Fundamentals of Data Visualization Final Project

DataSet is from https://www.vizforsocialgood.com/join-a-project/2022/9/15/undp-accelerator-labs-network that wants wants to increase acces to affordable and clean energy around the world. The dataset contains 359 grassroots energy solutions from across different regions, demographics and energy sources. It also contains a page per region specifiyng which country belong to which region. The regions are Africa, Asia Pacific, Arab States Europe and the Commonwealth of Independent States and Latin America.

Each energy project is a row and has 35 columns of information (after removing the images columns), such as the country the energy solution is from, what kind of energy source does it use, If it's already a product or a prototype, the cost of the solution, if the solution is DIY, what sustainable goals its addresing, if its a clean cooking aplication and so on.

Questions to ask the data.

Viz4social good already gives the questions to ask the data, they are many and broad so I will focus on a couple of them, as they are good questions and what the organization needs from the visualization. Such questions are:

- Where are the solutions coming from? What is their distribution per country & per region?
- What type of energy source is more prevalent, what is less? Are there differences per region, and why?
- What are global commonalities across solutions; what are typical applications & use cases for solutions? Are there patterns that emerge when looking at the distribution per country & per region?
- What overall challenges are the solutions addressing or contributing to overcome?
- Which Sustainable Development Goals are the solutions advancing in particular, and how?
- Looking at the use case of clean cooking solutions, what is their prevalence, distribution, and source of energy?

If applicable / information is available*:

- How are the solutions solving a specific community problem? What are the solutions addressing or contribute to solving? (if applicable)
- How can we display the more quantitative information (ratio of IP vs DIY solutions; Prototype vs Product) in an appealing way that signifies the availability of solutions in country? (if applicable)

The presentation with visualizations ends in October 30 so there are no current visualizations for this data available yet.

Tasks

The tasks pusued will be:

1. Identifying the geographic areas the solutions are comming from and the types of energies used geographically.

    - The goal of the task is learning from which areas are more solutions comming and from which less, maybe to see where would it be best to invest in such solutions and where to help develop new solutions.
    - The means or how would be comparing the number of solutions per country and regional area.
    - The task seek to learn how many solutions exist per region.
    - It operates on the data gathered by the mappers that says where each solution comes from.
    - Workflow would be to be able to zoom in by region to see solutions by country.
    - The task would be executed by a general user.
2. Identifying if they are general patterns of the type of energies used geographically for Clean Cooking solutions.

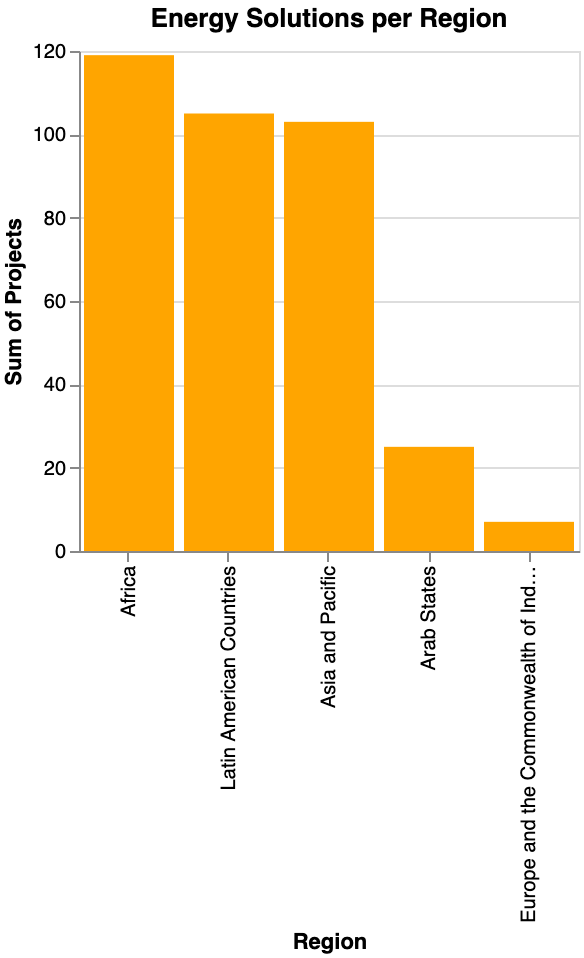
    - To see if some energies are more prevalent in different regions and what countries in the region could need more of it.
    - By looking at the types of energies in a map.
    - Where are energies more prevalent.
    - Same data as before plus the data column specifing the kind of energies.
    - The workflow
    - The task is also executed by the general user.

