

# **CS171 Process Book**

Global Health Infrastructure

Nishanth Nayakanti

## **Table of Contents:**

- Overview and Motivation:
- 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 visualization tool to analyze global health infrastructure and a general health indicator like Life Expectancy.

## **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>

2. Physicians per 1000

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

3. Hospital Beds per 1000

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

4. Life expectancy at birth:

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

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

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

### **Scraping method:**

The below 5 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 is scraped from a website. I implemented data scraping/processing in **Python and BeautifulSoup**

[https://github.com/nishanthDS/cs171-pr-  
GlobalHealthInfrastructureAndMedicalTourism/tree/master/data%20scraping](https://github.com/nishanthDS/cs171-pr-GlobalHealthInfrastructureAndMedicalTourism/tree/master/data%20scraping)

Here is the sample data (/data/Hospital\_Rank.csv)

<b>Rank</b>	<b>Hospital</b>	<b>Country</b>
689	Bermuda Hospitals Board	Bermuda
26	Alberta Health Services	Canada
98	Hospital for Sick Children	Canada
109	Santé Montérégie Centre for Addiction & Mental	Canada
113	Health	Canada
133	Vancouver Coastal Health	Canada
144	London Health Sciences Centre	Canada
177	University Health Network	Canada

## Data Structure

```
Console | Search Emulation Rendering

    ↳ 0: Object
    ↳ 1: Object
    ↳ 2: Object
      ↳ geometry: Object
        ↳ id: "VEN"
      ↳ properties: Object
        ↳ LE: Object
        ↳ gdp: Object
          ↳ name: "Venezuela"
        ↳ p1000: Object
        ↳ pcap: Object
        ↳ __proto__: Object
        ↳ type: "Feature"
      ↳ __proto__: Object
    ↳ 3: Object
  

```

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.

## Exploratory Data analysis:

I checked if sufficient number of countries has data for top hospitals. 134 countries have data for top hospitals. So, I decided to use it for visualization.

```
In [10]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

In [23]: hospitals = pd.read_csv('data/Hospital_Rank.csv')
print hospitals.loc[:,['Country','Hospital']].groupby("Country").count().describe()
hospitals.loc[:,['Country','Hospital']].groupby("Country").count()

      Hospital
count    134.000000
mean     91.738806
std     340.255298
min      1.000000
25%     2.000000
50%    10.000000
75%    58.000000
```

- Design Evolution

### Design choice 1:



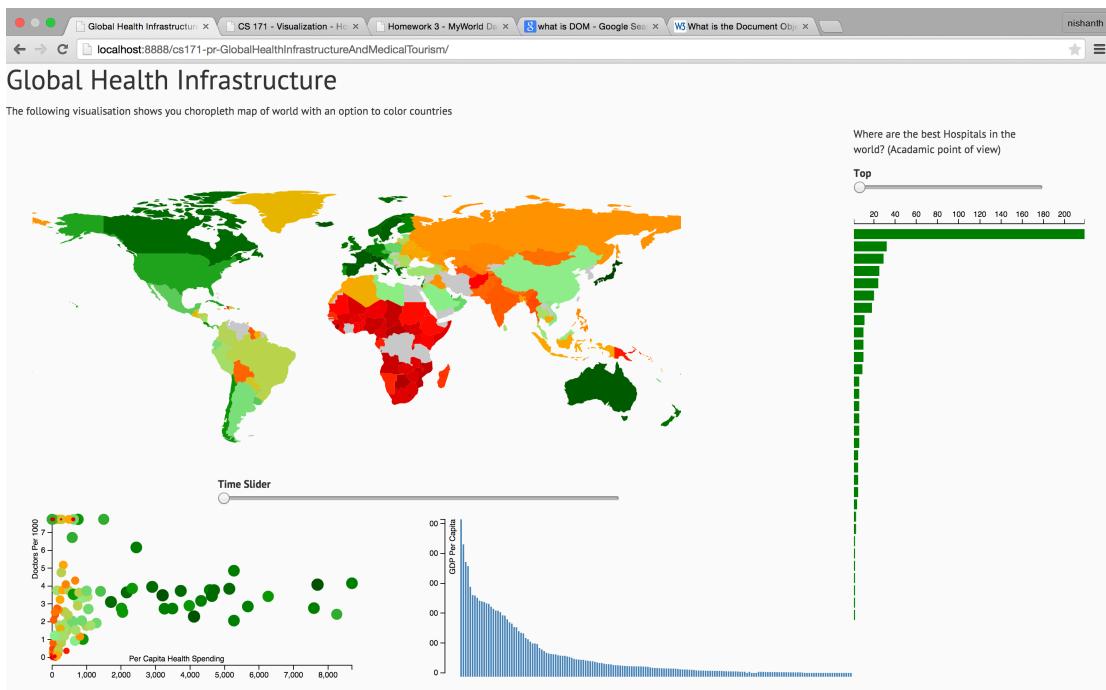
1. Choropleth map with color coding on top
2. Line graph with year/time on XAxis on bottom
3. Ranking of countries based on top hospitals on right

