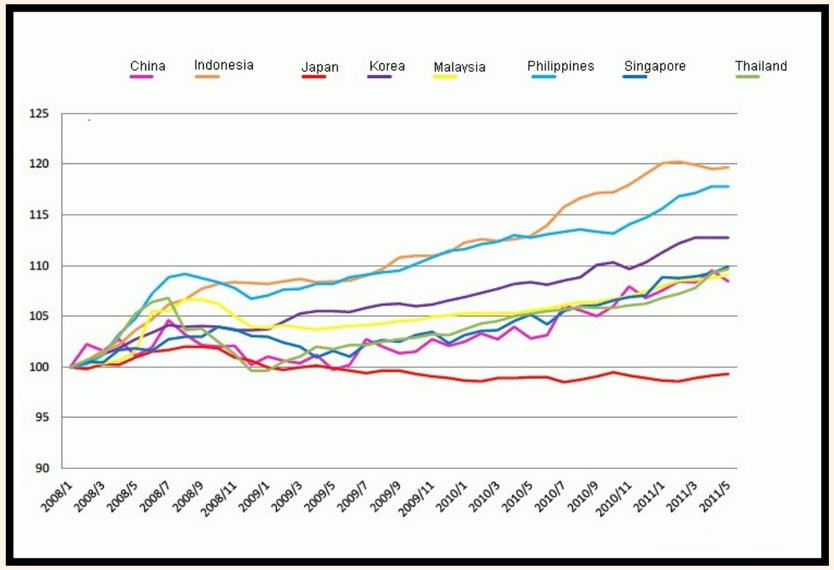
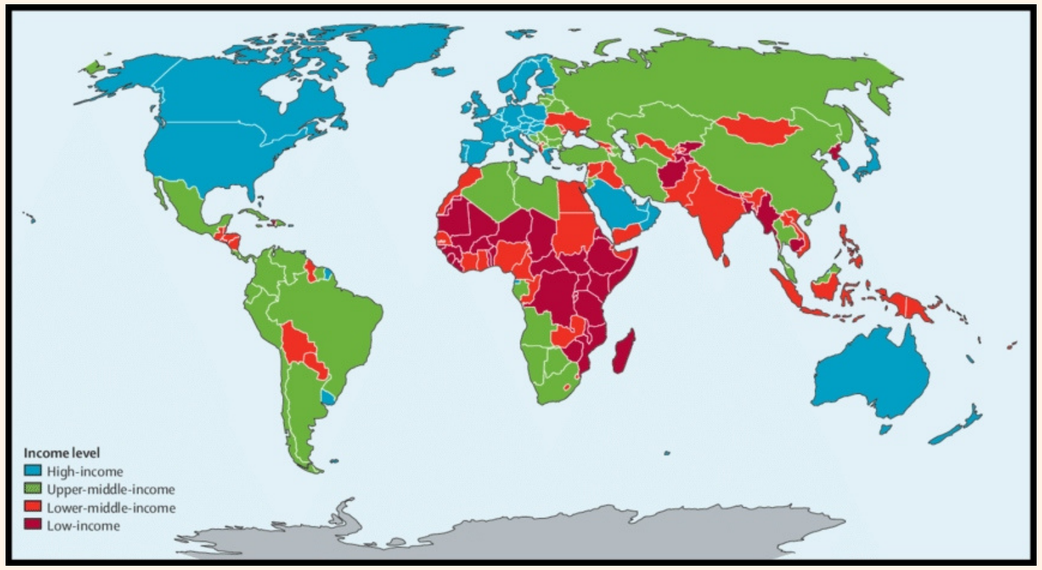
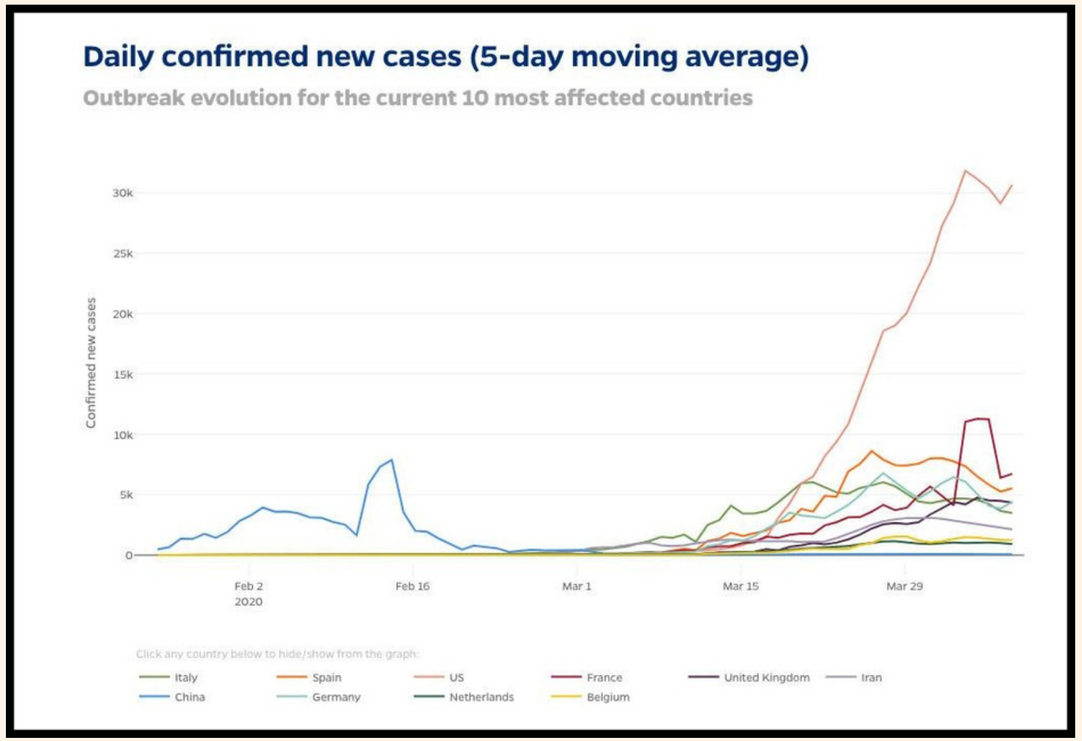
A Visualization Tool: Comparison and Clustering of Countries