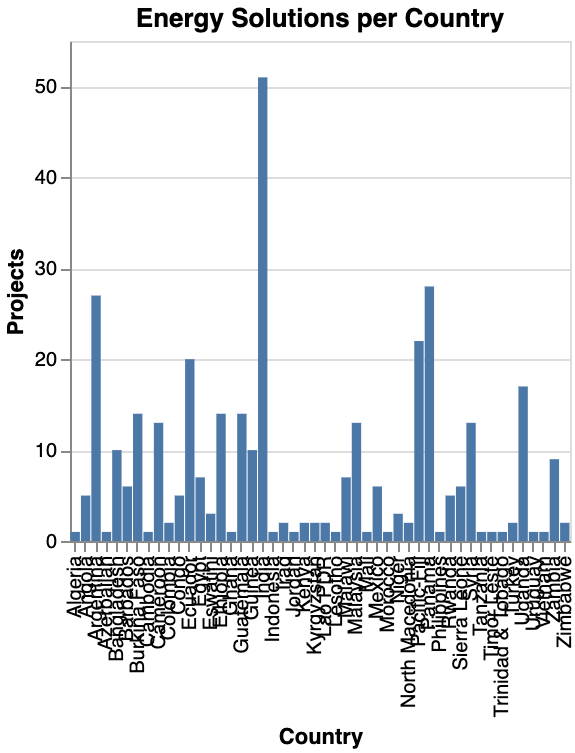
1. Looking at the ratios of DYI vs IP and Prototype vs Product by region and country.

    - To see what kind of projects are more prevalent by region and country.
    - The means would be looking at the ratios of both things by country in a scatter plot.
    - Worlflow is to see if there is any trend by regions.
    - The task would be executed by general user.

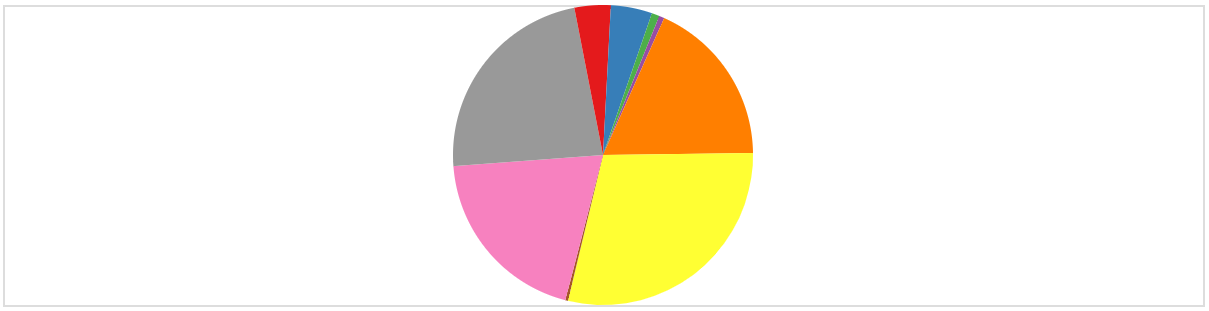
    A. Look at sustainable goals approached by the solutions.

    - Goal. Observe what sustainable goals are more approached by region.
    - The means by a scatter plot.
    - Workflow you can check by region or country.
    - Task is done by general user.

