Internal Displacement and Migration patterns: a zoom-in on Australia

Project Process Book

Link to mercury
https://mercury.swin.edu.au/cos30045/s103496628/final_project/

COS300045 Data Visualisation
Semester 1 2023
Tutorial Monday 12:30pm

Group 11
Duc Tung Nguyen 103496628
Tevy Tunsay 103139978

Word Count: 5214 words

Table of Content

1. Introduction

1.1. Background and Motivation

Migration is a natural and inevitable aspect of human history. In the current globalized world, international migration has become more common. In 2020 alone, according to the World Migration Report, there were 281 million international migrants or 3.6 percent of the world's population. There are both push factors and pull factors resulting in cross-border migration. On the former, internal displacement has been one of the keys. Promising economic benefits and living conditions are high on the list of pull factors (IOM, 2021).

For similar reasons, people from other continents have migrated to Australia. The World Migration Report 20202 indicates that the country is of the 10 top destinations for international migration (IOM, 2021). According to the data from the Australian Bureau of Statistics (ABS), in the year ending 30 June 2022, overseas migration contributed a net gain of 171,000 to Australia's population. This represents a very large increase in net overseas migration in the 2020-21 financial year (ABS, 2022).

A good understanding of migration patterns, causes, and effects is important for a country's effective policymaking in areas that can affect and be affected by international migration. Those include, for example, population policy, economic policy, and international relation agenda (IOM, 2021).

However, many countries, including Australia, suffered from natural disastrous events that occur almost every year, consisting of both less and heavily severe conditions. Such scenarios forced the population to move, to unwillingly leave their familiar home for the sake of salvaging their living condition that was sabotaged by natural disasters.

In this assignment, we will use available quantitative data and create visualizations on two migration-related questions where readers can expect to:

1. identify the major trend and patterns of migration to Australia: the migrant's origin country, their visa categories, and their destination to any state of Australia.
2. Understand the push factor of Internal Displacement which is Natural Disasters: comparing Australia with other countries among types of disasters that lead to internal displacement.

1.2. Visualisation Purpose

In this assignment, two diagrams are produced to visualize two specific phenomena relating to cross-country migration and within-border migrations, with a specific focus on Australia's metrics:

1) Natural Disasters as a cause of internal migration (Australia vs. Other countries)

2) The pattern of migration (to Australia).

Our purpose is to aim towards the general public and inform such audiences with relevant information regarding both internal and external migration trends within Australia, as well as its relationship with other countries around the globe.

1.3. Project Schedule and deliverable

As we move towards developing a fully programmatic data visualization, we aim to follow an iterative approach in both data processing and visualization designs. Furthermore, both processes should also be performed in parallel to one another, iteratively progressing to achieve the set requirements.

Week	Tasks Scheduled
Week 7: Topic selection and design	○ Data researching ○ Clean up the data
Week 8: Work on data & Design	○ Finalize topic selection ○ Sketching on Data visualisation ○ Sketching on Website Visualisation
Week 9: Data conforming	○ Clean up the data ○ Finalise design decisions
Week 10: Programming	○ Start programming/implementing the visualisation. ○ Applying necessary data and design changes (iterative method)
Week 11: Visualisation design	○ Re-design iterative and implementation into the code
Week 12: Finalize	○ Finalize the website functionalities.
Week 13: Usability	○ Finalize Process Book ○ Perform Usability Testing

2. Data

2.1. Data Sources

We obtained data from two different sources for the two different viewpoints, which will be implemented into two different visualisations.

Global internal displacement database

The database is made publicly available on the website of the Internal Displacement Monitoring Center (IDMC). The database provides comprehensive information on internal displacement worldwide. It covers all countries and territories for which IDMC has obtained data on situations of internal displacement, and provides data on situations of:

- Internal displacement associated with conflict and generalized violence (2009-2022);
- Displacement associated with sudden-onset natural hazard-related disasters (2008-2022)

The dataset used is obtained from the following links:

https://www.internal-displacement.org/countries/australia

Permanent Migration Pattern to Australia

The data used to visualize the pattern of migration to Australia is retrieved from the public repository of the Australian Bureau of Statistics (ABS), latest releases on “Permanent migrants in Australia: Characteristics of permanent migrants who arrived in Australia between 1 January 2000 and 10 August 2021.” from the following link, under “Data downloads” section:

https://www.abs.gov.au/statistics/people/people-and-communities/permanent-migrants-australia/latest-release

Both are renowned and official data sources that provide sufficient data records to assist us in developing our data visualisation project that is also up-to-date, authentic and of high integrity.

2.2. Data Processing

While the datasets collected from previously mentioned organisations ensure high integrity and timeliness, necessary data cleaning and re-organization are to be performed on the original datasets to allow ensure the desired level of “conformity” required for visualization in D3.js. The data processing process takes the necessary separate steps for two different retrieved datasets, as each required a level of pruning, pivoting, and aggregating. Upon design selection lock-in, the dataset will then be filtered to fit the requirements of the data visualization design, which will be the milestone of “conformity”, stating that the data is ready to be visualized.

The Internal Displacement Data

The originally downloaded data were included in a Comma Delimited Value file. As working with a majority amount of quantitative data, integrity must be ensured in order to provide precise representations of the data as we implement them into the visualisation development stage.

The original dataset contains all specific data related to recent internal displacement events, type of event, number of affected populations and the region in which the event took place (as well as dates). As each record here represents a single natural disaster event displacement in sequentially exact dates and includes multiple missing "Event Name" fields, our team aims to aggregate the data, grouped by specific country Name and Hazard Type by summing all Internal Displacement counts within such group.

Name	Year	Start Date	Event Name	Hazard Type	Internal Displacements
Abyei Area	2018	1/07/2018	Abyie: Flood - 01/07/2018	Flood	2
Abyei Area	2019	1/06/2019	Abyei: Flood - southern parts - 01/06/2019	Flood	40000
Afghanistan	2008	1/01/2008	0	Extreme temperature	0
Afghanistan	2008	17/04/2008		Earthquake	3250
Afghanistan	2008	1/08/2008		Flood	180
Afghanistan	2009	1/01/2009		Flood	25185
Afghanistan	2009	1/01/2009		Earthquake	3250
Afghanistan	2009	1/01/2009	0	Flood	0
Afghanistan	2009	1/01/2009	0	Mass movement	0
Afghanistan	2010	1/01/2010		Flood	70000
Afghanistan	2010	1/01/2010	0	Flood	0
Afghanistan	2010	1/01/2010	0	Mass movement	0
Afghanistan	2010	1/01/2010		Earthquake	1000
Afghanistan	2010	1/01/2010	0	Mass movement	0
Afghanistan	2011	1/01/2011		Flood	3000
Afghanistan	2012	1/01/2012		Mass movement	238
Afghanistan	2012	1/01/2012		Flood	1500
Afghanistan	2012	1/01/2012		Flood	1872
Afghanistan	2012	1/01/2012		Flood	12589
Afghanistan	2012	1/01/2012		Extreme temperature	714
Afghanistan	2012	1/01/2012		Earthquake	5082
Afghanistan	2012	1/01/2012		Extreme temperature	637
Afghanistan	2012	1/01/2012		Mass movement	546
Afghanistan	2012	1/01/2012		Extreme temperature	1656
Afghanistan	2012	1/01/2012		Flood	4685

Figure 1: Snapshot of the original IDP dataset

The first step of cleaning resulted in a completely aggregated and pivoted dataset, listing each country and different Internal Displacement Populations for each specific Hazard Type for the relevant country, as seen in Figure 2.

A	B	C	D	E	F	G	H	I
Country Name	Flood	Extreme Temperature	Earthquake	Mass movement	Storm	Drought	Volcanic eruption	Wildfire
China	43247080	1009612	18742239	138565	29595677	0	0	50157
Philippines	7320804	315	1055053	33617	44546444	5405	722177	94
India	39531602	0	166510	8086	13646865	63404	0	89
Pakistan	15397726	16104	1092728	15764	61978	266	0	0
Bangladesh	5651728	0	100	74524	9795760	0	0	0
United States	535655	19871	683	2411	7075786	0	2800	2866089
Indonesia	4235667	0	2223525	51441	171559	0	866548	2731
Cuba	22638	0	0	0	6773795	0	0	0
Nigeria	6075069	0	0	0	19709	0	0	0
Myanmar	3363385	0	28992	12053	2599798	0	0	46
Viet Nam	579000	0	68	3112	4507250	0	0	720
Japan	504233	116	707465	407742	2992893	0	363	43
Chile	44286	2431	3995701	31001	393	0	24425	29478
Nepal	910511	0	2669209	49465	23667	0	0	1299
Colombia	3453794	0	2667	39956	22562	0	61550	683
Somalia	2027558	0	0	0	91611	1379459	0	4644
Sri Lanka	2644765	0	0	23170	679684	0	0	231
Thailand	3020898	288	0	310	167465	0	0	185
Brazil	2209694	1422	206	86200	523977	9815	0	166
Ethiopia	2072928	0	0	1458	7600	705960	3900	11447
Mexico	1361918	306	200896	1319	900076	0	470	6875
Haiti	35233	0	1728363	0	491552	0	0	0
Dem. Rep. Congo	1190778	0	4577	13795	95835	0	599448	7211
South Sudan	1892856	0	0	0	8200	5000	0	1166
Niger	1859631	0	0	0	0	0	0	0

Figure 2: Aggregated and Pivoted Internal Displacement Data

There are eight hazard types included in the revised dataset. They are:

1. Mass movement (majority as wet mass movement)
2. Earthquake
3. Volcanic Eruptions (including volcanic activity, as all records of volcanic activity are volcanic eruptions)
4. Storm
5. Flood
6. Wildfire
7. Extreme Temperature
8. Drought

Note: Human-induced are unclassified natural disasters and will not be included in the final dataset.

As each category aggregates to different totals, some may be significantly less than others, necessary pruning and filtering is to be done in order to ensure that only the most significant records are visualised. In addition to this process, as a great gap could be observed between the maximum and minimum values of IDPs between countries, visualising the data using raw population data would cause great usability issues, as when high-value data positioned next to low-value data, the visualised proportion would not achieve the desired purpose (which will be discussed later within functionalities). Because of this, necessary data alterations are processed.

The data alteration step includes two specific goals:

- Normalise the dataset onto a common scale (percentile ranking) which will allow a more intuitive and user-friendly representation of the data.
- Filter-to-focus on specific countries of highest IDP counts.

Given these steps, the dataset is pruned to only keep the top country records where the average IDP counts over all Hazard Types are greater than 10,000 people. Following the filtering, each IDP records of each Hazard Type, for each country will be ranked based on its current position relative to other countries, within the specific Hazard Type, resulted in a normalised dataset with percentile ranking data in place of raw IDP counts.

