Voice integrated Visual Analysis of Global Migration

Kaushik JAGANATHAN | kaushikj.2017@mitb.smu.edu.sg

Priyadarsan SHANKAR | priyadarsans.2017@ mitb.smu.edu.sg

SONG Xuejing | xjsong.2017@mitb.smu.edu.sg

Abstract—With the world shrinking by the day with technology and humans being more than ever ready to scale distances in quests to follow a career path or to escape political persecution and war or in search of better quality of life or to be closer to family or friends, Migration and the resulting ethnic and racial diversity are amongst the most emotive subjects in contemporary societies. In recent times, the political salience of migration has strongly increased due to uproar of natives and other factors. For origin societies, the departure of people raises concern about the 'brain drain' on the one hand. For receiving societies, the settlement of migrant groups and the formation of ethnic minorities can fundamentally change the social, cultural, economic dimensions and cause lack of resources for the natives.

Using R's geospatial and quadrant analysis techniques, we aimed to delve into the migrant numbers data over the years from country to country conjoined with the remittances according to the world bank by them to the origin locations. Migration and remittance numbers will be analysed grouping the locations on continent, region and socio-economic status scales to identify relations between the factors separating the dominating and the dominated societies. A balance model was applied to analyze the migration flows between countries of the same development level.

The visualizations are linked between each other and wrapped on to an interactive application which can be controlled by user voice input's, we will be using a Javascript library interfaced to R shiny to control the application's select, filter and switch inputs.

Index Terms—Global migration, migrant remittance, GDP, bilateral balance, gender bias, quadrant analysis, push and pull factors.

---

1 INTRODUCTION

Migration is a key feature of our increasingly interconnected world. It has also become a flashpoint for debate in many countries, which underscores the importance of understanding the patterns of global migration and the economic impact that is created when people move across the world's borders. Migration and remittance will be analysed grouping the locations on continent, region and socio-economic status scales to identify relations between the factors separating the dominating and the dominated societies.

2 OBJECTIVE AND MOTIVATION

Migration and the resulting ethnic and racial diversity are amongst the most emotive subjects in contemporary societies. While global migration rates have remained relatively stable over the past half a century. the political salience of migration has strongly increased.

For origin societies, the departure of people raises concern about the 'brain drain' on the one hand, but it also creates the hope that the money and knowledge migrants gather abroad can foster human and economic development. For receiving societies, the settlement of migrant groups and the formation of ethnic minorities can fundamentally change the social, cultural, economic and political fabric of societies, particularly in the longer run.

Our application integrates voice into the RShiny application. The main motivation behind using voice to control a dashboard is to make the app hands-free so the application can be easily used in tablets while travelling or in a situation when a user is eating and not able to place hands on the dashboard.

Our research aims to incorporate voice and visual analytics for better insights on global migration flows using the rich data of migration and remittance from World Bank. This research aims to:

(a)   Create a user-friendly and interactive visualization platform for data exploration that can be used to view migration flows from all possible angles and see the relationship between migration and remittance.

(b)   Integrate voice into the RShiny application to control different parameters and alter visualizations accordingly.

3 PREVIOUS WORKS

Our project is a voice-driven RShiny app. We found a past project which integrates voice with RShiny using Annyang, a JS library which lets users control websites using voice. The project was done by Yihui Xie and used voice to change the basic parameters in a scatterplot like title, colour and size.
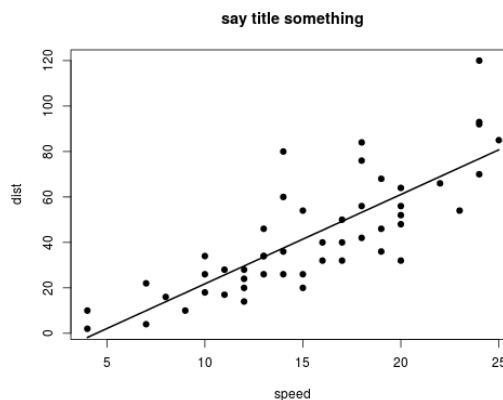
Figure 1: Voice input can change the color and size of points

