**Heart Disease Death Rate Analysis**

This report presents the leading causes of death in United States for the time period of 2011 - 2014. It further discuss deep about one of the leading causes, Heart Disease with respect to various factors affecting it. The analyses include sex, race, education, income, employment status, residence and region in the national level. The significant risk factors such as alcohol, diabetes, high blood pressure, obesity and smoking are also discussed at state level. The report briefly walk-through some of the techniques to determine the significant factors among all those listed above.

**Data Consideration:**

The data is primarily obtained from CDC – Centres for Disease Control and Prevention. This website provides trend data on various health aspects. The data tables include information for various ethnic groups, education and income, gender, age and other socio-demographical data. It is presumed that the data is appropriate apart from some noisy and missing data.

**Challenges faced in data collection:**

Data is collected after an intensive research from various resources. Collecting data with same units and time period was challenging as it is difficult to obtain the relative data for all the factors under consideration. There could be some deviation in the outputs of the results due to these pitfalls.

**Data Processing:**

The collected data was very noisy and there were lot of missing data. There was a need to clean up the data by removing the empty values and variables that are not of interest for the analyses. The cleaned data had also need to be mapped to view the correlations and other important patterns.

**Analysis:**

1. **Leading Causes of Death:**

As per the National Vital Statistics Report on Deaths, 2013, volume 64 issued by CDC, a total of 2,626,418 resident deaths were registered in United States. The age adjusted death rate was 724.6 deaths per 100,000 U.S standard population. The leading causes of death were:

1. Disease of heart
2. Malignant neoplasm (cancer)
3. Chronic lower respiratory diseases
4. Unintentional Injuries
5. Cerebrovascular diseases
6. Influenza and pneumonia
7. Diabetes mellitus

The graph in Fig 1.1 precisely shows the leading causes of death among adults aged from 15 years to 65 years and above.

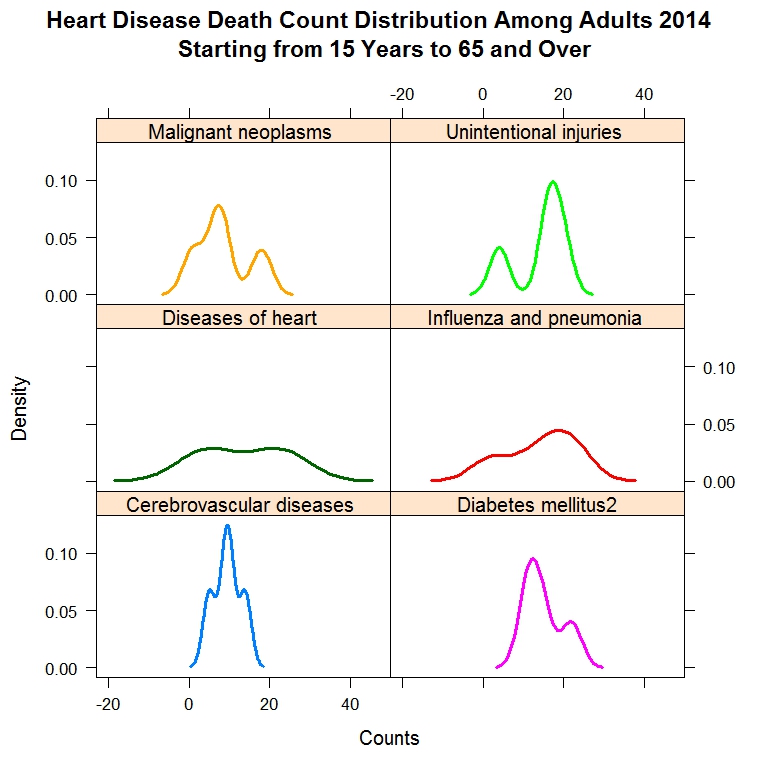


Fig 1.1 Heart Disease Death count Distribution

This graph explains the distribution of diseases among adults. Malignant Neoplasm is prevalent in the adults aged between 20 to 55 years. Unintentional injuries are frequent among the adults aged from 30 to 60. Heart disease is almost evenly distributed among all the adults from 15 to 65 years and above. Influenza and Pneumonia is widely seen among the adults aged from 30 to 65 years. Cerebrovascular diseases are more prevalent in the adults starting from 25 to 55 years. Diabetes Mellitus is more frequently seen in people aged from 25 to 65 years.