Country Name	Flood	Earthquake	Mass movement	Storm	Wildfire	Volcanic Eruption
China	1	1	0.987	0.987	0.913	0
Philippines	0.962	0.925	0.901	1	0.506	0.987
India	0.987	0.851	0.777	0.975	0.493	0
Pakistan	0.975	0.938	0.827	0.604	0	0
Bangladesh	0.938	0.53	0.962	0.962	0	0
United States	0.641	0.592	0.679	0.95	1	0.876
Indonesia	0.925	0.962	0.938	0.74	0.765	1
Cuba	0.086	0	0	0.938	0	0
Nigeria	0.95	0	0	0.456	0	0
Myanmar	0.901	0.777	0.802	0.901	0.469	0
Viet Nam	0.691	0.518	0.703	0.925	0.629	0
Japan	0.604	0.913	1	0.913	0.456	0.802
Chile	0.148	0.987	0.888	0.172	0.901	0.925
Nepal	0.753	0.975	0.925	0.493	0.703	0
Colombia	0.913	0.654	0.913	0.481	0.617	0.962
Somalia	0.839	0	0	0.629	0.802	0
Sri Lanka	0.876	0	0.876	0.839	0.555	0
Thailand	0.888	0	0.469	0.716	0.543	0
Brazil	0.864	0.567	0.975	0.827	0.518	0
Ethiopia	0.851	0	0.617	0.32	0.851	0.888
Mexico	0.777	0.864	0.604	0.864	0.827	0.814
Haiti	0.123	0.95	0	0.814	0	0
Dem. Rep. Congo	0.765	0.691	0.814	0.641	0.839	0.975
South Sudan	0.827	0	0	0.333	0.679	0
Niger	0.814	0	0	0	0	0

Figure 3: Normalised Dataset, with less significant results pruned/filtered

At this current progress, only six most significant hazard types are kept for the normalised dataset (filtering out the bottom two least significant hazard types of “Drought” and “Extreme Temperature”), due to the prevalent number of countries not experiencing such hazards (where their IDP value holds at 0.)

At this point, these two versions of the dataset will be maintained for the purpose of visualisation (one normalised, the other with raw IDP values). The data structure of both versions is the same, except the data type of each “Hazard Type” record.

Variable names	Types	Description	Values
Country Name	String	Name of the subject country	"Viet Nam", "Australia", etc.
Flood	Integer (or Floating points for *normalised* version)	Records of Internal Displacement Population caused by Flood. The data type will be different between normalised and unnormalized versions. It should hold IDP counts or percentile ranking in decimals	200134, 1232004, 23340 (or 0.934, 0.790, 0.124 in normalised versions)
Earthquake	Integer (or Floating points for *normalised* version)	Same function as above for Earthquake category	200134, 1232004, 23340 (or 0.934, 0.790, 0.124 in normalised versions)
Mass movement	Integer (or Floating points for *normalised* version)	Same function as above for Mass movement category	200134, 1232004, 23340 (or 0.934, 0.790, 0.124 in normalised versions)
Storm	Integer (or Floating points for *normalised* version)	Same function as above for Storm category	200134, 1232004, 23340 (or 0.934, 0.790, 0.124 in normalised versions)
Wildfire	Integer (or Floating points for *normalised* version)	Same function as above for Wildfire category	200134, 1232004, 23340 (or 0.934, 0.790, 0.124 in normalised versions)
Volcanic Eruption	Integer (or Floating points for *normalised* version)	Same function as above for Volcanic Eruption category	200134, 1232004, 23340 (or 0.934, 0.790, 0.124 in normalised versions)

Table 1: Internal Displacement Cleaned Dataset Data Dictionary

The de-normalised version of the dataset will be kept for "detailing" purpose within the visualisation and will not be visualised onto the intended visualisation medium.

The Migration Pattern Data (Australia)

The migration pattern data was taken from ABS in multiple tabular (and pivoted) forms. As we are working mainly with quantitative data for the second time, necessary cleaning steps should be taken in order to ensure its integrity. As the retrieved data cube from ABS consists of 10+ sub-datasets of different Level of Details, for the purpose of the assignment, only two sub-sets are of note. They are data that related to:

- Table 2: Greater Capital City Statistical Areas by visa stream, permanent migrants
- Table 5: Country of birth by visa stream, permanent migrants

These two sub-tables come in human-readable format and must be further broken down and pruned in order to "conform" them for data visualization.

Australian Bureau of Statistics

34170DO001_2021, Permanent migrants in Australia, 2021

Released at 11.30am (Canberra time) Wed 29 Mar 2023

Table 2 Greater Capital City Statistical Areas by visa stream, permanent migrants

	Skilled	Family	Humanitarian	Other permanent	Total
	no.	no.	no.	no.	no.
New South Wales					
Greater Sydney	481,569	312,781	86,398	396	881,146
Rest of New South Wales	55,313	41,779	9,353	49	106,498
Total (a)	*537,121*	*354,728*	*95,797*	*448*	*988,099*
Victoria					
Greater Melbourne	465,129	249,190	86,426	255	800,996
Rest of Victoria	31,869	22,875	7,827	60	62,625
Total (a)	*497,157*	*272,262*	*94,342*	*305*	*864,067*
Queensland					
Greater Brisbane	175,022	89,450	26,896	102	291,466
Rest of Queensland	103,903	60,514	9,021	76	173,510
Total (a)	*279,214*	*150,164*	*35,958*	*170*	*465,512*
South Australia					
Greater Adelaide	102,198	41,976	23,454	62	167,690
Rest of South Australia	5,867	4,160	778	4	10,817
Total (a)	*108,138*	*46,186*	*24,261*	*69*	*178,661*
Western Australia					
Greater Perth	245,115	95,759	23,675	189	364,730
Rest of Western Australia	21,975	10,916	799	4	33,695
Total (a)	*267,299*	*106,795*	*24,521*	*193*	*398,808*
Tasmania					
Greater Hobart	11,743	4,130	1,903	-	17,771

Figure 4: Example of sub-dataset Table 2, representing flows of migrant towards different State/Territory in Australia

For "Table 2", different Australian States/Categories were aggregated by each Visa Streams, and the data structure is then unpivoted. Due to ABS's integrity, data alteration was not necessary. The resulting dataset could be seen in the figure below:

Visa Stream	State and Territory	Permanent Res
Skilled	New South Wales	537121
Family	New South Wales	354728
Humanitarian	New South Wales	95797
Other permanent	New South Wales	448
Skilled	Victoria	497157
Family	Victoria	272262
Humanitarian	Victoria	94342
Other permanent	Victoria	305
Skilled	Queensland	279214
Family	Queensland	150164
Humanitarian	Queensland	35958
Other permanent	Queensland	170
Skilled	South Australia	108138
Family	South Australia	46186
Humanitarian	South Australia	24261
Other permanent	South Australia	69
Skilled	Western Australia	267299
Family	Western Australia	106795
Humanitarian	Western Australia	24521
Other permanent	Western Australia	193
Skilled	Tasmania	17491
Family	Tasmania	7879
Humanitarian	Tasmania	3841

Region of Birth	Visa Stream	Permanent R
South-eastern Asia	Skilled	237654
South-eastern Asia	Family	220187
South-eastern Asia	Humanitarian	30862
South-eastern Asia	Other permanent	62
Eastern Asia	Skilled	264255
Eastern Asia	Family	177419
Eastern Asia	Humanitarian	5821
Eastern Asia	Other permanent	165
Western Asia	Skilled	41052
Western Asia	Family	53137
Western Asia	Humanitarian	89798
Western Asia	Other permanent	56
Central Asia	Skilled	2267
Central Asia	Family	2012
Central Asia	Humanitarian	2077
Central Asia	Other permanent	0
Southern Asia	Skilled	535155
Southern Asia	Family	156996
Southern Asia	Humanitarian	69227
Southern Asia	Other permanent	50
Latin America and the Caribbean	Skilled	26473
Latin America and the Caribbean	Family	23367
Latin America and the Caribbean	Humanitarian	595
Latin America and the Caribbean	Other permanent	46
Northern America	Skilled	25942

Figure 5: Table 2 and 5 after aggregation and unpivoting

As similar notion could be applied onto “Table 5” sub-dataset, yielding provided results.

Prior to aggregation, in “Table 5”, as the data represents “Country of Birth”, which includes a multitude of country records. This does not align to the purpose of the project, as we aim to show specific regions data, not to country. Given this, region conversion steps are provided using Excel’s VLOOKUP and matching between an external record and our original dataset. This eventually results in the provided dataset in the figure above.

Country of Birth	Citizenship	Permanent Resident	Region	Region	name
India	Australian citizen	282773	Southern Asia	Southern Asia	Afghanistan
India	Not an Australian citizen	156186	Southern Asia	Southern Asia	Å...land Islands
India	Not stated	770	Southern Asia	Southern Asia	Albania
China (excludes SARs and Taiwan)	Australian citizen	121965	Eastern Asia	Eastern Asia	Algeria
China (excludes SARs and Taiwan)	Not an Australian citizen	212279	Eastern Asia	Eastern Asia	American Samoa
China (excludes SARs and Taiwan)	Not stated	623	Eastern Asia	Eastern Asia	Andorra
England	Australian citizen	182466	Northern Europe	Northern Europe	Angola
England	Not an Australian citizen	94398	Northern Europe	Northern Europe	Anguilla
England	Not stated	664	Northern Europe	Northern Europe	Antarctica
Philippines	Australian citizen	114261	South-eastern Asia	South-eastern Asia	Antigua and Barbuda
Philippines	Not an Australian citizen	52620	South-eastern Asia	South-eastern Asia	Argentina
Philippines	Not stated	482	South-eastern Asia	South-eastern Asia	Armenia
South Africa	Australian citizen	86989	Sub-Saharan Africa	Sub-Saharan Africa	Aruba
South Africa	Not an Australian citizen	30895	Sub-Saharan Africa	Sub-Saharan Africa	Australia
South Africa	Not stated	342	Sub-Saharan Africa	Sub-Saharan Africa	Austria
Vietnam	Australian citizen	49782	South-eastern Asia	South-eastern Asia	Azerbaijan
Vietnam	Not an Australian citizen	32349	South-eastern Asia	South-eastern Asia	Bahamas
Vietnam	Not stated	275	South-eastern Asia	South-eastern Asia	Bahrain
Australia External Territories	Australian citizen	54255	Australia and New Zeala	Australia and New Zealand	Bangladesh
Australia External Territories	Not an Australian citizen	21403	Australia and New Zeala	Australia and New Zealand	Barbados
Australia External Territories	Not stated	270	Australia and New Zeala	Australia and New Zealand	Belarus
Iraq	Australian citizen	40127	Western Asia	Western Asia	Belgium
Iraq	Not an Australian citizen	32302	Western Asia	Western Asia	Belize
Iraq	Not stated	261	Western Asia	Western Asia	Benin
Malaysia	Australian citizen	27167	South-eastern Asia	South-eastern Asia	Bermuda
Malaysia	Not an Australian citizen	41928	South-eastern Asia	South-eastern Asia	Bhutan
Malaysia	Not stated	99	South-eastern Asia	South-eastern Asia	Bolivia (Plurinational State of)
Sri Lanka	Australian citizen	47995	Southern Asia	Southern Asia	Bonaire, Sint Eustatius and Saba
Sri Lanka	Not an Australian citizen	19555	Southern Asia	Southern Asia	Bosnia and Herzegovina
Sri Lanka	Not stated	156	Southern Asia	Southern Asia	Botswana

Figure 6: Region mapping each country for Table 5 dataset.

The resulted datasets conform the following structure:

Variable names	Types	Description	Values
For Table 5			
Region of Birth	String (Categorical)	Geological Region of their origin.	"South-eastern Asia", etc
Visa Stream	String (Categorical)	Visa Stream category where a migrant population yielded their permanent resident	"Skilled", "Humanitarian", "Family" or "Other permanent"
Permanent Resident (value)	Integer	Number of migrants that falls under specific category of Region of Birth, within specific Visa Stream	200134, 1232004, 23340
For Table 2			
State and Territory	String (Categorical)	Destination State and Territory in Australia where PRs migrated to	"Victoria", "New South Wales", "Northern Territory", etc
Visa Stream	String (Categorical)	Visa Stream category where a migrant population yielded their permanent resident	"Skilled", "Humanitarian", "Family" or "Other permanent"
Permanent Resident (value)	Integer	Number of migrants that falls under specific category of State and Territory, within specific Visa Stream	200134, 1232004, 23340

Table 2: Data Dictionary for Migration Pattern cleaned dataset.