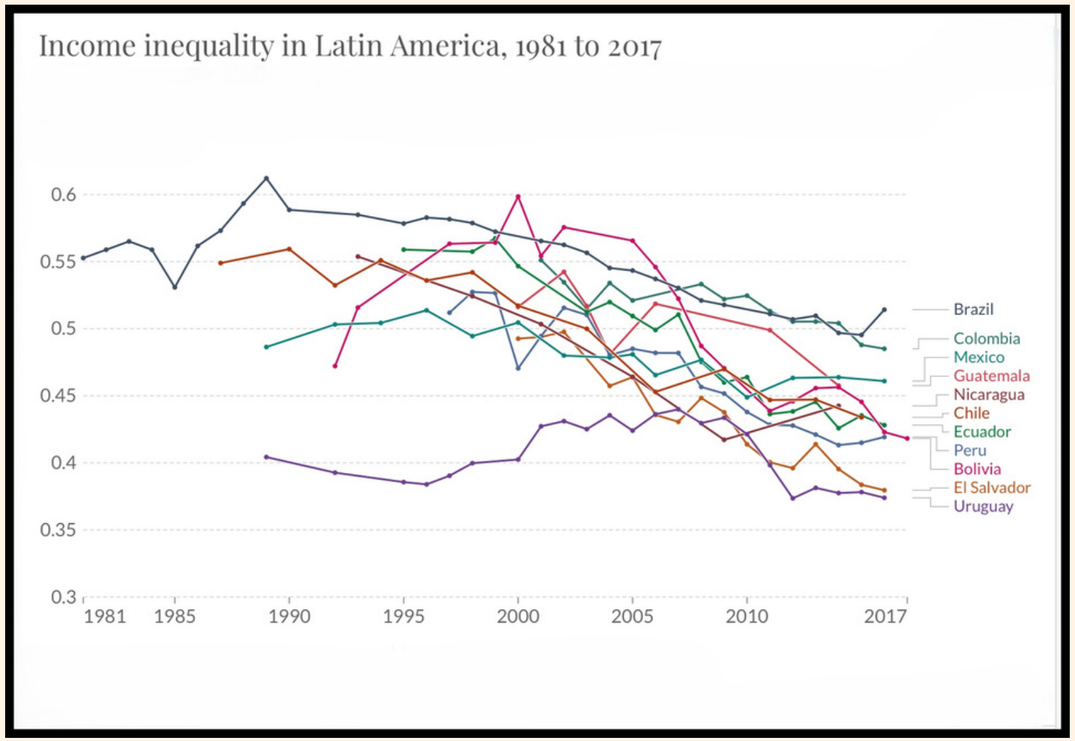




Mehmet Kadri GOFRALILAR
Hasan Ali ÖZKAN
Zeynep Filiz EREN

Time Series Data



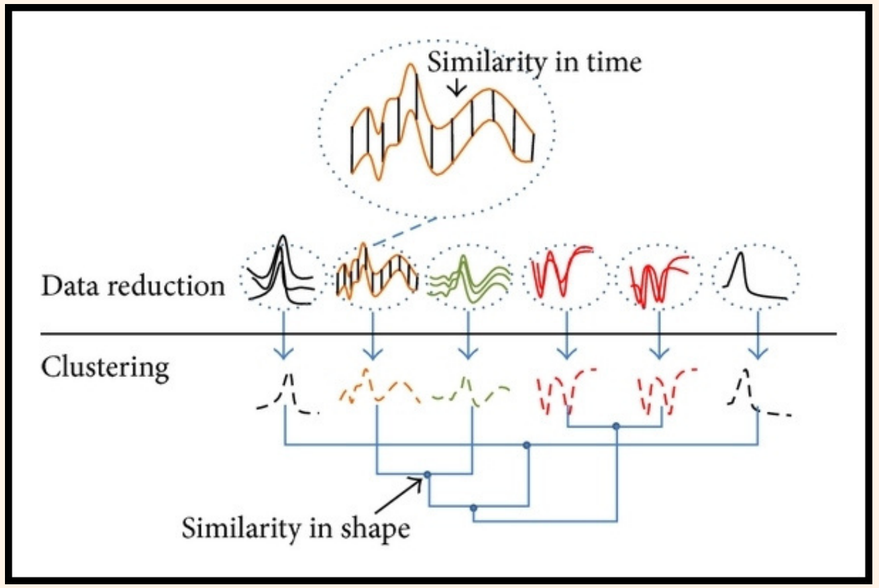


Common Problems

- Varying Length
- Outliers
- Unbalanced Cluster Size

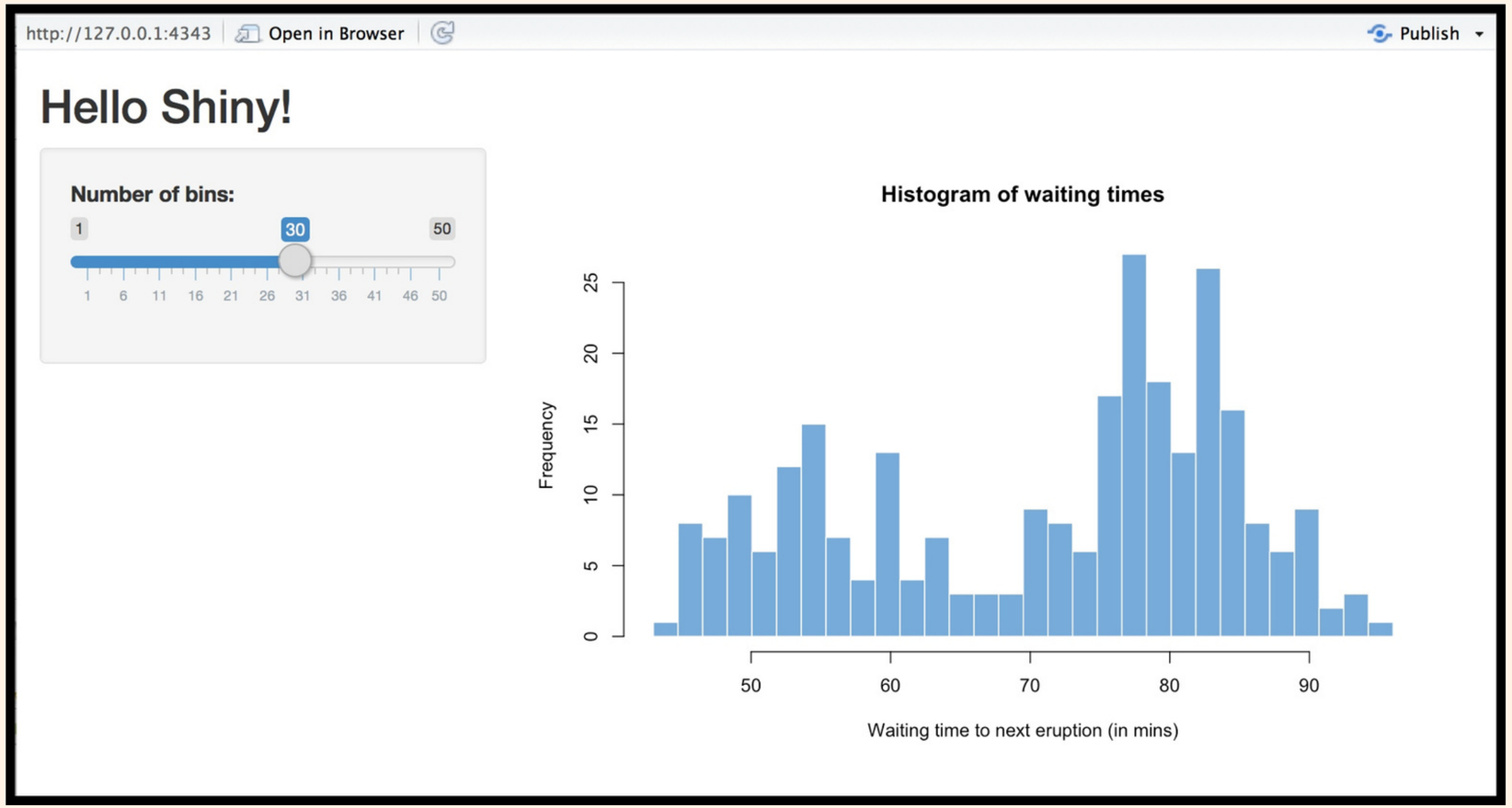
Time Series Clustering

- Partitioning Clustering (k-Means, etc.)
- Hierarchical Clustering

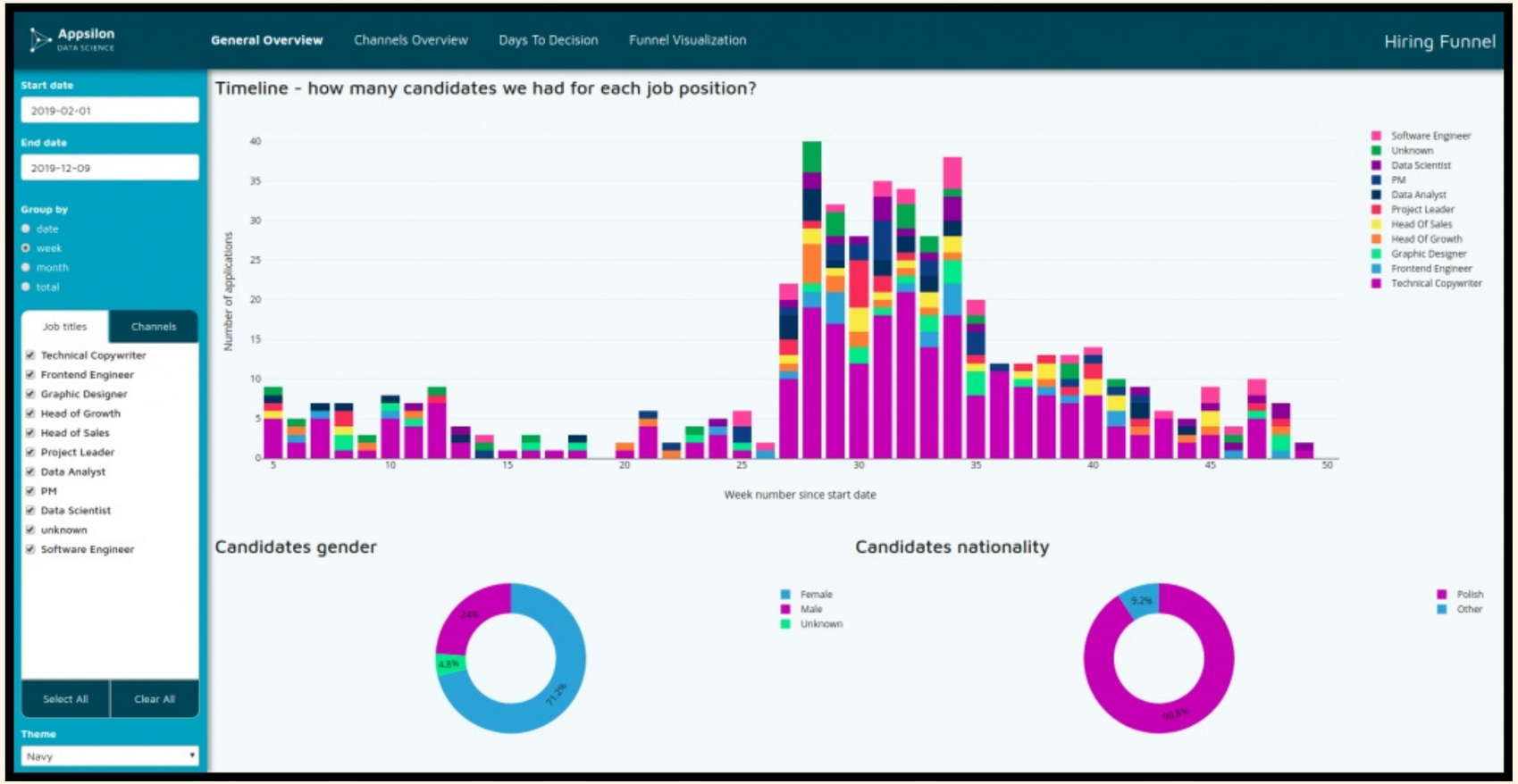


- Euclidean Distance
- DTW (Dynamic Time Warping)

