AI Investment of 62 Countries: An RShiny App

Ishita Sarraf

2024-03-09

Academic Honesty Statement

Name(s) of all authors: Ishita Sarraf

Assignment name: Individual Project

Assignment due date: March 11, 2024

Written/online sources used: Mentioned in resources

Help obtained (Acknowledgments): Prof. Jimenez, Elijah (mentor)

"I/we confirm that the above list of sources is complete AND that I/we have not talked to anyone else about the solution to this problem."

Contents

Project Overview

The *purpose* of this RShiny app was created to investigate the AI investment by 62 countries. As AI becomes more and more powerful and relevant in today's world, it is important to look at its growth in various countries. My *goal* is to create visualizations that would help us look at different factors of a country and their investment in AI to identify more powerful countries.

There are 3 different graphs and 1 world map as visualizations.
Graph 1: Bar Plot
This visualization gives information about the countries' geographical regions and their income group, political power, and political regime.
Graph 2: Bubble Plot
This visualization gives us information about the countries' investment in R&D for AI, with the bubble size determined by various quantitative variables.
Graph 3: Spider Plot
This visualization gives us information about each country's numerical scores for the quantitative variables.
World Map:
The interactive world map allows us to see the *AI Global Rank* (total score) and income group for each country.

As seen from all the visualizations, USA seemed to have gained the highest total score for AI investment.

Audience: The audience for this Shiny app can be any student or adult interested in learning about AI Investment by these 62 countries.

Data Description

There are 2 datasets used in this project. The first is the dataset that consists of the AI investment for the 62 countries and the second is the dataset that has the spatial coordinates needed to create the map.

Dataset 1:
This dataset (62 rows x 13 columns) was collected from a Kaggle source [1] and consists of 13 indicator variables, out of which 8 are quantitative variables and 5 are categorical variables. The 8 quantitative variables were calculated by Tortoise Media [2] via weighting and summarizing 143 other indicators.

Variable Name	Variable Type	Description
Country	character	name of country
Talent	double	indicator focuses on the availability of skilled practitioners for the provision of artificial intelligence solutions
Infrastructure	double	indicator focuses on the availability of skilled practitioners for the provision of artificial intelligence solutions
Operating.Environment	double	indicator focuses on the regulatory context, and public opinion surrounding artificial intelligence

Variable Name	Variable Type	Description
Research	double	indicator focuses on the extent of specialist research and researchers; investigating the amount of publications and citations in credible academic journals
Development	double	indicator focuses on the development of fundamental platforms and algorithms upon which innovative artificial intelligence projects rely
Government.Strategy	double	indicator focuses on the depth of commitment from national government to artificial intelligence; investigating spending commitments and national strategies
Commercial	double	indicator focuses on the level of startup activity, investment and business initiatives based on artificial intelligence
Total.score	double	the AI global rank
Region	character	region the country belongs to
Cluster	character	how powerful is the country politically
Income.group	character	how much income the country has
Political.regime	character	what political policy does the country have

Here are the first 3 rows of the dataset.

```
##                       Country Talent Infrastructure Operating.Environment
## 1   United States of America 100.00          94.02                 64.56
## 2 People's Republic of China  16.51         100.00                 91.57
## 3             United Kingdom  39.65          71.43                 74.65
##   Research Development Government.Strategy Commercial Total.score       Region
## 1   100.00      100.00               77.39     100.00      100.00     Americas
## 2    71.42       79.97               94.87      44.02       62.92 Asia-Pacific
## 3    36.50       25.03               82.82      18.91       40.93       Europe
##                Cluster Income.group  Political.regime
## 1         Power players         High Liberal democracy
## 2         Power players Upper middle  Closed autocracy
## 3 Traditional champions         High Liberal democracy
```

Dataset 2:

This dataset is a huge dataset (242 rows x 169 columns) that consists of all the country names in the world with their respective geographical spatial coordinates. It was collected by Natural Earth Data [3]. I used this dataset and performed a `left_join()` function to join the spatial data with the 62 countries I have in order to create the interactive world map using `leaflet()`.

Insights

I wanted to see what trends exist between countries and their different AI investments in the various quantitative and categorical variables. Based on the visualizations we can see the following:
1. USA has the highest value of investment in Research and Development (R&D), followed by China.
2. If a country has is a high income.group, it is not necessary that it has a high total score. For example- Italy comes under a `high` income group but its total score is 24.45 and India comes under a `lower middle` income group and has a total score of 30.36.

Data Cleaning

For creating of the map, I had to clean the data such that I could join the 2 datasets together, created color palettes for the map, created labels and popups that would make the map interactive, created polygon geometries, and finally, made the dataset reactive in order to appropriately use it with shiny. Using leaflet made it more easy for a user to zoom and move around in the map.

For creating the bar graph, I simply used the user input to plot the x-axis and used different `theme()` functions in `ggplot` to make the bar plot more easy to read.

For creating the bubble plot, I created the sizes using user inputs and added a label to make it informative for the user, and added different options for the bubble sizes for the user to make it more animated. I converted the bubble plot to an interactive plot using `plotly` so that a user could zoom and play around with the plot.

For creating the spider plot, I created plot using `radarChart()` so that I could accurately create the spider web and then edited the font sizes and colors to make it more readable.