Due to the complex nature of migration, any attempt to create a visualization of migration data would encounter many difficulties and challenges. An example of an existing visualization to look at migrant flows can be seen in Figure 1 which was originally published by Ann Jackson at http://www.jacksontwo.com/. This flow diagram showcases the migrant movements on a continent level or a country to continent level. Our application aims to examine migration flows between continents, regions within a continent and between countries too.
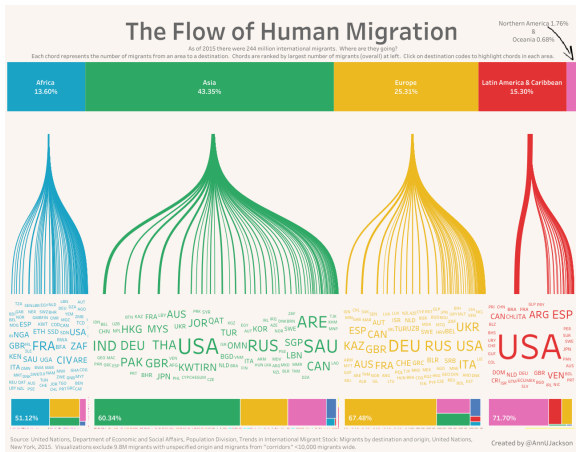

Figure 2: Flow of human migration

In the past, migration flows have typically been represented using chord diagrams. A report on "Quantifying Global International Migration Flows" by Nikola Sander, Guy J. Abel & Ramon Bauer published in Science magazine on 28 Mar 2014 shows how chord diagram (Figure 2) was used to illustrate the flow of migrants. The chord diagram can be found at http://www.global-migration.info/.
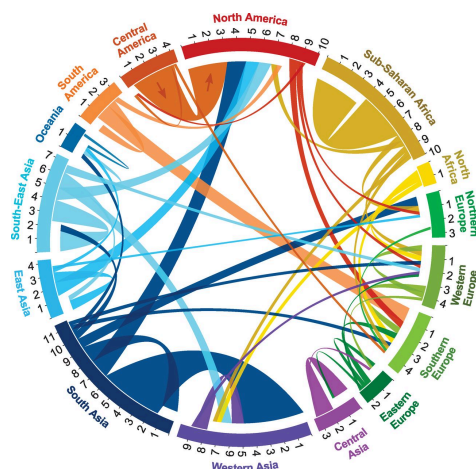

Figure 3: Chord diagram showing flow of migrants

4 DATASET AND PREPARATION

The dataset for building the visual models for this analysis have been obtained from World Bank Open Data repository. We have taken the bilateral matrix of migrant stock counts and the remittance amounts contributed by the migrants for the years 1990-2015 in 5-year intervals. To added dimensionality to the analysis we also obtained the GDP per capita (USD), GDP (USD), Population, Migrants as a % of population from the World Bank's WDI (World development Index) series data. Lastly, we have used a GeoJSON file with polygon data for all countries of the world to visualize the spatial distribution for analysis. The data was fairly clean and the cleaning steps involved only replacing the cells that were filled with dots in cells indicating no migrant stock from the row origin to column destination country to blanks for readability into R. The rows containing the sub-region, continent, world level aggregation of migrant stock to the countries were removed. The other data preparation involved reshaping the data to suit creating the visualizations. For the chord diagram, a bilateral matrix with non-zero cells of migrant stock numbers and zeroed diagonal matrix was created. For the choropleth map, the migrant stock data was binned to percentiles based on the origin/destination country. For performing the quadrant analysis, the migrant stock inverse percentiles are calculated to each origin country dynamically for every destination. For the corrplot, the bilateral balance index model was applied to the data by categorizing the countries based on their development levels.
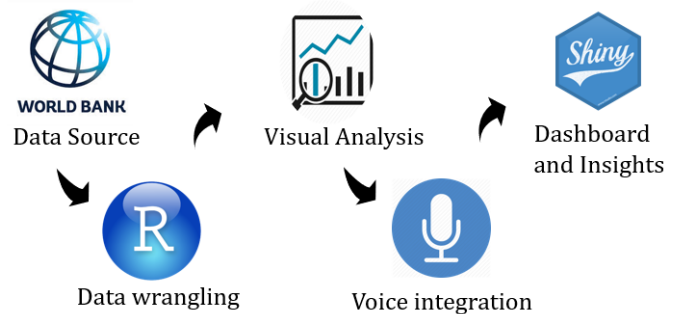
5 DESIGN FRAMEWORK AND METHODOLOGY

Figure 4:Process Flow

Process Flow steps:

1. Obtain the global migration and remittance data from the World bank website.
2. Perform data wrangling to transform data to match the input format of various visual models
3. Build visual models to analyse global migration flows and to study the relationship between migration and remittance
4. Use Annyang and SpeechKITT to integrate voice into the application
5. Visualise the data in the RShiny dashboard and obtain the insights.