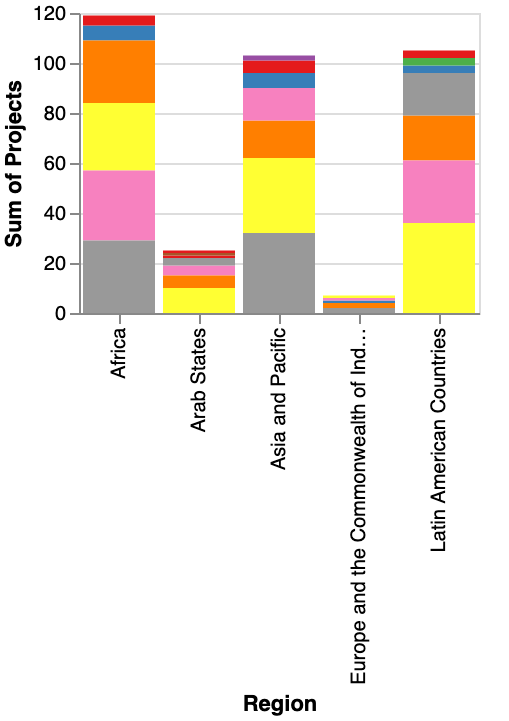

Energy Solutions per Region
Sum of Projects
120
100
80
60
40
20
0
Africa
Latin American Countries
Asia and Pacific
Arab States
Europe and the Commonwealth of Ind…
Region