Pivoting or Unpivoting (discussed after design lock-in)

A common discussion within the team's data processing process is the decision to whether pivot or unpivot a dataset. While a common computing machine would prefer processing in "unpivoted" and record-based data, which is necessary to our visualization process as we will be developing a data visualization powered by D3.js. However, within D3 process, we also understand how the program interprets each data records that will be passed in: "as JavaScript data objects". A data object should yield multiple attributes, some of each will serve the purpose of identifying the object, others should hold necessary values. As the processed data will be stored in a CSV format, D3.js will inject the data, row by row, into an array of record objects.

Within this reason, it is necessary to pivot the data to ensure that D3 will be able to retrieve necessary data correctly. However, network-based visualisations such as a network map, parallel graph or Sankey chart, where each visualized components are records of nodes and links, unpivoted datasets help maintain consistency in determining which columns are “target”, which is” source”, and what is the defined “value” between those two. Conclusively, as data processing for data visualization comes to an “on-demand” standpoint, it is necessary that the team takes different approaches to different scenarios.

3. Requirement

3.1. Must-Have Feature

To help reader conveniently understand the data visualisation, the following must-have feature are included:

1. **Interactive visualization: Interactive elements can greatly enhance the user experience and understanding of data visualisation.**
    a. Hover effect:
        i. For Spider chart, hover effect was added to show detailed data, and explanation on Axis label.
        ii. For Sankey chart, hover effect was added to highlight possible pathways and show detailed data
    b. Drill-Down:
        i. For Spider Chart, different description box tabs were shown and display How-to-read section and summary/background section. Also, Detailed data table was displayed.
    c. Selectors:
        i. For Spider Chart, select button were added to allow user to choose which country they want to compare the data with Australia.
2. **Additional information: Providing sufficient context is key to understanding a visualization.**
    a. Additional label:
        i. In spider chart, additional label is displayed as a grouping for types of hazards.
        ii. In Sankey chart, additional label is displayed to indicate the direction of flow.
3. **Clear data label:** Proper labeling ensures users understand what the data represents.
4. **Colour contrast:** Different colours are used to represent different data points for readers’ optical convenience.

3.2. Optional feature

1. **Tooltip implementation:** Tooltips provide immediate context and prevent the need to remember information. When a user hovers over a specific data point or object in the visualization, a tooltip can provide additional information. While unable to implement this functionality, we are still able to provide necessary alternative that serves similar purposes.
2. **How to Read section:** With a complex visualization, a necessary "manual" should be provided to increase user-friendliness.

4. Visualisation Design

We approach our visualisation designs within four stages, each for both visualisation development. They are:

- **Conceptualise:** where we generate and brainstorm our rough sketches for the visualisations and decide on the visualisation medium.
- **Visualise:** where we provide a more completed look for the conceived sketches
- **Materialise:** where we decide on the styling, colouring and elements of the final design
- **Implementation:** where we apply the design onto d3.js programs and propose iterative changes.

4.1. Conceptualise (Design Iteration) stage:

Within the first early stage of the visualization design process, we aim to provide a simple, yet interactive layout that would suit our targeted audiences as general public. While working with categorized ordered data, we take different approach to different visualisations, while keeping them consistent to ensure a streamlined experience for the users.

For the first visualisation: Migration Pattern, we draw rough sketch to show our ideas about what we want to show the reader, without looking thorough on the data we have found. Since we aim to visualise the flow of immigrants to Australia via different visa streams, we focused on "Network" type visualisations, and consider our options between parallel chart and Sankey charts.

In the figure below, the sketch illustrates our attempt to implement a "Sankey Chart", which is most effective for use to visualise categorised ordered data that suggests flows and proportions. is used to illustrate the pattern of people who migrate to Australia, while also showing their visa categories.

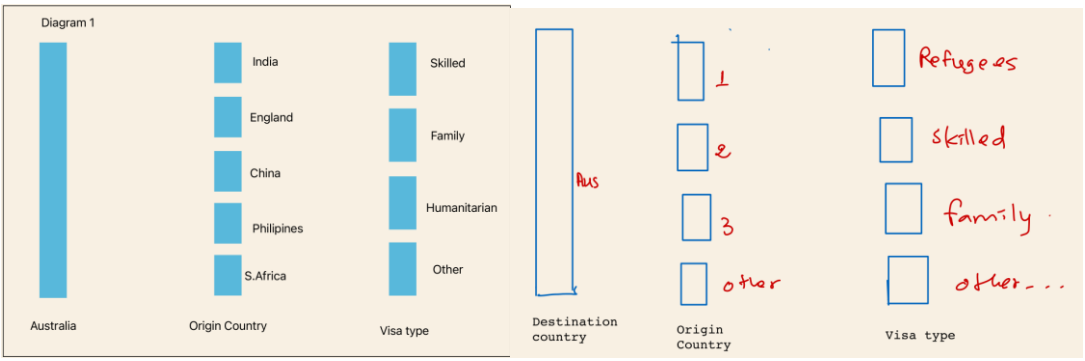


Figure 7: Rough sketches towards "Sankey diagram" visualisation

While we have not yet decided on the idea of which data category to be visualised, we have a clear view of showing the user a “flow” experience while interacting with the viz. One of the better candidates for this specific design is “parallel chart”, which highlights the exact pathway of which a record coming from an origin could follow through their destinations.

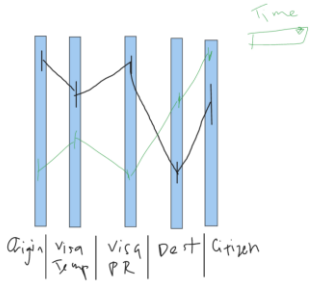


Figure 8: Another iteration of design using "parallel chart" to show "pathways" of migrations

Different approaches were also considered, which in turns create a different experience to the users. The Circular packing sketch below were conceptualized to show the proportion of migrant

population, which could be coupled with the “Dendogram” visualization to represent flow of migration from and to places.

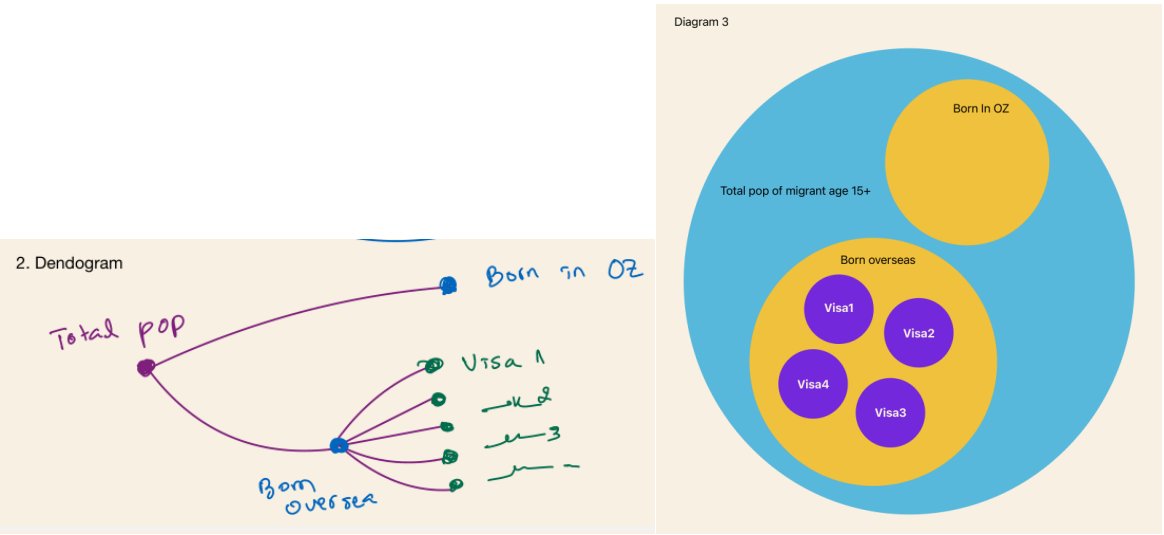


Figure 9: Dendogram & Circular Packing alternatives that could have a different take to our approach

However, by the end of the conceptualization stage, we have decided to implement Sankey charts, due to its duality characteristics as it both represent proportion and flow within one diagram, as well as better visibility due to the simplicity of its core design.

For Internal Displacement dataset, we have fixated with the idea of representing such data on a Radar chart. While there are some controversial takes arose within the modern usage of data visualization towards Radar chart due to its “harder to read” 2D multi-category plane (Holtz 2023) and the possible differences that axis positioning could implied to the meaning of such visualization, we believe that a Radar/Spider chart is the best possible way to visualize a “profile” of a country’s IDP population. Here, we take advantage of the influential axis-positioning characteristic to attach different shapes of the radar path to different meanings.

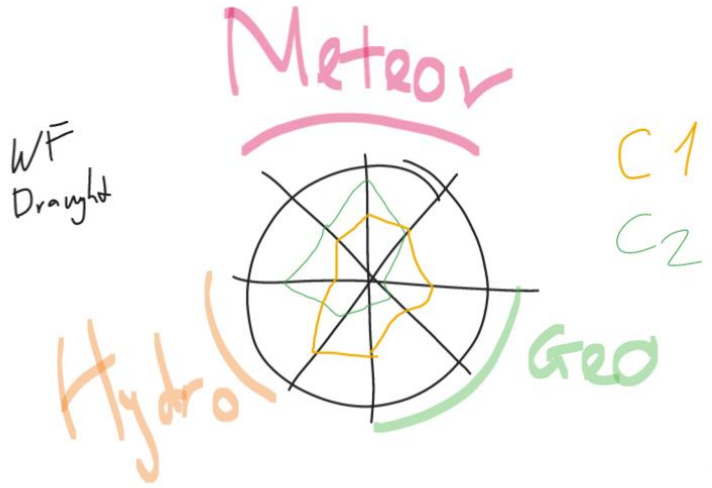


Figure 10: Iterative rough sketch on positioning "categories" that allows a meaningful profile to be interpreted

The sketch above shows how positioning two Hazard Types of a same category next to one another, within a set of six types, the visualization could out-right depict a necessary trend of a specific country's profile. For example, if we look at the green paths above, we could interpret how this profile leans toward the category "Meteor" more than it does toward other categories. While no data is provided, the user could immediately interpret our purpose within matters of seconds. Furthermore, with correct usage of hues, different paths could be differentiated and immediately compared, as they sit on top of one another, perfectly serves our purpose of visualizing the IDP data between Australia and other countries.

4.2. Visualise Stage

Upon locking into our design choices, we now generate a set of higher in fidelity designs that helps us shape the idea of these visualisations in context. For the Migration Pattern visualization, we are interested in seeing a justified, centered look where different sections represents different field of categories, instead of solely visualizing them as “nodes” and “links”.