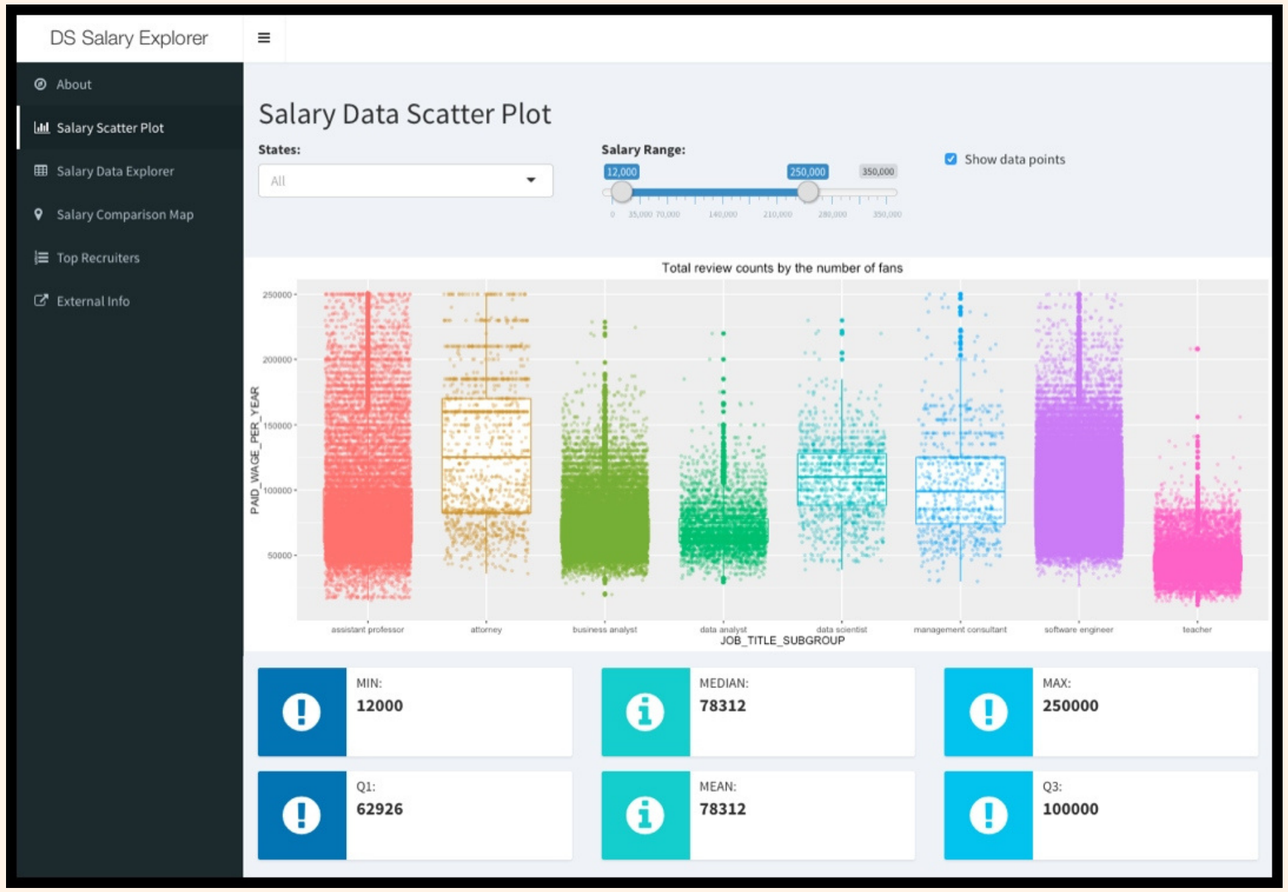
Shiny


http://127.0.0.1:4343
Open in Browser
Publish
Hello Shiny!
Number of bins:
1
30
50
1 6 11 16 21 26 31 36 41 46 50
Histogram of waiting times
Frequency
0 5 10 15 20 25
50 60 70 80 90
Waiting time to next eruption (in mins)

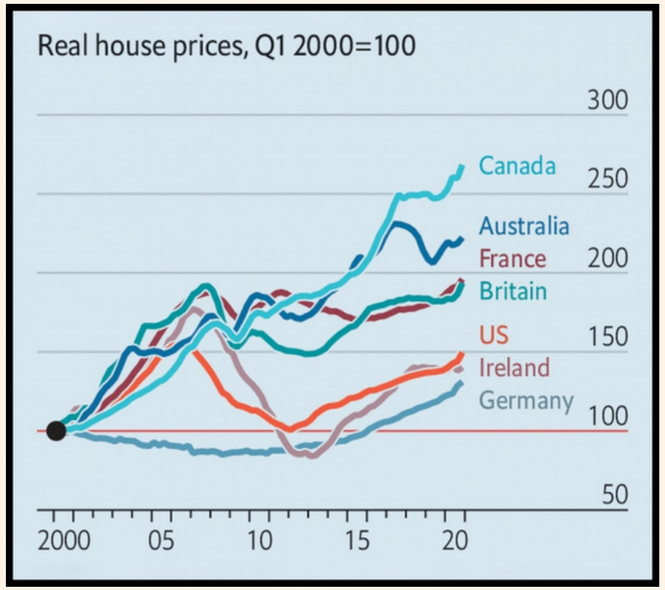
Shiny

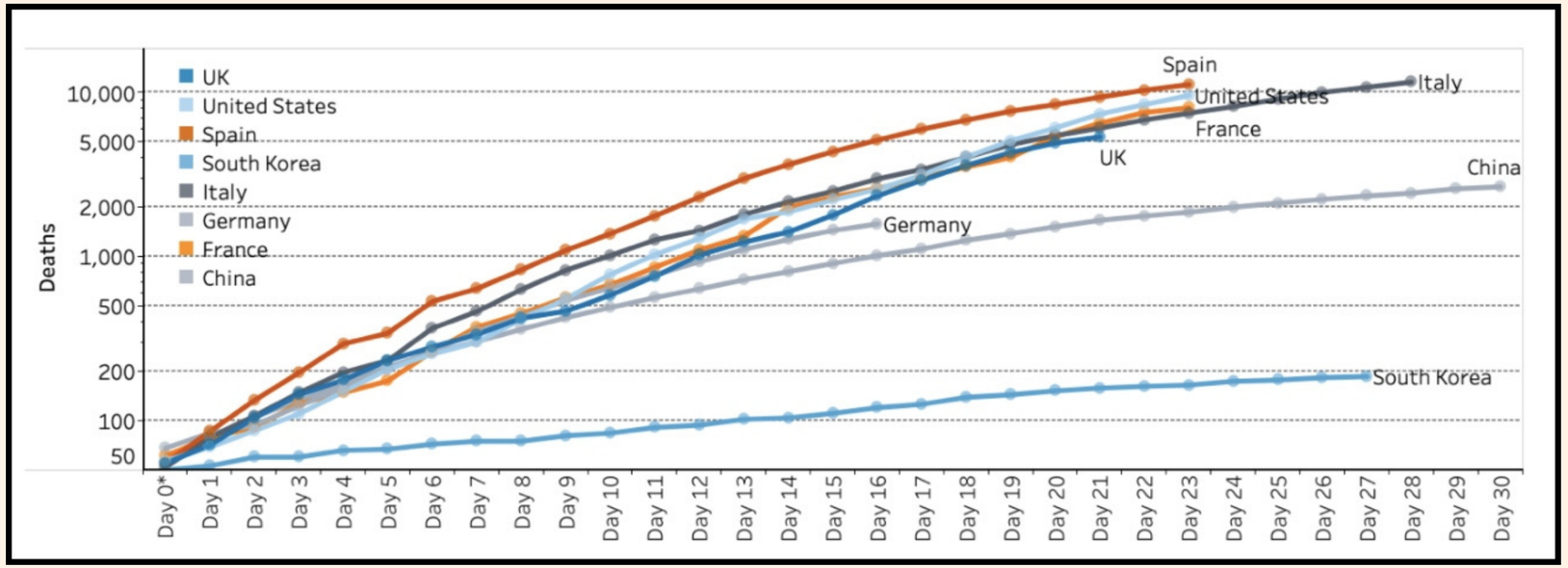

Appsilon
DATA SCIENCE
General Overview
Channels Overview
Days To Decision
Funnel Visualization
Hiring Funnel
Start date
2019-02-01
End date
2019-12-09
Group by
date
week
month
total
Job titles
Channels
Technical Copywriter
Frontend Engineer
Graphic Designer
Head of Growth
Head of Sales
Project Leader
Data Analyst
PM
Data Scientist
unknown
Software Engineer
Select All
Clear All
Theme
Navy
Timeline - how many candidates we had for each job position?
Number of applications
Week number since start date
Software Engineer
Unknown
Data Scientist
PM
Data Analyst
Project Leader
Head Of Sales
Head Of Growth
Graphic Designer
Frontend Engineer
Technical Copywriter
Candidates gender
Female
Male
Unknown
24%
4.8%
71.2%
Candidates nationality
Polish
Other
9.2%
90.8%