Since heart disease is one of the evenly distributed diseases among the adults, this report further dives deep into heart disease, prevalence statistics and its associated risk factors.

1. **Heart Disease:**

Heart Disease is the top most cause of death in United States from 2011 to 2014. There are many frightening and some surprising factors associated with this.

The graph in Fig.2.1 shows the death rate in United States due to heart disease in from 1950 to 2015.

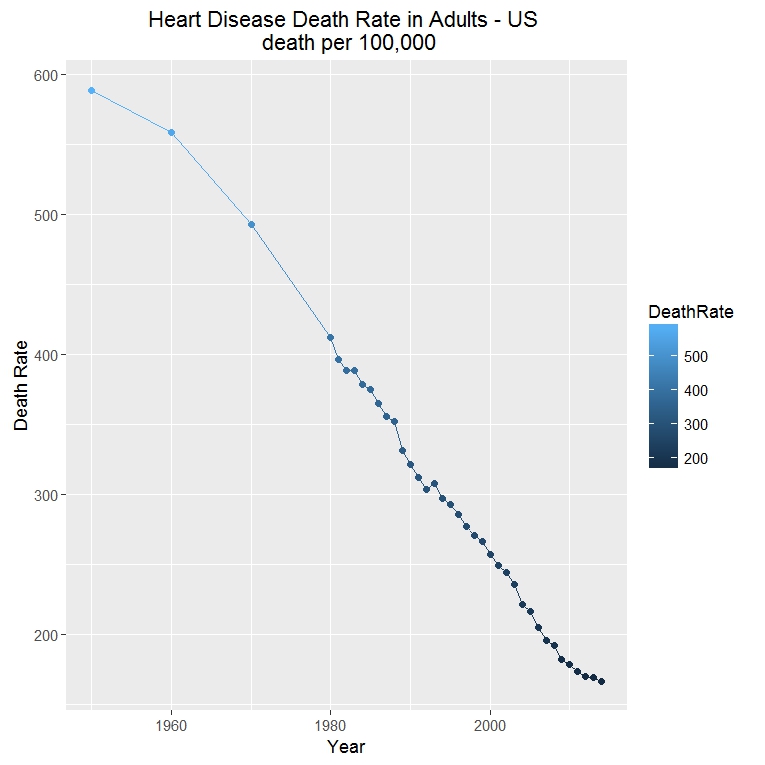


Fig 2.1 Heart Disease Death Rate in Adults - US

As we could see that the death rate has constantly reduced from 1950 to 2015. This may be due to the advancements in the medical field and awareness among the people. But even then the heart disease is one of the dominating factors affecting death. This report will take in consideration of some of the significant reasons behind this.

**Heart Disease Death Pattern:**

East coast shows high prevalence of risk factors compared to west except California and Nevada as shown in Fig 2.2.

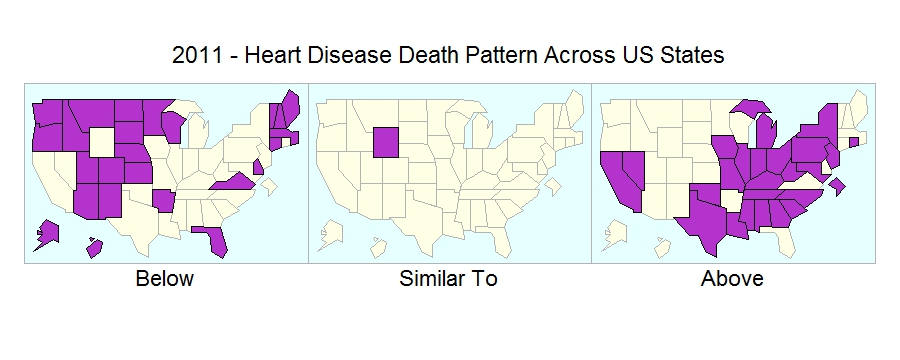


Fig 2.2 Heart Disease Death Pattern in United States – 2011

1. **Prevalence of Risk Factors associated with Heart Disease:**

Major factors affecting heart diseases are smoking, physical inactivity, obesity, high blood pressure, diabetes and alcohol. Below are the two factor designs depicting the prevalence of these factors among various sectors.

