information on the location of the top will also be shown.

What

There are two datasets used for the visualization, the first dataset was obtained from Webometrics and the other from Kaggle. The first dataset contains details on the number of universities, split by rank ranges, for each country in the current year 2021. The second dataset shows data for several ranking methodologies between 2011 and 2016, with details on metric scores as well as university name, location, and rank. For the second dataset, I chose to focus mainly towards the data on rankings based on the Times Higher Education methodology.

Name: Tah Wen Zhong
StudentID: 29940672

Why and how

There are three sections in the visualization, the first section focuses on the location of the top universities. For this section, I chose a choropleth map due to the properties of the data available. In essence, all the data available are connected to enumeration units, that is country, and can be normalized to obtain the percentage of top-ranking universities in each country. With this, the data is compatible for this idiom. Furthermore, the idiom effectively shows the quantitative data classes through color luminance on geographical areas within the map.

The second section shows the relation between scores and world ranks. For this visualization, a line chart is used as it effectively portrays the progression of total score as rank increases. Furthermore, it also allows users to easily interpret the magnitude of the score differences between ranks. In addition, the chart and data were very suited to apply interactive features offered in Vegalite, mainly the flexible changes to the line's properties based on the filter. As such, the line chart can now include multiple lines using the Vegalite interactive features. The filter combine with varying opacity will prevent additional lines to affect the readability of the main line.

Lastly, the last section is focused on the metrics used to determine a University's score. There are two idioms used here, those being a stacked bar chart and the boxplot. Both idioms share the same data but reveals different information which complements one another. The stacked bar chart shows the average score of each metric, and an interesting fact is that the sum of these series of metrics equates to the total score. So, this idiom allows the total score progression to be analyzed while also showing the weight distribution of total score on the metrics. However, this idiom does not show the score distributions for the individual metrics. This led to my decision to include the boxplot as it covers the items which the stacked bar chart lacks. The boxplot reveals how the metric scores are distributed and allows outliers to be identified and analyzed.

Name: Tah Wen Zhong
StudentID: 29940672

Design

The visualization's layout was structured well where there is a consistent row pattern that each section follows, excluding the introduction. Throughout the visualization, there are clear sightlines shown which makes the overall visualization look neat. Symmetry and balance were also achieved due to how the components were positioned and resized.

For color, I chose a blue as the main color theme as it fits the theme of university. The charts used color schemes which fits its usage, mainly color saturation for quantitative and color hue for qualitative. The boxplot and stacked bar chart share the same qualitative data, so the same color scheme was used to show a form of connection between the two charts.

For this visualization, figure-ground techniques were frequently applied to highlight importance of certain items. Interesting parts within a text are bolded to provide a level of importance. Additionally, texts are color coded when they share connections between items, mainly the metrics for the third section. Additionally, there is a clear size difference between text, titles, and section titles to express a level of importance for each item.

For typography, there are two typefaces used throughout the visualization. These two typefaces being Lato and Roboto. Lato was used for most of the text due to its clarity. When a text is bolded using Lato, the text becomes highlighted. However, its weight does not affect its readability, making it an ideal font choice. Roboto was used for the title as its contrast in design, weight, and size with Lato creates a form of importance.

For storytelling, sections and items were positioned to follow the sequence which they should be read on. As such, this design ensures the reader will read the items following the correct order. Enclosure and similarity techniques were used as well for user to classify the different parts of the visualization. One example would be the enclosure on the section titles, this shows a clear difference between other items which helps users for item classification.

Name: Tah Wen Zhong
StudentID: 29940672

University Rankings

An insight on the methodologies used to for ranking universities

There are many known University ranking organizations, each specializing on different scales with varying ranking methodologies. The elements which rankers use for ranking Universities are based on its effectiveness for evaluating the quality of Universities. While there is no definitive method for ranking, ranking systems have been acknowledged to play an important role in the higher education system (Thakur, 2007). The **Times Higher Education World University Rankings** is one of the more popular ranking system that evaluates Universities in an International scale, and will the be ranking system that the following sections are based on.

Location of the Top Universities

From choropleth map below, we can see that most of the Top ranking Universities are located in the **America**. However, out of the 3216 Universites in **America**, only 7.87% of the Universities are within the Top Rankings. The **United Kingdom** on the otherhand has 280 Universities, but 28.21% of them are within the Top Rankings.