Energy Solutions per Country
Projects
50
40
30
20
10
0
Country

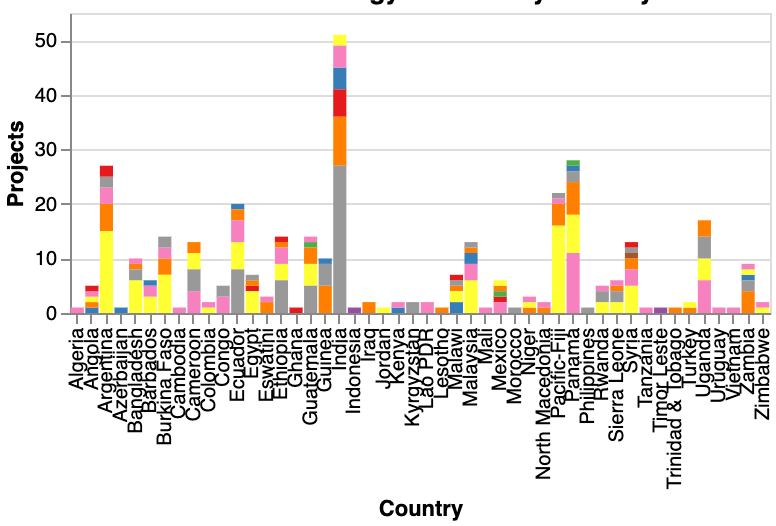
Percentage of Solutions Energy sources



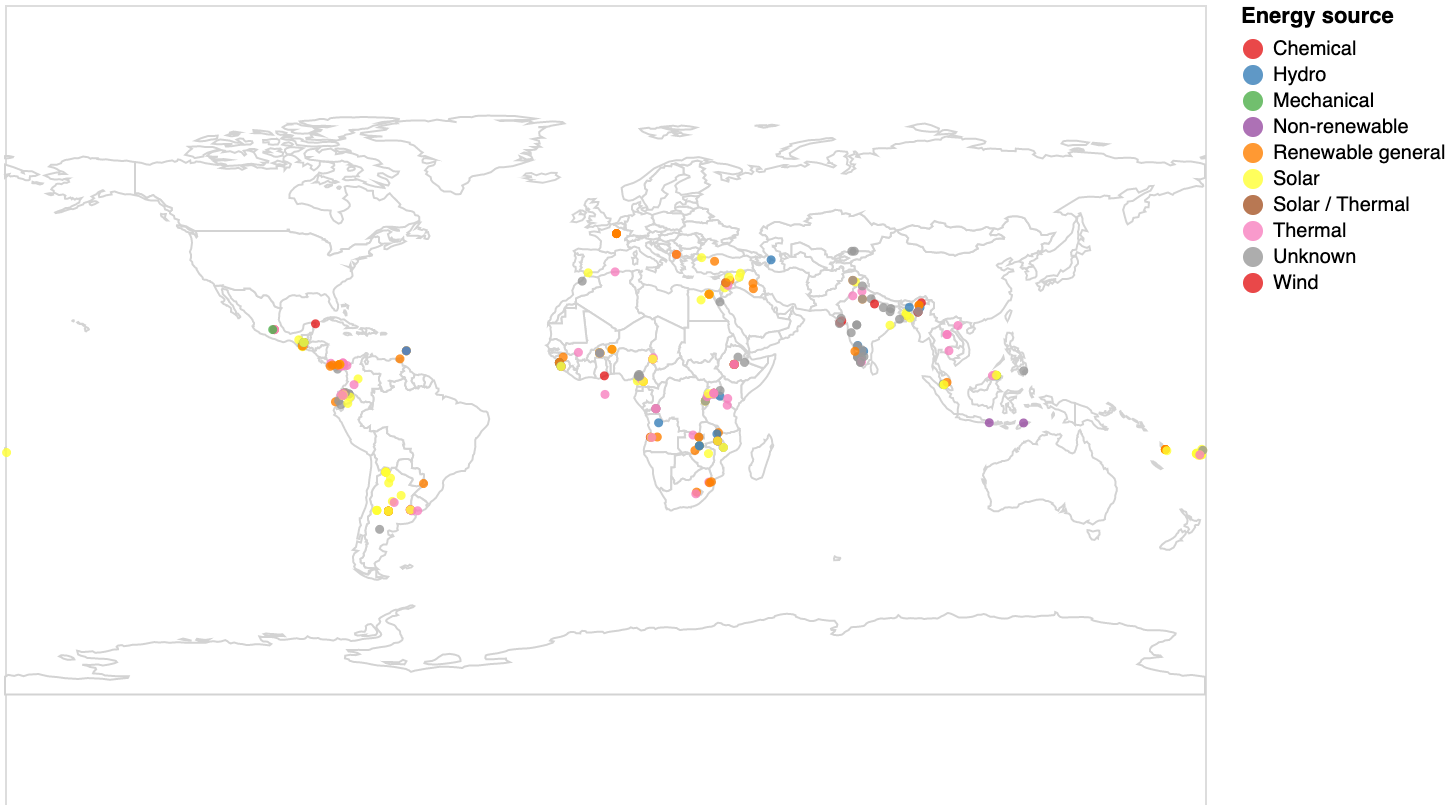