Shiny


DS Salary Explorer
About
Salary Scatter Plot
Salary Data Explorer
Salary Comparison Map
Top Recruiters
External Info
Salary Data Scatter Plot
States:
All
Salary Range:
12,000
250,000
350,000
Show data points
Total review counts by the number of fans
PAID_WAGE_PER_YEAR
250000
200000
150000
100000
50000
assistant professor
attorney
business analyst
data analyst
data scientist
management consultant
software engineer
teacher
JOB_TITLE_SUBGROUP
MIN:
12000
MEDIAN:
78312
MAX:
250000
Q1:
62926
MEAN:
78312
Q3:
100000

Cross Country Comparison

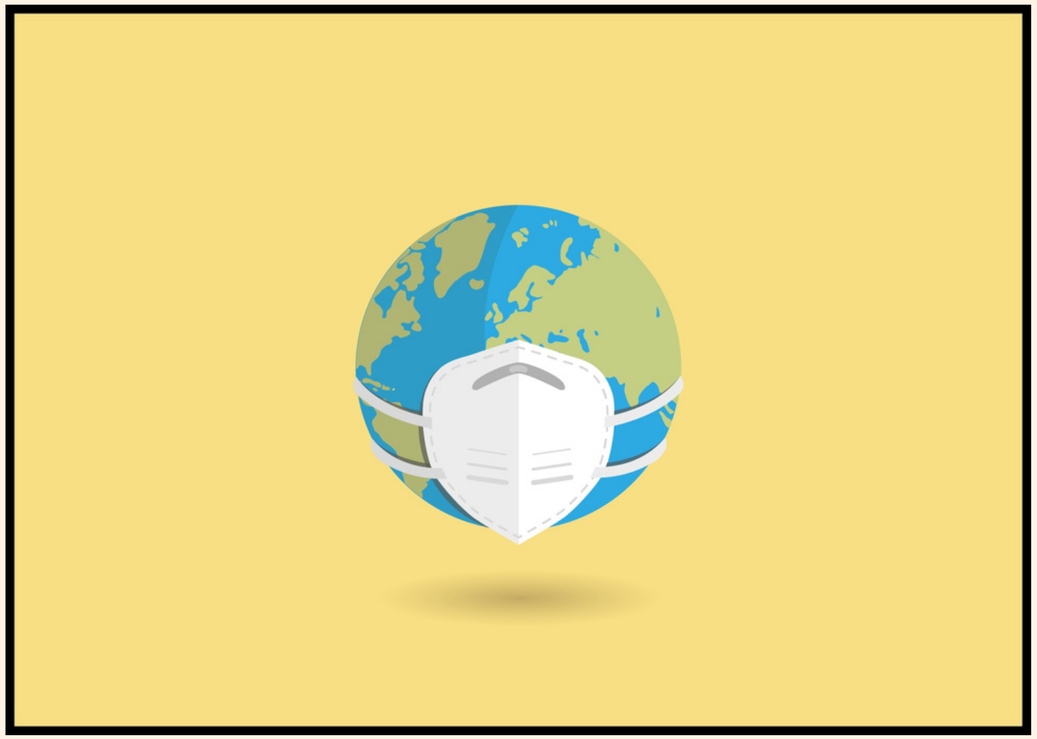




Applicable Areas

Finance

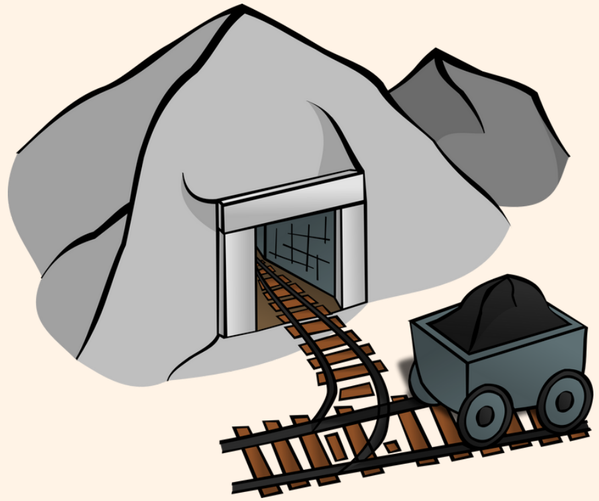
Applicable Areas



Disease

Applicable Areas

Literacy



Valuable Ores

Recources
(Oil etc.)