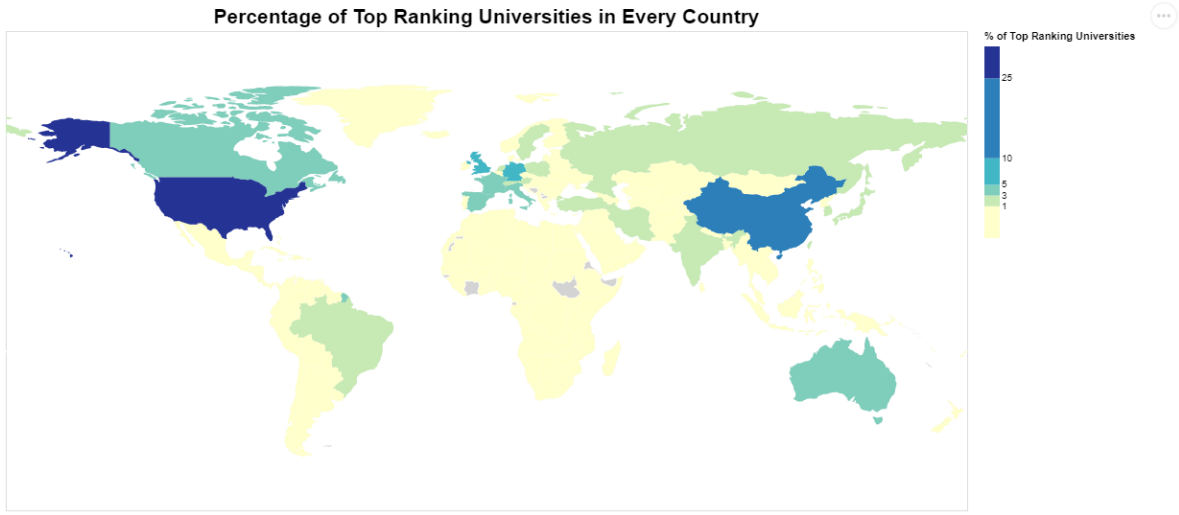


Figure 1: Introduction and first section of the visualization

Name: Tah Wen Zhong
StudentID: 29940672

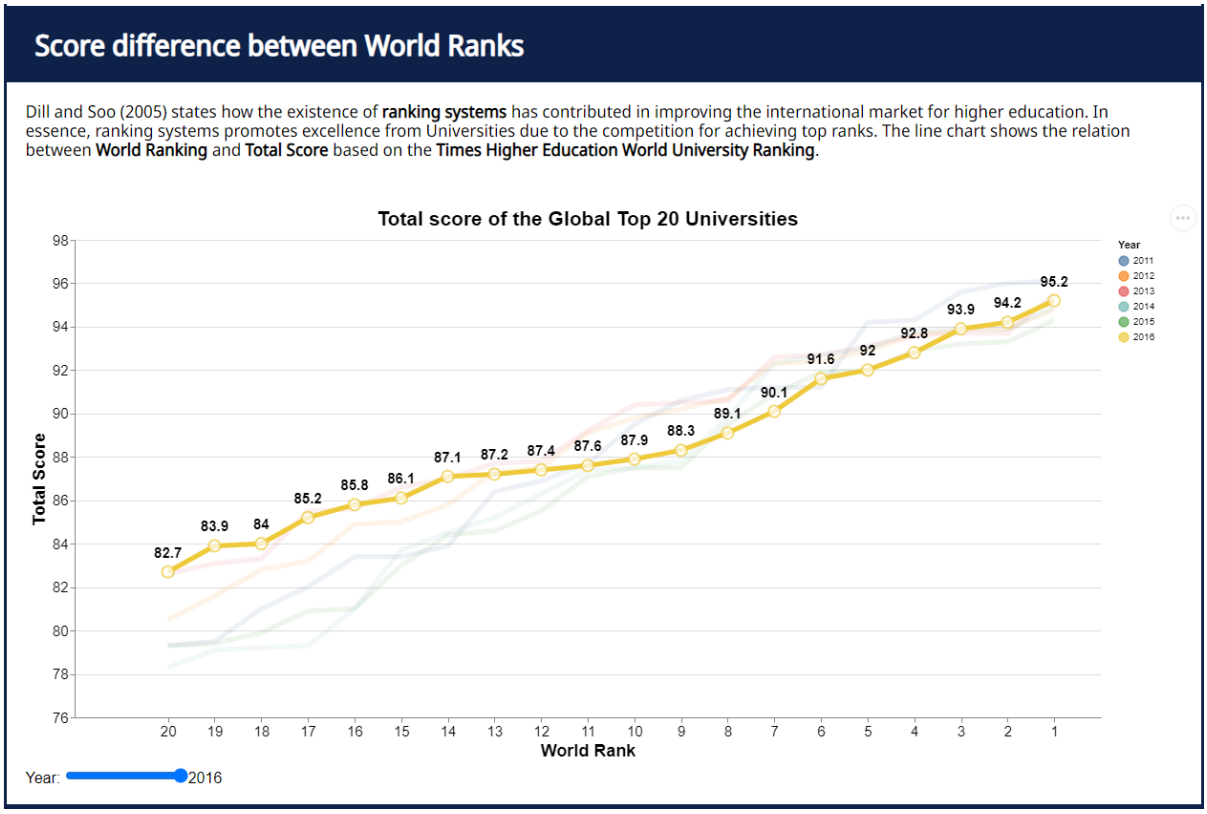


Figure 2: Second section of the visualization

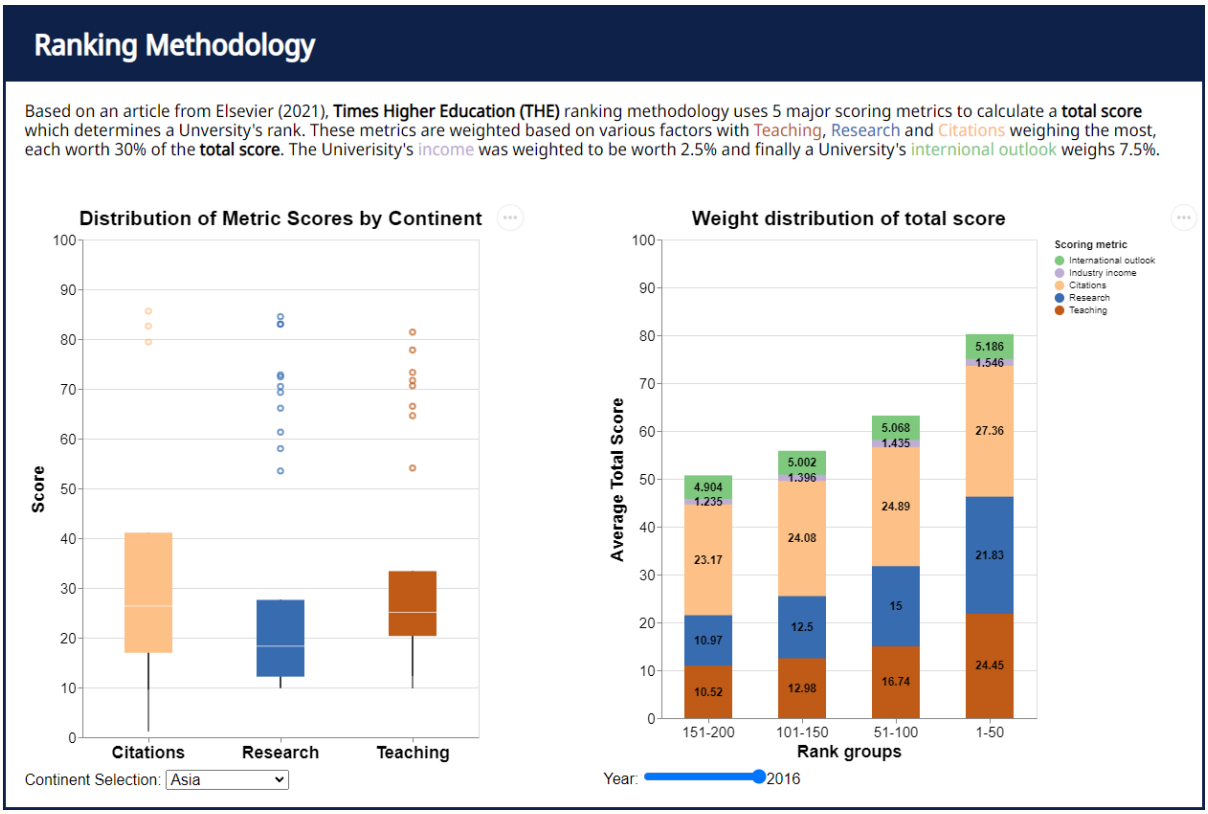