### Cons:

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

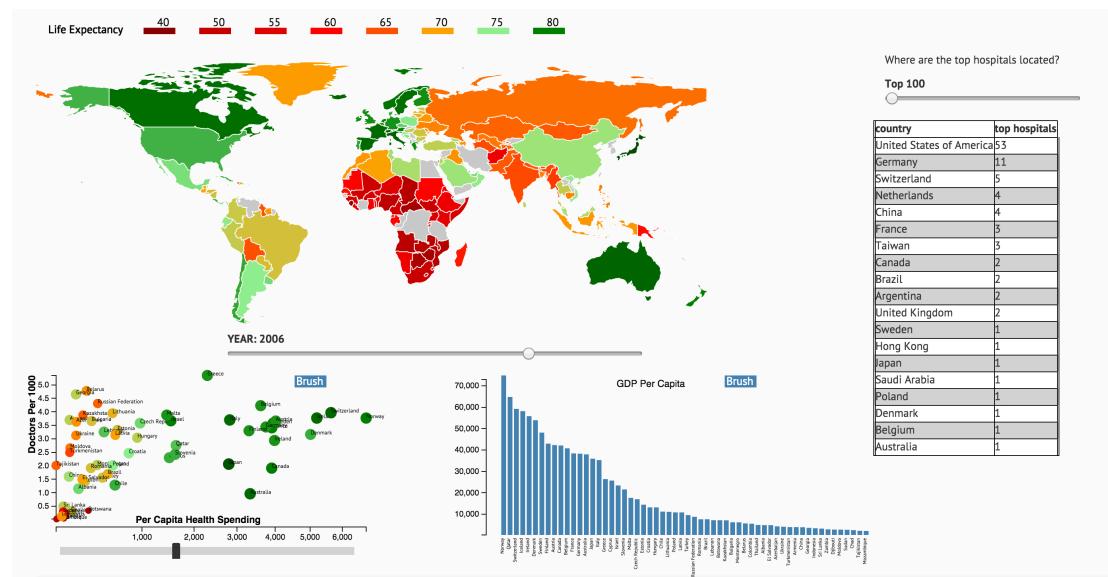
All the indicators/parameters will increase with time, there is not much to visualize or analyze

## Design choice 2



1. Choropleth map with color representing Life Expectancy
2. Scatter Plot with Per capita 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. Bar chart with each rectangle representing a country and per capita GDP on yaxis. Brush is implemented to select a range of countries.
4. Time Slider for visualizing the change in 1,2,3 over time (1995 to 2012)
5. Bar chart representing the distribution of best hospitals. Each bar represents a country.