* 1. **Social Factors:**

***Risk Factors by Sex:***

Alcohol consumption and smoking is more among men than women whereas physical inactivity is more prevalent among women that in men. Obesity, high blood pressure and diabetes prevalence is about the same among men and women.

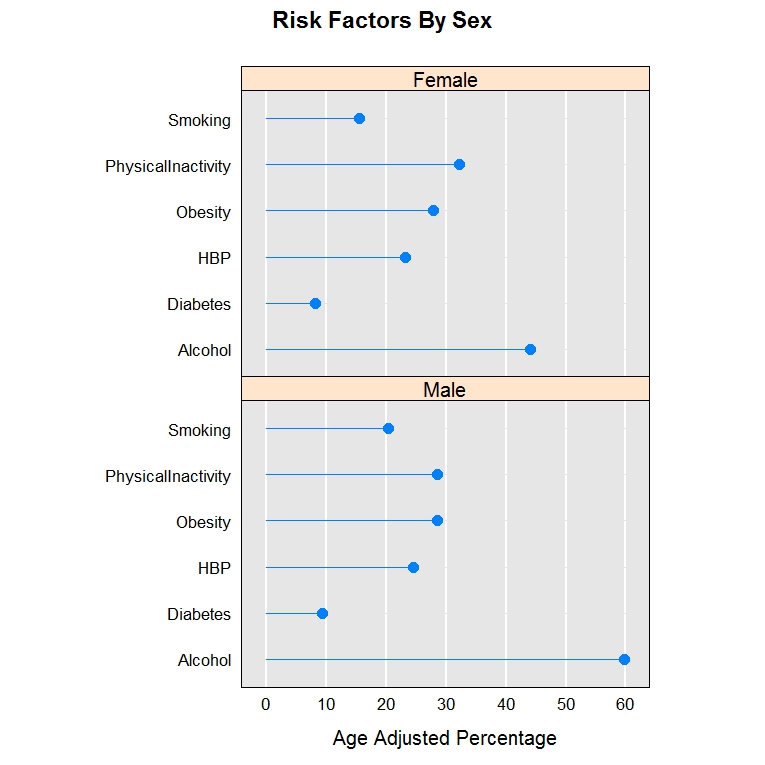


Fig 3.1 Risk factor by Sex

The statistics in Fig 3.1 shows that obesity, high blood pressure and diabetes seem to be pretty good factors affecting heart disease among men and women.

***Risk Factors by Race:***

The factors vary widely across different races. For example alcohol consumption is the most prevalent factor among Black people followed by Native American, Asian and White. At the same time black people have very low diabetes rate compared to other race. There is no one such factor that strongly determines the heart disease except obesity. On an average obesity is almost evenly distributed across all the races under consideration.