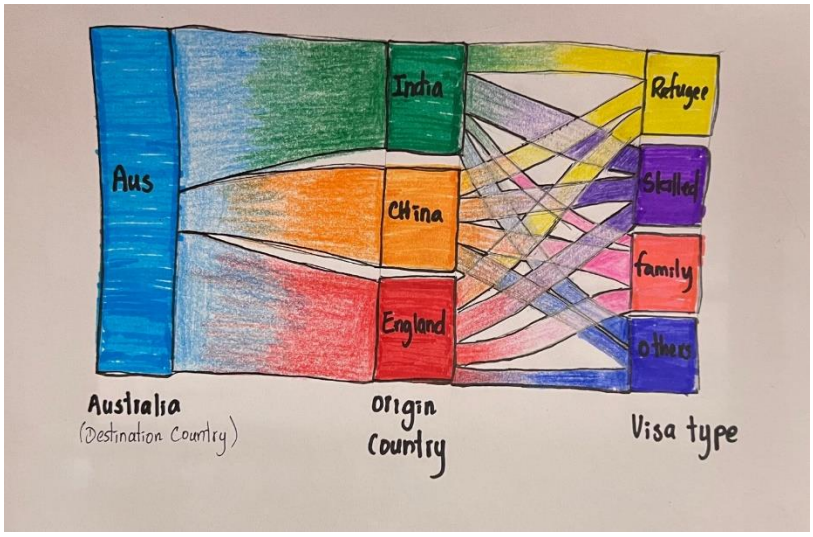


Figure 11: A colored and more visualisable sketch for Migration Pattern viz.

For the Internal Displacement viz, while processing the dataset, we have noticed the difficulty in positioning different numbers of axes where the IDP profile could be presented in the most accurate, yet non-redundant way. We have figured out that by visualizing six most impactful Hazards types, countries’ IDP profiles could be represented best, as well as it aligned with the category positioning discussed earlier.

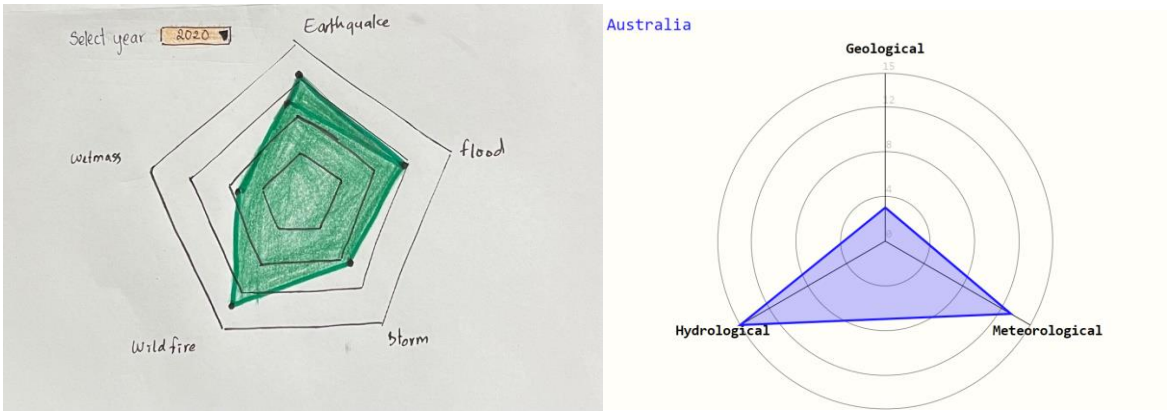


Figure 12: Deciding whether circular or penta/hexagonal designs are more approrpiate

4.3. Materialise Stage:

Upon materialization stage, we started to implement appropriate colors that could represent data variation best, and not to confuse the users. For Migration Pattern, while handling with more than ten categories caused considerable challenges to our design, we have decided to implement D3.js default coloring schemes which would be of most appropriate for a visualization this complex.

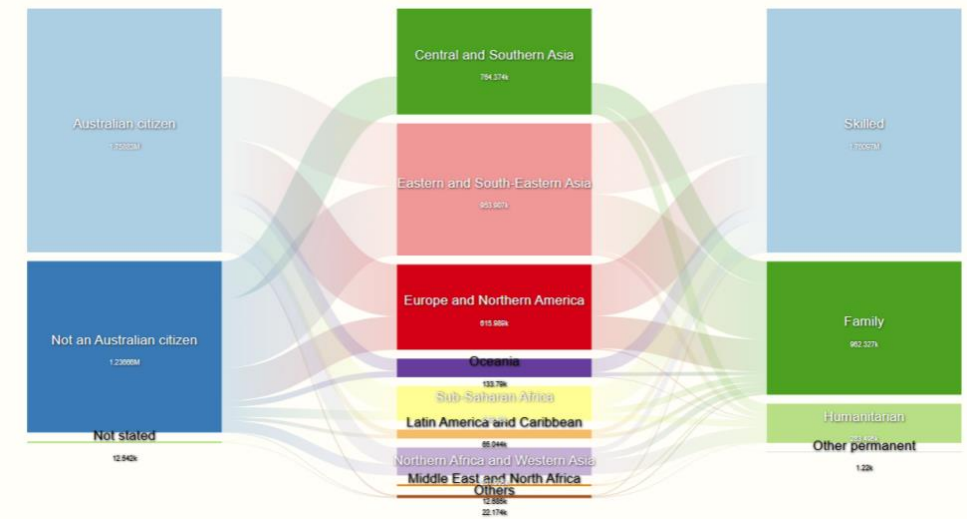


Figure 13: Digitised "sketch" to determine layout and coloring for each Sankey node

For Internal Displacement dataset, we have decided to follow a circular radar design, with two contrasting hue "red" and "green" which helps differentiate between two IDP profiles in comparison.

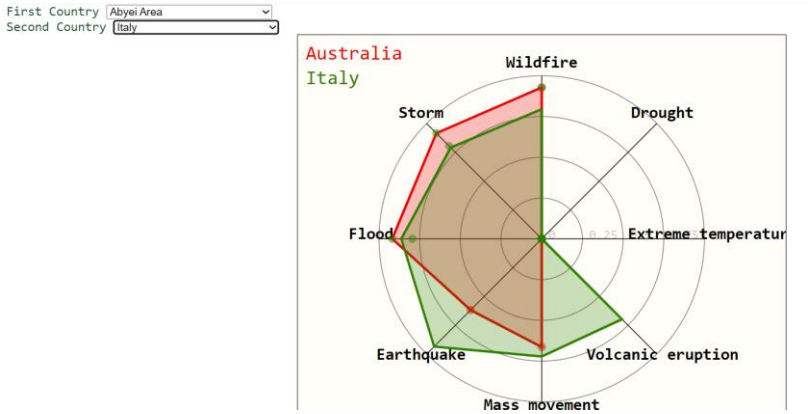


Figure 14: Testing out hue contrasts between red, green differentiation.

4.4. Implementation Stage:

Upon finalising the styling and providing additional tweaks and turns, we have successfully implemented the basic structures (so-called “skeletons”) of the intended visualisations. Within implementation stage, we aim to finalise the set out functional requirements that would serve as “interactive features” for the visualisation to communicate its meanings.

Firstly, as we planned to provide pop-up tooltip on hover feature, we conceptualised how it would be an appropriate way to provide “detail-on-demand” without losing the user’s attention (having to look away or at another position to view details.)

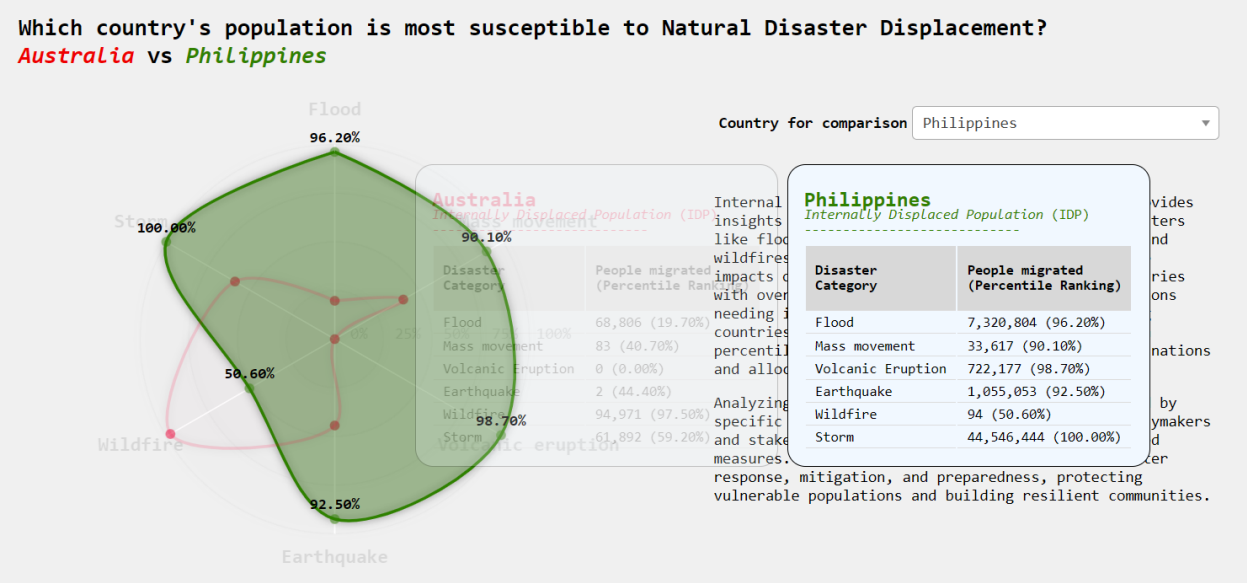


Figure 15: Implementation of a now-deprecated tooltip to show details.

The hovering tooltip “detail table” was soon replaced by a distinct table within the details tab, which now allows the user to view both normalised and raw data at an instance, without crowding the main visualisation areas

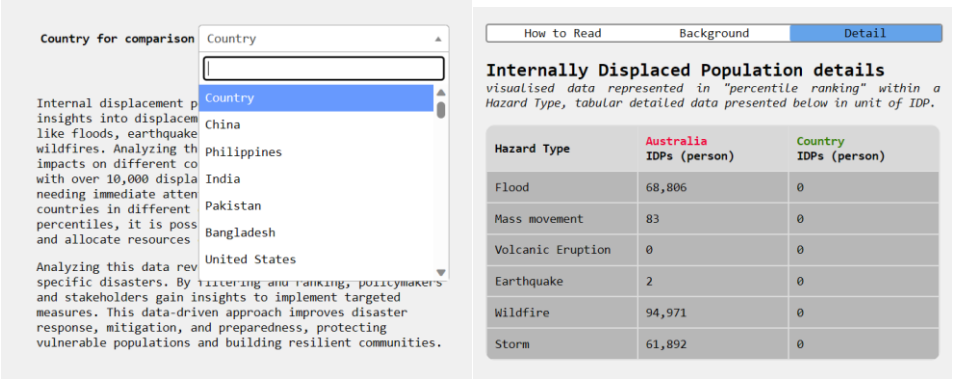


Figure 16: Search bar and table of details to replace previously implemented "tooltip."

For Migration Pattern, as discussed above, we have implemented “Tableau 10” color scheme which resulted in a high visibility design for the Sankey chart. Sankey links are also highlighted on-hover, which allows necessary filtering and focusing for the user, where they could use their cursor to follow set out paths to better understand the relationships between the “nodes”.