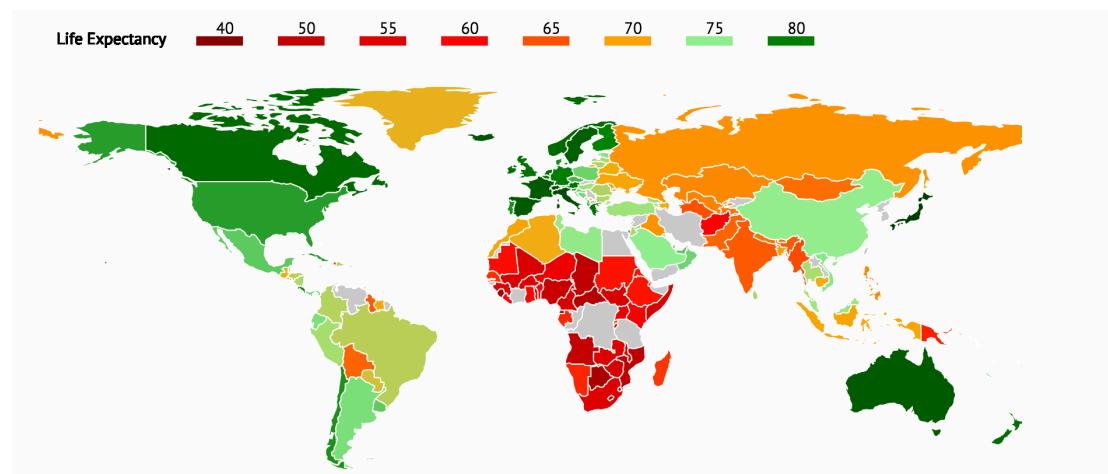
## Final Design:



The only change from design 2 is using table for showing distribution of top hospitals among countries.

## Implementation:

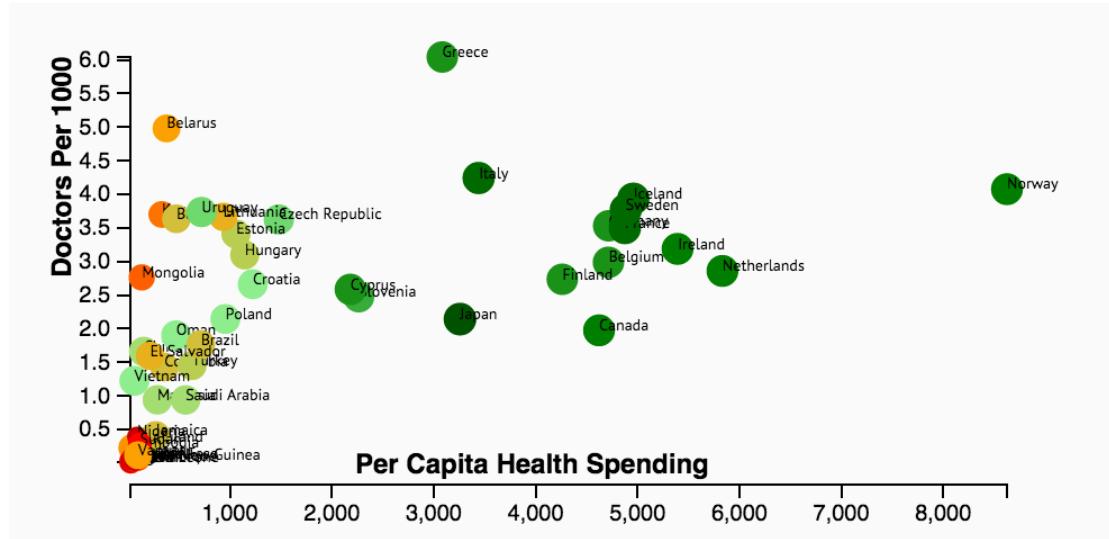
### Choropleth map color coded with Life Expectancy