Figure 3: Third section of the visualization

References

Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher education*, *49*(4), 495-533.

Elsevier. (n.d.). *University Rankings Data: A closer look for research leaders*. Elsevier.com. Retrieved October 17, 2021, from https://www.elsevier.com/research-intelligence/university-rankings-data.

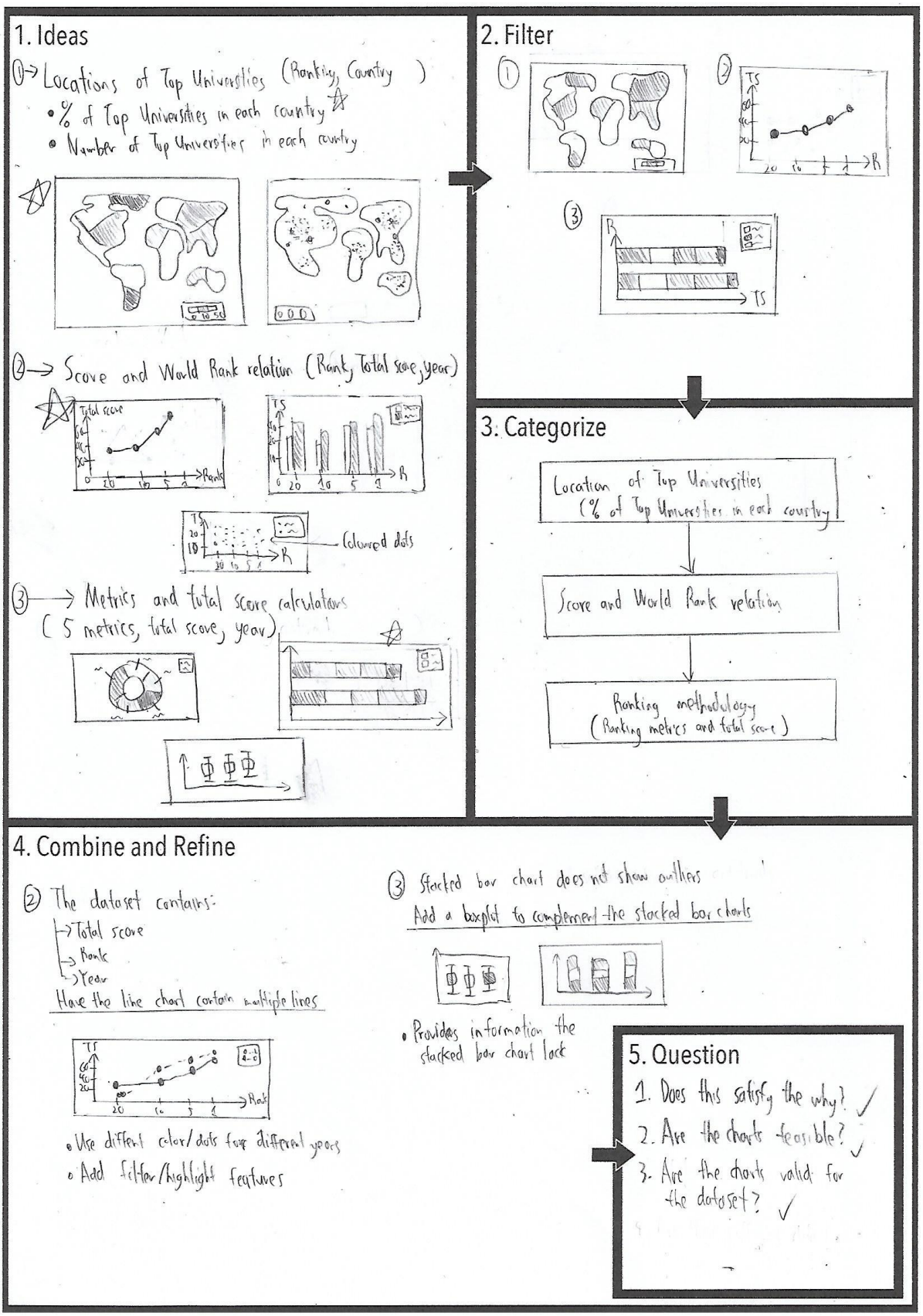
O'Neill, M. (2019). *World University Rankings* [Data file]. Retrieved from https://www.kaggle.com/mylesoneill/world-university-rankings