Solutions Energy Sources by Region

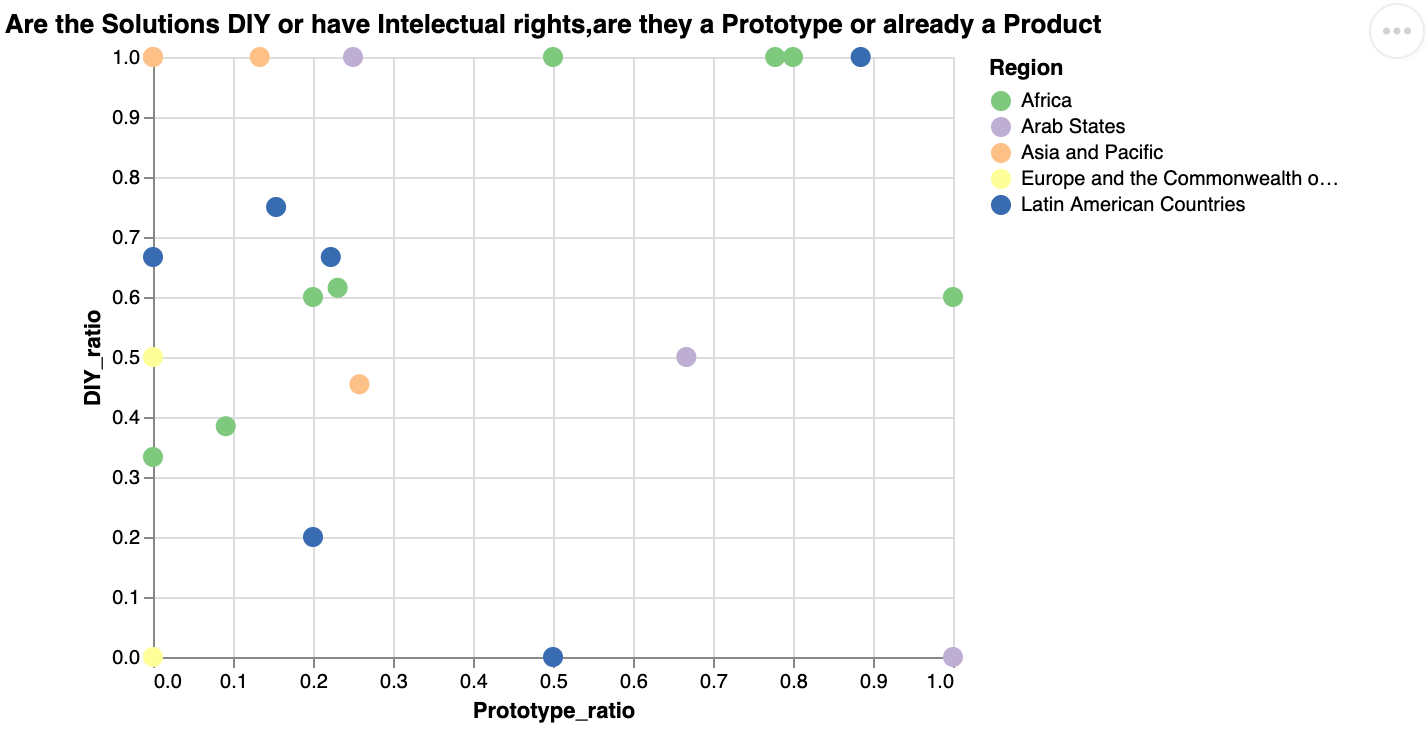


Solutions Energy Sources by Country

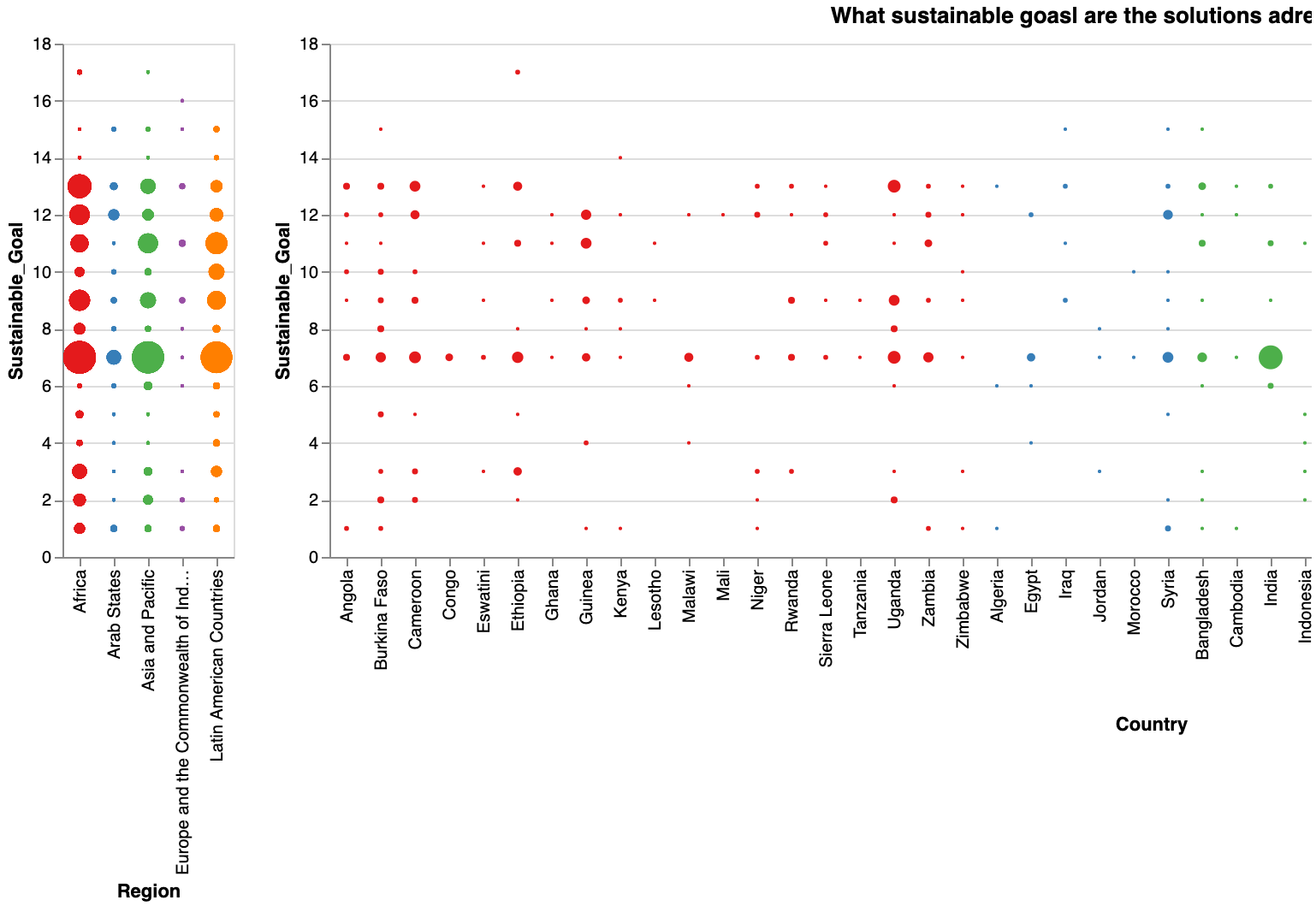