Slider to select year from 1995 to 2010

YEAR: 2003

### Scatter Plot with brush selection



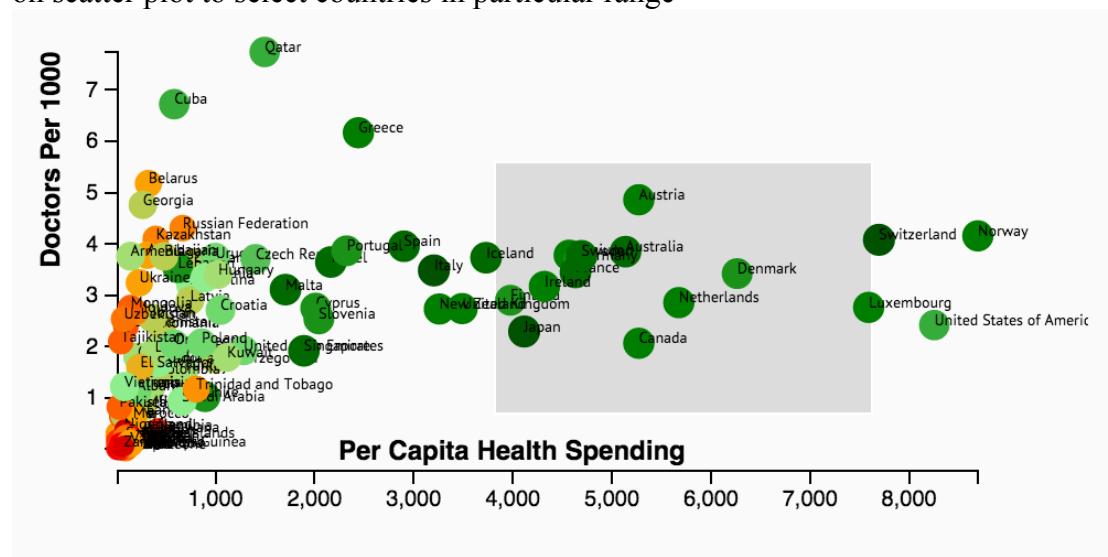
XAxis: Per Capita Health Spending

YAxis: Physicians per 1000

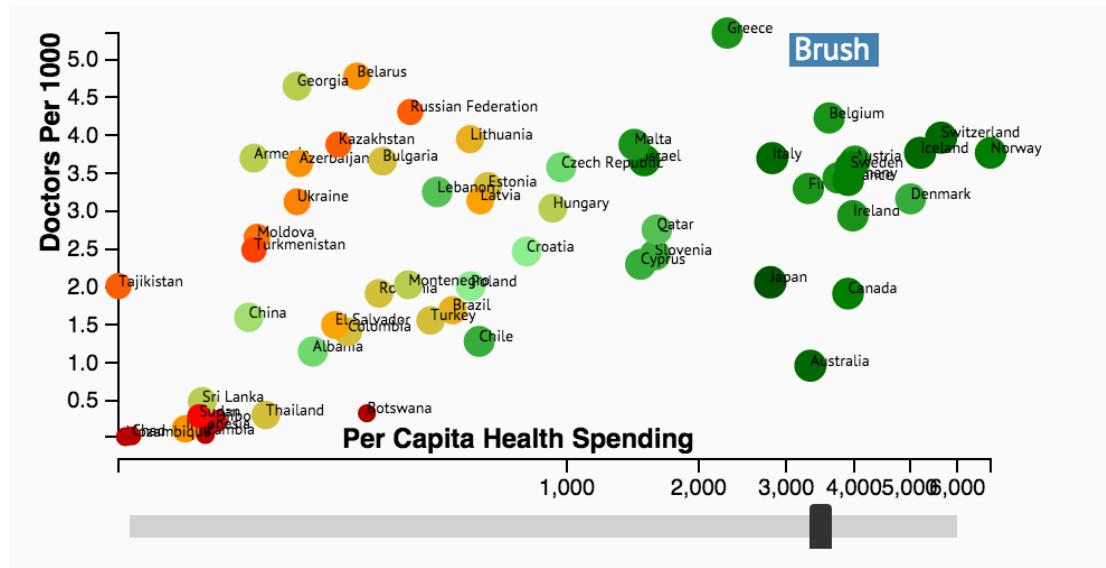
Life Expectancy is coded both in color and radius of the circle.

There are about 200 countries, So the scatter plot is very messy. We need a way to select countries in a particular range of GDP.