Thakur, M. (2007). The impact of ranking systems on higher education and its stakeholders. *Journal of Institutional Research*, *13*(1), 83-96.

Webometrics (2021). *Countries arranged by Number of Universities in Top Ranks* [Data file]. Retrieved from https://www.webometrics.info/en/distribution_by_country

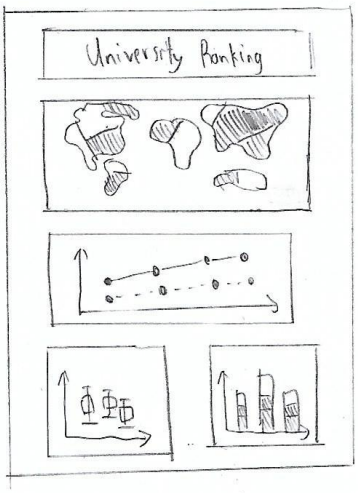
Name: Tah Wen Zhong
StudentID: 29940672

Appendix – 5 Design Sheets



Name: Tah Wen Zhong
StudentID: 29940672

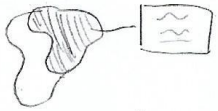
Layout



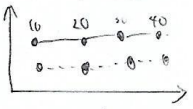
Title: University Rankings
Auther: Tah Wen Zhang
Date: 4/10/2021
Sheet: 2
Task: Chart_focused layout

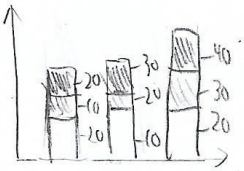
Operations

Tooltips to show information of the focused area

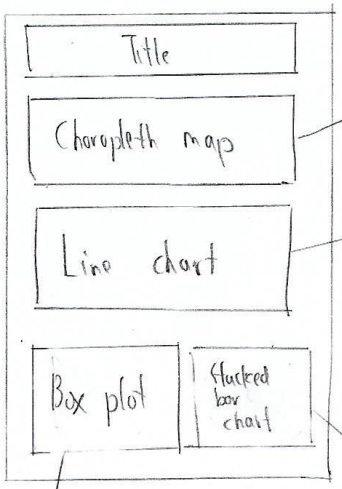


In text annotations for line chart and stacked bar chart





Focus



Choropleth maps to show the percentage of top ranking Universities in each country

Line chart to analyze the relation between ranks and total score

Stacked bar chart to show the changes in metrics value and its effects toward the total score

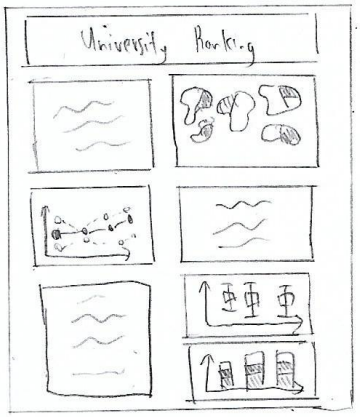
Boxplot to show distribution between the metrics for total score calculation

Discussion

+ Simple and compact
+ Chart focus will induce users to utilize the operation available
+ Straightforward
- Lack of text removes explanations for what each section represents
- The stacked bar chart and boxplot may have irregular sizes compared to other charts
- Line chart and map may not scale well on width

Name: Tah Wen Zhong
StudentID: 29940672

Layout

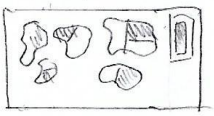


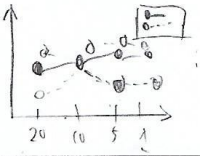
Title: University Rankings
Auther: Tah Wen Zhong
Date: 4/10/2021
Sheet: 3
Task: Balanced layout