Extincting
Species

Agriculture

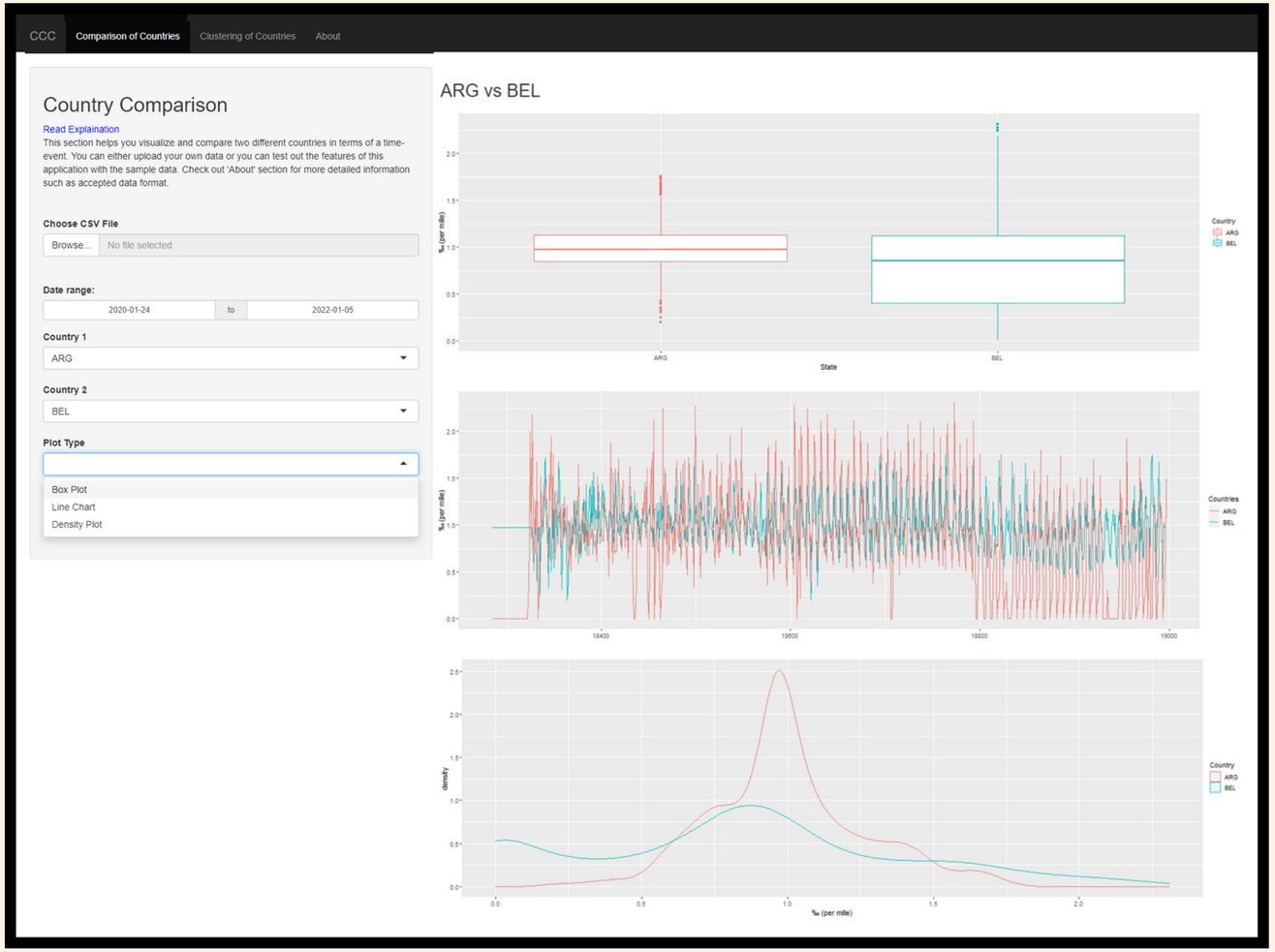
Example Data

COVID-19 GROWTH RATE DATA

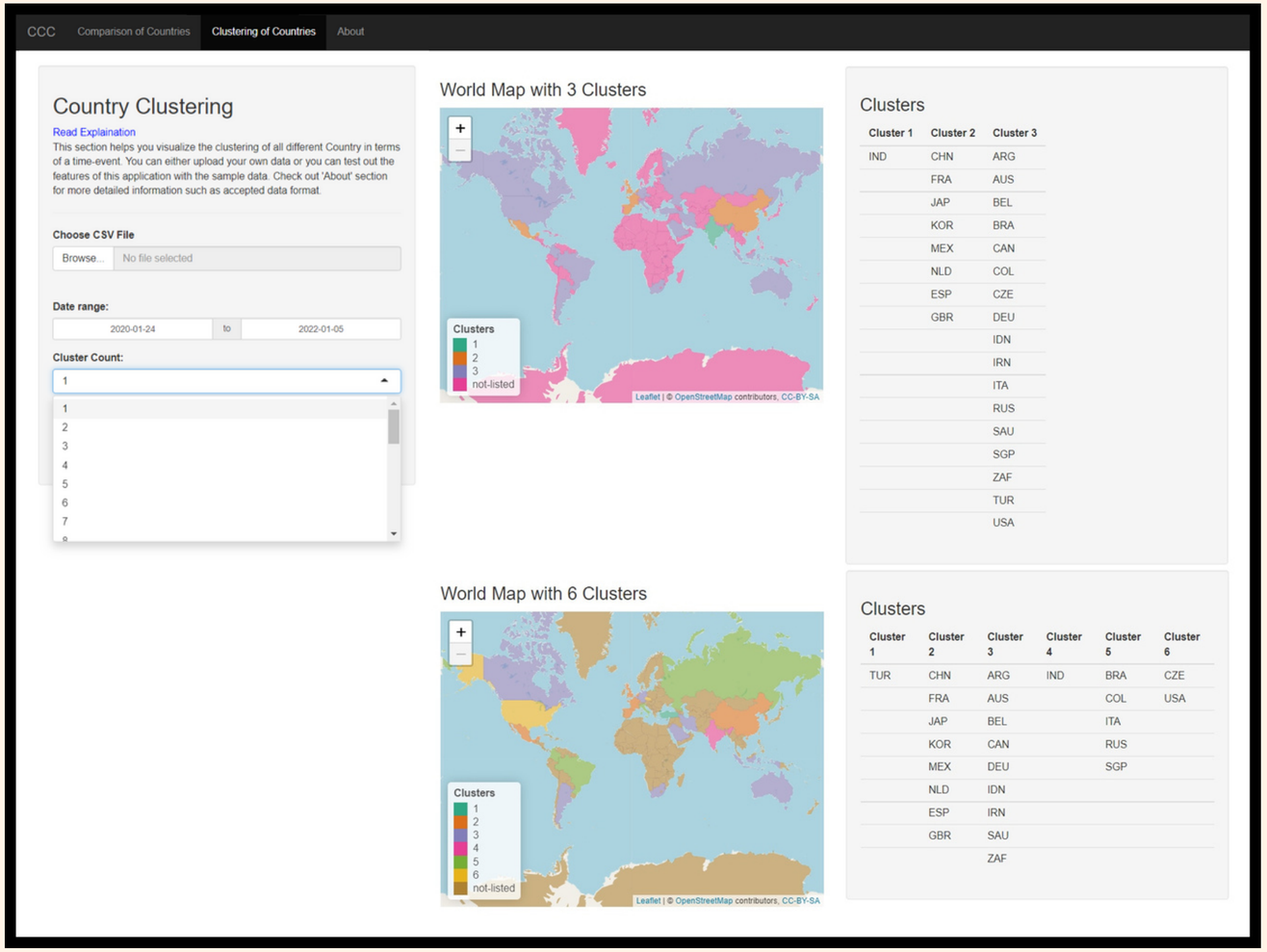
- 26 Countries
- 2 Years (713 Days)

Date	ARG	AUS	BEL	BRA	CAN	CHN	COL	CZE	FRA	DEU	IND	IDN	IRN	ITA	JAP	KOR	MEX	NLD	RUS	SAU	SGP	ZAF	ESP	TUR	GBR	USA
1_24_202	0	0	0	0	0	2.91579	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
1_25_202	0	0	0	0	0	1.75451	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_26_202	0	0	0	0	0	1.37654	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_27_202	0	0.25	0	0	0	1.1988	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
1_28_202	0	0	0	0	0	3.2818	0	0	0	3	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
1_29_202	0	0	0	0	0	0.2196	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_30_202	0	3	0	0	0	3.55363	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1_31_202	0	0	0	0	0	0.80867	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
2_1_2020	0	0	0	0	0	1.25768	0	0	0	3	0	0	0	0	1.25	0.14286	0	0	0	0	1	0	0	0	0	0
2_2_2020	0	0	0	0	0	2.26855	0	0	0	0.66667	0	0	0	0	0	3	0	0	0	0	0.66667	0	0	0	0	0
2_3_2020	0	0	0	0	0	0.65119	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2_4_2020	0	0	0	0	0	1.29326	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2_5_2020	0	0	0	0	0	0.93535	0	0	0	0	0	0	0	0	0.5	3	0	0	0	0	0.66667	0	0	0	0	0
2_6_2020	0	0	0	0	0	0.84302	0	0	0	0	0	0	0	0	0	1.33333	0	0	0	0	0	0	0	0	0	0

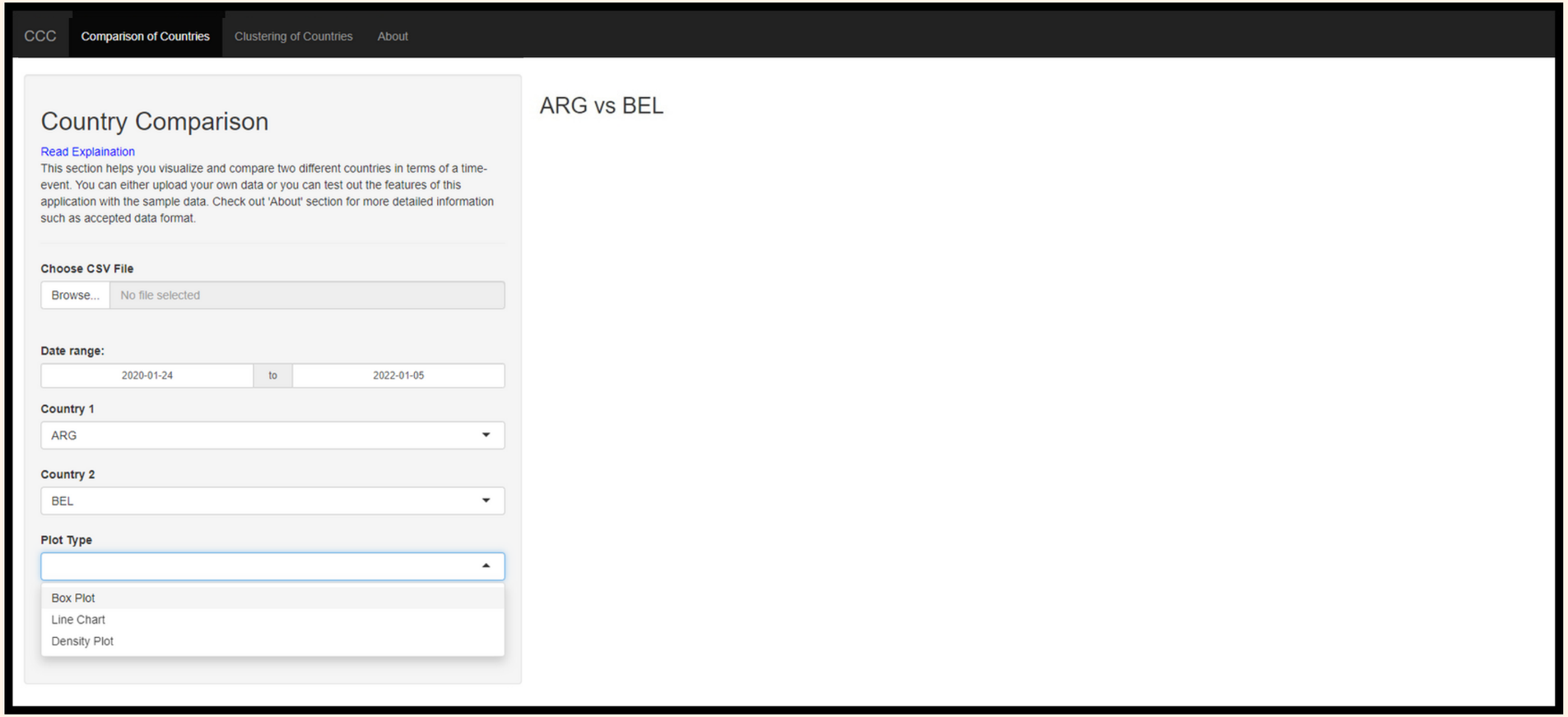
A Shiny App :
Comparison and Clustering of Countries