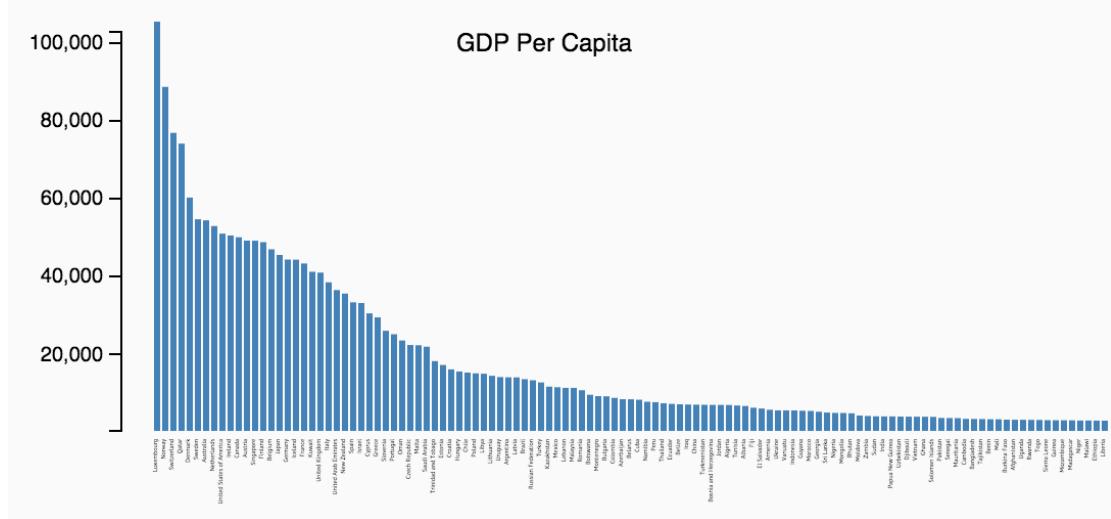
All other views will also display around 200 countries so a 2 D brush is implemented on scatter plot to select countries in particular range



### **Slider to implement Power scale for Scatter Plot:**



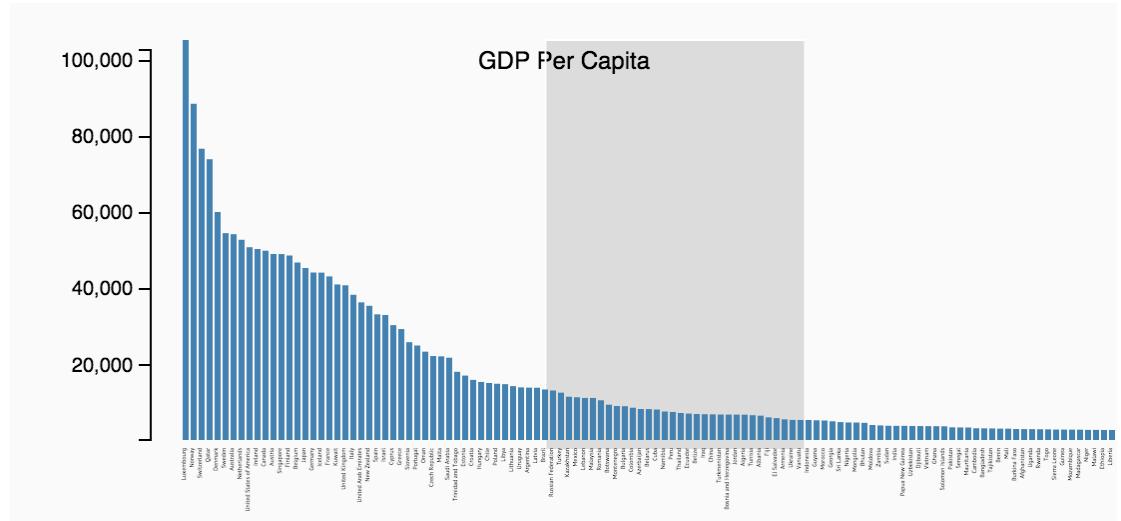
## Bar chart with Countries GDP with Brush Selection:



**YAxis:** GDP per capita

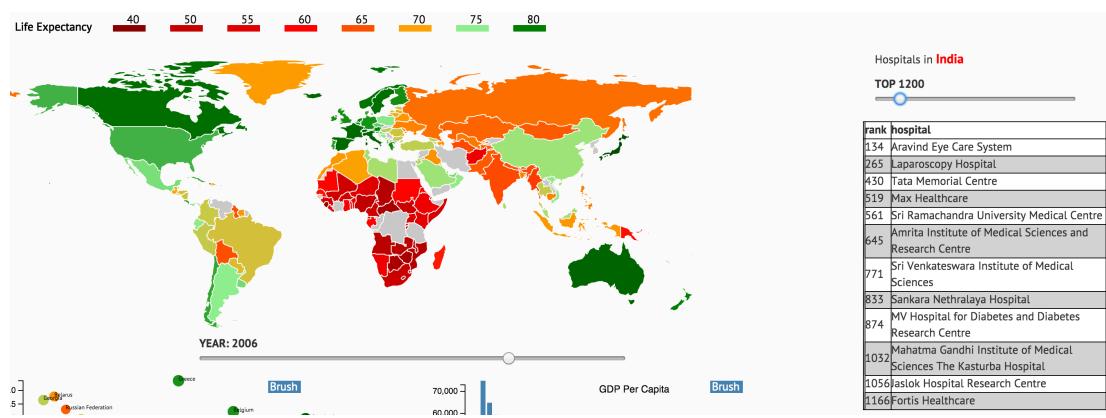
Each rectangle represents a country sorted in ascending order. Brush selection can be used to select a set of countries in a particular range of Per capita income.

**Brush Selection:** All other views will be updated to display just the selected countries



### mouseover feature:

On hovering over a country( choropleth, scatter and histogram) top hospitals in that country will be displayed.



## Distribution of top hospitals

Where are the top hospitals located?