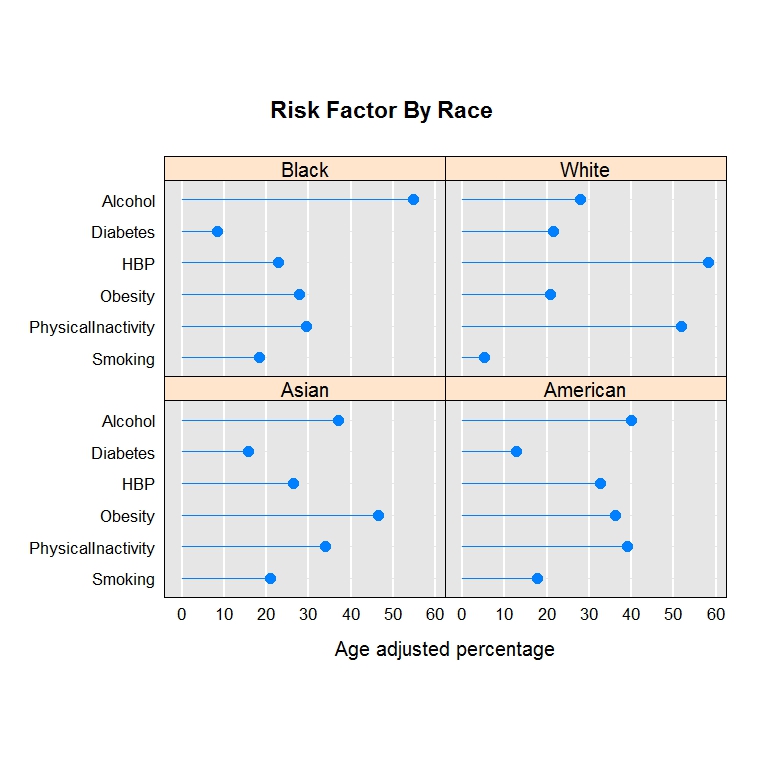


Fig 3.2.1 Risk Factors by Race

* 1. **Economic Factors**

Economic factors include education, income, employment status and residence.

***Risk Factor by Education:***

The analyses mostly go by the general assumptions about the factors and education status. Highly educated people have less risk factor than the people who are not well educated except for alcohol.

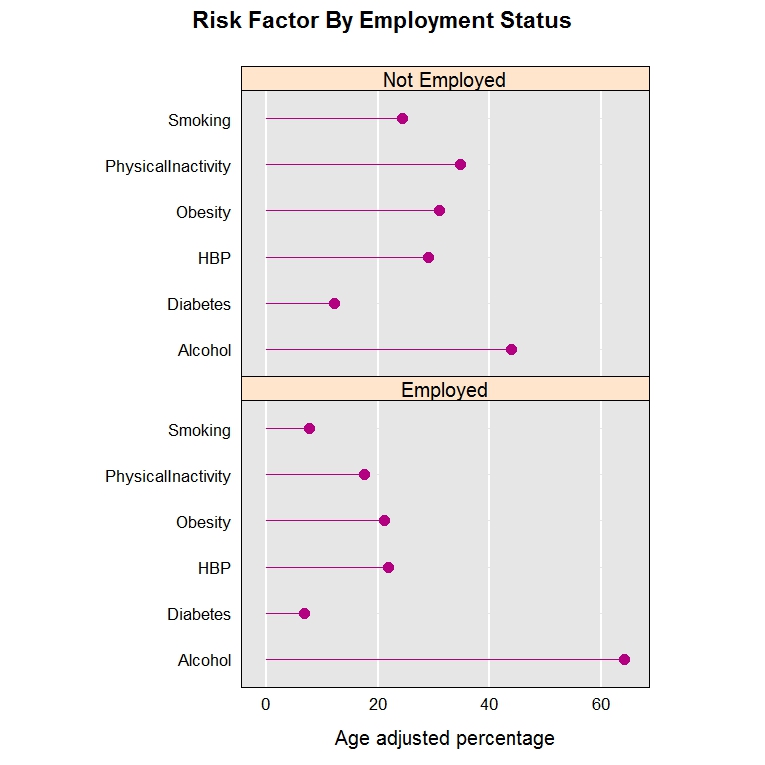
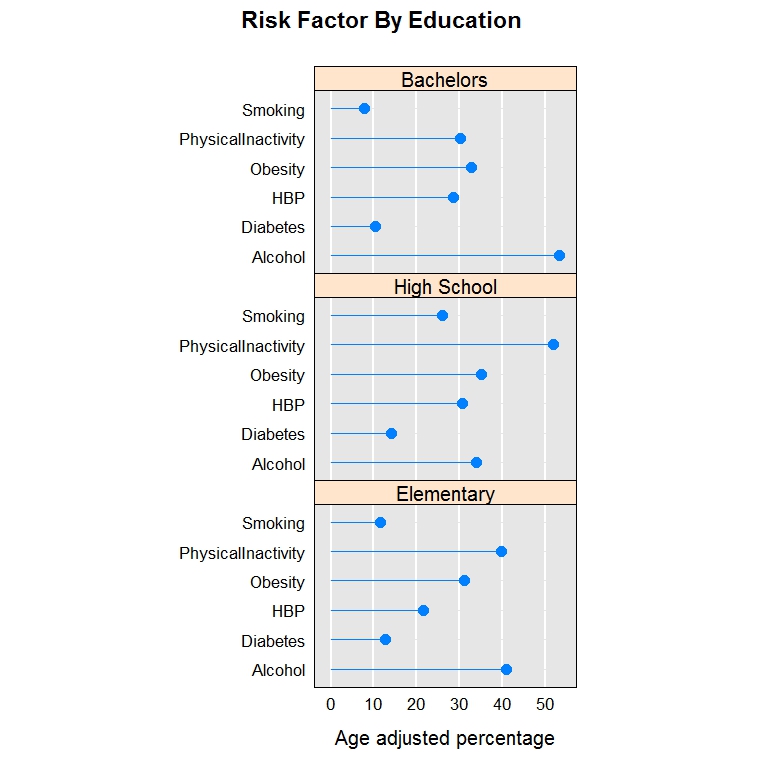