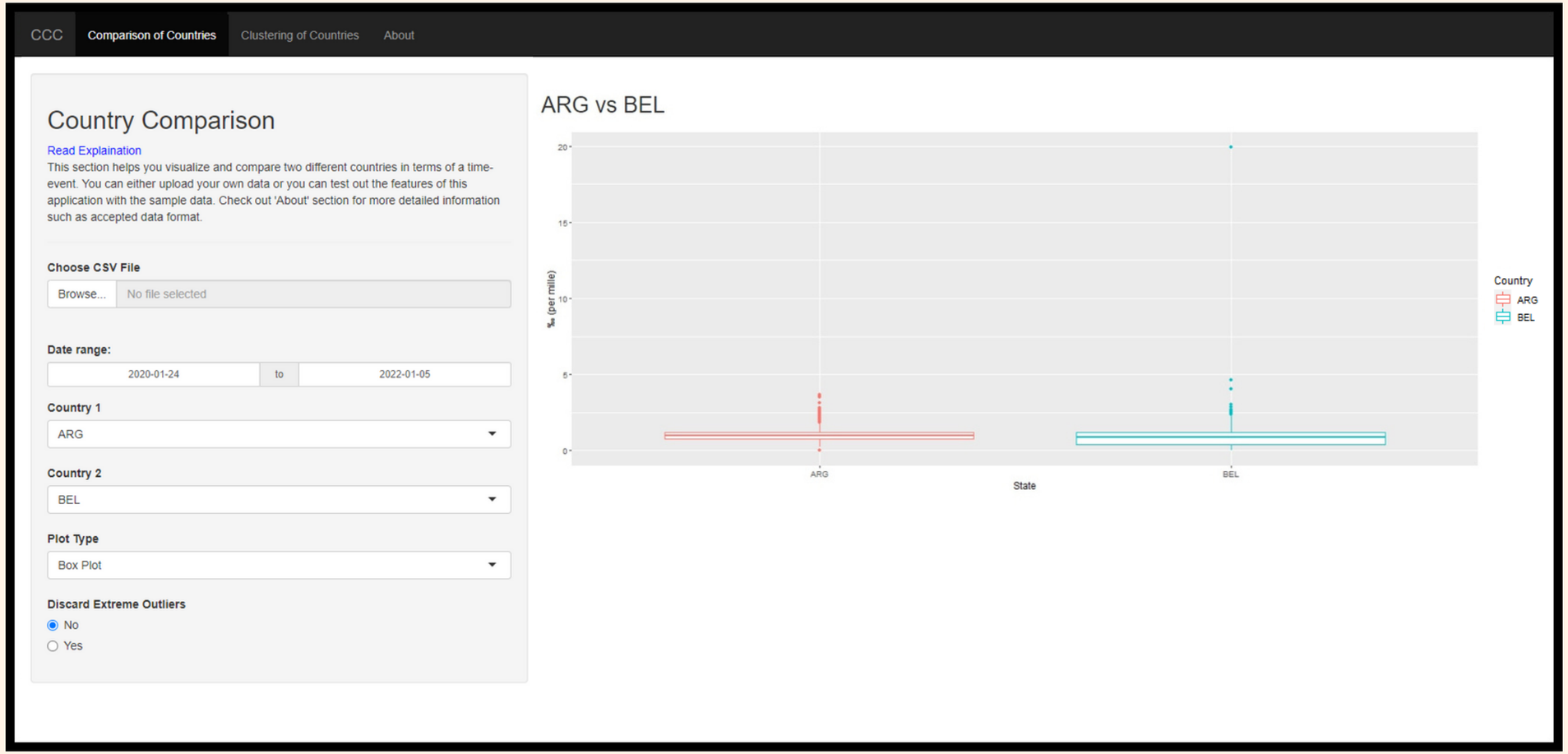
A Shiny App :
Comparison and Clustering of Countries



Comparison

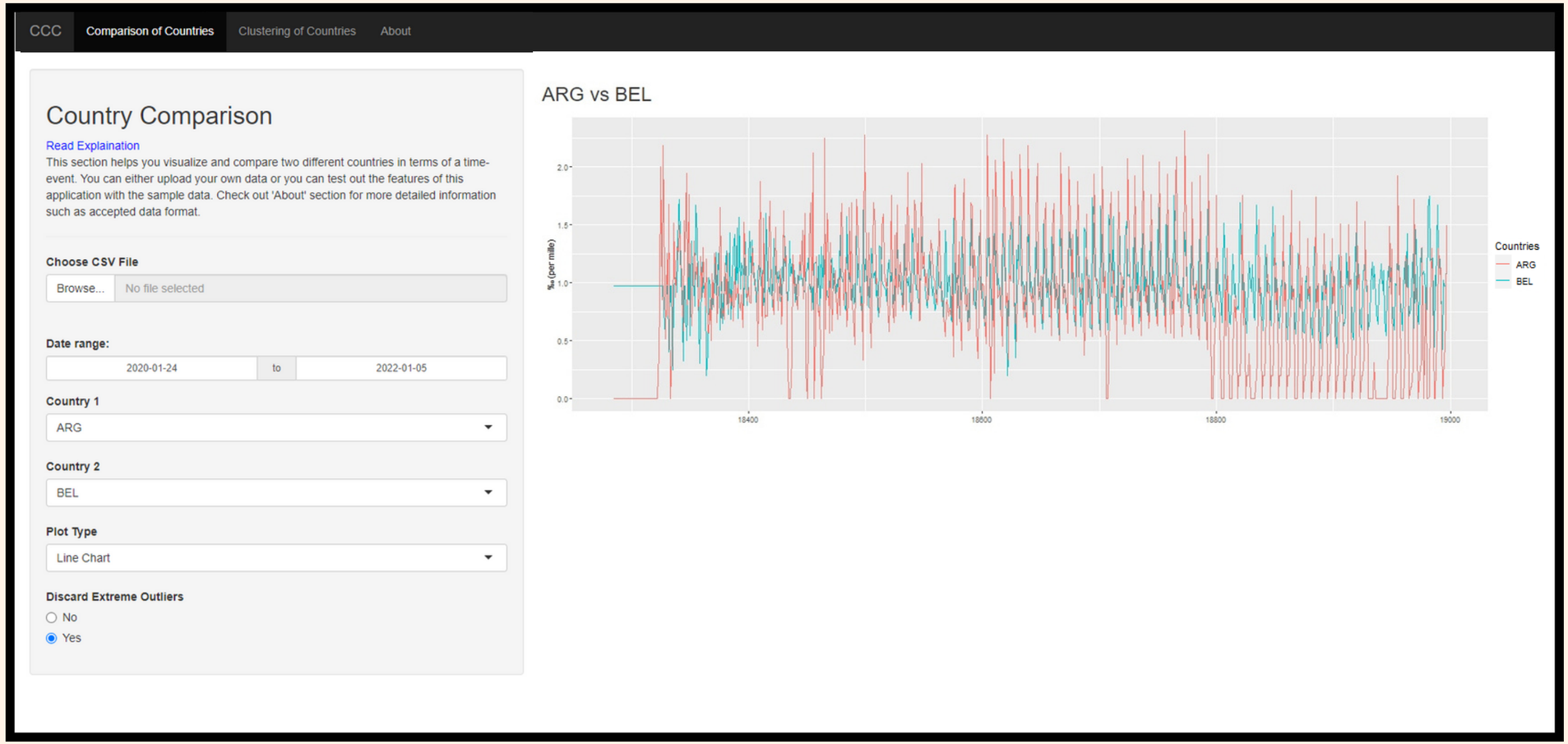

CCC
Comparison of Countries
Clustering of Countries
About
Country Comparison
Read Explaination
This section helps you visualize and compare two different countries in terms of a time-event. You can either upload your own data or you can test out the features of this application with the sample data. Check out 'About' section for more detailed information such as accepted data format.
Choose CSV File
Browse...
No file selected
Date range:
2020-01-24
to
2022-01-05
Country 1
ARG
Country 2
BEL
Plot Type
Box Plot
Line Chart
Density Plot
ARG vs BEL

Comparison (Box plot)