**Top 100**

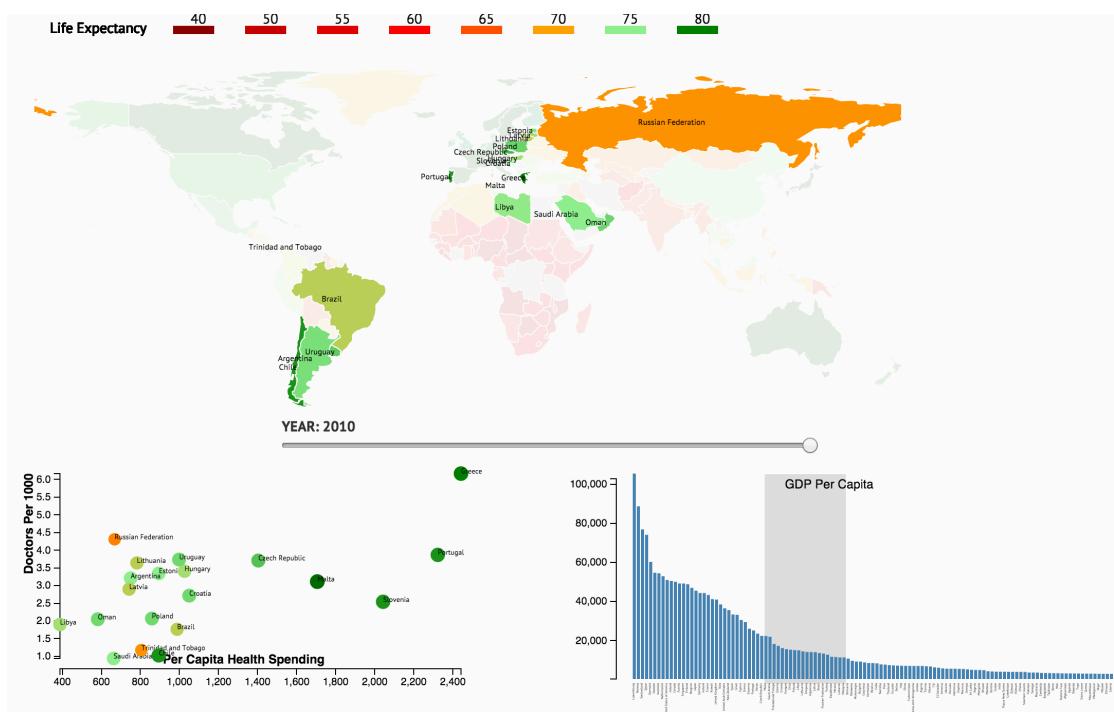


country	top hospitals
United States of America	53
Germany	11
Switzerland	5
Netherlands	4
China	4
France	3
Taiwan	3
Canada	2
Brazil	2
Argentina	2
United Kingdom	2
Sweden	1
Hong Kong	1
Japan	1
Saudi Arabia	1
Poland	1
Denmark	1
Belgium	1
Australia	1

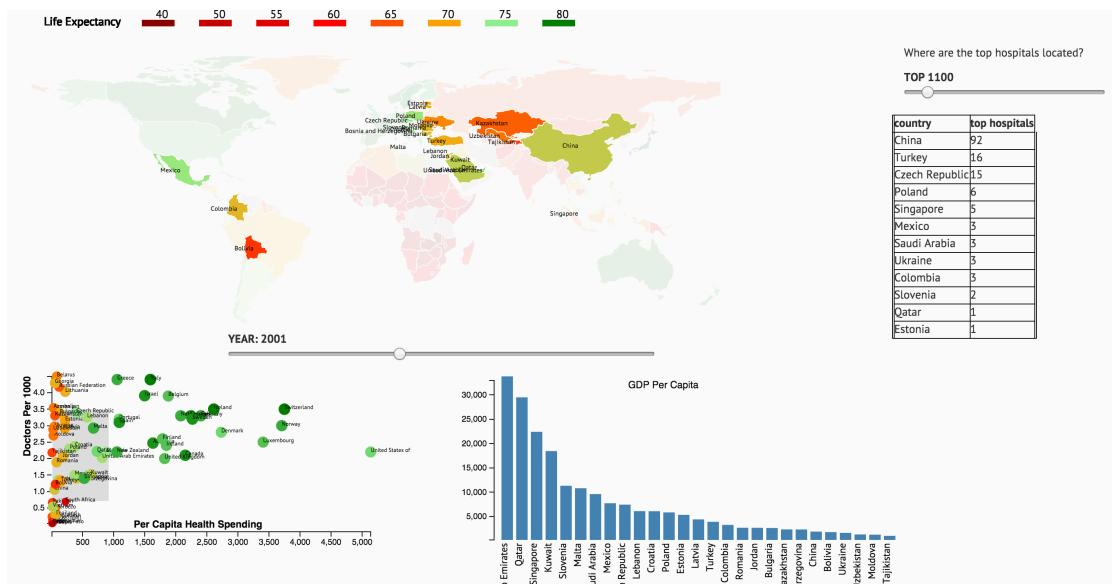
This view shows the distribution of best hospitals among the countries.  
There are 16000 top hospitals in the data set with ranking 1 to 16000.

We might be interested in knowing the distribution of maybe top 100 or top 1000 etc.  
So, I implemented a slider to select top N hospitals.