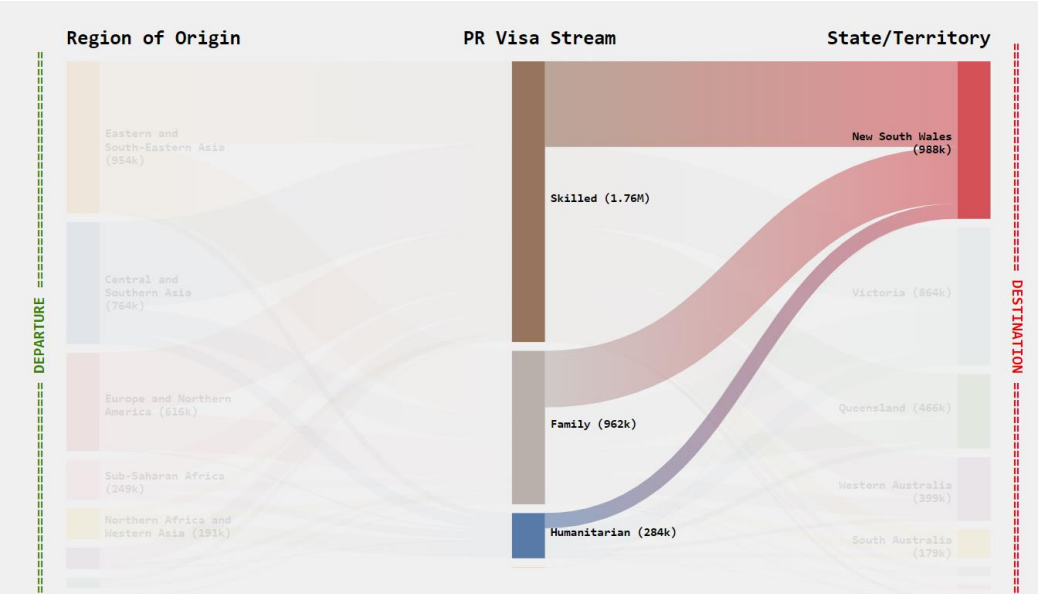


Figure 17: Color choice made for maximized visibility, with hover actions implemented.

4.5. Design Principles Adherence

At the final stage of design, we would also aim to present specific design principles that were implemented at the final iterations of the Data Visualisations.

Clear colour utilisation

Specifically, the IDP Profile visualisation depicts clear and consistent use of specific hue that embeds data into its appearance. For example, **#ee0a3a** and **#4e9759** were specifically and consistently used to represent Australia and the compared country, respectively.

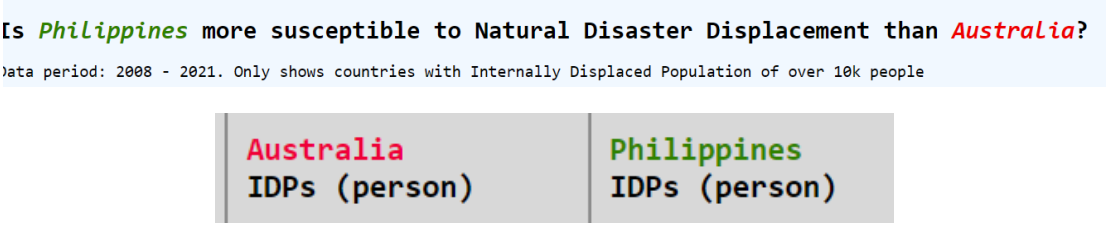


Figure 18: Colors were used consistently throughout the layout of the visualization

A similar practice could also be seen with “Hazard Categories” representation:

Hazard Type	Australia IDPs (person)	Philippines IDPs (person)
Flood	68,806	7,320,804
Mass movement	83	33,617
Volcanic Eruption	0	722,177
Earthquake	2	1,055,053
Wildfire	94,971	94
Storm	61,892	44,546,444

Figure 19: Similar consistent color practice was also seen in Hazard Category

Reachable on-demand details that provide comparison.

In order to avoid “Change Blindness” as well as to provide details “on-demand” (Munzner 2014), visible legends and labels, as well as the comparative interaction could be accessed via on-hover functionalities, which directly depicts the necessary data variation without having to change its views.

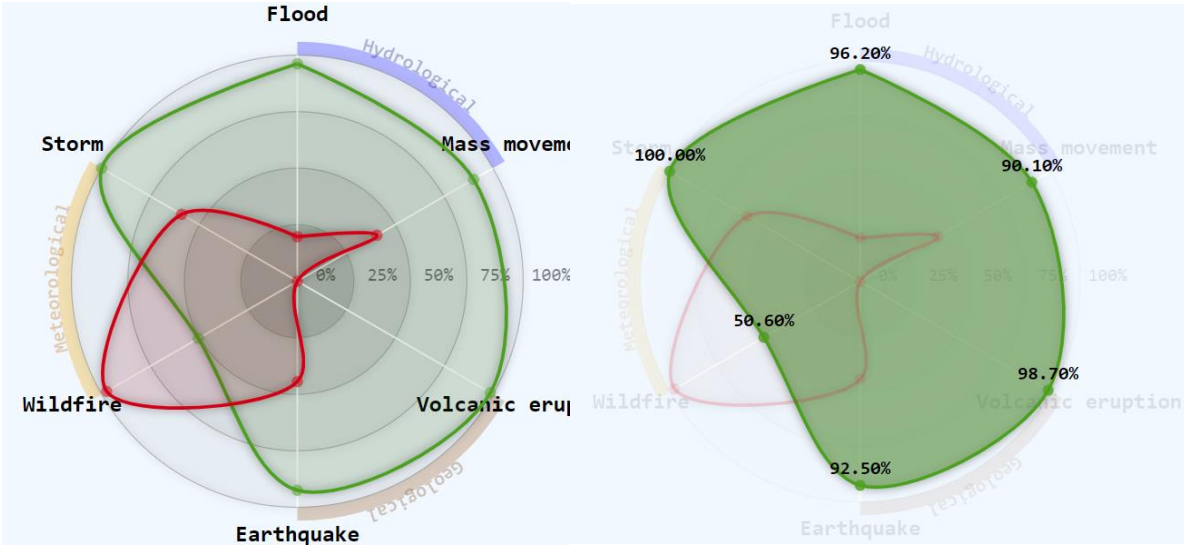


Figure 20: In-viz change and hover functions to avoid change-blindness

Appropriate Data Proportion

The representation of values, as physically measured on the surface of the graphic itself, is directly proportional to the numerical quantities represented, relative to other elements.

Multi-functionality that reduces non-data ink

The Sankey chart provides multi-functionality that both represents proportion of data as well as data relationships (links).

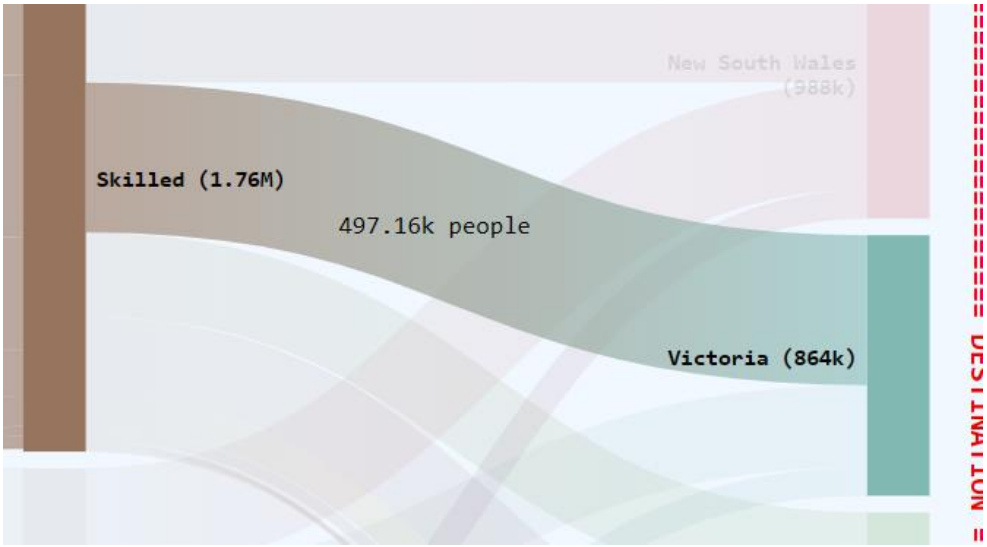


Figure 21: Data proportion was presented alongsides with data relationship

5. Final design

The final iteration of our data visualization set is presented below, which could also be accessed via provided mercury link, or https://s4ppyh4t.github.io/test-repo/

