**CS171 Process Book**

Global Health Infrastructure

Nishanth Nayakanti

**Table of Contents:**

* Overview and Motivation:
* Related Work:
* Questions:
* Data: Source, scraping method, cleanup, etc.
* Exploratory Data Analysis:
* Design Evolution:
* Implementation:
* Evaluation:

**Overview and Motivation:**

Motivation of the project is to build a tool to analyze relationship between health infrastructure and general health indicators of all the countries over time.

**Questions:**

1. Is overall health (for example life expectancy) correlated to Health infrastructure of country (health spending, Health Personnel, Hospitals)?
2. How did overall health of the countries change over time?
3. How did health infrastructure of all countries change over time?
4. Compare health spending, infrastructure and life expectancy of the countries over time?
5. Which countries have best Hospitals?

* **Data:**

**Source**

1. Health expenditure per capita

<http://data.worldbank.org/indicator/SH.XPD.PCAP>

1. Physicians per 1000

<http://data.worldbank.org/indicator/SH.MED.PHYS.ZS>

1. Hospital Beds per 1000

<http://data.worldbank.org/indicator/SH.MED.BEDS.ZS>

1. Life expectancy at birth:

<http://data.worldbank.org/indicator/SP.DYN.LE00.IN>

1. Best Hospitals in the world (Academic point of view)

<http://hospitals.webometrics.info/en/world>

**Scraping method:**

The below 4 modules doesn’t need any processing, they are well formatted datasets available at <http://data.worldbank.org>

1. Health expenditure per capita
2. Physicians per 1000
3. Hospital beds per 1000
4. Life expectancy at Birth
5. GDP per capita

5. Best Hospitals in the world:

Hospital ranking data should be scraped from a website. Processing for this data will be done as part of data scraping. I am planning to implement data scraping/processing in **Python.**

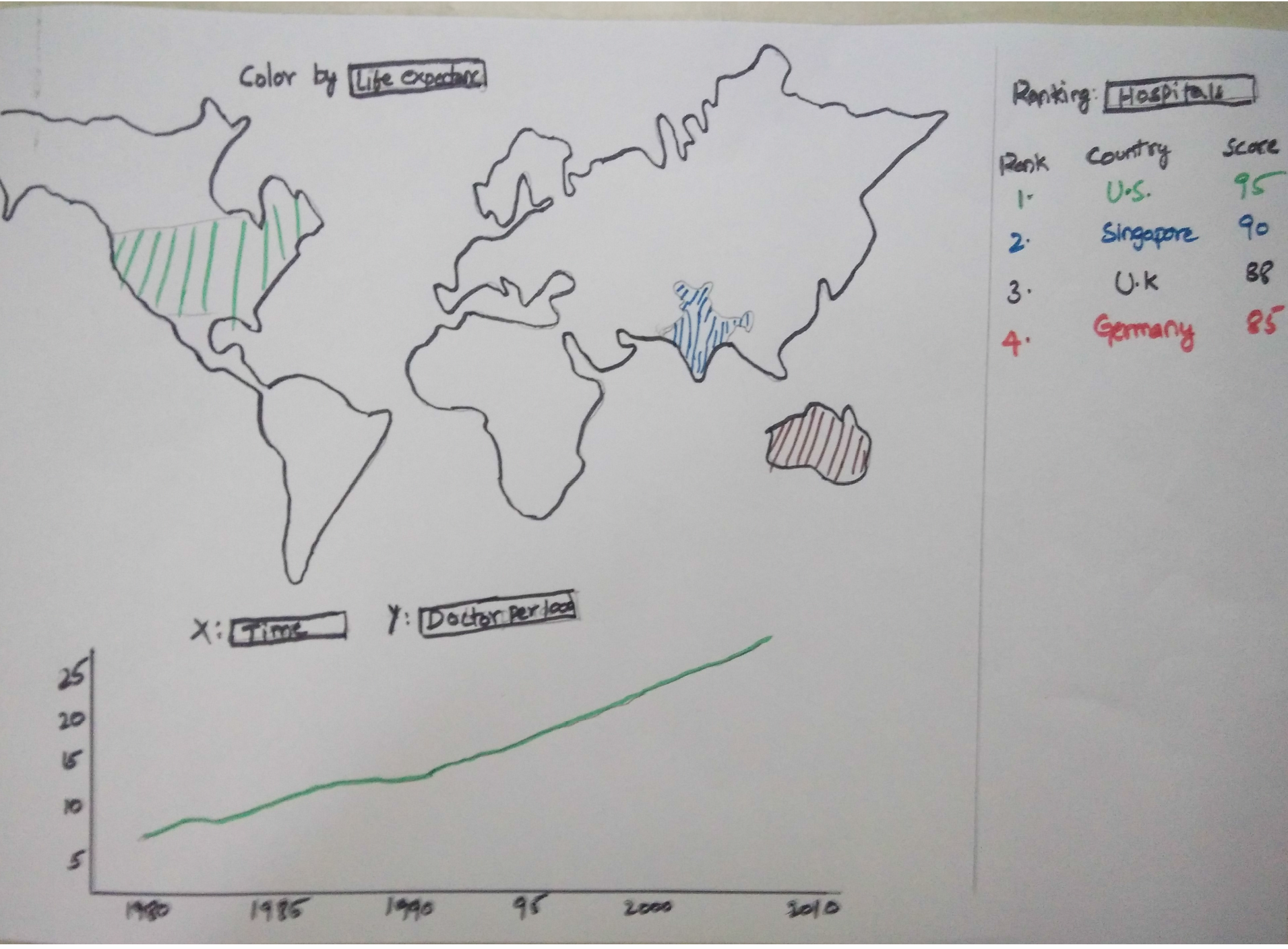
I would like to get the data in the below format

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Institution | Country | Continent |
| 1 | Cleveland Clinic | U.S | North America |

Data scraping for Best Hospitals is implemented in Python and BeautifulSoup.

* **Design Evolution**

**First design in Project proposal:**

****

1. Choropleth map with color coding on top
2. Line graph with **year/time** onXAxis on bottom
3. Ranking of countries based on top hospitals on right

**Cons:**

**Line graph with year/time on Xaxis and parameters like (Life Expectancy, Pcap health spending, Physicians per 1000):**

All the indicators/parameters will increase with time, there is not much to visualize or analyze

**Ranking of Countries based on hospitals:**

Country rankings based on hospitals is not enough to know if a country is good for medical tourism.

**Current Design choice:**

Dividing the project into two parts/visualizations

1. **Change of Global Health Infrastructure and Life expectancy over time**

1. Is overall health (for example life expectancy) correlated to Health infrastructure of country (health spending, Health Personnel)?

2. How did overall health of the countries change over time?

3. How did health infrastructure of all countries change over time?

4. Compare health spending, infrastructure and life expectancy of the countries over time?

**Data:** Doctors per 1000, Percapita Health spending, Percapita GDP, Life Expectancy

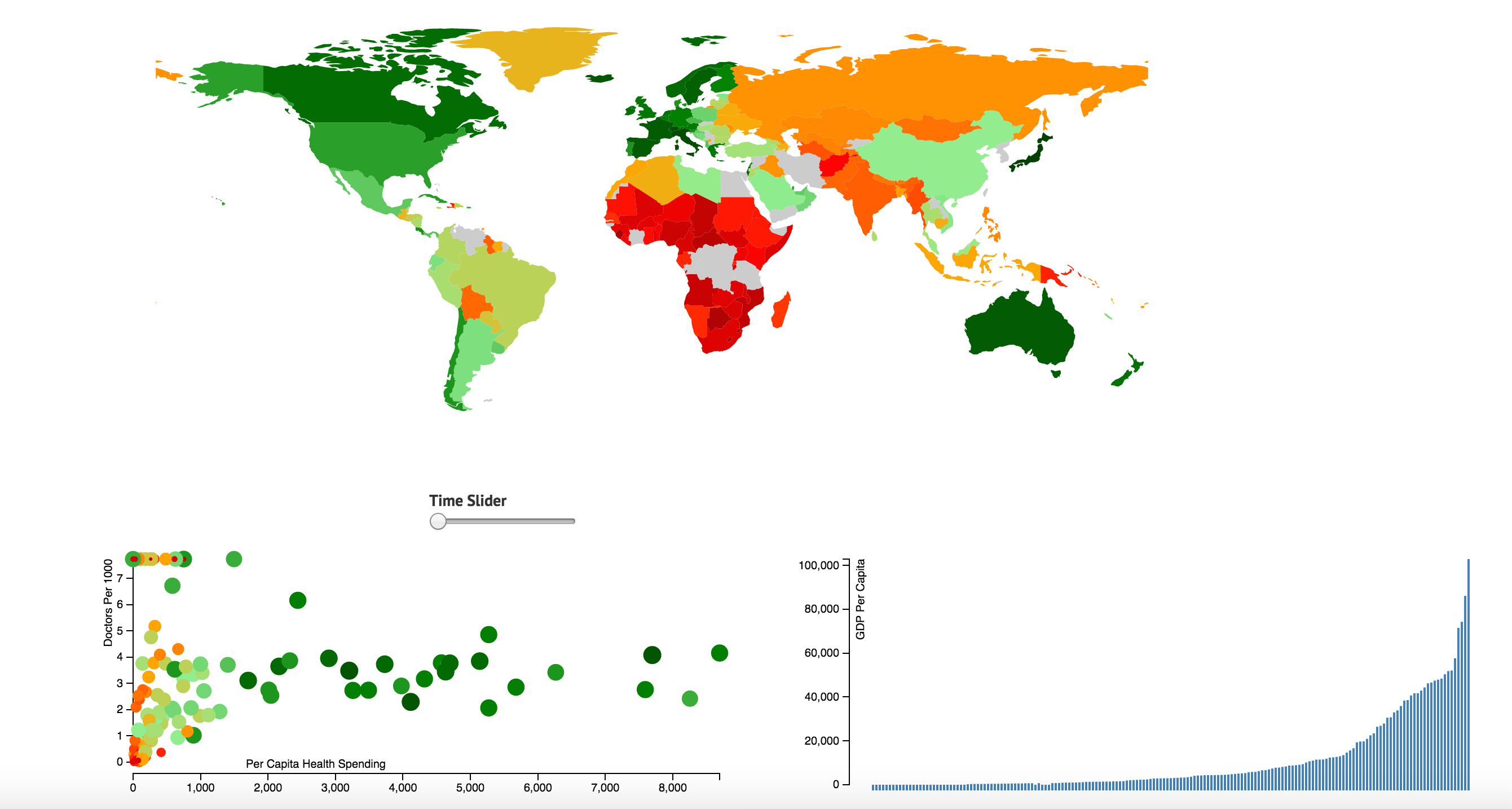
1. **Hospital Ranking**

1. Which countries have best hospitals

**Data:** Hospital Ranking

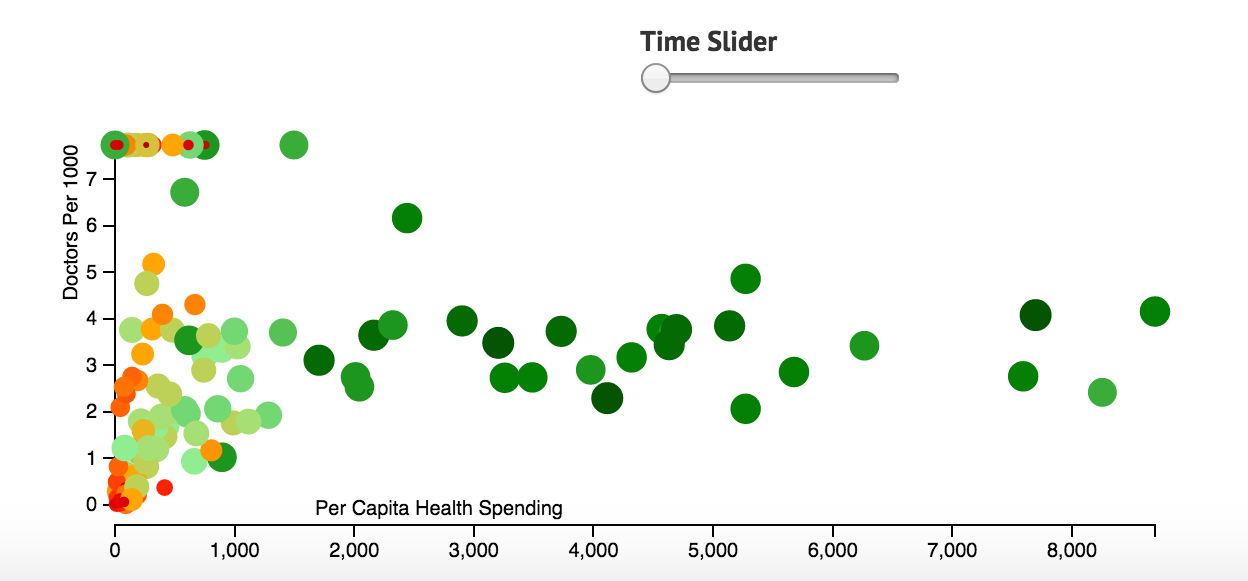
**Part a:**

Design:



1. Choropleth map with color representing Life Expectancy
2. Scatter Plot with Percapita Health spending on xaxis and Physicians per 1000 on yaxis. Life Expectancy is coded in both Radius and color of the circle. Same color coding used as choropleth map
3. Histogram with each rectangle representing a country and percapita GDP on yaxis. Brush is implemented to select a range of countries.

**Scatter Plot:**

****

**XAxis:** Per Capita Health Spending

**YAxis:** Physicians per 1000

Life Expectancy is coded both in color and radius of the circle.

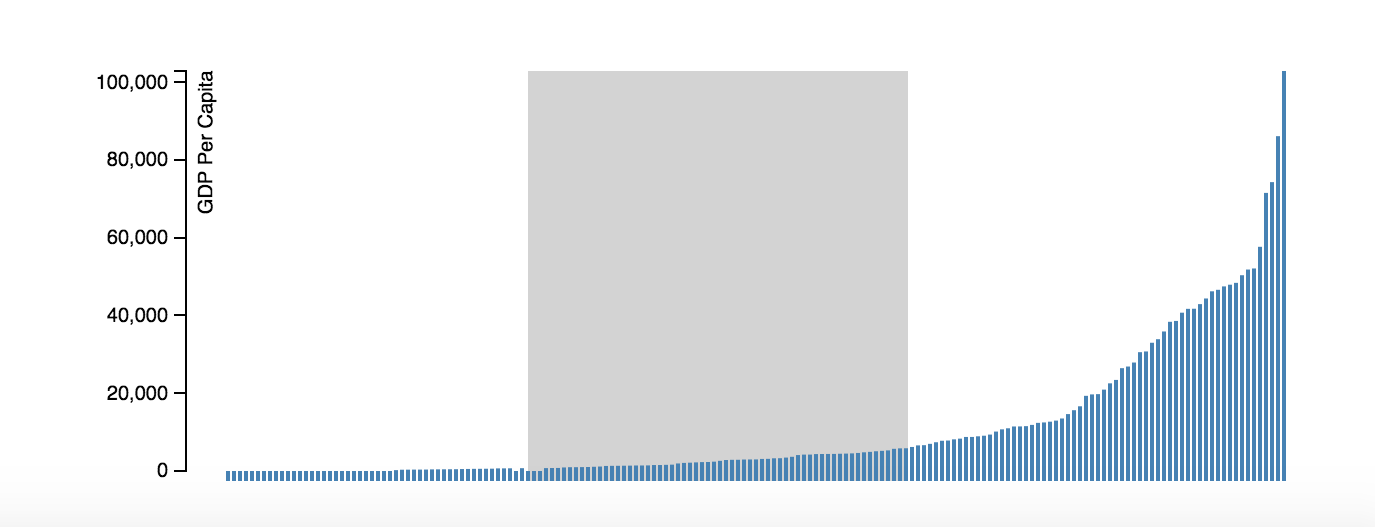
**Time Slider:**

Time Slider is used to visualize the changes over time in all the 3 views.

**Challenge:**

There are around 200 countries, so the scatter plot is messy. We need a way to select few countries

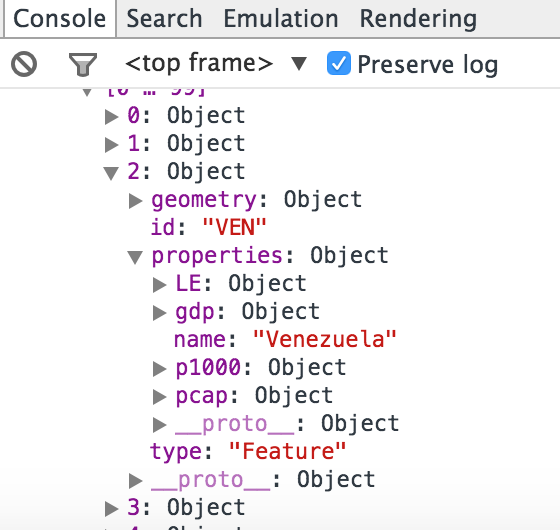
**Histogram with Brush Selection:**

****

**YAxis:** GDP per capita

Each rectangle represents a country sorted in ascending order. Brush selection can be used to select a set of countries in a particular range of Percapita income.

**Data Structure for part a:**



Each Object represents a Country.

* Geometry: data of geojson for choropleth map
* LE: Life expectancy data for each country for all the years
* Gdp: GDP per capita income for all the years
* P1000: Physicians per 1000 people for all the years
* Pcap: Per capita health spending for all the years.