Operations

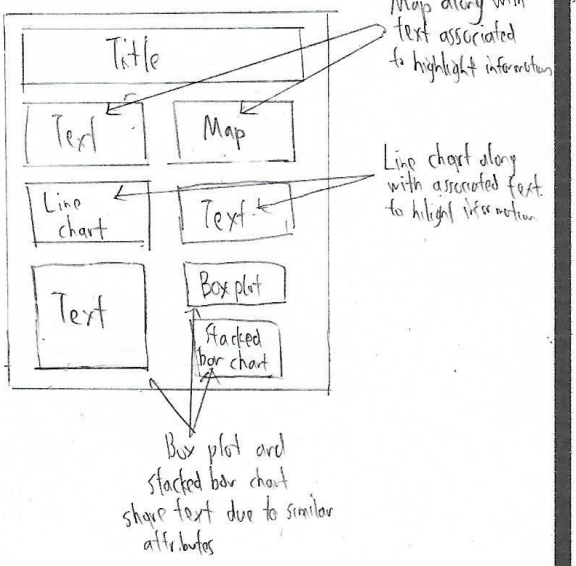
Use vegalite interactive features to adjust opacity to highlight main line

Legends for line chart and map





Focus

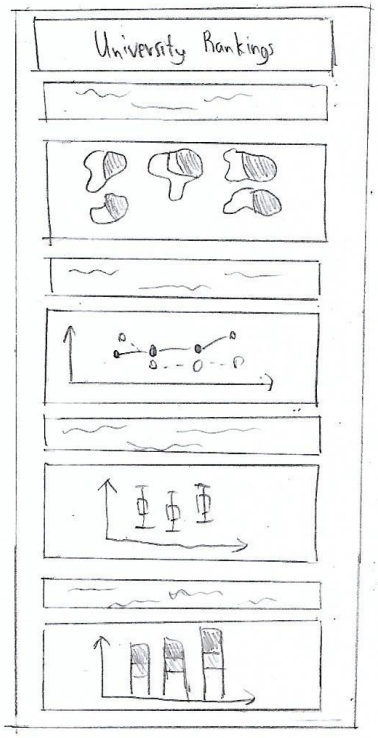


Discussion

+ Text can highlight interesting information
+ Sections are equally balanced
+ Clear sightlines
- Map size reduced by half, harder to analyze
- Chart heights are dependant on text size and vise versa
- Boxplot and stacked bar chart may be unbalanced due to tight vertical space

Name: Tah Wen Zhong
StudentID: 29940672

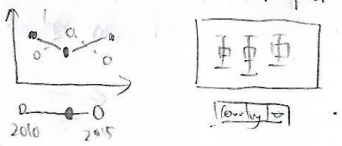
Layout



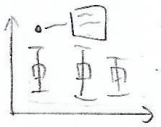
Title: University Rankings
Auther: Tah Wen Zhong
Date: 4/10/2021
Sheet: 4
Task: Balanced Layout