5.1. Home Page:

Figure 22: Home Page Final

5.2. Internal Displacement Page:

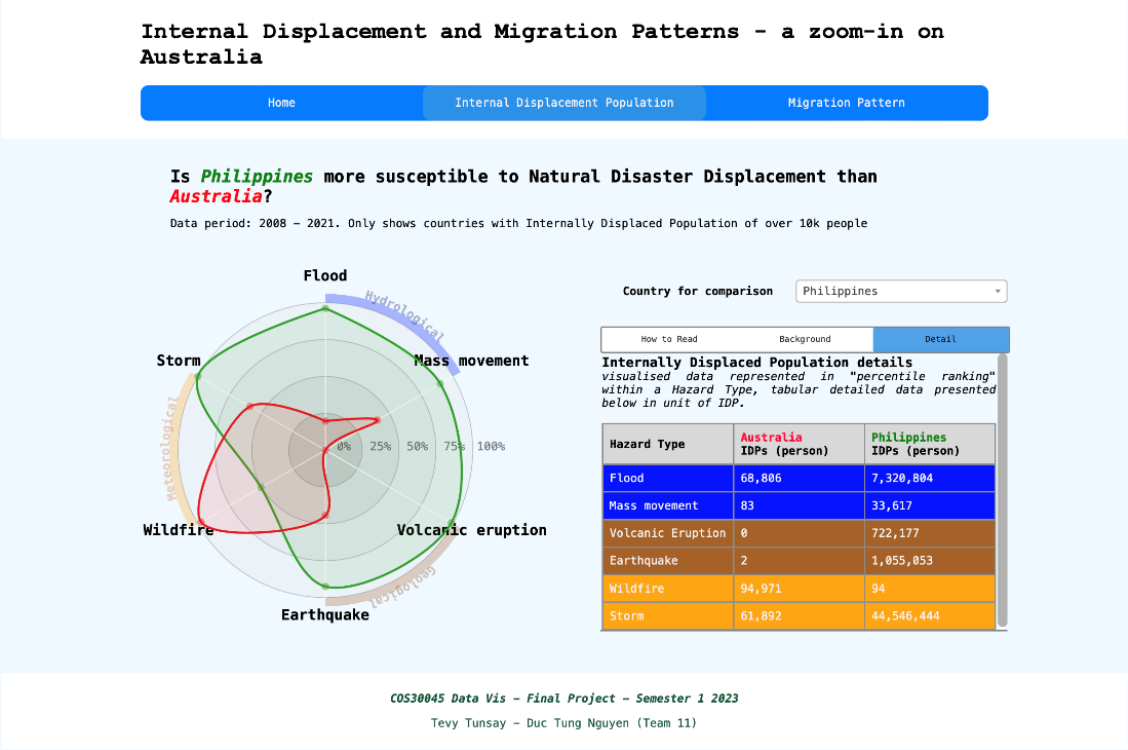


Figure 23: Internal Displacement - Details view

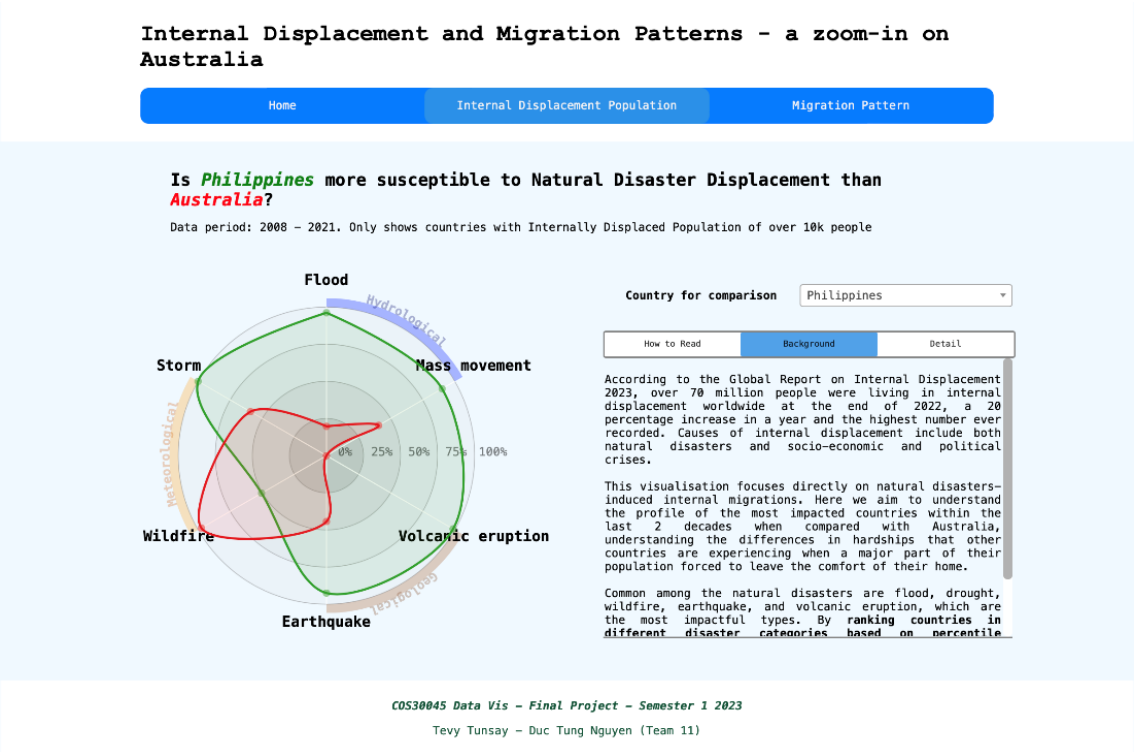


Figure 24: Internal Displacement - Background View

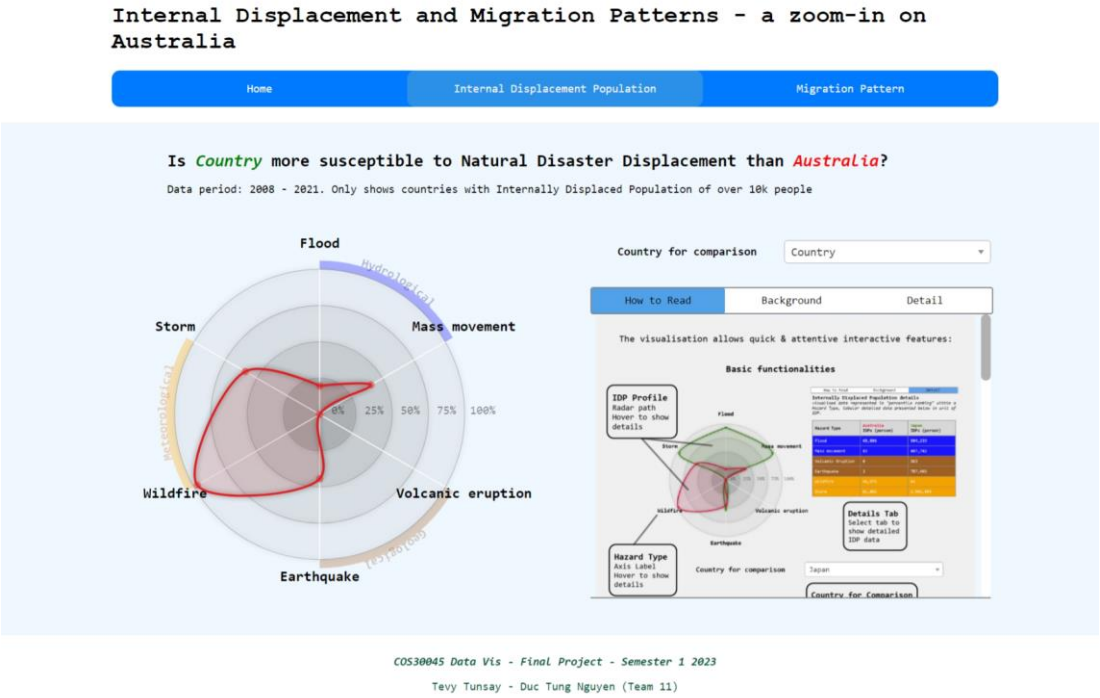


Figure 25: Internal Displacement - How to Read view

5.3. Migration Patterns Page

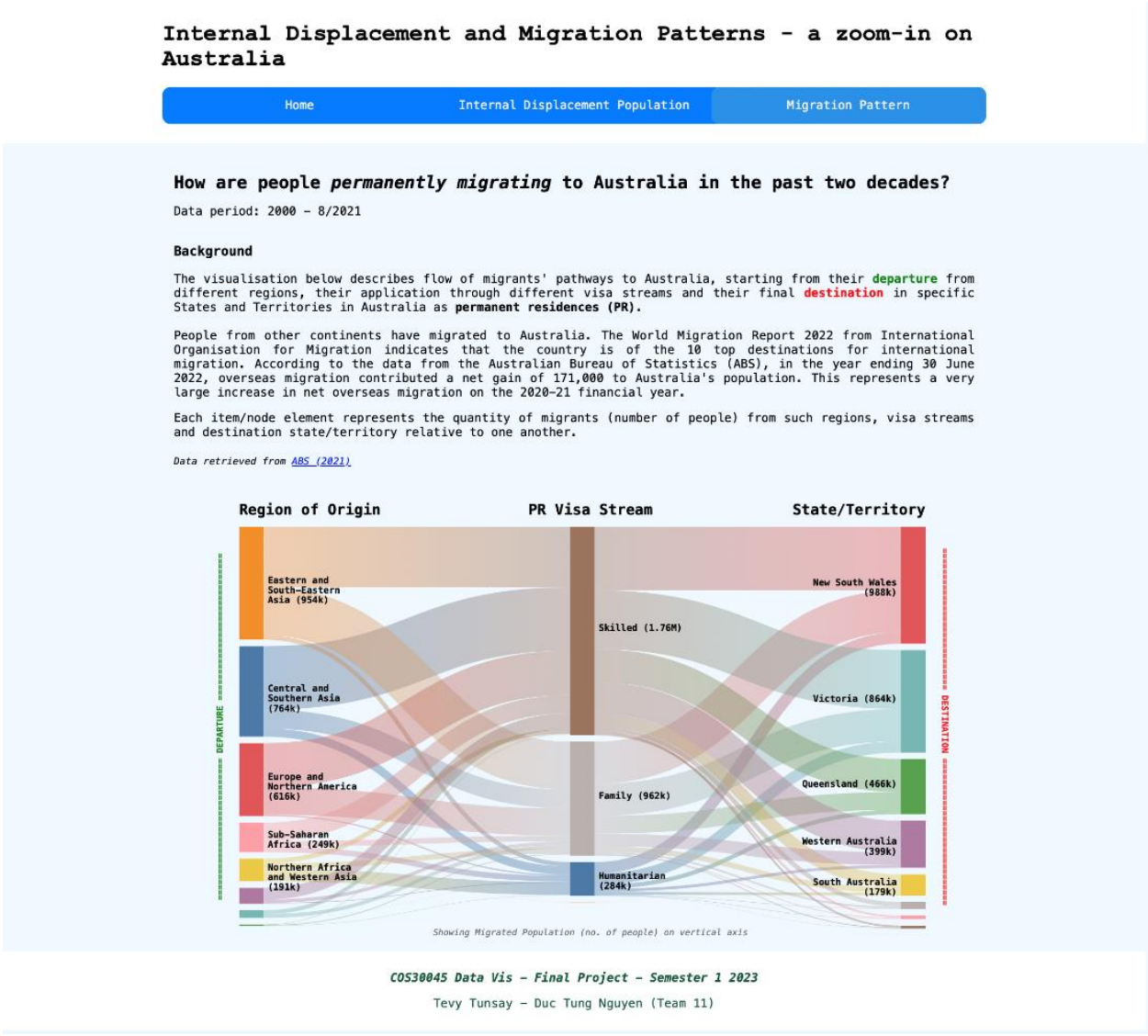


Figure 26: Migration Pattern Final

6. Validation

This section is about Usability Testing to ensure that the visualization effectively communicates information, is user-friendly, and meets the needs of the intended audience. For this purpose, 5 people were selected to be the testers of the visualization.

6.1. Target criteria value

Task	Unassisted Task Completion Rate	Errors per user	Task Completion Time (seconds)	SUS Score
Task 1: Find the highest natural disaster type leading to displacement in Australia	95%	Less than 1	10 second	
Task 2: Find whether Australia more susceptible to hydrological natural disasters than Viet Nam	90%	Less than 2	15 second	
Task 3: Find how many people in Nigeria had to displace/migrate internally due to Flood	95%	Less than 1	15 second	
Task 4: Find what do the percentages represent in this visualisation	90%	Less than 3	30 second	
Task 5: Find the definition of *Mass movement* hazard type defined in this visualisation	85%	Less than 3	30 second	
Task 6: Find how many Skilled Visa stream migrants moved to Victoria	95%	Less than 1	10 second	
Task 7: Find which region have the most people migrated to Australia	95%	Less than 1	10 second	
Task 8: Find what is the most PR Visa type people held when migrating to Australia	95%	Less than 1	10 second	
Task 9: Find what state in Australia received the most migrants	95%	Less than 1	10 second	
Task 10: Find If more people from Central & Southern Asia migrate via Skilled than Family Visas.	95%	Less than 2	10 second	
Total			**2minutes 30 seconds**	**75%***

* Note: SUS score for entire interface (not available for individual tasks)

6.2. Participant characteristics

Participants	Comfortable in navigating Web App	Use/interact with data visualisation in career	Comfortable in interpreting data	Tool used for data visualisation	Education/Training on data visualisation
12	Somewhat comfortable	No	Somewhat comfortable	Tableu;PowerBI;R;	Yes, formal education
P2	Very comfortable	Yes	Somewhat comfortable	Excel;	Yes, online training
123	Very comfortable	Yes	Very comfortable	Excel;Tableu;	Yes, formal education
Tester3	Very comfortable	Yes	Somewhat comfortable	Excel;PowerBI;	Yes, online training
1604	Neither comfortable nor uncomfortable	Yes	Somewhat uncomfortable	Python;Excel;	Yes, online training

6.3. Methods

The participants were put to the test via video conferencing. However, this test is not done using video recording, but we, the investigators, take notes on the whole session, on the seconds they took on each task, the error they made, etc.

They were told to make every task complete without assistance. Every issue encountered was the result of the visualization, not them and they were also instructed to think aloud

Testers were given inform consent and other forms of questionaries prior to the meeting. After they completed the informed consent, they would show their screen while doing the task and complete questionaries after that.

6.4. Performance Data

Task 1: Find the highest natural disaster type leading to displacement in Australia

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	10

P2	100%	0	10
123	100%	0	7
Tester3	100%	0	5
1604	100%	0	5

Task 2: Find whether Australia more susceptible to hydrological natural disasters than Viet Nam

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	10
P2	0	1	10
123	100%	0	7
Tester3	100%	0	5
1604	100%	0	5

Task 3: Find how many people in Nigeria had to displace/migrate internally due to Flood

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	10
P2	100%	1	15
123	100%	0	7
Tester3	100%	0	5
1604	100%	0	5

Task 4: Find what do the percentages represent in this visualisation

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	15
P2	100%	0	30
123	100%	0	10
Tester3	100%	0	10
1604	100%	0	5

Task 5: Find the definition of *Mass movement* hazard type defined in this visualisation

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	33
P2	100%	0	38
123	100%	0	27
Tester3	100%	0	23
1604	100%	0	25

Task 6: Find how many Skilled Visa stream migrants moved to Victoria

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	25
P2	100%	1	40
123	100%	0	30
Tester3	100%	0	15
1604	100%	0	25

Task 7: Find which region has the most people migrated to Australia

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	7
P2	100%	0	11
123	100%	0	10
Tester3	100%	0	9
1604	100%	0	6

Task 8: Find what is the most PR Visa type people held when migrating to Australia

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	10

P2	100%	0	12
123	100%	0	10
Tester3	100%	0	10
1604	100%	0	7

Task 9: Find what state in Australia received the most migrants

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	10
P2	100%	0	9
123	100%	0	12
Tester3	100%	0	12
1604	100%	0	7

Task 10: Find If more people from Central & Southern Asia migrate via Skilled than Family Visas.

Participants	Unassisted Task Completion Rate in %	Errors	Task completion Time (seconds)
12	100%	0	10
P2	100%	0	13
123	100%	0	14
Tester3	100%	0	11
1604	100%	0	8

6.5. Attitude Data

Based on the SUS calculator (StaurtAffect, 2020), the SUS score given by tester, the lowest is 50 % and the highest is 92.5%. However, the average of the SUS is 71.5%, which was below the aimed score of 75%.

Participants	Task-1 Difficulty Rating	Task-2 Difficulty Rating	Task-3 Difficulty Rating	Task-4 Difficulty Rating	Task-5 Difficulty Rating	Task-6 Difficulty Rating	Task-7 Difficulty Rating	Task-8 Difficulty Rating	Task-9 Difficulty Rating	Task-10 Difficulty Rating	SUS Score
12	3	4	3	4	5	4	1	1	1	1	55
P2	1	1	1	3	4	2	1	1	1	1	92.5
123	2	1	2	2	3	1	1	2	1	2	77.5
Tester 3	1	2	3	3	3	1	1	1	2	2	82.5
1604	3	3	3	3	3	3	3	3	3	3	50

7. Conclusion

Migration, especially cross-country one, has become very common in our globalized world. Both push and pull factors are behind the trend. Among them, internal displacement due to natural disasters and the desire for better economic opportunities top the list.

To illustrate the above phenomena, we extracted, processed, analyzed and visualized two data sets from 1/ Internal Displacement, and 2/ Migration patterns into Australia.

We chose 2 different types of diagrams. Spider Chart is used to show and compare the different causes of internal displacement and Sankey Chat is used to show the places of origin, types of visas and destinations of migration to Australia. The data were plotted using the D3.js library of JavaScript. For both visualisations, the must-have feature includes interactive visualisation, additional background info, a clear data label, and contrasting color of the different data points.

Finally, usability testing was conducted to ensure that the visualization effectively communicates information, is user-friendly, and meets the needs of the intended audience.

8. References

ABS 2022, Overseas Migration, viewed 19 May 2023, https://www.abs.gov.au/statistics/people/population/overseas-migration/latest-release#cite-window2.

Harsh 2018, Home / Moving To Australia / Which City Is The Best To Live In Australia?, viewed 10 May 2023, https://www.aussian.com/city-best-live-australia/.

Holtz, Y 2023, The Radar chart and its caveats, Data-to-viz.com, viewed 4 June 2023, https://www.data-to-viz.com/caveat/spider.html.

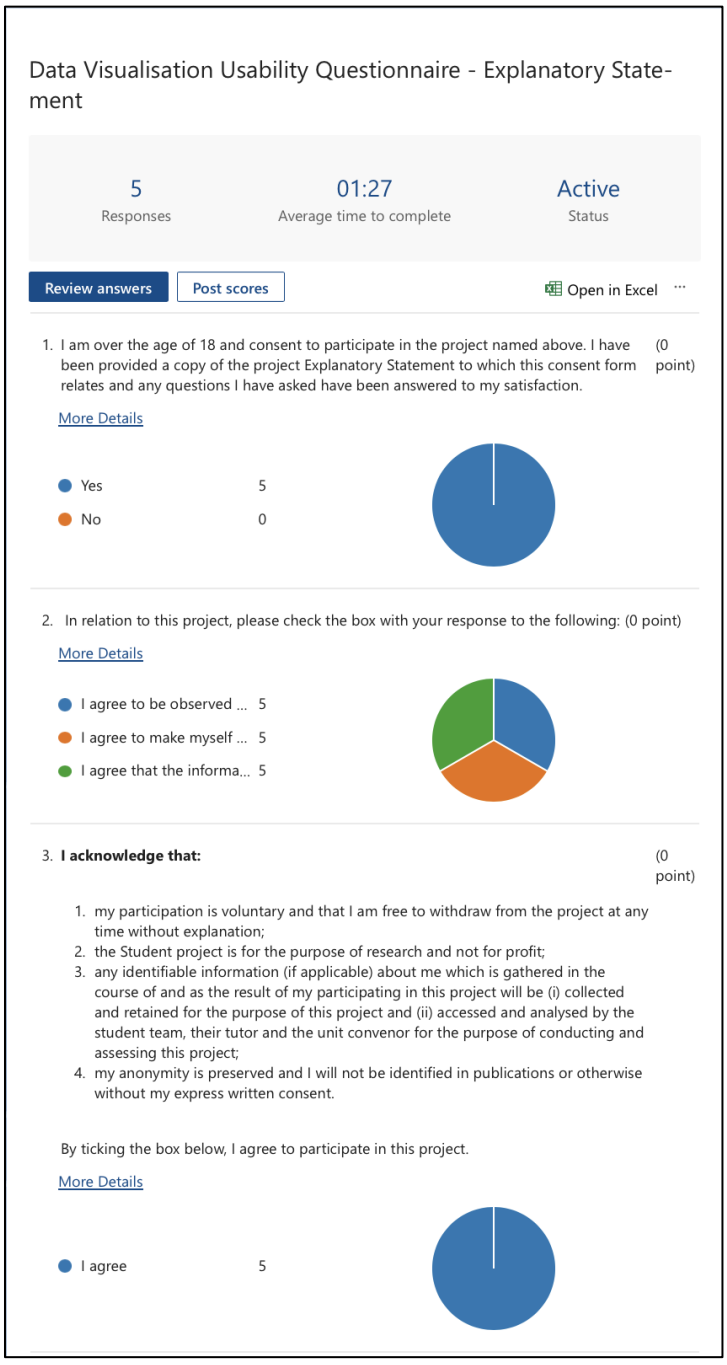
IOM 2021, World Migration Report 2022, Geneva.

Munzner, T 2014, Visualization analysis and design, CRC press.

Vaghri, Z., Tessier, Z. & Whalen, C. 2019, Refugee and Asylum-Seeking Children: Interrupted Child Development and Unfulfilled Child Rights, viewed 10 May 2023, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6915556/.

9. Appendix

Inform Consent:



Data Visualisation Usability Questionnaire - Explanatory Statement

5 Responses | 01:27 Average time to complete | Active Status

Review answers | Post scores | Open in Excel ···

1. I am over the age of 18 and consent to participate in the project named above. I have been provided a copy of the project Explanatory Statement to which this consent form relates and any questions I have asked have been answered to my satisfaction. (0 point)

More Details

Yes 5
No 0

2. In relation to this project, please check the box with your response to the following: (0 point)

More Details

I agree to be observed ... 5
I agree to make myself ... 5
I agree that the informa... 5

3. **I acknowledge that:** (0 point)

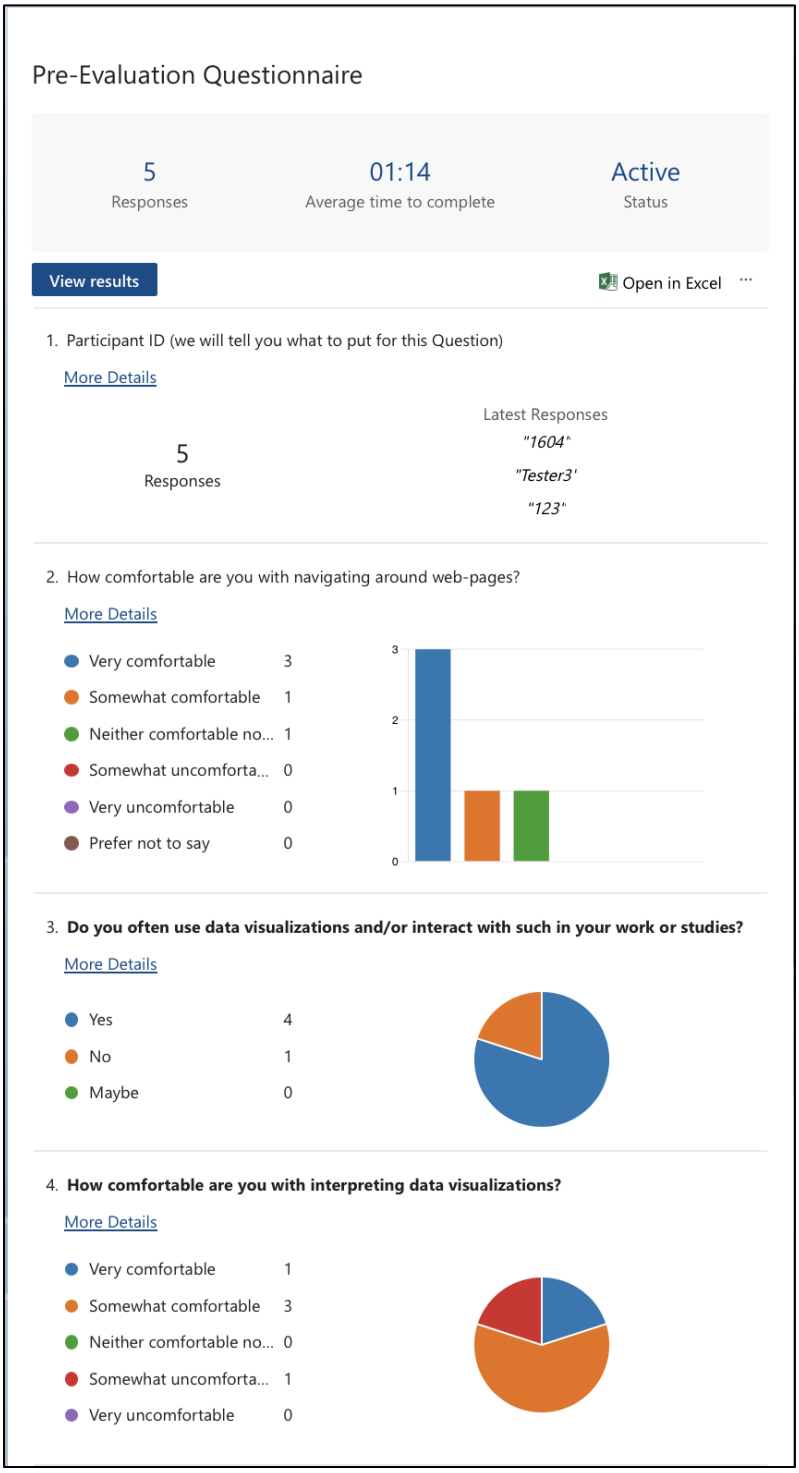
1. my participation is voluntary and that I am free to withdraw from the project at any time without explanation;
2. the Student project is for the purpose of research and not for profit;
3. any identifiable information (if applicable) about me which is gathered in the course of and as the result of my participating in this project will be (i) collected and retained for the purpose of this project and (ii) accessed and analysed by the student team, their tutor and the unit convenor for the purpose of conducting and assessing this project;
4. my anonymity is preserved and I will not be identified in publications or otherwise without my express written consent.