Where are the solutions geographically Located



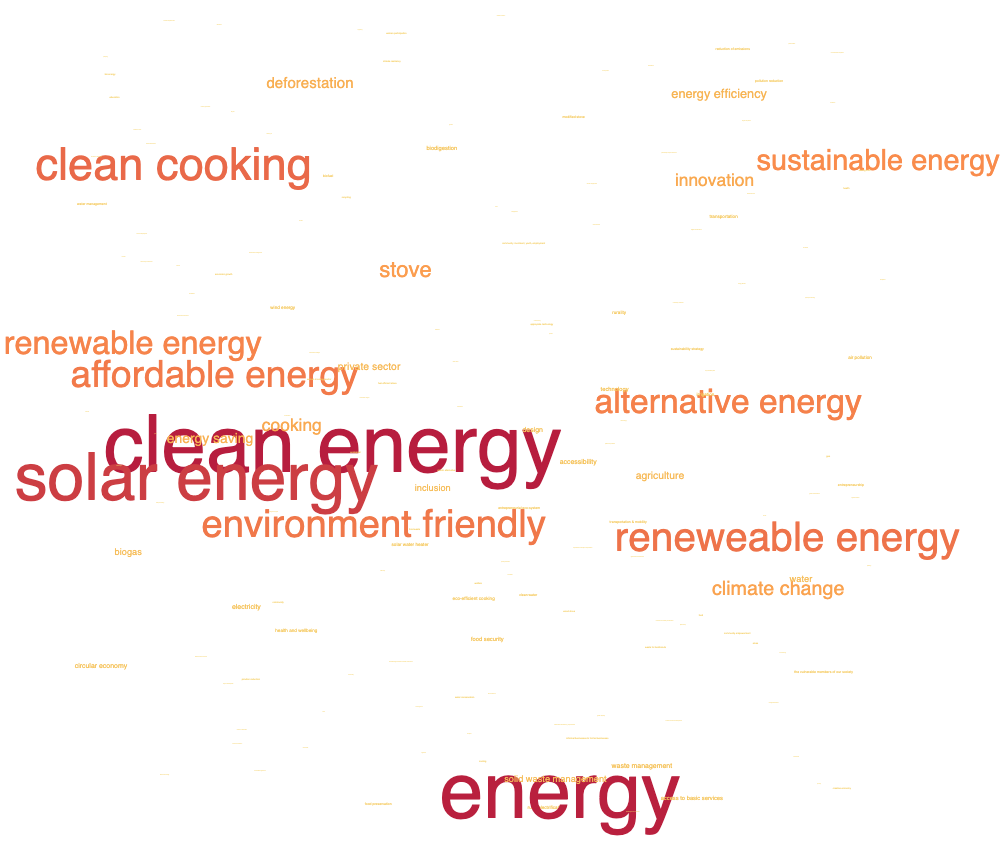


What sustainable goal are the solutions addresing the most by country and region?