Operations

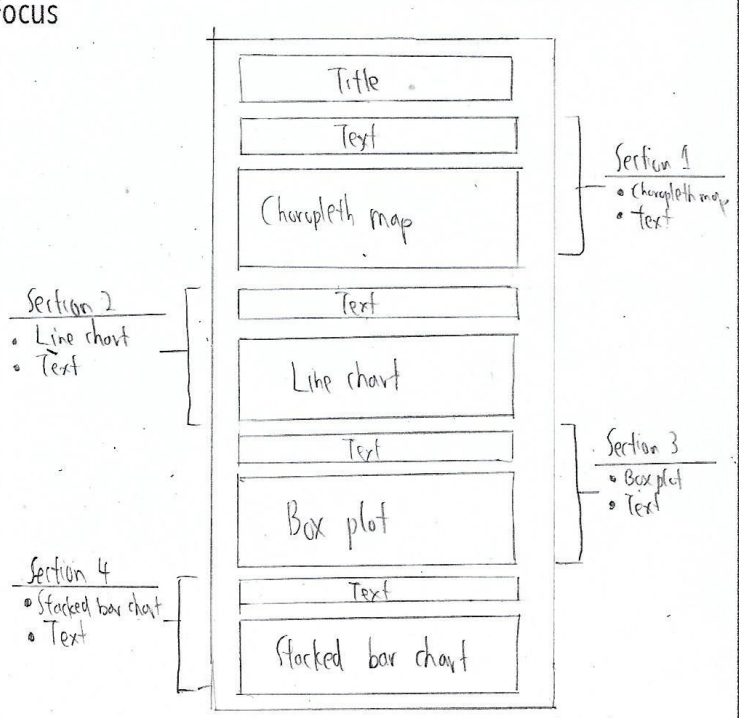
Filter for line chart and boxplot



Tooltip for outliers in boxplot



Focus

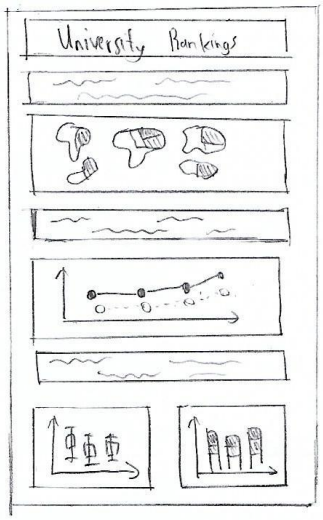


Discussion

+ Text can highlight useful information
+ Good symmetry
+ Useful operations for more in-depth analysis
+ Spacious for all charts
- Seperate sections for box plot and stacked bar chart will lose the intuitive factor of their connections
- Higher possibility of redundant text
- Reduced height of stacked bar chart for symmetry may reduce its effectiveness

Name: Tah Wen Zhong
StudentID: 29940672

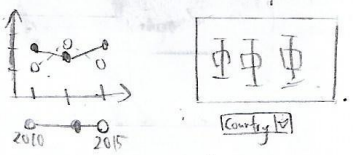
Layout



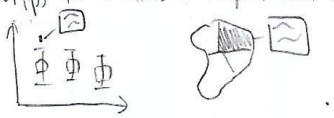
Title: University Rankings
Auther: Tah Wen Zhong
Date: 4/10/2021
Sheet: 5
Task: Final Implementation Design

Operations

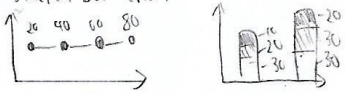
Filter for line chart and boxplot:



Tooltips for outliers in boxplot and map

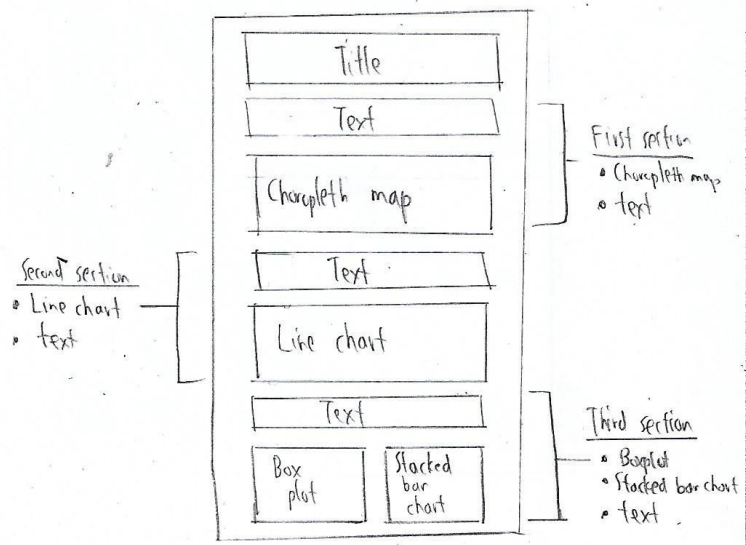


In text-annotations for line chart and stacked bar chart



Vegalite interactive features for adjusting opacity to focus on selected line in line chart

Focus



Detail

- Dataset type : csv
- Extensions for map : TopoJson
- Software to build visualisation : Vegalite
- Desired completion date : 13/10/2021 (prior to critique session)
- Use percentage instead of count for choropleth map
- Use mean for most attributes
- Utilize Vegalite interactive features