Fig 3.2.1 Risk Factors by Education Fig 3.2.2 Risk Factors by Employment

***Risk Factor by Income:***

Smoking, Obesity, High Blood Pressure and alcohol consumption is more in the high income range whereas the physical inactivity is more in the low income sector and is obvious.

***Risk Factor by Employment Status:***

Unemployed people have high prevalence of risk factors except alcohol and are similar to the risk by education.

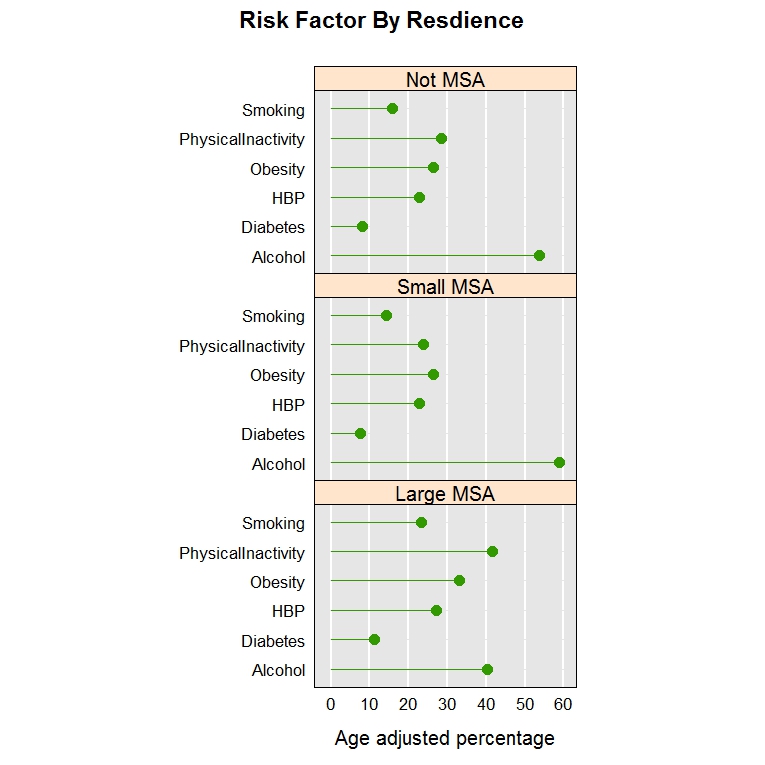
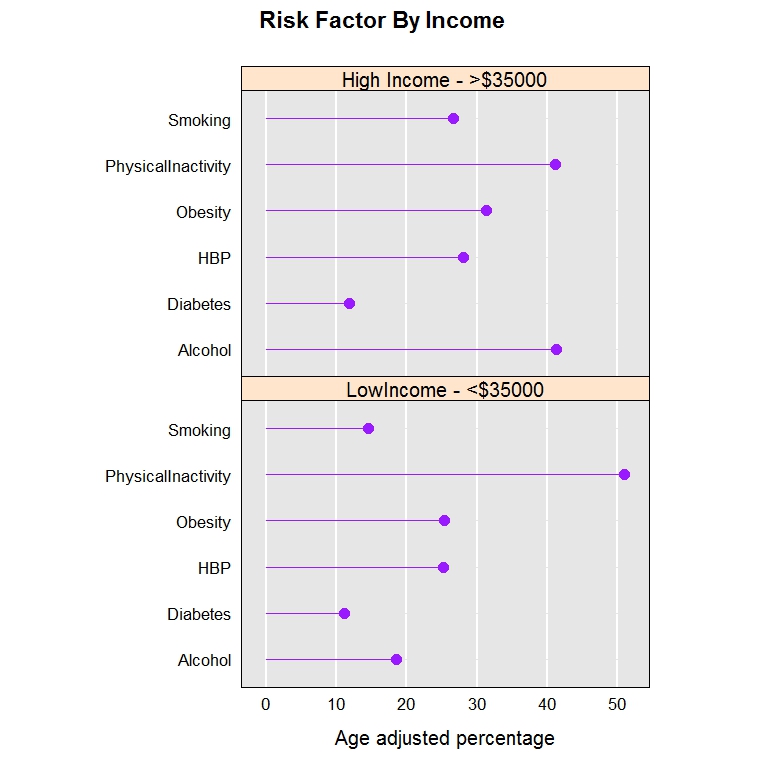