I then combined all these visualizations and used them to create a shiny app [3].

Visualization Designs

To describe the visualizations I will use the `What-Why-How` Framework described in the book by Tamara Munzer [4].

1. Interactive Map
What- Spatial Geometry
Why - Query: Compare
This idiom was used to compare if the total score of a country can have any connection with the income group.
How - Using color (channel) and spatial region (mark)

2. Bar Plot
What- Flat Table
Why - Analyze: Present
This idiom helps to present the information of region with income group, political power, and political regime.
How - Using color hue (channel) and area (mark)

3. Bubble Plot
What- Multidimensional table
Why - Analyze: Discover
This idiom helps to discover the research and development of various countries and then see the effect of various elements - Talent, Operating Environment, Government Strategy, Commercial, Total.score
How - Using color saturation (channel) and area (mark)

4. Spider Plot
What- Flat Table
Why - Analyze: Enjoy
This idiom is just to enjoy the look of a spider plot and to see the numeric variables displayed in a different way for each country.
How - Using area (channel) and containment (mark)

Possible Additions

1. More visualizations that help us draw more insights into the AI Investment, for example - a graph that can depict which `regions` have `higher total scores`.

2. A better CSS framework that would make the website look more attractive.

References

[1] K. Meleshenko, May 2023, "AI Global Index," Source. [Online]. Available: https://www.kaggle.com/datasets/katerynameleshenko/ai-index

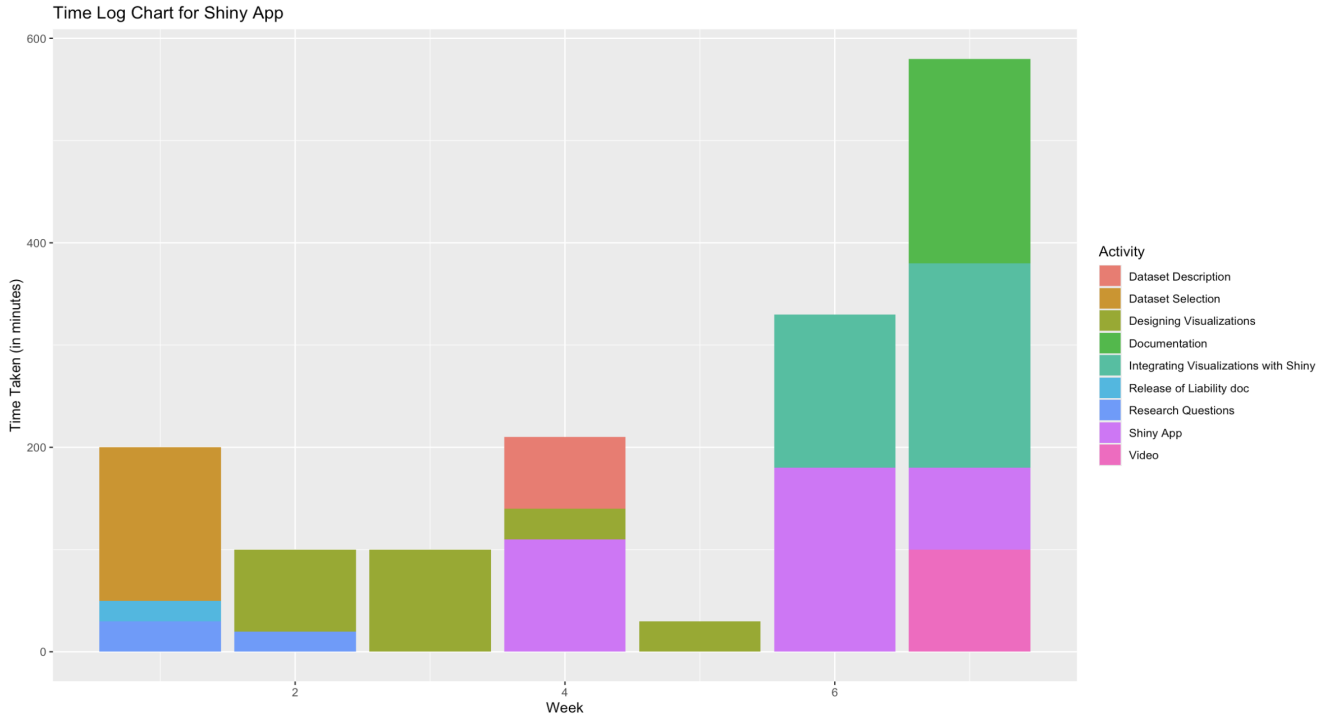
[2] Natural Earth, 2009-2024, "Admin 0- Countries," Source. [Online]. Available: https://www.naturalearthdata.com/downloads/50m-cultural-vectors/50m-admin-0-countries-2/

[3] C. Beeley, *Web Application Development with R Using Shiny*, 2nd ed. Birmingham, UK: Packt Publishing, 2016.

[4] T. Munzer, *Visualization Analysis & Design*, New York, NY, USA: CRC Press, 2015.

Appendix

1. Time Log Chart



2. Required for 100%

i) A particularly nice piece of code, for example, animations, extensions.
My code for the interactive map and the bubble plot

ii) Combine various datasets
I combine my AI dataset with the spatial dataset for creating the map by performing a left_join.