FIT5147 – Data Exploration and Visualization

Data Visualization Project – Report

Topic: The Influence of Anime

Done by

Harisriguhan Sivakumar

Student ID: 31386709

Course: Master of Data Science

Tutorial: 4

Tutor: Angel Das

Introduction

In our project we are going to create a webpage for the anime fans , our target audience. Our web page will narrate the market share of Anime in various countries, how various factors impact the quality of an anime and the total views as well as total users around the years. The reason behind the narration is to show how much anime as a form of animation has grown outside of its origin country (i.e: Japan) as well as what makes it so special and amazing [1]. It also gives the new fans of the medium some insight on what kind of anime to watch or get heavily invested in. It also tells then the places where Anime is very popular and it has a good market.

Design

When it comes to designing our webpage the first thing we need to do is to brainstorm a bunch of ideas where for our project. These ideas include the graphs and charts that we can use here. If we refer to sheet 1 of our 5 design sheet given in the appendix we can see that we have come up with around 15 different visualizations where we have used a different kinds of graphs in each case to show some variety. Once we have done so, we will now filter out the visualization ideas that we do not need at all. For instance, we do not require the word cloud, Gantt chart and lollipop chart where it was out of context. Then we will categorize the ideas based on what they are conveying. We are categorizing based on 4 criteria which are market share, User demographic, Influence of factors and views/Users. Then we will move on to combine the ideas that are very similar and can be done in one type of graph. Charts like bar, horizontal bar, scatter and pictorial bar can be represented in a simple bar graph. We than see if these answer our questions or not. As it does so, we will move on with the proposed layouts.

We will have 3 potential designs that may end up being our final design. When it comes to the color background we have decided to keep it white as we have decided to use the concept of ‘bleaching’ [2] where a white background is used in order for the users to focus on the visualizations. The color palettes that we will be implementing for the maps will be the Dark2 palette. And for the bar chart and sunburst chart we have decided to use the default colors displayed by plotly. For the line charts we will use the custom colors for the line graph.

Figure 1: An example of 'bleaching' where white is used for us to emphasise on the Ken Kaneki the protagonist of Tokyo Ghoul.

The first design is where we have a map where it shows the market share of each country and next to it is a horizontal bar chart that will show the factor importance. These two chart will be interactive as clicking on them will change the chart next to them which shows the market share of that country based on the factor chosen. Underneath is a 3 tier sunburst chart where we will show the user demographic based on the hierarchy selected in the drop down menu below. A line chart displaying the average score based on the factor chosen is next to it. We will also have a multiline graph where we will show the users as well as the views over the years. Appropriate tooltips and popups that will give more detail about the selected element is given. Here the layout is based on user interactivity. The User will have a lot of interaction in this design but as an exchange to it will bring a lot of complexity to the webpage. Also many users may find it difficult to understand the interactions.

The second design is where we will initially have a map where it shows the market share of each country but it can switch to a bar graph if chosen by the user and a 2 tier sunburst chart where we will show the user demographic based on the hierarchy selected in the radio buttons given above. Here when a user clicks on the country (or a bar in the bar chart), the sunburst chart will show the user demographic of that country. Then in the second row we will have a horizontal bar chart that will show the factor importance. It will be followed by the filters for the Factor Vs Average Score/Market Share graph

that will be aside it. A checkbox to show only the top 10 results will also be placed. We will also have a line + scatter graph where we will show either users or the views over the years. We will also have a slider where the user can choose a year range. Appropriate tooltips and popups that will give more detail about the selected element is given. The reason why the option to navigate between bar and map was given is because there are some users that may not be comfortable with the map. Here the complexity of the layout is not as much as the first layout but it is still considerable. As the visualizations are not too complex the user will be able to understand the message that we are trying to convey.

The third design is where we will have a bar graph and a Factor Vs Average Score/Market Share graph that will be aside it. The Factor Vs Average Score/Market Share graph will have many filters. Here when a user clicks a bar in the bar chart, the Factor Vs Average Score/Market Share graph will show for that country. Underneath is a 3 tier sunburst chart where we will show the user demographic based on the hierarchy selected in the drop down menu below. We will also have a line graph where we will show either users or the views over the years. We will also have a slider where the user can choose a year range. Appropriate tooltips and popups that will give more detail about the selected element is given. Here the main problem is the complexity factor. The graph is too simple for this project even though we are able to convey our message there are very few visualizations.

In the final sheet of out five design sheet we have decided to use our second layout with some minor aesthetic changes. The reason for choosing this layout is because of the following reasons.

- Gives users a lot of filters and viewing options
- Narrative aligns perfectly with what we want to convey
- User are able to understand given done appropriate signposting.
- It has a decent level of complexity.

Implementation

With regards to how we need to pull this off, we need to begin by first importing the libraries we need the libraries that we will be using are as follows.

- **Shiny**: As we are going to do our project in shiny it is obvious that we need to import its library
- **Shinythemes**: A library that is used to assign a theme for our shiny application. Used basically for aesthetic purposes.
- **Tidyverse**: used to access, manipulate and modify datasets
- **Rpart**: is used to create a decision tree model in order for us to determine the importance of the independent variables on the dependent variable.
- **Leaflet**: used to create interactive maps.
- **Maps**: used to get the polygons for the leaflet
- **Ggplot2**: used for simple graphs.
- **Plotly**: used to make interactive graphs.

When it comes to our data we need to do extensive wrangling, more than what we did for the data exploration project as the size of our data [3] is around 4 GB, we need to reduce the size significantly. This is due to the fact that there will be instances in our implementation where we will be transforming our data according to the user inputs.

We will have 5 datasets which require all the data we need for our project. They are as follows.

- **GlobalMktShare.csv**: Has data about the market share of the 37 countries that have significant market share.

location	Location_I	percent	mktshare
ARG	Argentina	1.13	Low Market Share (less than 2% of the Global Market)
AUS	Australia	4.38	Medium Market Share (between 2 to 5% of the Global Market)
BEL	Belgium	1.3	Low Market Share (less than 2% of the Global Market)
BGR	Bulgaria	1.49	Low Market Share (less than 2% of the Global Market)
BRA	Brazil	5.36	High Market Share (more than 5% of the Global Market)
CAN	Canada	6.05	High Market Share (more than 5% of the Global Market)
CHL	Chile	0.98	Low Market Share (less than 2% of the Global Market)
DEU	Germany	5.26	High Market Share (more than 5% of the Global Market)
DNK	Denmark	1.2	Low Market Share (less than 2% of the Global Market)

Figure 2:GlobalMktShare.csv

- **AnimeStatsRefined.csv**: Has data about the details of the anime as well as their market share has 41242 records. Here we have converted all the multivalued attributes to a single value

A	B	C	D	E	F	G	H	I	J	K	L	M
anime_id	title	title_englis	type	source	episodes	status	Duration_i	rating	score	studio	genre	mktshare
1	Cowboy B	Cowboy B	TV	Original	13 - 30	Finished Ai	21 - 30	R - 17+ (vi	8.81	Sunrise	Sci-Fi	0.124304
1	Cowboy B	Cowboy B	TV	Original	13 - 30	Finished Ai	21 - 30	R - 17+ (vi	8.81	Sunrise	Comedy	0.124304
1	Cowboy B	Cowboy B	TV	Original	13 - 30	Finished Ai	21 - 30	R - 17+ (vi	8.81	Sunrise	Action	0.124304
1	Cowboy B	Cowboy B	TV	Original	13 - 30	Finished Ai	21 - 30	R - 17+ (vi	8.81	Sunrise	Space	0.124304
1	Cowboy B	Cowboy B	TV	Original	13 - 30	Finished Ai	21 - 30	R - 17+ (vi	8.81	Sunrise	Adventure	0.124304
1	Cowboy B	Cowboy B	TV	Original	13 - 30	Finished Ai	21 - 30	R - 17+ (vi	8.81	Sunrise	Drama	0.124304
5	Cowboy B	Cowboy B	Movie	Original	0 - 12	Finished Ai	0 - 10	R - 17+ (vi	8.41	Bones	Action	0.038846

Figure 3:AnimeStatsRefined.csv

- **AnimeViews.csv**: Here we will have the number of views in that particular year.

my_start_	views
1990	248
1991	228
1992	359
1993	340
1994	438
1995	678
1996	743
1997	1008
1998	1473

Figure 4:AnimeViews.csv

- **UsersJoin.csv**: Has information about the number of new users that year.

join_date	users
2004	3
2005	38
2006	340
2007	10790
2008	34481

Figure 5:UsersJoin.csv

- **UsersClean.csv**: Contains the user information. It has about 52000 records.

A	B	C	D	E
gender	location	Location_I	age	
Female	PHL	Philippines	18 - 29 years	
Male	JPN	Japan	18 - 29 years	
Female	CAN	Canada	18 - 29 years	
Female	ITA	Italy	13 - 17 years	
Male	RUS	Russia	18 - 29 years	
Female	CAN	Canada	30 + years	
Male	HUN	Hungary	30 + years	
Male	NOR	Norway	18 - 29 years	

Figure 6: UsersClean.csv

When it comes to our actual implementation we will divide it into 3 subsections and they are as follows.

- Map/Bar and sunburst
- Factor Vs Average Score/Market share
- Total Users/Views

Map/Bar and sunburst:

In the first subsection we will be looking at the first row which will be the most difficult and complex part of our visualization. A small mind map of what we are planning to in this subsection is given below.

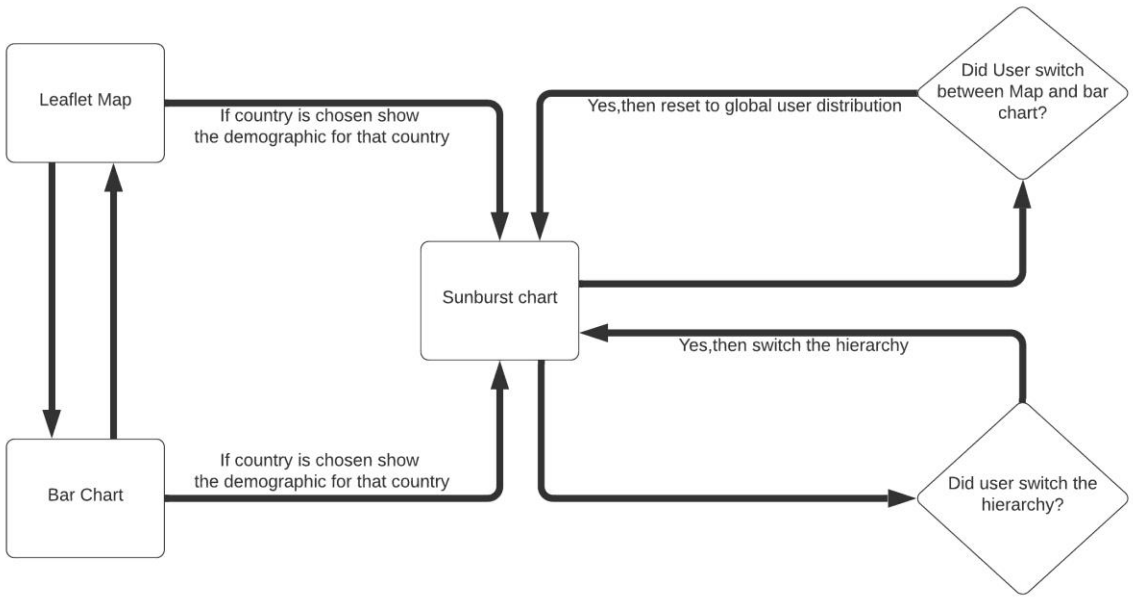


Figure 7: Mind map od Map/bar and Sunburst Implementation

The preliminary step is to first create the leaflet and bar chart. Creating the leaflet was doable but the main challenge was making the bar chart. The main reason behind this is the fact that we need to send the name of the country selected when we click on the bar or map. Thanks to the thread regarding click event [4] we were able to do so comprehensively. But as Plotly does not have such feature we needed to rely on ggplot2 and do it manually. With the help of a discussion regarding it [5], we were able to pull it off. So with the name of the country in our hand we will now have to send it to the user defined function which would do the following.

- If country is given, get the users of that country only.
- Based on the hierarchy chosen, change the data to a form suitable for sunburst plot.

With the help of reactive variables, we will be updating the sunburst chart as soon as the country is selected. Also if the user switches their view from map to bar or vice versa it will reset the sunburst chart to global demographic.

Factor Vs Average Score/Market share:

Here we will have 3 main filters for our Factor Vs Average Score/Market share bar plot.

- The factor (X- axis) in the form of a drop down list
- The Statistic in the form of a radio button
- The top 10 check box if the user wants only the top 10

After we have all the inputs from the user we will create a user defined functions that would use these values and produce the appropriate dataset for the graph.

Aside from that we will have a normal plotly bar chart showing us the factor importance. We will find it out using a rpart model.

Total Users/Views:

For this we will have 2 filters where one is a radio button that will allow the user to switch from Total users and the total views and a slider for the year range. We just have to read the data in filter it according to the year range.

After that we will be done with our implementation. A few challenges that I faced during it is given below.

- The wrangling required to bring the required data was quite cumbersome
- Implementing linked interactions was very complex.

User Guide

As our layout does not contain any sidebar all out content is displayed in one single page. Below we will explain how to use different parts of the Page

Global Market Share of anime:

Our default view for this is a leaf let map but we can switch to a bar chart that shows only the top 10 countries with the radio button as shown in the picture below. This radio button can be used to reset the user distribution chart to global.

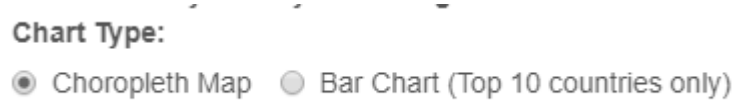


Figure 8:Radio button for chart type

In the Choropleth Map we can see the countries divided by 3 categories of market share (High, Medium and Low).

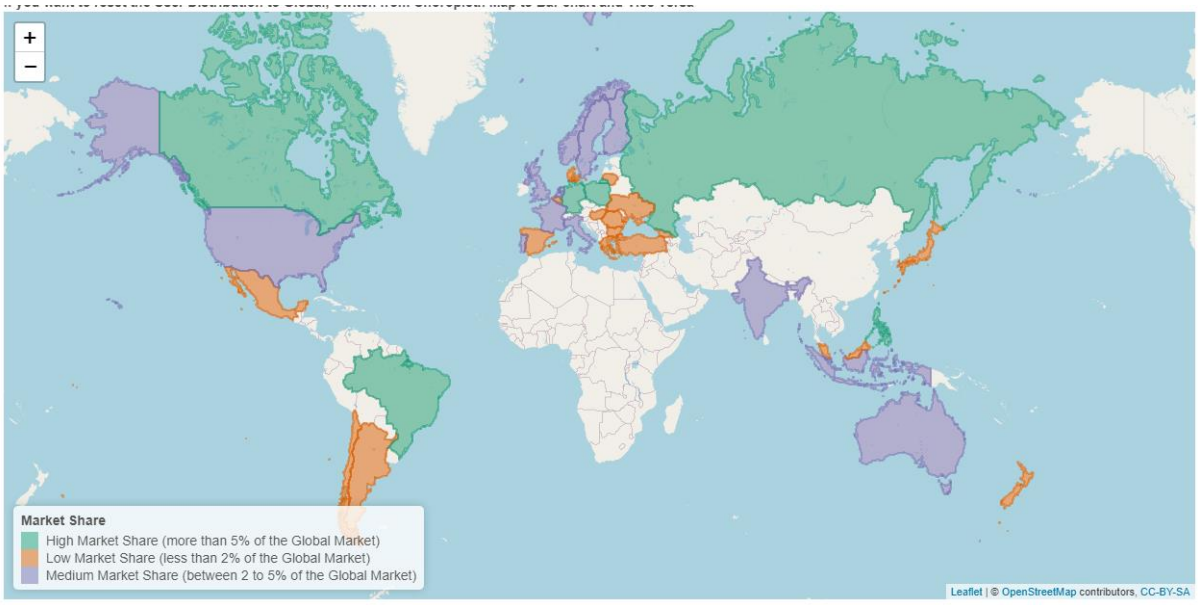


Figure 9:Choropleth Map

If we hover over any country, we will get the country name.

Figure 10:Leaflet tooltip

If we click on the country, we will get more details about it.

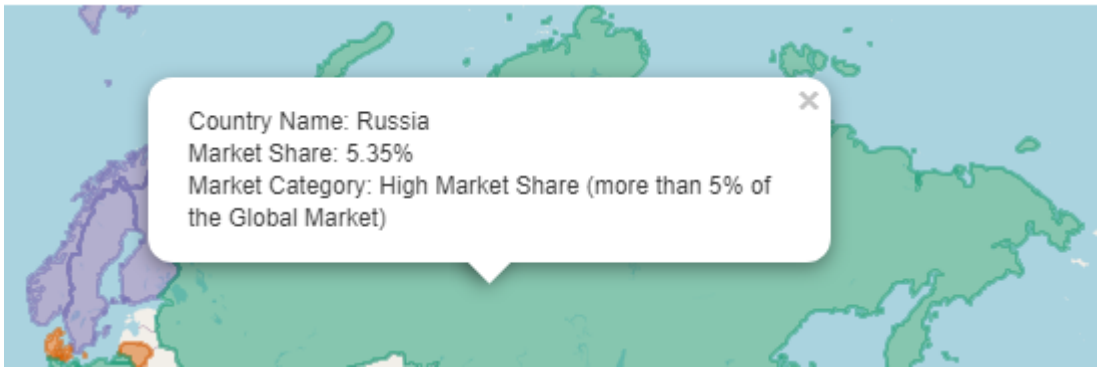


Figure 11: Leaflet Popup message

Now when it come to the bar chart all the relevant information is displayed on the chart itself. But if we click on a bar it will get highlighted. Unfortunately, due to the nature of ggplot we are not able to produce any on mouse over action but if we click on a bar it will definitely get highlighted.

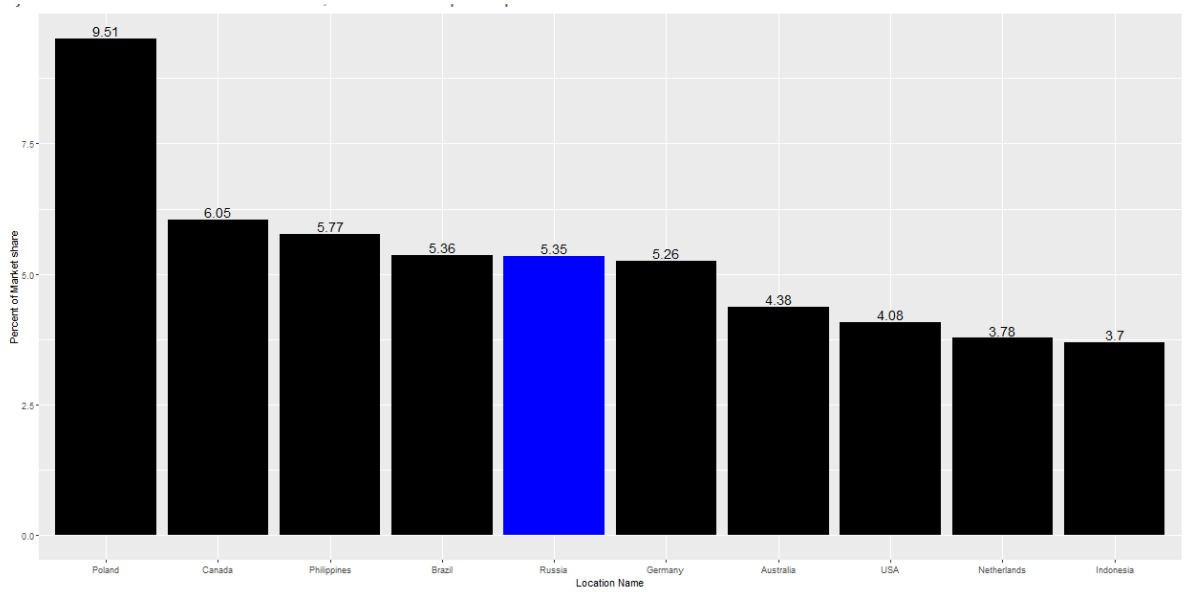


Figure 12: Global Market share bar graph

User Distribution:

Here we are given a sunburst chart where we can change the hierarchy of the chart based on what the user chooses. The user can choose it through a radio button.

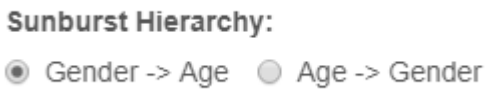


Figure 13:Sunburst hierarchy

Based on the radio button selected we get the sunburst diagram for the world according to the hierarchy chosen.

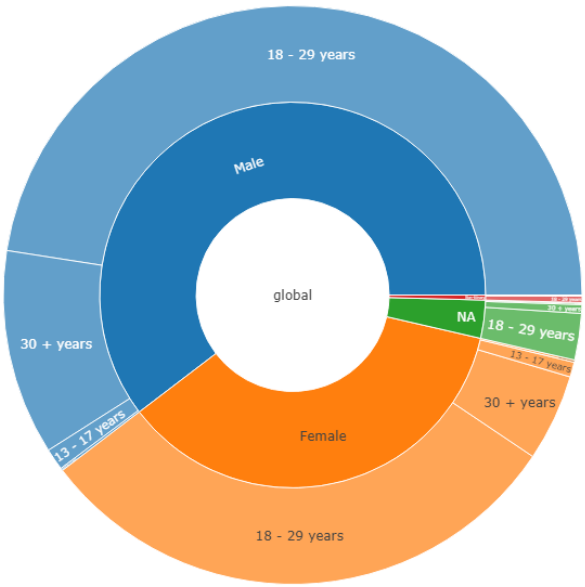


Figure 14: Gender-age hierarchy

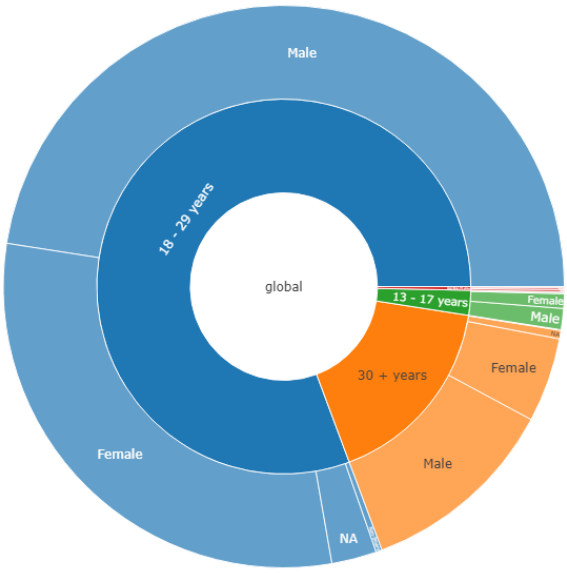


Figure 15: Age - Gender hierarchy

Also when we click on a country in the map or a bar in the bar chart (depending on the user's preferred view) we will get the user distribution for that specific country.

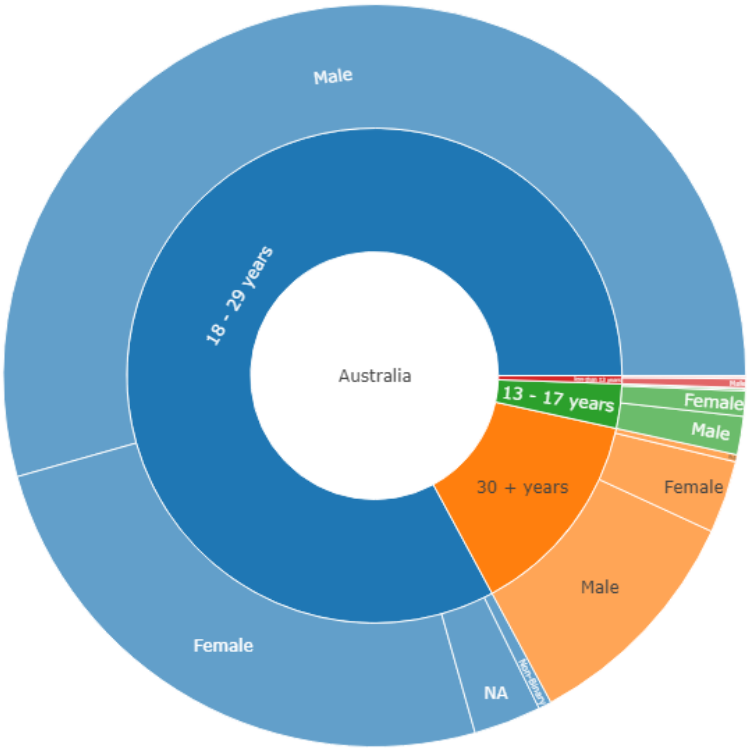


Figure 16: User Distribution of Australia

If we click on the any sector of the inner circle we will get a detailed view of the selected sector. Also if we hover over it we will get a tooltip with the label and the percent value.

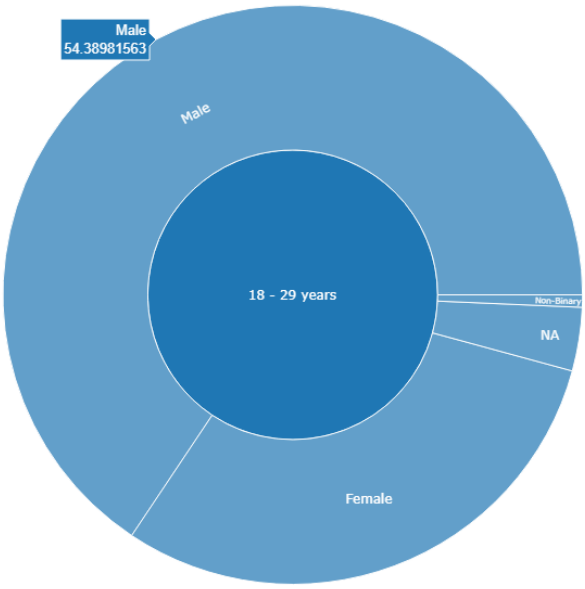


Figure 17: gender distribution of people between 18 and 29 years

Factors determining the market share and quality of an Anime:

Here we are using the plotly horizontal bar chart where it shows how a factor contributes to a success of an anime. Basically this is a chart for the user to get an idea of what are the important factors.

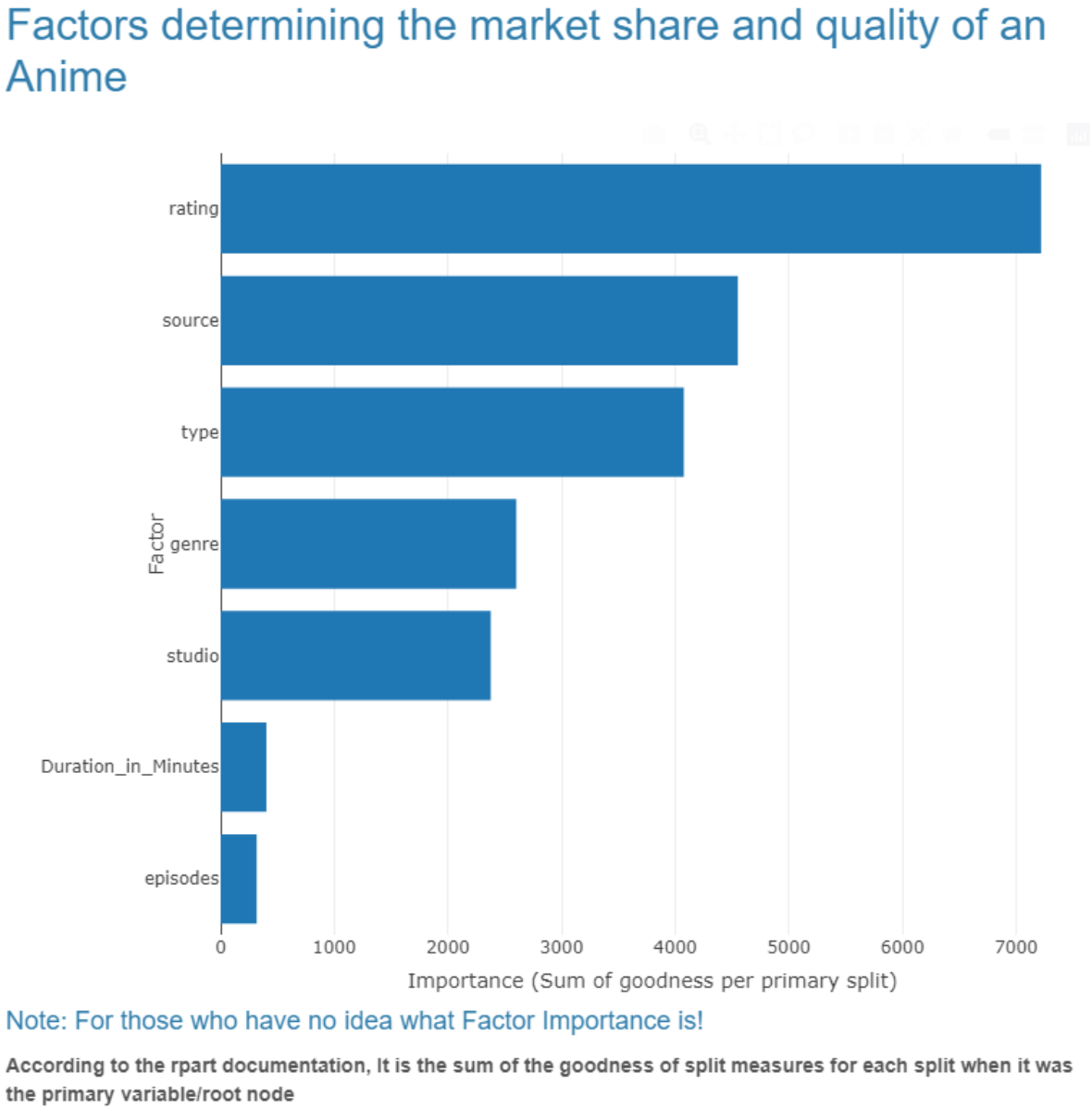


Figure 18:Factors determining the market share and quality of an Anime

Factor Vs Average Score/Market Share:

We have a few filters where the user can choose the factor, statistic and toggle between showing only the top 10 values or not.

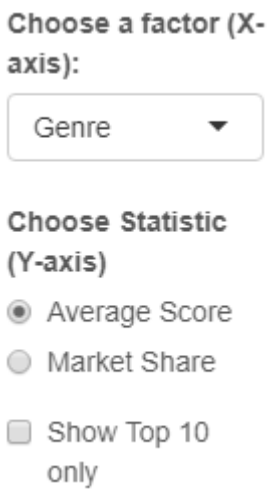


Figure 19:Factor Vs Average Score/Market Share Filters

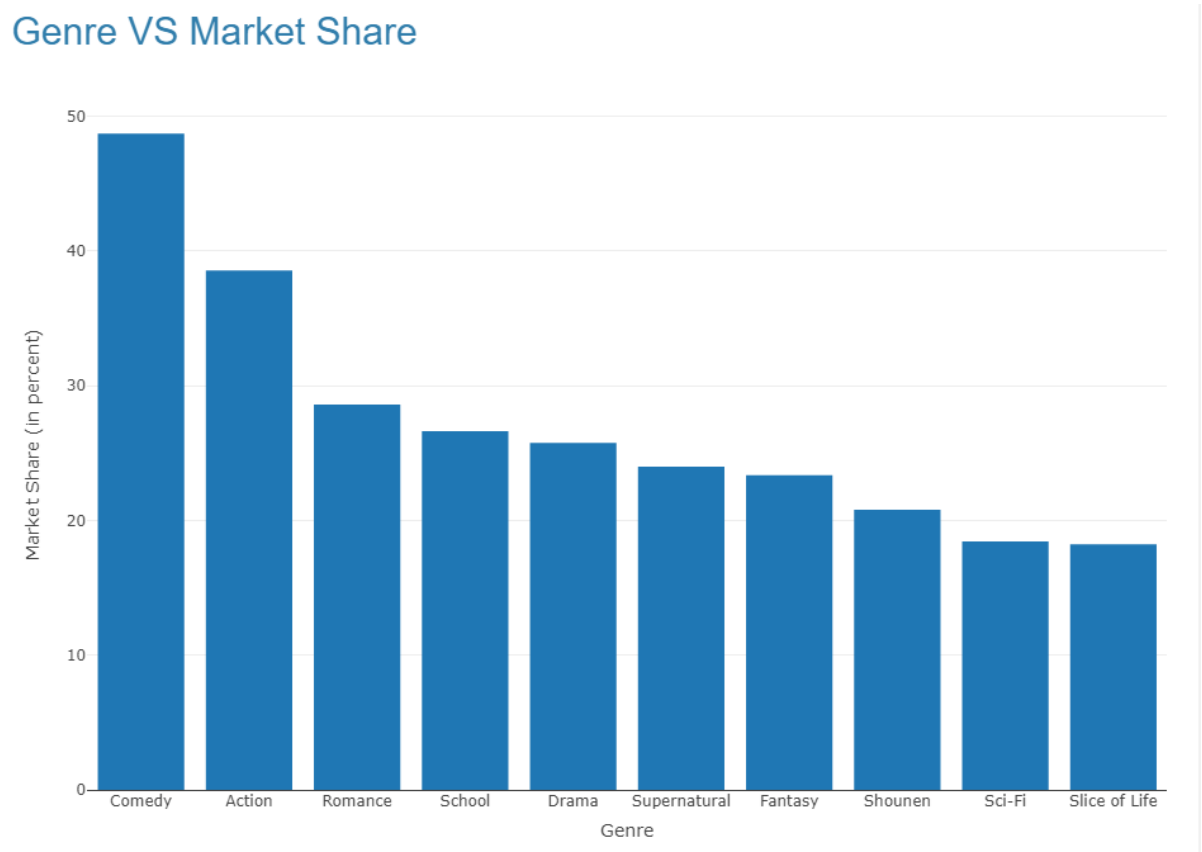


Figure 20:Factor Vs Average Score/Market Share

Total Views/Users:

We have a radio button to switch between the y-attribute and a slider to determine the limits of the x axis.

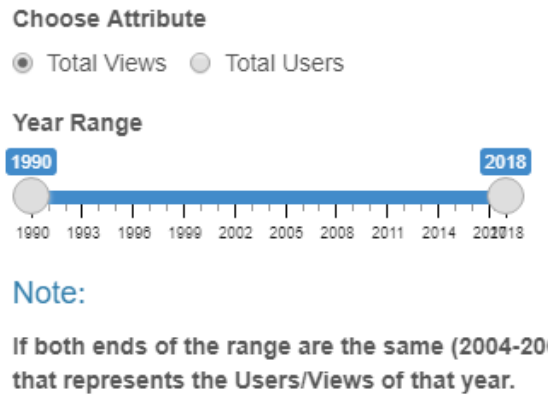


Figure 21: Filters for total user/views

According to the input given by the user the line + scatter plot will change.

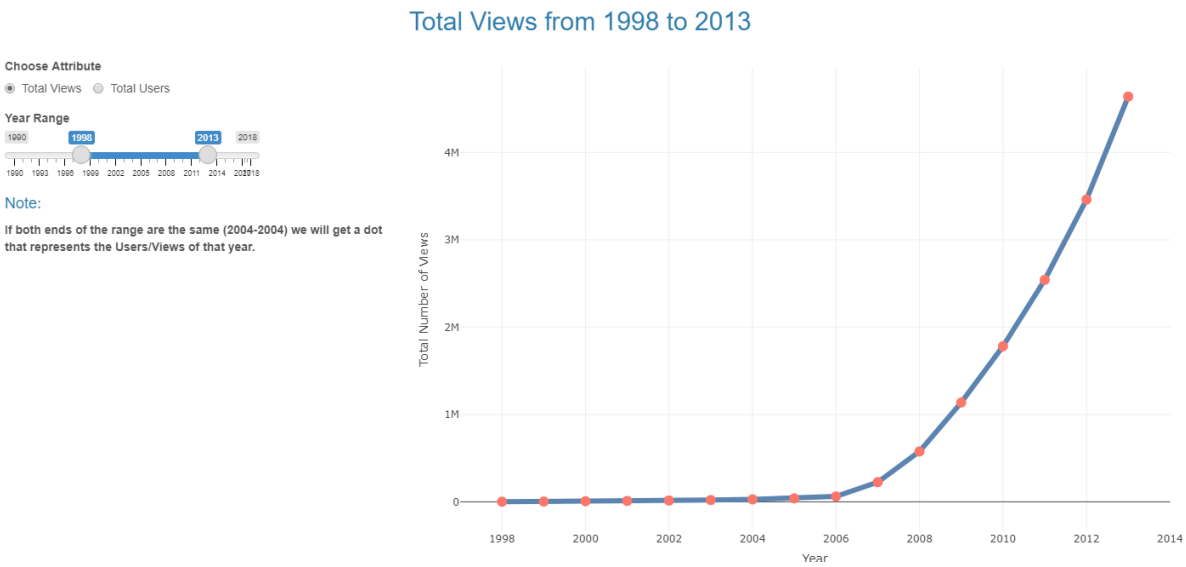


Figure 22:total user/views

As usual we would also get the x,y values if hover over it.

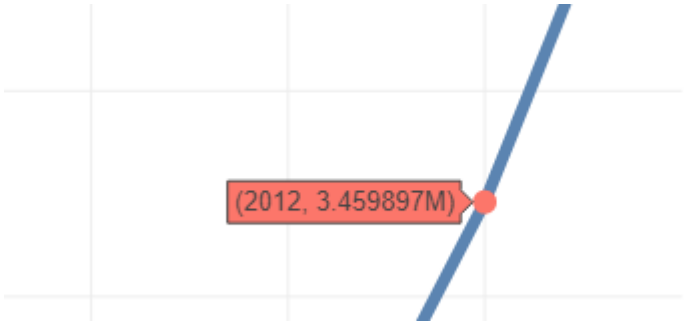


Figure 23: Tooltip of total user/views

Conclusion

From our visualization project we have managed to find out the following.

- The Global Market share of Anime
- The countries that contribute to that market share
- The User demographic of anime fans as well as fans in a specific country.
- The factors that have an impact on an anime's success.
- The average score and market share based on those factors.
- The growth of the users and views throughout the years.

We have achieved the following when it comes to our narrative visualization.

- We told our target audience how much the genre has grown over the years
- Have given our Anime fans (otakus) information about the what are the best kinds of anime
- We have shown our target audience where anime is popular and where it greatly celebrated

In hindsight I would have loved to do the following

- Added more anime aesthetics into our web page in order to stay true to our context
- Add more animation that would be preferred by our anime fandom.
- Had a much cleaner data set that would have made our wrangling processes much easier as well as would have given us more data for our webpage.

When it comes to future work on this project, it is something I personally need to think about as there are not proper data sets for us to implement that. Maybe after a few year I may want to rework on this when I get better data. The time constraint was also the problem in this case. As this is an assignment there were many things that I personally couldn't implement. Once I have completed my course and come to a point in life where I would have free time, I am sure that I would revisit this project.

Bibliography

[1] Henerson, Evan,"The Influence of Anime",Keyframe Magazine. https://keyframemagazine.org/2020/08/14/the-influence-of-anime/ (Accessed 2020).

[2] C.Cousins, "Tips on Using White Backgrounds in Website Design", Designmodo. https://designmodo.com/white-backgrounds/ (Accessed Jan. 07, 2014)

[3] Račinský, Matěj ,"MyAnimeList Dataset." Kaggle, 2018, doi: 10.34740/KAGGLE/DSV/45582. URL: https://www.kaggle.com/azathoth42/myanimelist .

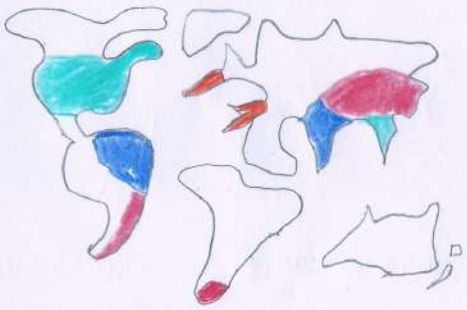
[4] "shiny leaflet ploygon click event",Stack Overflow. https://stackoverflow.com/questions/42798377/shiny-leaflet-ploygon-click-event .

[5] "How to make bars in ggplot2 barplot interactive in shiny?", Stack Overflow. https://stackoverflow.com/questions/31961622/how-to-make-bars-in-ggplot2-barplot-interactive-in-shiny

Appendix

1. IDEAS

① Choropleth map



MARKET SHARE

>50% 25-50% 10-25%

<10% 0%

② PIE CHART

AGE GROUP

<12 13-18 18-30

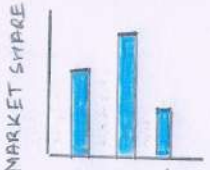
30+

③ SEMICIRCLE DONUT CHART

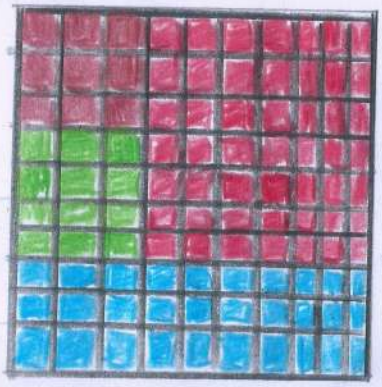
COUNTRY DIST. FOR GENDER X

US UK AUS

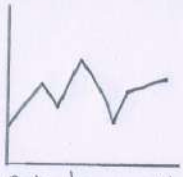
④ BARCHART



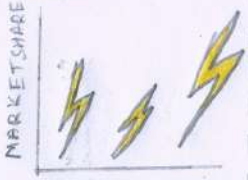
⑤ WAFFLE CHART



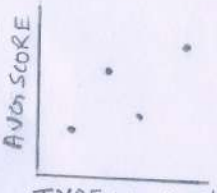
⑥ line chart



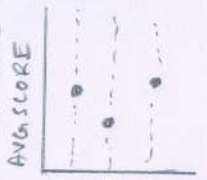
⑦ PICTORIAL BAR chart



⑧ Scatter plot



⑨ Dot chart

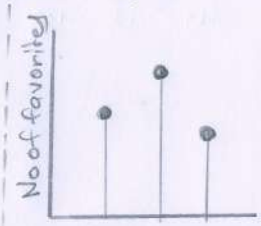


⑩ Word cloud

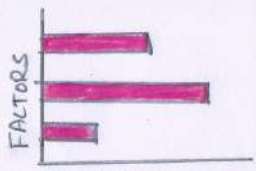
FAVOURITE ANIME

NARUTO AOT
ONE PIECE
SAO BLEACH

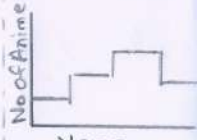
⑪ Lollipop chart



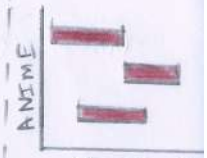
⑫ HORIZONTAL BAR chart



⑬ STEPPED LINE GRAPH



⑭ Gantt chart



⑮ Multiline graph

2. FILTER

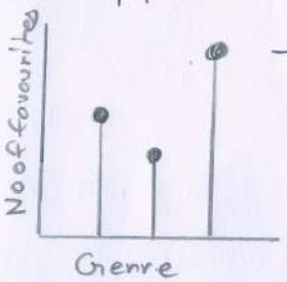
⑩ WORD CLOUD

FAVOURITE ANIME

NARUTO
ONE PIECE
BLEACH

→ EVEN though Aesthetically nice, this is completely irrelevant to our project as we are trying to find out the factors that make an anime good not the best anime

⑪ Lollipop chart



→ Similar to ⑩ this is also irrelevant as we do not care about what genre is the most popular as it does not contribute to the anime's quality

⑬ Line graph (stepped)

↳ No of anime produced in a year doesn't give us the market share

⑭ Gantt chart

↳ The airing of an anime doesn't determine its success

3. CATEGORIZE

① ④ ⑦ → Deals with market share

② ⑤ ③ → Focuses on the Gender Demographic

⑥ ⑨ ⑫ ⑧ → emphasizes on the average score and the factors that influence it

⑮ - Looks into views and total users over the years

4. Combine AND Refine

④ and ⑦ can be combined to ~~form~~ form a bar chart with a filter to choose the x axis

⑥, ⑨ and ⑧ can merge to ~~form~~ form a line chart with a selector for the x axis

②, ③ and ⑤ can be joined to create a sunburst diagram where the hierarchy can be swapped

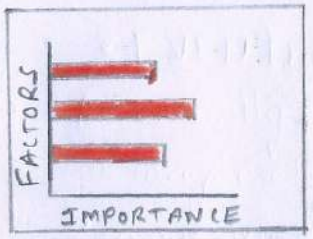
QUESTIONS

★) What is the market share of anime globally?

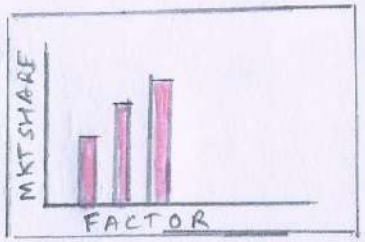
★) What are the factors that make an anime good?

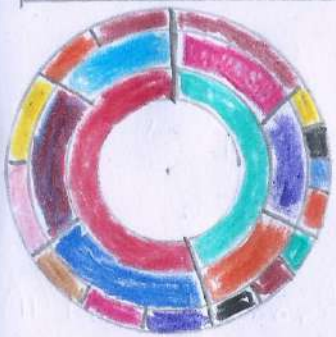
★) Has anime grown over the recent years?

LAYOUT



SCORE ANALYSIS

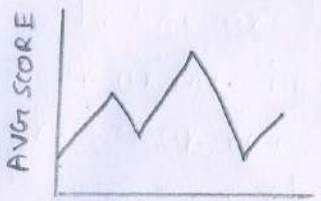




DEMOGRAPHIC HIERARCHY

GENDER -> COUNTY -> AGE

USER DEMOGRAPHIC



CHOOSE FACTOR

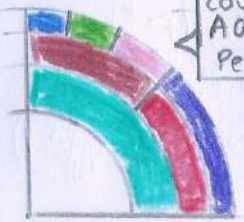
FACTOR

TOP 5 ☑ ALL ☐

YEAR

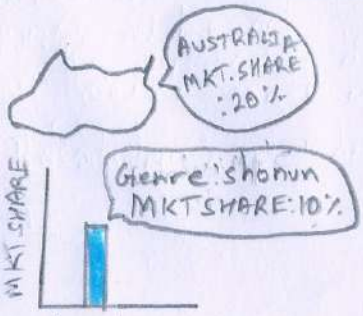
ANIME OVER THE YEARS

FOCUS

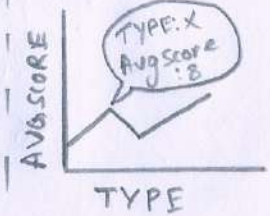


Gender: Male
County: AUS
AGE: 30+
Perc: 20%

If we hover over a part of the sunburst diagram it will show the ~~data~~ details and percentage of the hovered part



When we hover on a bar or country we will get the market share and the country / factor name



When we hover over a point in any of the line graphs we get the details of the x and y axis values

TITLE: INFLUENCE OF ANIME

AUTHOR: HARISRI GUHAN SIVAKUMAR

DATE : 8th October 2021

SHEET: 2

TASK : Visualization Layout

OPERATIONS

- Clicking a country will only give details of that country only. This can be done by clicking on a country on the map
- Click on a factor gives us the factor vs market share bar chart. If a country is ~~clicked~~ clicked it will only show for that country
- Drop downs are available for hierarchy and factor selection
- Slider is provided to give the year range

DISCUSSION

Pros :-

Very creative & interactive

Has a decent amount of filters

Cons :- As the information we get for factor vs market share depends on what we click on other graphs it is pretty difficult

- Multiline graphs is not suitable as Users is a number, and Market share is in %.

LAYOUT

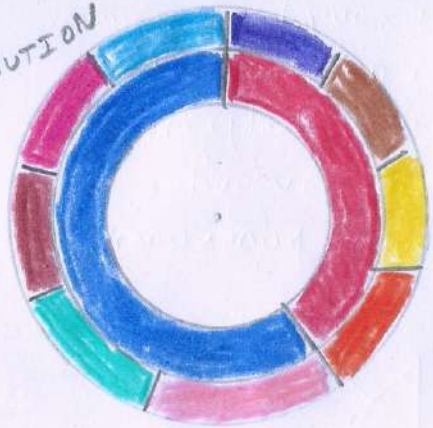
GLOBAL MARKET SHARE
⊙ MAP ○ BAR CHART (TOP 10)

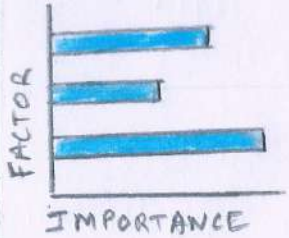
MARKET SHARE
- >50%
- 25-50%
- 10-25%
- <10%
- 0%

AREA: WORLD

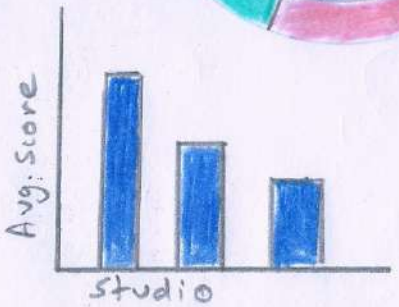
HIERARCHY
⊙ Gender-Age
○ Age-Gender

USER DISTRIBUTION



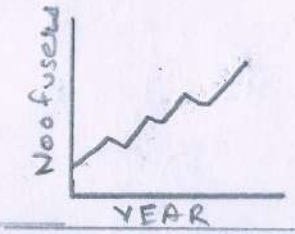


FACTOR
STUDIO ▼
⊙ Avg. Score
○ Marketshare
□ Top 10



YEAR RANGE

⊙ Users ○ No of Views



TITLE : THE INFLUENCE OF ANIME

AUTHOR : HARISRIGUHAN SIVAKUMAR

DATE : 8th October 2021

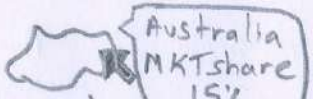
SHEET : 3

TASK : Visualizing Layout

OPERATIONS

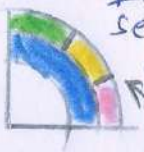
- Clicking on a country in the market share map will give us the user distribu-tion for that specific country
- Can toggle between a map & a barchart for global market share. Similarly many radial buttons are present to choose between options.
- Hovering over any part of any graph gives detailed information on it.
- A Year slider helps us set a range of years
- Top 10 has a checkbox to switch between All and top 10

FOCUS



→ If we click on a country the user demo-graphic for that country will be displayed

If we hover over a country we get the details of its market share



If we hover over a sector in the sunburst diagram we will get some details

Area: world Age: 30+
Gender: male

If we hover over a bar or line graph we will get the x and y values

⊙ Avg Score
○ MKT share
→ They allow us to choose the required parameters

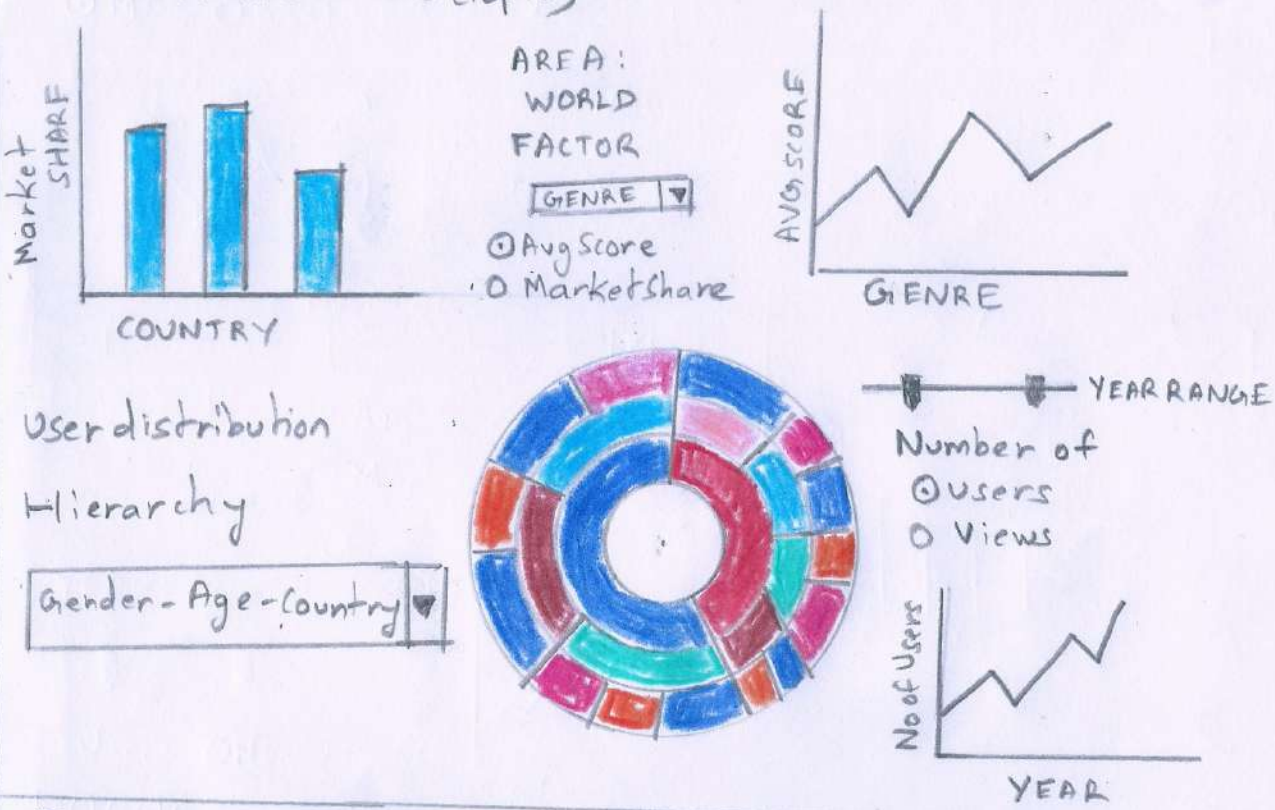
helps us select the year range

DISCUSSION

Pro :-
- Lots of filters, so the visualization will not be clumsy
- Sun burst diagram has only 2 levels so can understand it better.
- Filter options are given clearly without any confusion
- Line graph more clear as we can choose a y axis value

CONS :- A lot of filter can be cumbersome

LAYOUT



TITLE : THE INFLUENCE OF ANIME

AUTHOR : HARISRIGUHAN SIVAKUMAR

DATE : 8th October 2021

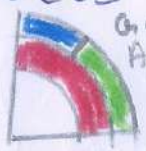
SHEET : 4

TASK : Visualization Layout

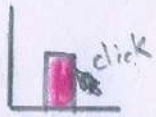
OPERATIONS

- Dropdown for Factor and Hierarchy
- ~~Hovering~~ Clicking on a bar in the global market share graph will display the factor vs score/market for that country
- Dual slider for year range
- Hovering will give us the details as a tool tip
- Radial buttons are provided for Yaxis attributes

FOCUS



Hovering over the sunburst chart gives us detailed info about it.



If we click on a bar we will get information for that country in the graph next to it

If we hover over a line graph we get the exact x and y values

slider lets us assign the year range

Avg Score
MKt share

clicking on one of them changes the Y axis

DISCUSSION

Pros

:- Bar graph for Global market share makes it easier to understand.

Cons :-

Absence of a map reduces data complexity

3 tier sunburst chart is ~~difficult~~ difficult to understand for some

LAYOUT

GLOBAL MARKET SHARE

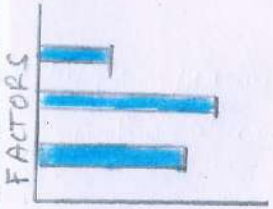
⊙ MAP ○ BAR CHART (TOP 10)

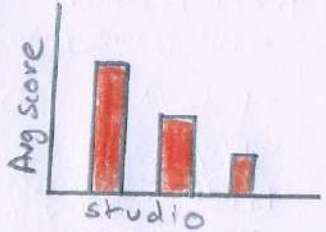
FACTOR

STUDIO

⊙ Avg score
○ Market share

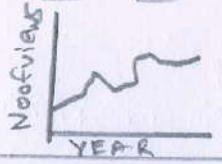
☐ Top 10





YEAR RANGE

⊙ Users ○ No of views



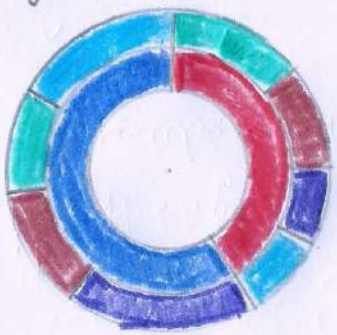
USER DISTRIBUTION

SELECTED AREA: WORLD

HIERARCHY: ⊙ Gender-Age
○ Age-Gender

Gender
- MALE
- FEMALE

Age:
- 30+
- 20-29
- 13-18
- <13



TITLE: THE INFLUENCE OF ANIME

AUTHOR: HARISRIGUHAN SIVAKUMAR

DATE: 8th October 2021

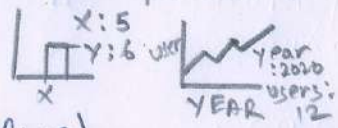
SHEET: 5

TASK: Visualization Final Layout

OPERATIONS

- Clicking a country on the market share map/bar will give us the distribution of users for that country
- Many toggle/radial buttons are provided to chose between many options. This is available on most graphs.
- Hovering any part of a graph gives the details about it
- A slider to choose the year range
- A checkbox to enable/disable the top 10 feature

FOCUS



If we hover over any graph we will get the details as a tooltip

If we click on a country in the global market share map or bar chart It will change the User Distribution for that specific country

⊙ Score ○ Market

A few radial buttons are kept to choose one between them

☐ Top 10

The top 10 checkbox displays only the top 10 of the visualisation

slider for us to decide the range

DETAIL

Algorithms: linear Regression

dependencies: R, shiny, HTML, CSS, lm, tidy verse, broom

DATASETS: Multiple subsets of the data sets about my anime list data set will be used