By ticking the box below, I agree to participate in this project.

More Details

I agree 5

Pre-Questionaries



Pre-Evaluation Questionnaire

5 Responses | 01:14 Average time to complete | Active Status

View results | Open in Excel

1. Participant ID (we will tell you what to put for this Question)

More Details

5 Responses

Latest Responses
"1604"
"Tester3"
"123"

2. How comfortable are you with navigating around web-pages?

More Details

Very comfortable	3
Somewhat comfortable	1
Neither comfortable no...	1
Somewhat uncomforta...	0
Very uncomfortable	0
Prefer not to say	0

3. **Do you often use data visualizations and/or interact with such in your work or studies?**

More Details

Yes	4
No	1
Maybe	0

4. **How comfortable are you with interpreting data visualizations?**

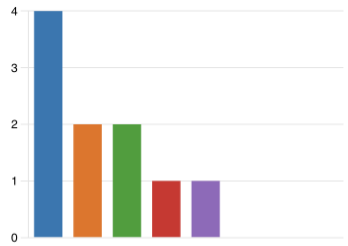
More Details

Very comfortable	1
Somewhat comfortable	3
Neither comfortable no...	0
Somewhat uncomforta...	1
Very uncomfortable	0

5. **Which data visualization tools do you use regularly?** (Check all that apply)

More Details

Excel	4
Tableu	2
PowerBI	2
R	1
Python	1
D3.js	0
Google Data Studio	0
Other	0

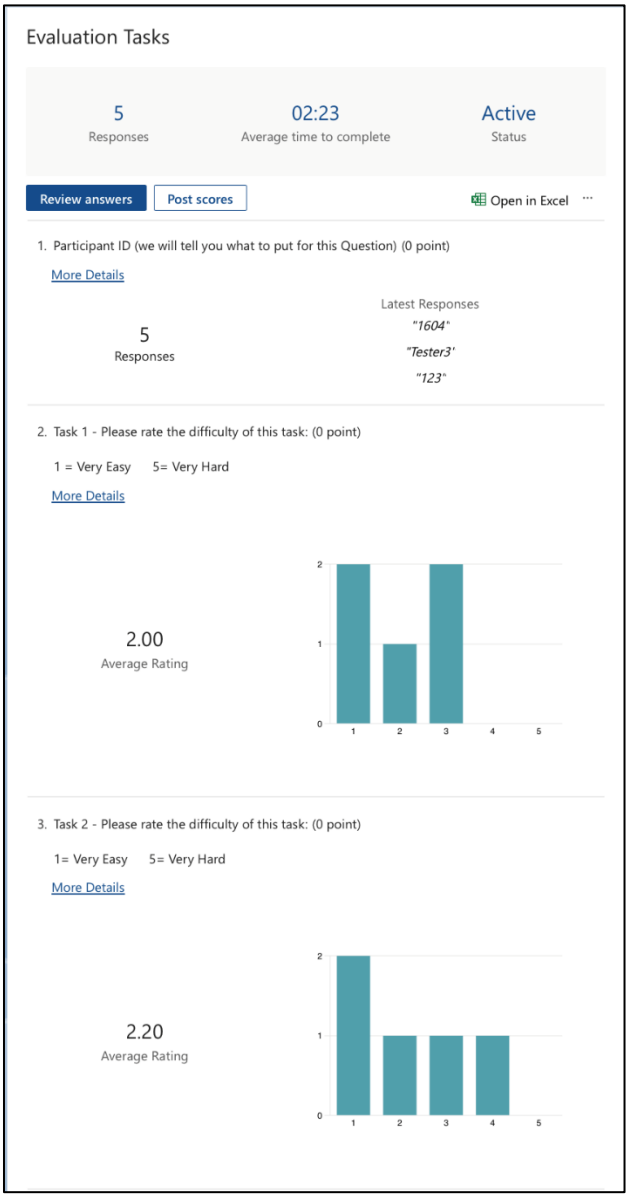


6. **Have you ever completed any formal education or training on data visualization?**

More Details

Yes, formal education	2
Yes, online training	3
Yes, workplace training	0
No	0
Other	0

Evaluation Task:

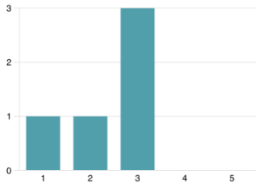


4. Task 3 - Please rate the difficulty of this task: (0 point)

1 = Very Easy 5 = Very Hard

More Details

2.40
Average Rating

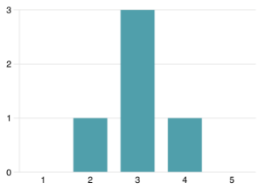


5. Task 4 - Please rate the difficulty of this task: (0 point)

1 = Very Easy 5 = Very Hard

More Details

3.00
Average Rating

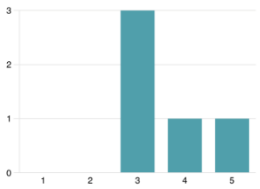


6. Task 5 - Please rate the difficulty of this task: (0 point)

1 = Very Easy 5 = Very Hard

More Details

3.60
Average Rating

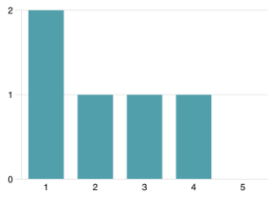


7. Task 6 - Please rate the difficulty of this task: (0 point)

1 = Very Easy 2= Very Hard

More Details

2.20
Average Rating

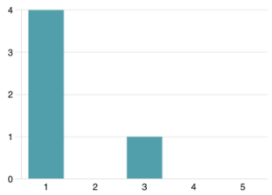


8. Task 7 - Please rate the difficulty of this task: (0 point)

1 = Very Easy 2= Very Hard

More Details

1.40
Average Rating

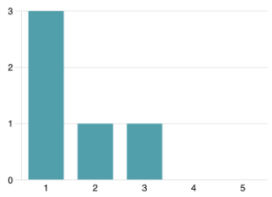


9. Task 8 - Please rate the difficulty of this task: (0 point)

1 = Very Easy 2= Very Hard

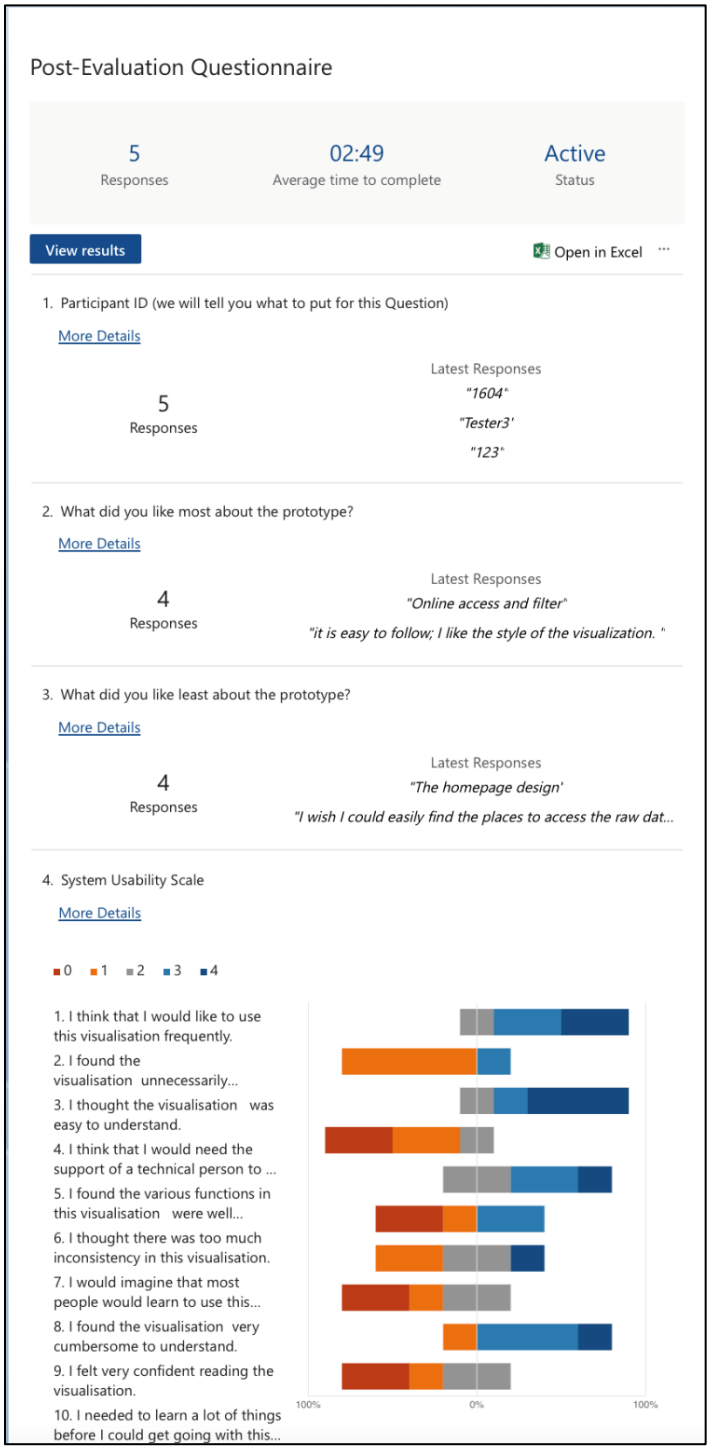
More Details

1.60
Average Rating




10. Task 9 - Please rate the difficulty of this task: (0 point)
1 = Very Easy 2= Very Hard
More Details
1.60
Average Rating
3
2
1
0
1
2
3
4
5
11. Task 10 - Please rate the difficulty of this task: (0 point)
1 = Very Easy 2= Very Hard
More Details
1.80
Average Rating
2
1
0
1
2
3
4
5

Post-Questionaries:



Post-Evaluation Questionnaire

5 Responses | 02:49 Average time to complete | Active Status

View results | Open in Excel

1. Participant ID (we will tell you what to put for this Question)

More Details

5 Responses

Latest Responses
"1604"
"Tester3'
"123"

2. What did you like most about the prototype?

More Details

4 Responses

Latest Responses
"Online access and filter"
"it is easy to follow; I like the style of the visualization. "

3. What did you like least about the prototype?

More Details

4 Responses

Latest Responses
"The homepage design'
"I wish I could easily find the places to access the raw dat...

4. System Usability Scale

More Details

0 1 2 3 4

1. I think that I would like to use this visualisation frequently.
2. I found the visualisation unnecessarily...
3. I thought the visualisation was easy to understand.
4. I think that I would need the support of a technical person to ...
5. I found the various functions in this visualisation were well...
6. I thought there was too much inconsistency in this visualisation.
7. I would imagine that most people would learn to use this...
8. I found the visualisation very cumbersome to understand.
9. I felt very confident reading the visualisation.
10. I needed to learn a lot of things before I could get going with this...

100% 0% 100%