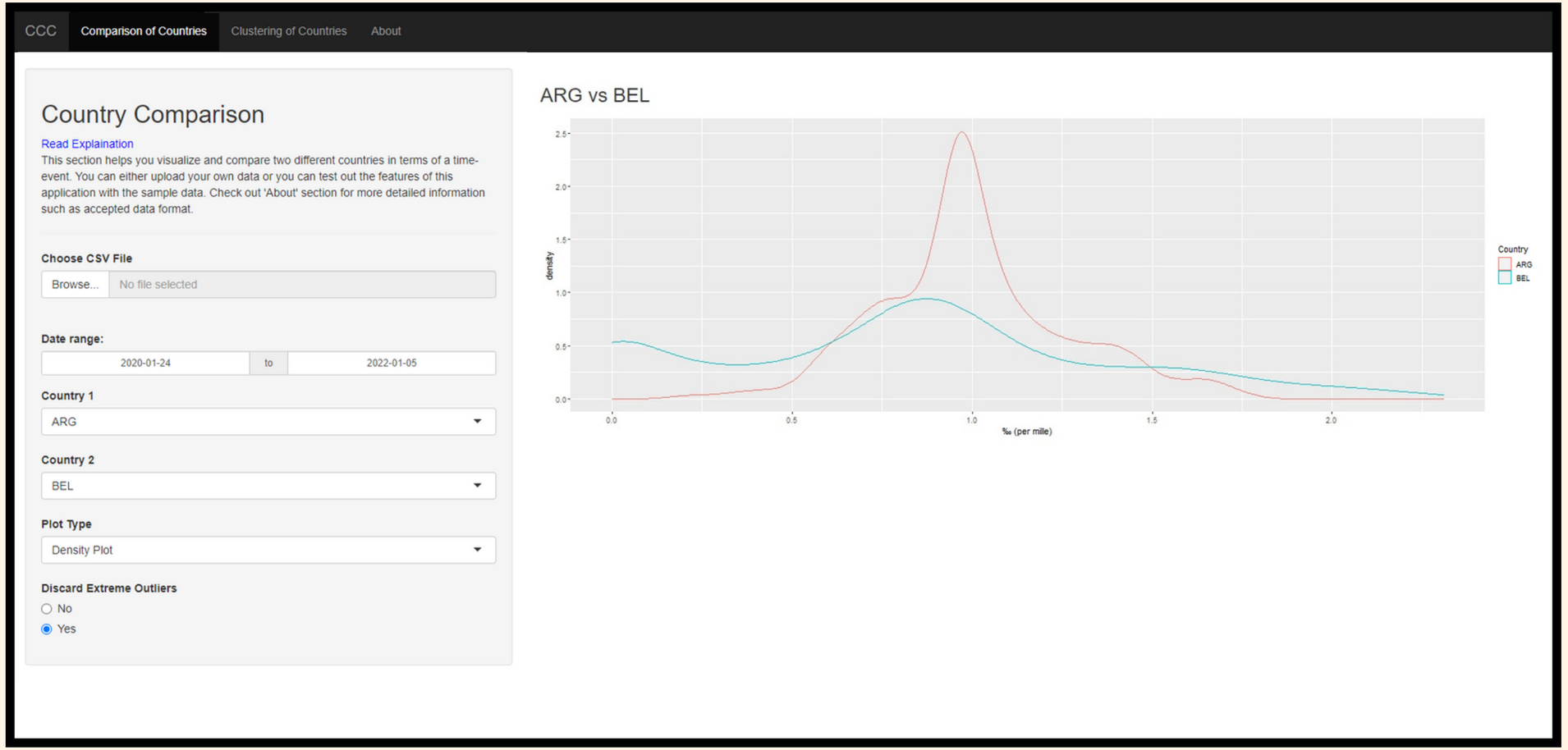


Comparison (Box plot, outliers removed)

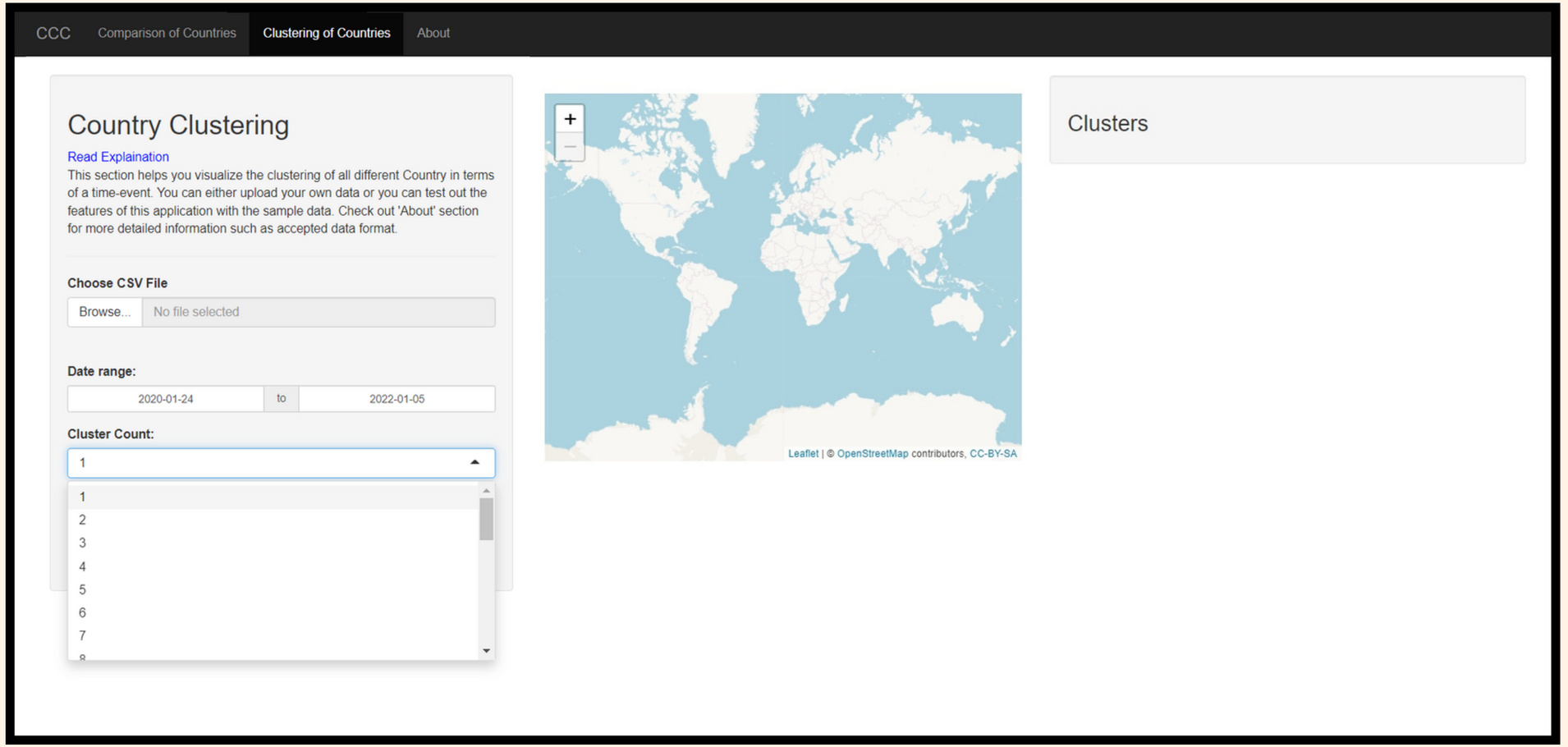
Comparison (Line chart, outliers removed)



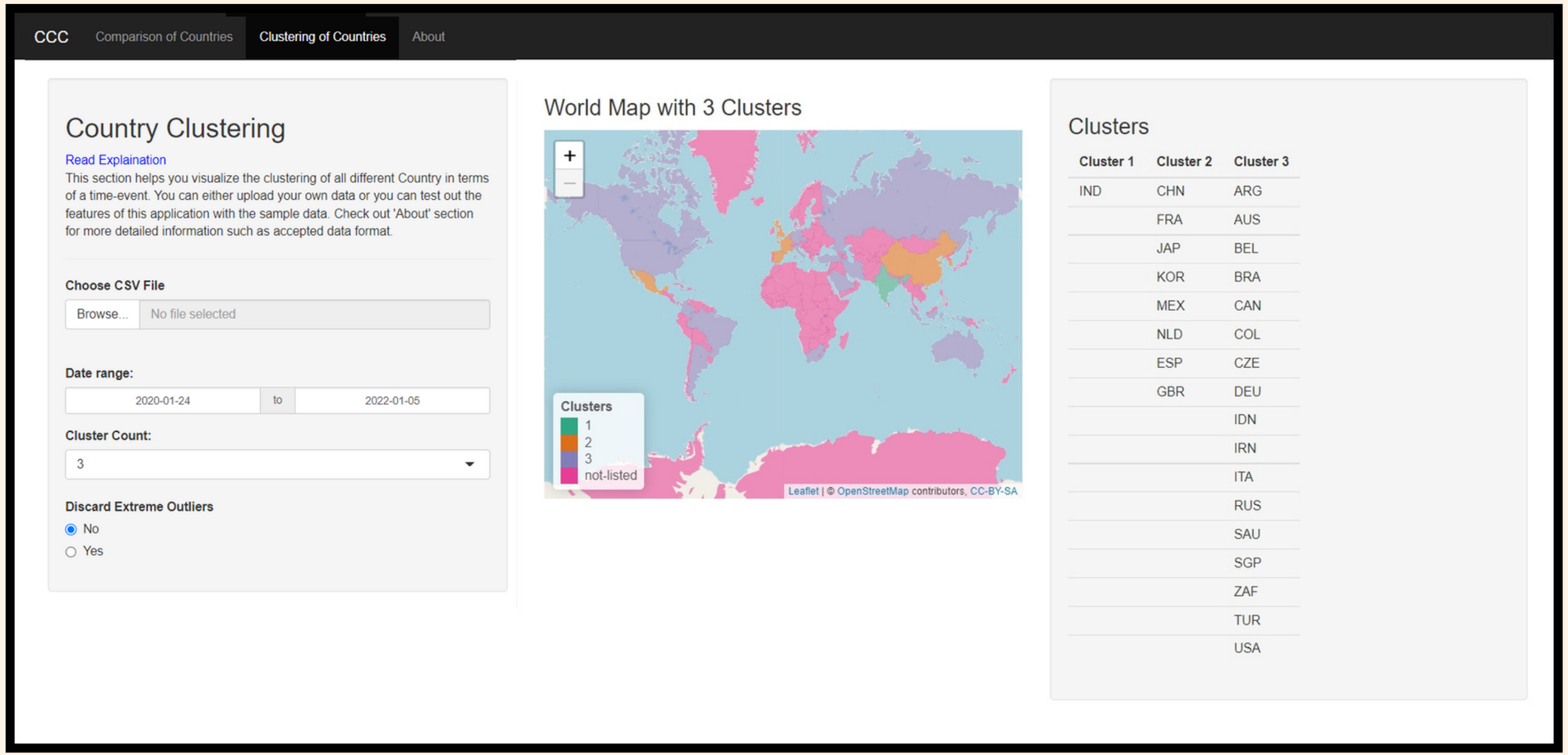
Comparison (Density plot, outliers removed)



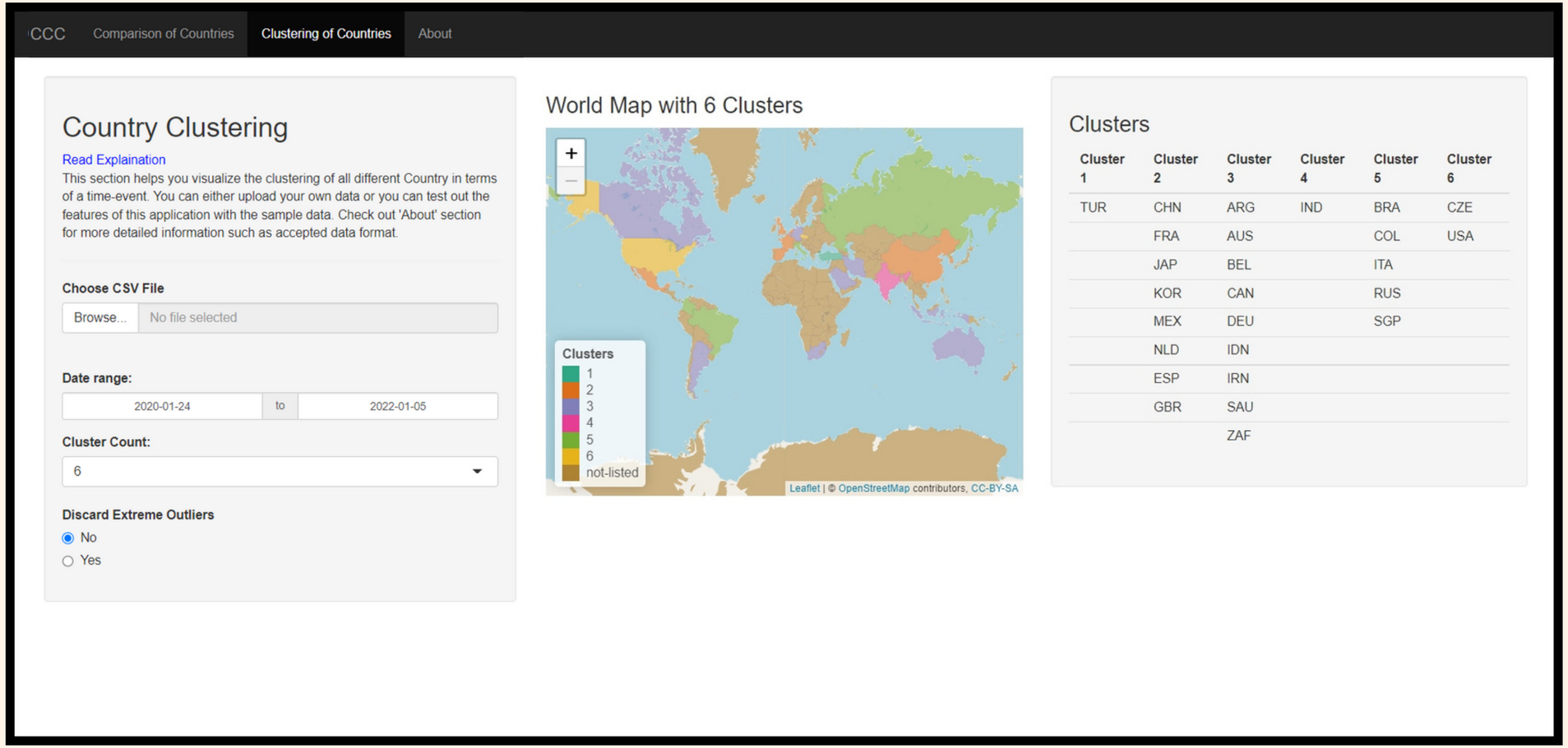
Time Series Clustering



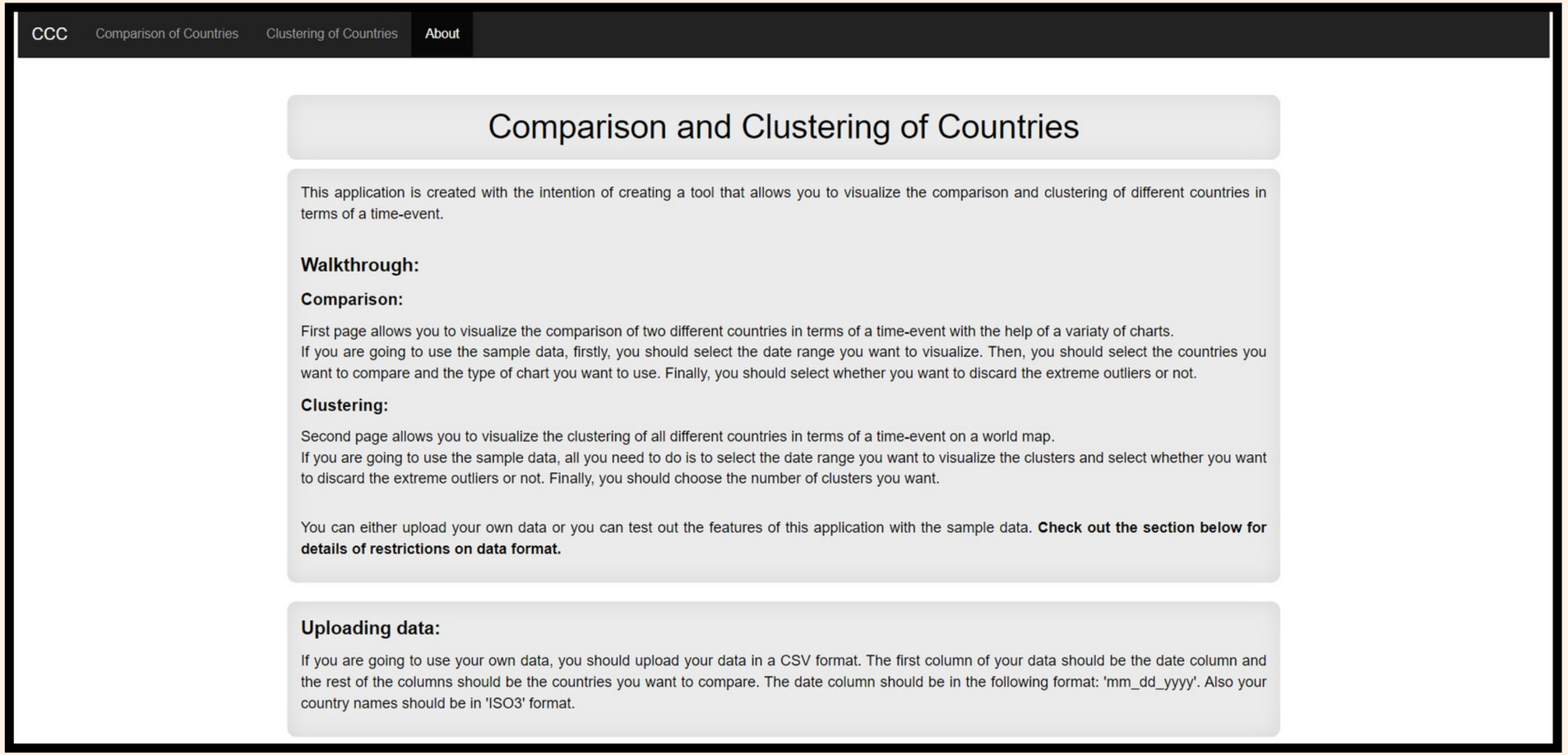
Time Series Clustering (3 clusters chosen)



Time Series Clustering (6 clusters chosen)



About


CCC
Comparison of Countries
Clustering of Countries
About
Comparison and Clustering of Countries
This application is created with the intention of creating a tool that allows you to visualize the comparison and clustering of different countries in terms of a time-event.
Walkthrough:
Comparison:
First page allows you to visualize the comparison of two different countries in terms of a time-event with the help of a variaty of charts.
If you are going to use the sample data, firstly, you should select the date range you want to visualize. Then, you should select the countries you want to compare and the type of chart you want to use. Finally, you should select whether you want to discard the extreme outliers or not.
Clustering:
Second page allows you to visualize the clustering of all different countries in terms of a time-event on a world map.
If you are going to use the sample data, all you need to do is to select the date range you want to visualize the clusters and select whether you want to discard the extreme outliers or not. Finally, you should choose the number of clusters you want.
You can either upload your own data or you can test out the features of this application with the sample data. Check out the section below for details of restrictions on data format.
Uploading data:
If you are going to use your own data, you should upload your data in a CSV format. The first column of your data should be the date column and the rest of the columns should be the countries you want to compare. The date column should be in the following format: 'mm_dd_yyyy'. Also your country names should be in 'ISO3' format.

References

- https://shiny.rstudio.com/
- https://ubidi.shinyapps.io/covid19world/
- https://petolau.shinyapps.io/coronadash/
- https://shubhrampandey.shinyapps.io/coronaVirusViz/_w_59897ec7/
- https://nycdatascience.com/blog/student-works/us-gdp-shiny-app/

Thank you..