## Brush selection on GDP Histogram:



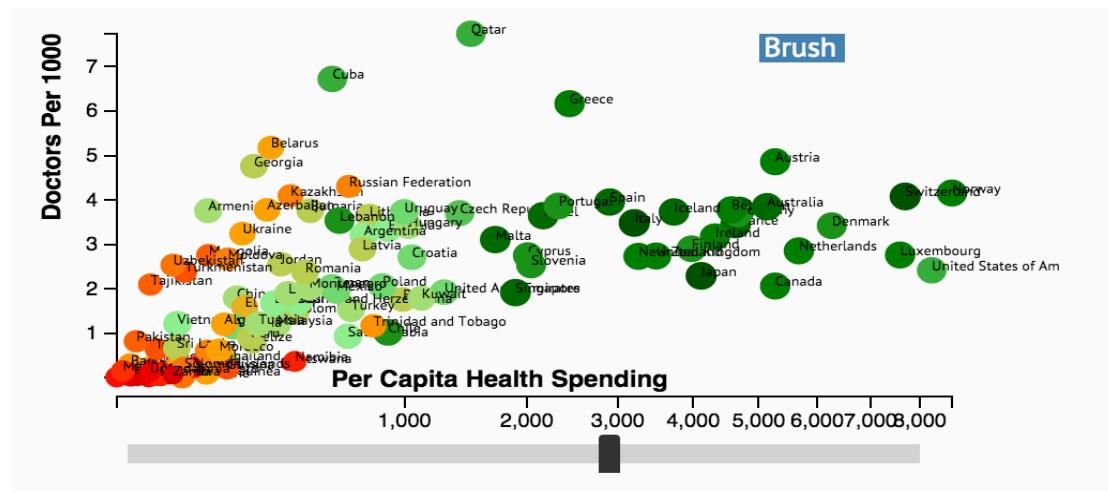
## Brush Selection on Scatter plot:



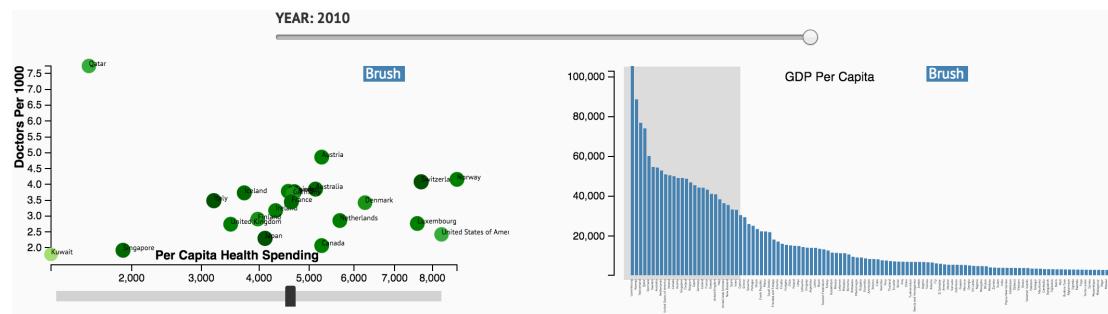
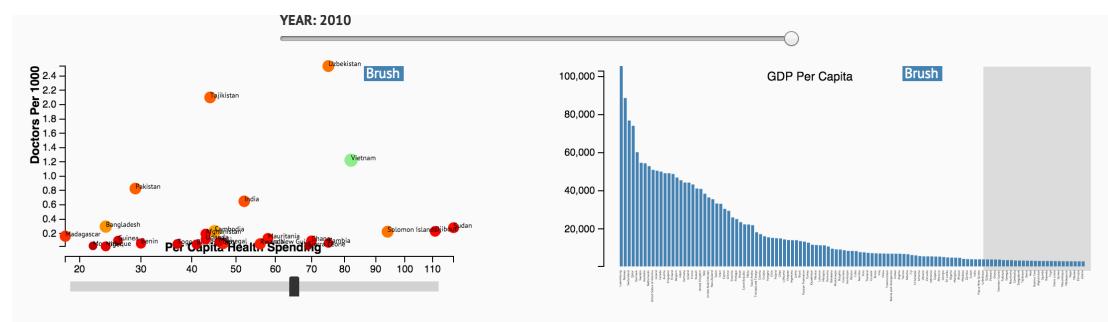
## Evaluation:

What did you learn about the data by using your visualizations?  
How did you answer your questions?

Life expectancy is heavily dependent on health infrastructure of the country



Health Infrastructure of the country is dependent on GDP.



## Life expectancy of many countries increased over time:

