**Write-up**

The dataset was taken from

<https://catalog.data.gov/dataset/center-for-medicare-amp-medicaid-services-cms-medicare-claims-data-79a6c>

* Understanding the Dataset

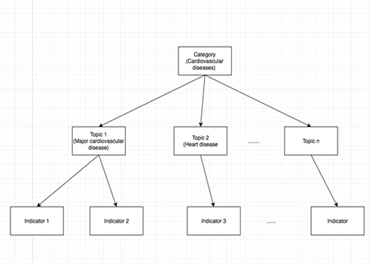
CMS studies claims data for Medicare and Medicaid patients of Category: Cardiovascular diseases to understand the public health burden of CVDs and the associated risk factors in the United States. By studying the claims data for patients of Category: Cardiovascular diseases, CMS would know the number of people being affected by Cardiovascular diseases and how the patients are doing, whether they are being treated or not

* Understanding the fields in the dataset

1. **Category, Topic and Indicators**

* The dataset includes patients who are suffering from Cardiovascular diseases. Cardiovascular diseases in general is a broad term. Under the cardiovascular diseases it has many “topics” for example major cardiovascular disease, heart disease etc. That is patients who are suffering from heart disease, stroke, heart failure, etc
* Now to know the condition of a state or a country for a particular year suffering from Cardiovascular diseases (major cardiovascular disease, heart disease etc) “indicators” are used.
* “Indicators” from this data source have been computed by personnel in CDC's Division for Heart Disease and Stroke Prevention (DHDSP).

The diagram for Category, Topic and Indicators will look like



* **What are indicators? How indicators work?**

Indicators for this dataset refer to explicit criteria by which the condition of a specific population of a state can be evaluated.

Indicators are calculated for a topic. For example, for topic Major Cardiovascular

Disease, the indicator calculated is ‘Prevalence of major cardiovascular disease

hospitalizations among all hospitalizations, US Medicare FFS beneficiaries (65+)’,

2. **Break-outs**

Breakout means that the population of the state is divided by Overall, Gender, Age and

Race. For when a data value of Gender: male is 12% for a particular topic and particular indicator let’s say ‘Prevalence of major cardiovascular disease hospitalizations among all hospitalizations, US Medicare FFS beneficiaries (65+)’, this means that the number of males that got hospitalized let’s say are 100 out of which 12 males got hospitalized for particular topic and indicator

3. **Data\_Value\_Units**

There are two units for our data value.

1. Percentage

Example: Let’s say we need to find the data value for a particular indicator for a particular year then 1)First the find the number of people from that state who got hospitalized for that year (let’s say 100) 2)Then find the number of people who got hospitalized for Major Cardiovascular Disease (let’s say 20). 3)Then divide 100 by 20 and then multiply the answer by 100 to know the percentage for that data value

b. Rate per 1000

Example:

So, if total number of males that got hospitalized for Coronary heart disease is 700 for a state. Total number of males are 25000 in that state. So, to find the rate per 1000

1. Divide the population by 1000. So 25000/1000=25. So, we have 25 occurrences of 1000 each.
2. Divide the total number that is number of males hospitalized by the occurrences. So, 700 divided by 25=28
3. So, there are 25 occurrences and in each of these occurrences the number of people getting hospitalized is 28
4. So, the rate per 1000 is 28

TASKS

1. The first task was to know the overall condition of each state for all indicators for all the years. The second task assigned was to know the overall condition of United States for all the indicators for all the years

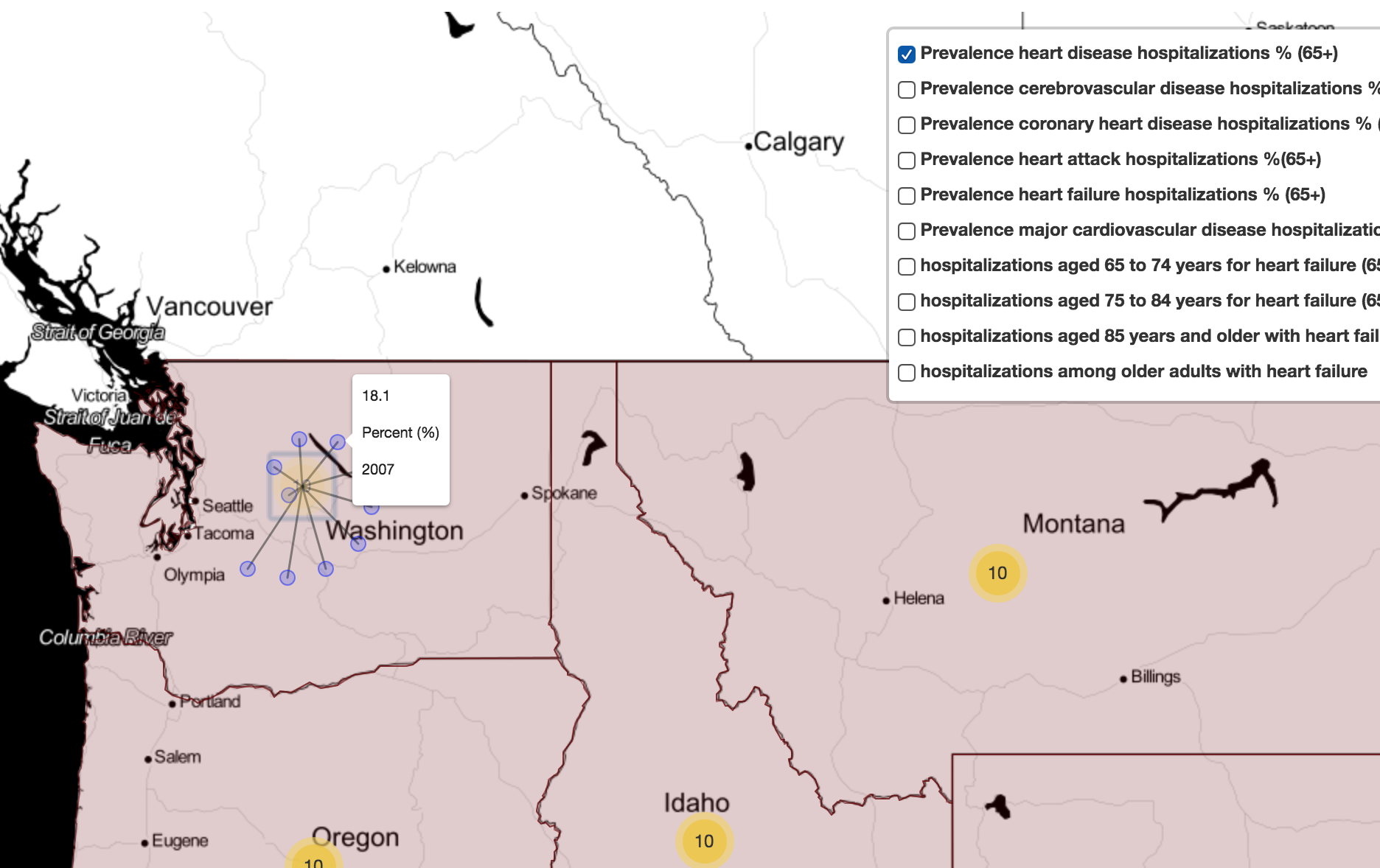
Since the indicator names were very big and they were causing an issue at the time of viewing a map,indicator names were reduced and shorter names were assigned.

The break\_out value considered was “Overall”. This is because here we need to know how a state is doing (that is without any breakouts like males, females, we need all the people, so the break\_out value is overall).

By using leaflet and downloading the correct Shapefile of the United States (since our dataset is about United States) two maps were created

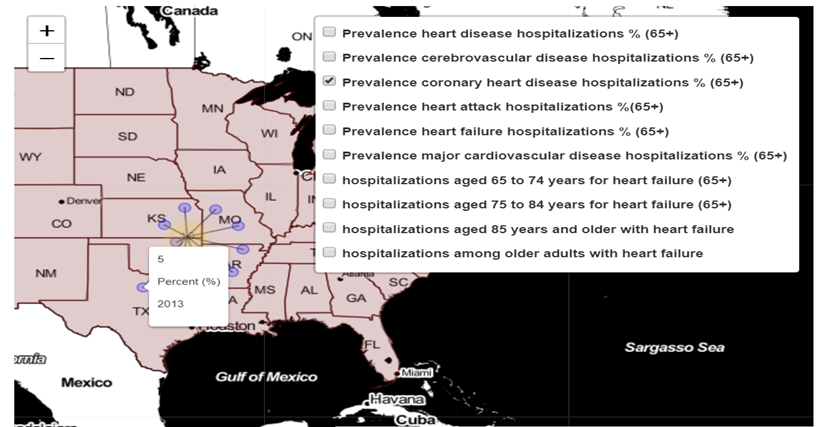
1. First map will tell us how a state is doing over the years (we can also compare like how one state is doing compared to other state for a indicator for a particular year)
2. Second map will tell how United States is doing for a indicator over the years

Image1:



The above image tells us how State Washington is doing over the years. The top right corner in the image are the filters(indicators). The indicator chosen is “Prevalence of Heart Disease Hospitalizations in %”. So, when we hover the mouse towards state Washington and click on the yellow circle, it creates 10 points (blue circles) each representing the value of the indicator for a particular year. In the above image it shows the value of “heart disease hospitalization in %” for year 2007(which is 18.1%)

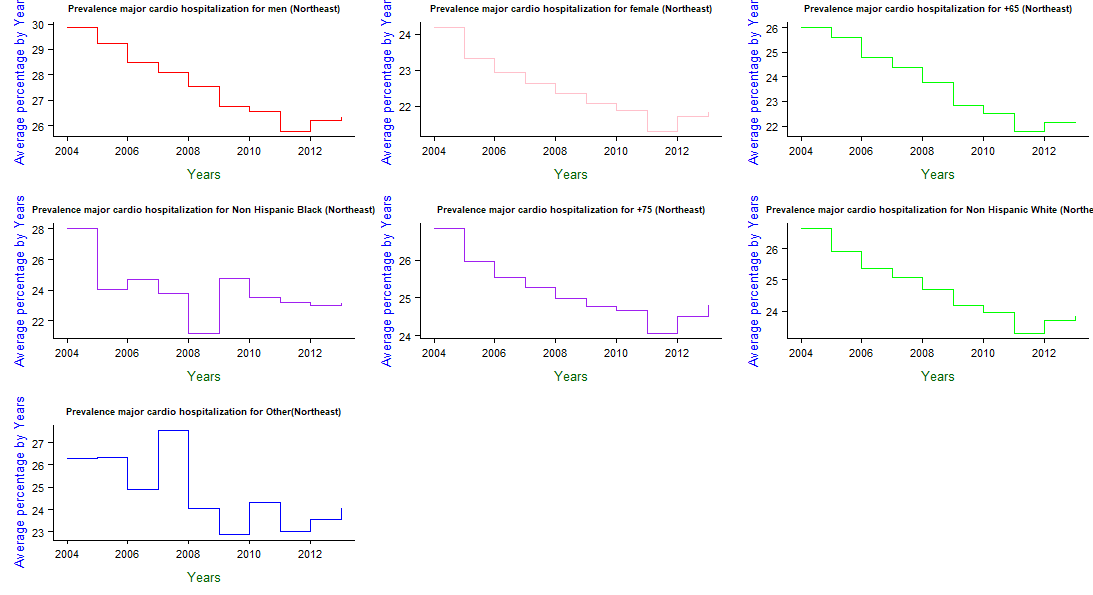
Image 2:



The above image shows the “% of Coronary Heart Disease hospitalizations in United States” for a particular year. The year chosen is 2013, so there were 5% hospitalizations for Coronary Heart Disease among all hospitalizations. The method for creating both the maps is same

1. Another part of visualization was to divide the dataset on the basis on region which were: 1. Northeast 2. Midwest 3. South 4. West.
2. First each state was assigned a region (This information as to which state falls in which region was checked on the internet)
3. The subset of the data containing all the Northeast region was created and worked upon. The regions were visualized for all the “Topics” (except Heart Failure) and analyzed using all the Break\_Outs

Image:



The above image gives us the information of all break\_outs (except for “Overall” because that part is already covered) for all years for all states in the northwest region for a particular topic which is “Major Cardiovascular Disease)”

Similarly, we can write for other break\_out categories

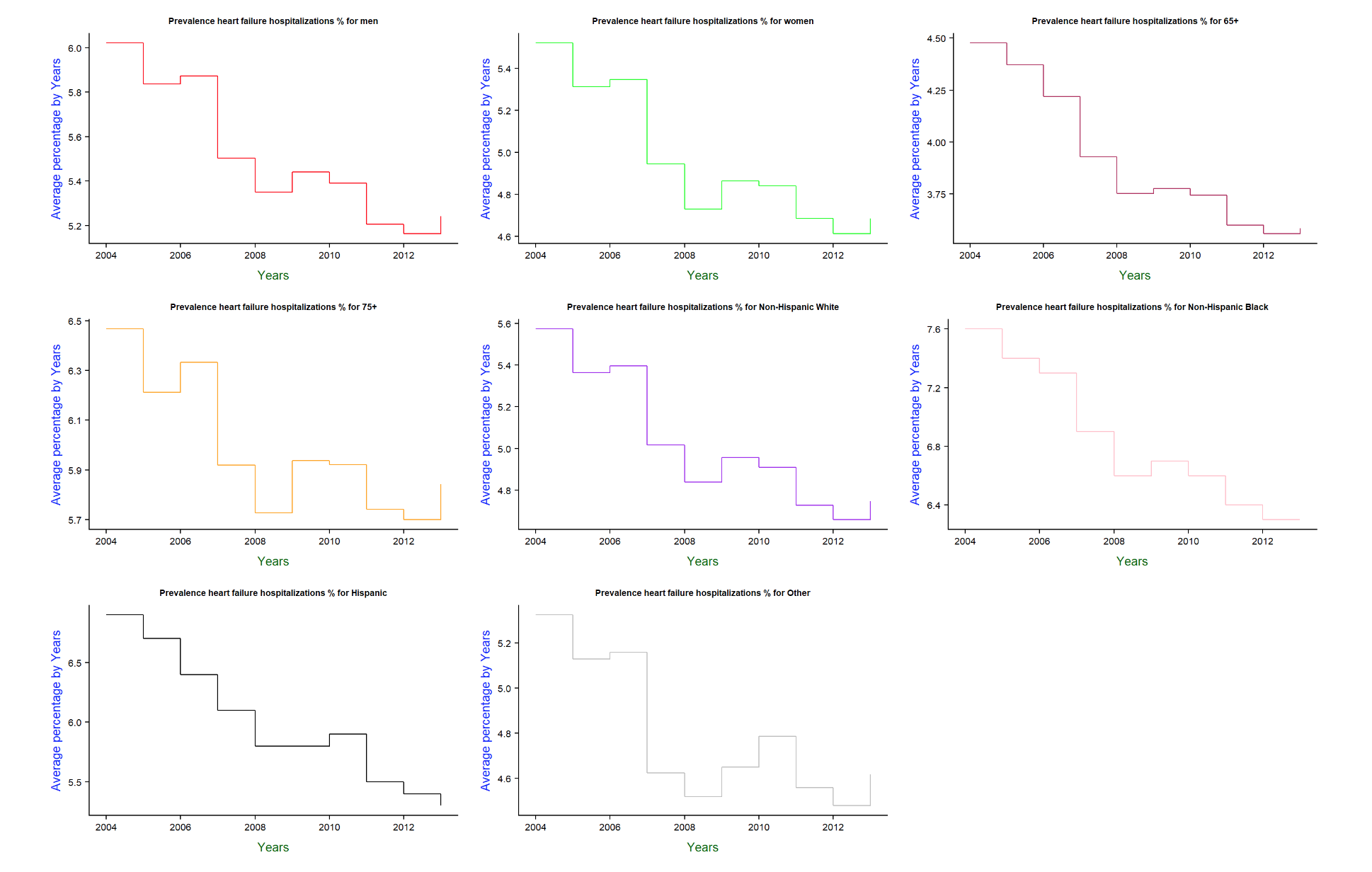
We can do the same for all indicators and for all breakouts for all the other three regions we created (that is Midwest, South, West)

As the Data\_Value\_Unit of the topic ‘Heart Failure’ has two variations namely ‘%’ and ‘rate per thousand’, it was not possible to directly work on this topic. So, we had to work upon the indicators of the Topic (Heart Failure) which were

1. Prevalence of heart failure hospitalizations among all hospitalizations, US Medicare FFS beneficiaries (65+) %
2. Rate of hospitalizations among older adults with heart failure as the principal diagnosis (among FFS Medicare beneficiaries (65+))
3. Rate of hospitalizations among adults aged 65 to 74 years with heart failure as the principal diagnosis (among FFS Medicare beneficiaries (65+))
4. Rate of hospitalizations among adults aged 75 to 84 years with heart failure as the principal diagnosis (among FFS Medicare beneficiaries (65+))
5. Rate of hospitalizations among adults aged 85 years and older with heart failure as the principal diagnosis (among FFS Medicare beneficiaries (65+))

Let’s consider the first indicator,

Image



With the above image we get to know how a specific breakout category (male, female etc) is doing over the years with respect to Heart Failure Hospitalizations (in %)

Results from the above image

(A) Men: We see that the average percentage of men in United States being hospitalized for heart failure among all hospitalizations were 6% in year 2004. This started decreasing to less than 5.4% in year 2008. After 2008 the average percentage increased to greater than 5.4. In Year 2010 it was 5.4% to eventually decrease to 5.2 in year 2012. In year 2013 it increased a bit.

(B) Women We see that the average percentage of women In United States being hospitalized for heart failure was greater 5.4%. This percentage for women was lesser than the average percentage for men in Year 2004. Over the years the average percentage started decreasing. It was less than 4.8% in year 2008. It was same in year 2009. After 2009 it started increasing to eventually decrease to less than 4.6. In year 2013 it increased a bit.

Etc etc

Similarly, we can get the information for the other break\_outs for the indicator “Prevalence of heart failure hospitalizations among all hospitalizations”

In the same way, we can get the information for all the other indicators of topic “Heart Failure” for all years and for all Break\_outs

We can also find out which disease (that is topic) is prominent for a break\_out for years 2004,2005,2011 and 2012.

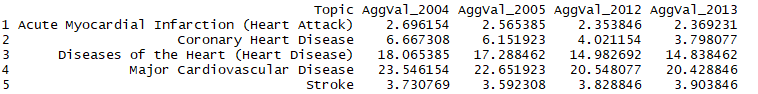
Once done we can find top 5 states where the value for that disease(topic) was high for the above-mentioned years. Similarly we can also find the bottom 5 states where the value was low for the mentioned years

Example:

Which disease is prominent in Women for years 2004,2005,2011 and 2012?

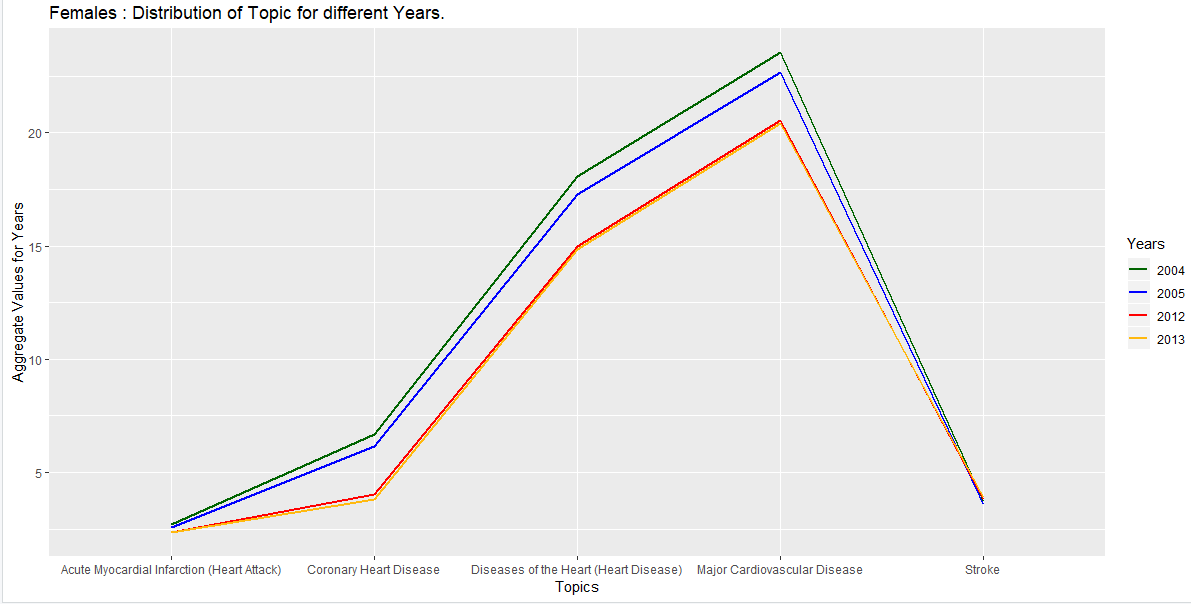
So, the result she got was

Image1:



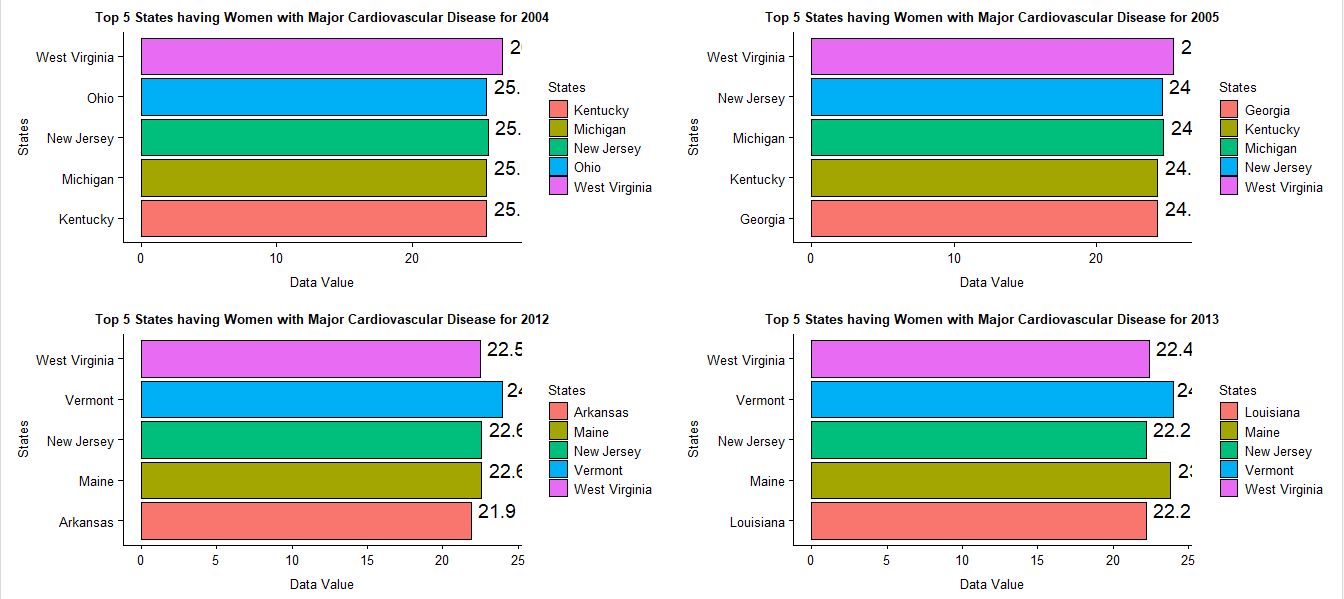
Plotting a graph for this

Image 2:

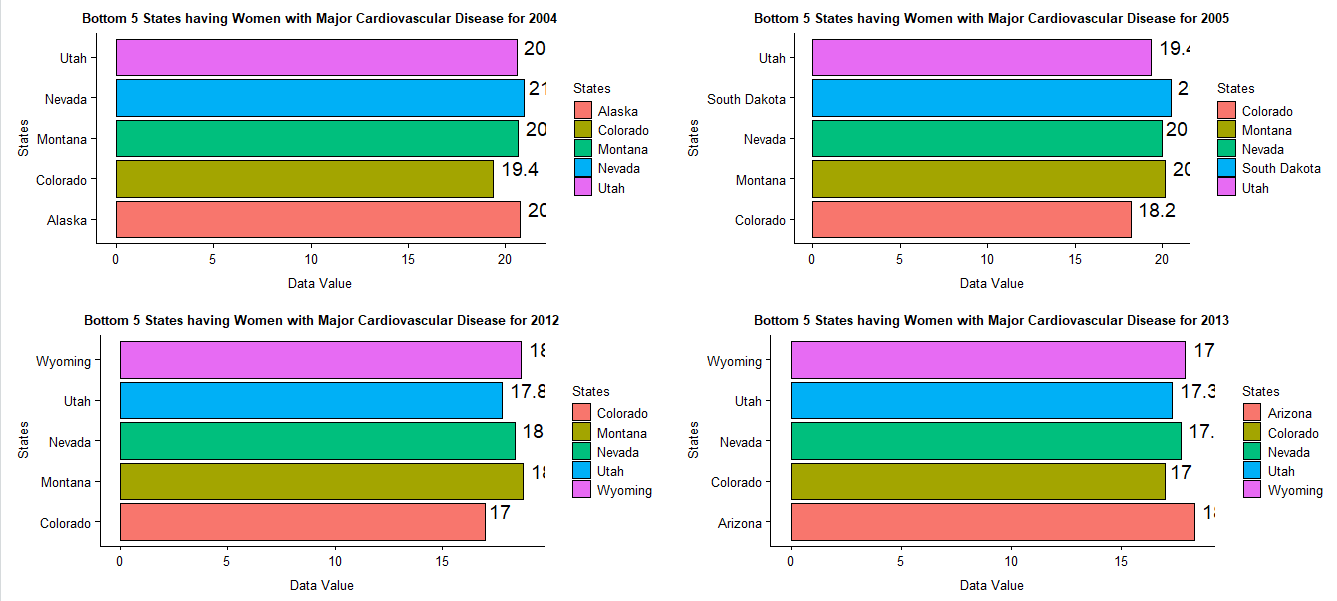


Finding the Top 5 states in USA where women suffering from Major Cardiovascular (because this was the highest) disease is prominent for the year 2004,2005,2012 and 2013

Image3:



For bottom 5 states



Similarly, we can do this for all the other break\_outs

We can also do the entire method for other topics as well for all the years