5.1 Voice integration - Annyang and SpeechKITT

The voice integration to Shiny process flow is as below:

Figure 5: Annyang and SpeechKITT

Annyang is a tiny JavaScript library for performing Speech Recognition and SpeechKITT is a GUI for the user to interact with Speech Recognition. Our project integrates both Annyang and Speech KITT using JavaScript to send the voice input to R Shiny dashboard for visual analysis.

```
var initVoice = function() {
if (annyang) {

//Define the voice commands to be used and their corresponding actions

  var commands = {
    'type *type': function(type) {
      Shiny.onInputChange('type', type);
    },

    'origin *origin': function(origin) {
      Shiny.onInputChange('origin', origin);
    },
                        'go to *gt': function(gt) {
      Shiny.onInputChange('gt', gt);
    },
```
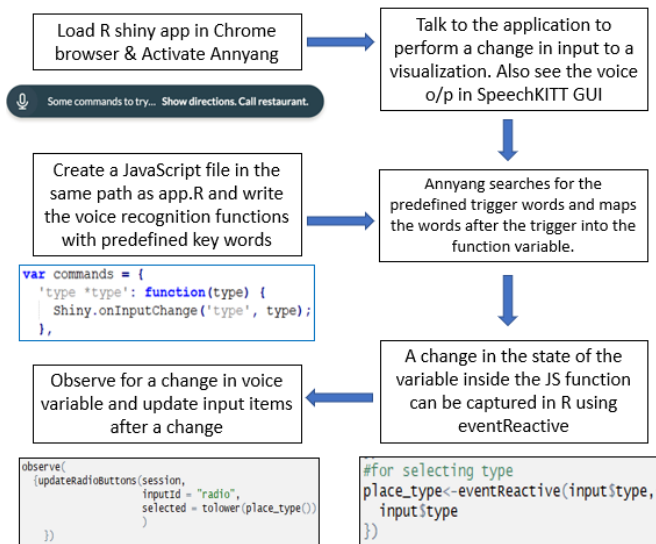
Figure 6: Annyang snippet

Figure 7: Flowchart for voice integration into app

Steps for voice integration in RShiny:

1. Create the Annyang JavaScript file with the voice recognition commands and include the Annyang and SpeechKITT libraries in R.
2. Speak to the app to change the input to visual model
3. Capture the change in speech by using the voice trigger variable inside a reactive function in R.
4. Use an observe function to look out for the change in the reactive voice variable and update the inputs in visual model accordingly.

5. Also, observe the voice input getting displayed in the SpeechKITT GUI.

5.2 Chord Diagram

The chord diagram visualises the inter-relationships between entities. The connections between entities are used to display that they share something in common. This makes Chord Diagrams ideal for comparing the similarities within a dataset or between different groups of data. Nodes are arranged along a circle, with the relationships between points connected to each other either using arcs or Bezier curves. Values are assigned to each connection, which is represented proportionally by the size of each arc. Colour can be used to group the data into different categories, which aids in making comparisons and distinguishing groups.

The 'chorddiag' package allows to create interactive chord diagrams using the JavaScript visualization library D3 from within R using the html widgets interfacing framework.
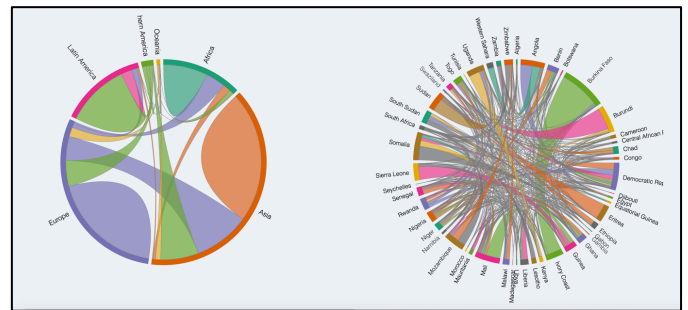

Figure 8: Linked chord diagrams for region flow and country flow

The above figure has 2 connected chord diagrams. The chord diagram on the left visualizes migration flows on a continent or a sub region level. Once a flow is selected, the second chord diagram reacts to the change and displays countries accordingly. The country flow migration flow can then be visualized.

5.3 Bidirectional Trend chart

The trend chart is a graphical representation of time series data showing the trend line that reveals a general pattern of change.
The chord diagram gives the bidirectional flow of migrants for a year, so we use a line chart to observe the trend of the migrant stock over the years. The input to the line chart is from the chord diagram when a flow line is selected. The line chart displays the bidirectional migrant stock trend for the two countries in the flow line selected.
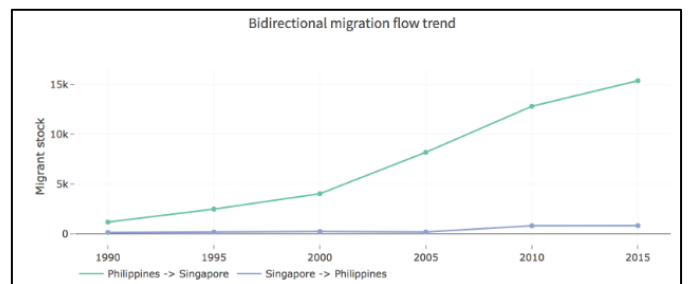

Figure 9: Trend of migrant stock over the years

5.4 Choropleth Map

A choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per-capita income. Choropleth maps provide an easy way to visualize how a measurement varies across a geographic area or show the level of variability within a region. A heat map is similar but does not use geographic boundaries.

Leaflet is one of the most popular open-source JavaScript libraries for interactive maps. 'leaflet' package makes it easy to integrate and control Leaflet maps in R. Choropleth Maps display divided geographical regions that are coloured in relation to a data variable. This provides a way to visualise values over a geographical area, which can show variation across the displayed location.
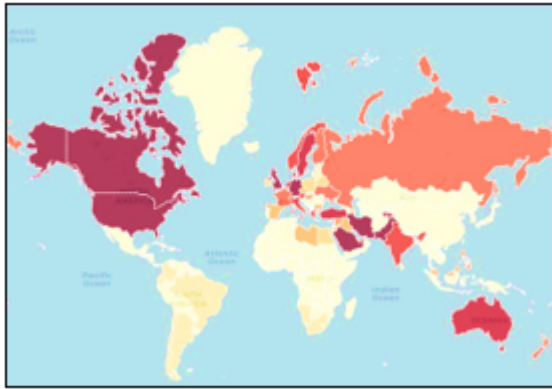
Figure 10: Choropleth map for migrant stock percentiles

In this case, we use choropleth map to display the migration and remittance data and filtering out the top 10 migrants/remittance countries after selecting the origin country using select input. When compared to chord diagram, which can only show the value of the migrants, in this map, we can also see the location of each country to investigate the detail of why people migrate. In this map, the darker the colour the higher the value. Therefore, we can find out the countries with large number of migrants, but with small amount of remittance.

5.5 Correlation Plot

Corrplot contains algorithms to do matrix reordering and is good at details, including choosing colour, text labels, colour labels, layout. The corrplot package is a graphical display of a correlation matrix, confidence interval.

Bilateral balance index is a measure of strength of balance in a bi-directional flow data. It ranges from 0 to 1, with 1 indicating an equally strong flow in both directions.

For any country pair i-j, the bilateral balance is calculated as:

$$2 \cdot \frac{min(m_{ij})}{m_{ij} + m_{ji}}$$

Mij = Migrant stock in j from i in a time period
Mji = Migrant stock in i from j in a time period

The corr plot is used to study the bilateral balance in the migration between countries of same development types. Darker cells in the matrix indicate higher bilateral balance between two countries.
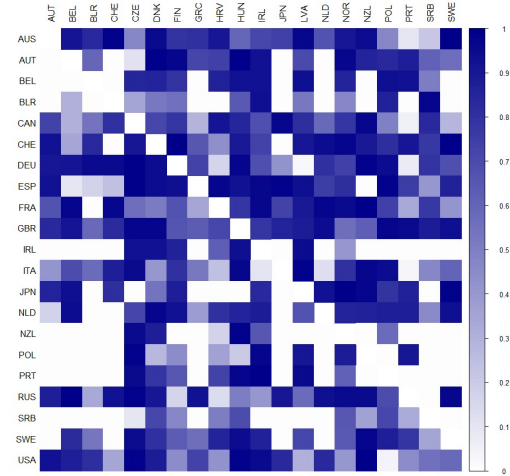

Figure 11: Corr plot showing the bilateral balance between countries

5.6 Quadrant Plot

Quadrant Analysis is used to see the relationship between two data variables analysed on four quadrants.

Plotly's R graphing library makes interactive, publication-quality graphs online. Plotly allows users to Zoom, Pan, and Hover Controls in Plotly Graphs. Also, when the user hovers over a point on a scatter plot, the tooltip shows information about this point. The Plotly output rendering can be fully customized, so we have removed all the non-data ink created by the ggplot wrapping onto native plotly creating multiple options for interactivity, we have removed the non-relevant options to each plot type on this regard. For example, for the bubble chart below, we have kept only the spike line, zoom, box select and the lasso select options.
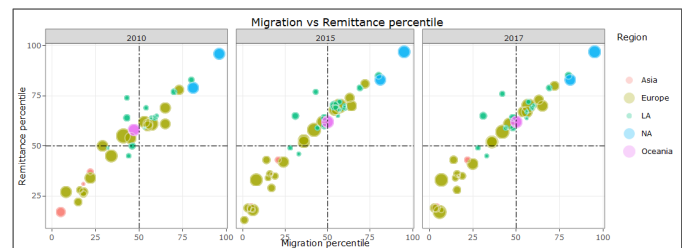
Figure 12: Plotly graph showing relationship between migration and remittance

This plot investigates the relationship between migration and remittance percentile for a selected origin country. The bubbles are coloured by region or income level or development type and sized by the GDP Per Capita of the destination country.

5.7 Bar chart

Bar charts are used to visualize the relationship between one categorical variable and one continuous variable. We have used it in

it's few forms to visualize part of whole, trend and migrant remittance relationships plotted using ggplot wrapped on to the plotly package, since we do not require any of the plotly filter features, we have removed all the unnecessary non-data ink.

The below 100% stacked bar chart shows the trend change in the part whole relationship of the world's total migrants as a proportion of their current country of residence by income level. There is an increasing trend in the migrants moving towards high income countries.
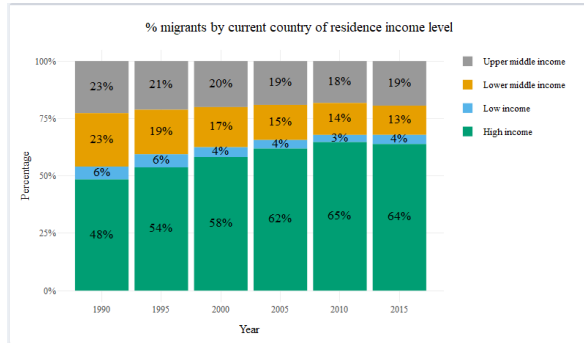

Figure 13 100% stacked bar chart of migrant %

The below dodged bar chart presents a gender bias analysis, when a flow and year is picked, bars indicating the trend of men and women migrant stock numbers over years is plotted. The selected year is highlighted in red, using ggplot's grammar of graphics by superimposing multiple layers by varying fill and border functions.
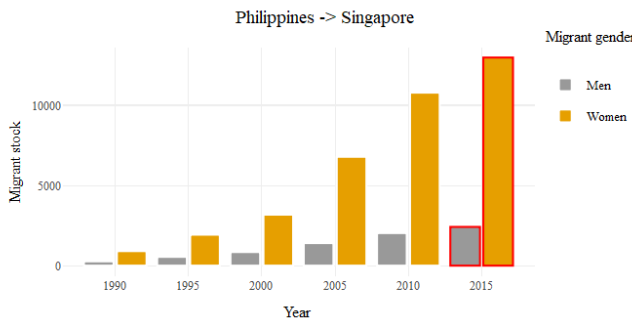

Figure 14 Dodged bar chart for Gender bias

The below dodged bar chart is implemented using plotly's event data method to captured a user selected subset of input data from an existing plotly plot to interactively visualize the migration and remittance numbers from an origin country across all selected destination countries.
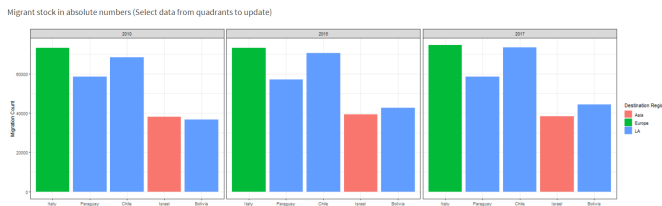

Figure 15 Bar chart for migrant stock numbers

6 INSTALLATION AND USER GUIDE

Online users would be able to find our application at the following website:
https://wiki.smu.edu.sg/1718t3isss608/Group10_Application

Simply click through the tabs to look at the different levels of analysis provided in our application.
Developers and coders would be required to install a list of packages used in our analysis for the application to run in your R/R Studio environment. The list of packages is available in Annex B.

REFERENCES

[1] World GeoJSON johan/world.geo.json.
     https://github.com/johan/world.geo.json (2018).
[2] Speech KITT GUI for Speech Recognition. TalAter/SpeechKITT.
     https://github.com/TalAter/SpeechKITT (2018).
[3] Annyang Javascript library for Speech recognition.
     https://github.com/TalAter/annyang (2018).
[4] Nogle, J. (1994). The Systems Approach to International Migration: An Application of Network Analysis Methods. *International Migration*, *32*(2), 329-342. doi: 10.1111/j.1468-2435.1994.tb00156.x
[5] Bertocchi, G., & Strozzi, C. (2007). *The age of mass migration*. London: Centre for Economic Policy Research.
[6] Castles, S., & Miller, M. J. (1998). The age of migration: International population movements in the modern world. New York: Guilford Press. Gheasi, M., & Nijkamp, P. (2017). A Brief Overview of International Migration Motives and Impacts, with Specific Reference to FDI. Economies, 5(3), 31.

ANNEX A – DATA DICTIONARY

Name	Description
Year	Year of data in YYYY format
Destination Region	Continent of destination country
Destination Sub region	Sub continent of destination country
Destination country	Destination country name
Destination Latitude	Destination country latitude
Destination Longitude	Destination country longitude
Destination Income level	Destination country's income categorization
Destination Region type	Destination continent's development categorization
Destination Country type	Destination country's development categorization
Origin Region	Continent of Origin country
Origin Sub region	Sub continent of Origin country
Origin country	Origin country name
Origin Latitude	Origin country latitude
Origin Longitude	Origin country longitude
Origin Income level	Origin country's income categorization
Origin Region type	Origin continent's development categorization

Origin Country type	Origin country's development categorization
Gender	Migrant gender
Flag	Migrant data type (Stock or Change)
Count	Migrant count
Destination Country Code	Destination country 3 byte code
Origin Country Code	Origin country 3 byte code
Remittance	Remittance amount in 1000 USD
Log Remittance	Log base 10 of remittance amount
Migration Percentage	Migrant count as a percentage of population of destination country
Remittance Percentage	Migrant remittance as a percentage of GDP of origin country
Origin GDPPC	Origin country's GDP per capita
Destination GDPPC	Destination country's GDP per capita
Destination GDP	Destination country's GDP
Origin GDP	Origin country's GDP per capita
Destination population	Destination country population
Origin Population	Origin country population
RemittanceOnOriginGDP	Migrant remittance as a percentage of GDP of origin country
RemittanceOnDestinationGDP	Migrant remittance as a percentage of GDP of destination country
TotalRemittanceOnOriginGDP	Total remittance received by origin country as a percentage of Origin country GDP

17	shinyjs
18	rsconnect
19	devtools
20	ggthemes
21	ggplot2
22	mapview
23	scales
24	leaflet.minicharts
25	shinydashboard
26	manipulateWidget
27	viridis
28	corrplot
29	shinyBS
30	shinyWidgets

[7]
ANNEX B – LIST OF PACKAGES USED

S/N	Name of Package
1	rgdal
2	geojsonio
3	shiny
4	readr
5	plyr
6	chorddiag
7	dplyr
8	shinythemes
9	plotly
10	tidyr
11	threejs
12	geosphere
13	leaflet
14	RColorBrewer
15	readr
16	leaflet.extras