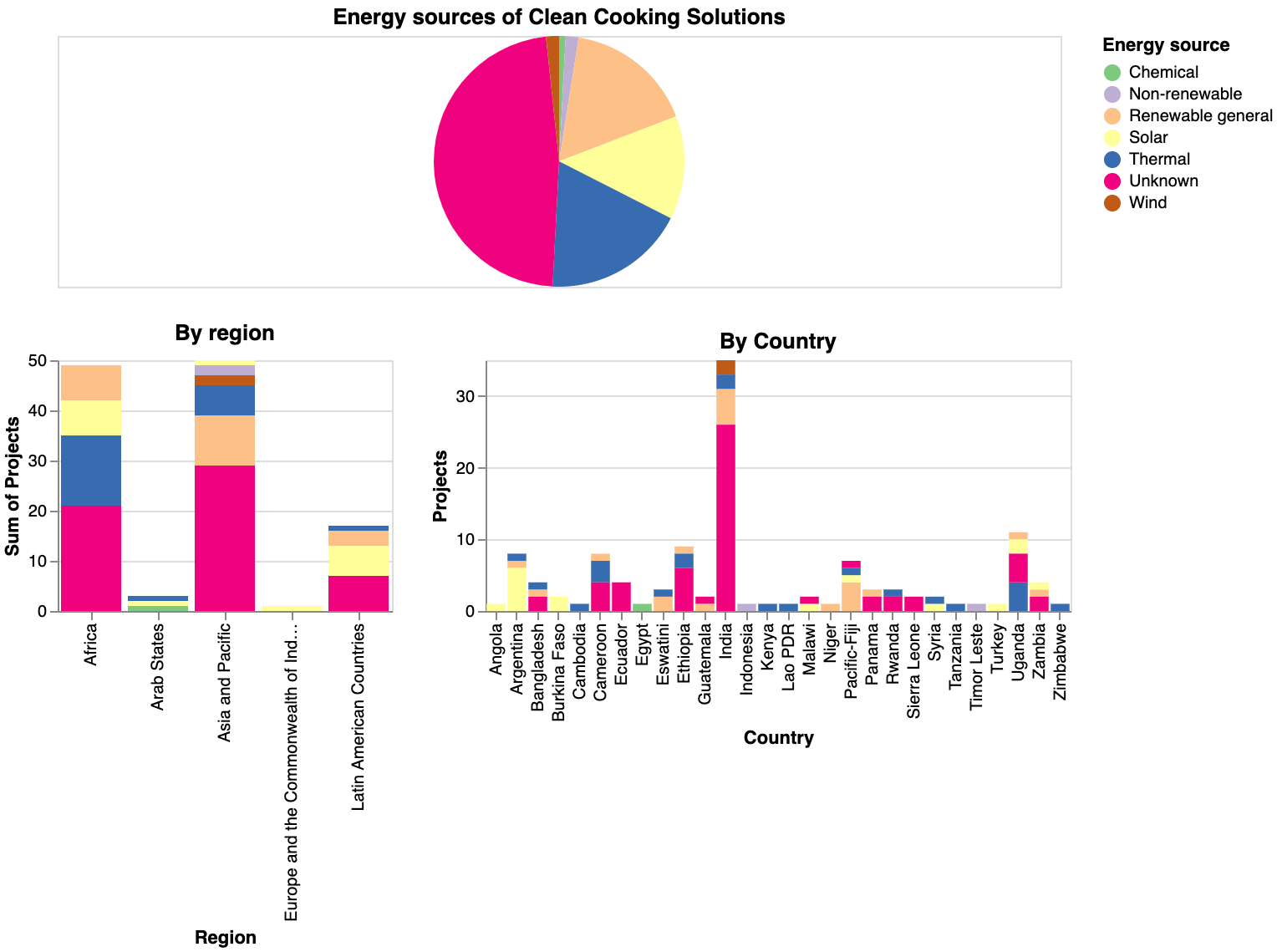


What are the key tags of the solutions for Sustainable Goal 7?

Sustainable Goal 7 is affordable and Clean Energy.



A third of the solutions are for Clean Cooking. What energy sources do the solutions use for such applications?



Key Elements of the Design

For the design I tried to focus in easy to understand plots. Used bar plots to compare the number of solutions with an interactive feature so you could see how the regions solutions decompose by country. At the start all the regions on the left are in orange and to the rights there is a bar for each country. If you click on the left plot on a region bar the right plot changes to only show the countries in such region. For each bar if you put the cursosr above it gives you the info of what country it is and how many solutions they registered.

To see what kind of energies where more used, I added stacked bars to the previousvisualization so the total length of the bar would still be the total number of solutions and the colors represent the fraction of the solutions for a given energy type. In this case the tooltip gives you both the total number of solutions and how many solutions of this specific contry. At the top of the two stacked bars I added a circle plot showing the total percentage of each solution and the toolbar tells you the kind of energy and percentage of such energy.

To complete this viz I added a map of the world with the points based on the latitued and longitude the solutions are registered, this in order to get a better approach of the geography of where the solutions are located.

I repeated the solution of the stacked bars to the data of the clean cooking solutions to be able to compare it with the general data.

For DIY and prototypes I choose a scatter plot as the medium and one point per country with the color representing the region and being able to select by region to see if there is any trend by region. The y axis represent the fraction of DIY/(DIY+Intelectual Property) while the y axis is the ratio of Prototypes/(Prototypes+Product). I choose this design to sse if there was any trend between thoose two aspects of the projects or any regional trend.

Finally to see which sustainable goals where more approached I did two scatter plots one by region that if you click would see the one by country. I use color to group by region and size of the mark to show how many solutions had the goal. From the plot it was clear number 7 was the most approached so I lloked at a word plot to see what kind of tags where more often added to solutions addresing sustainable goal number 7.

Evaluation Approach

For the evaluation approach as there wont be a lot of people evaluating I though the think aloud would be the best approach and to see what they think of the different ways the data is visualized and what insights they get from it. Three persons interacted with the visualizations, and I asked them to say everything they where thinking and discovering while they interacted with the visualizations.

There were 2 females and a male, two above 70 and the other one in her 40s. For the elder ones tha interaction wasn't intitive at all, they didn't click on the visualizations even when told and had a bit of a hard time with the colors. They still got the main insights of the graohs showing the numbers of solutions by country and region, all noticed that most of the solutions where in Africa and that India was the country with more solutions. They also noticed that there are large regions of the world with out any solutions reported and specifically that the solutions came only from Developing contries. When showed by type of energy one found the colors confusing as he couldnt tell between orange and pink, and that when all the countries were selected we couldnt tell between the countries. Other one appreciate tha solar was yellow, unknow was grey and hydro blue, but that didnt repeated in the clean cooking ones.

They all found the scatter ratios plot confusing and didnt see the point of it.

For the 17 UN Sutainable Goals plot all needed an explanation of what the y axis was as neither of them knew anything about the UN sustainable goals, so if the visualization is for the general public I would need to add a chart showing what this goals are and its number. One suggested that the color didnt help her understand the data and would be better if the color represnts the goal instead of the region of the country.

For the clean cooking the main message for the 3 of them was that there wasn't enough infromation about the energy sources as half of them dont have the energy source stated in the original data.

Overall the evaluation gave me a lot of points of how I could improve the visualizations.

For the selection, the older ones prefered the mouseover to get more info as the clicking wasn't intuitive fro them at all, however for the younger one the mouse over was annoying as when moving the coursor to focus on a detail the plot on the left would change without her wanting to do that.

Conclusions

What worked best of the visualizations where the bar plots showing the number of solutions and the focus of the country specially with the toolbar showing the data of each bar or piece of bar. The map gaves more of a geograpicall sense of where there is no data, but because the latitude and longitude data wasn't perfect in the data set it also made it a bit confusing.

Both scatter plots would need more context and explanation to be usefull for the general public. Probably the visualization should include the UN infographic about their sustainable goals, what each number is.

Given that the clean cooking data had too many unknows as energy sources showing the stacked bars wasnt as effective as the main insigith you get when seeing it is that more than half od the energy sources are unknow so the details of which regions and countries have more clean cooking solutions independently of the energy source got lost. To improve it I would change it to a simple bar plot only showing the number of clean cooking solutions or one showing the toal and with a color showing which of such solutions are clean cooking by country. So instead of having to check betwen the two visualizations, one for Total and one for cleaning, one single would have all the data.