Fig 3.2.3 Risk Factors by Income Fig 3.2.4 Risk Factors by Residence

***Risk Factor by Residence:***

Again except alcohol, the prevalence of risk factor decreases with decrease in metropolitan status. Large metros have high prevalence of risk factors.

* 1. **Demographic Factors**

***Region:***

In general, alcohol consumption is more in all the regions and the factors are evenly spread. Obesity remains close to 30% among all the regions. High blood pressure is close to 25% for all the regions except the Midwest where it touches 30%.

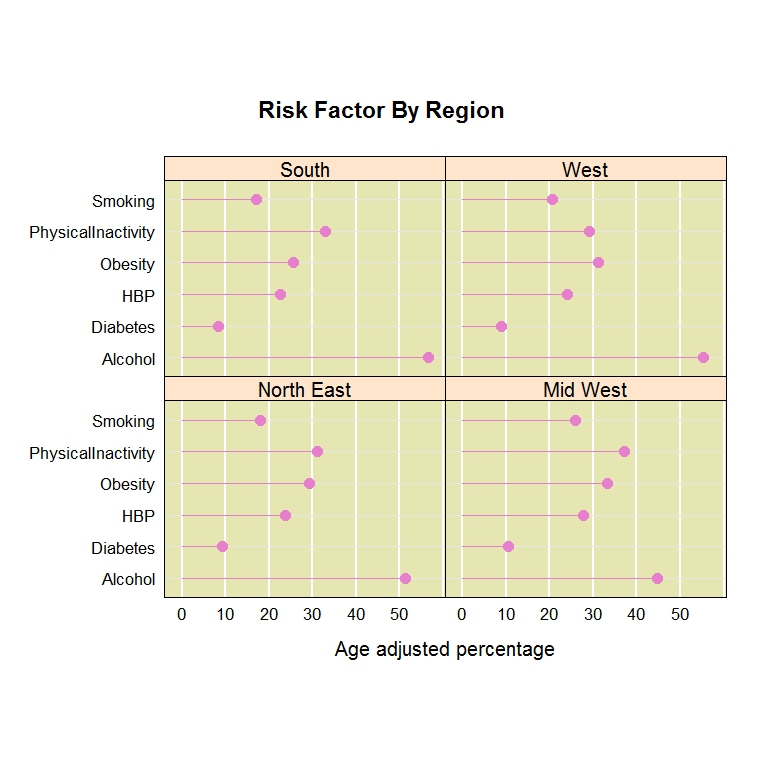


Fig 3.3.1 Risk Factors by Region

1. **State Wise Analysis of Risk Factors**

Micromaps are the best way to represent the state data. It makes it very easy to compare and contrast the factors across the states in United States. Also it provides confidence interval and lot of other options to make it more self descriptive.

The micromap shown in next page maps the heart disease death rate with various risk factors prevalence such as alcohol consumption, diabetes, obesity, high blood pressure and smoking.

The map shows that the alcohol has less effect on heart diseases. Diabetes shows good relevance but for few states it is not directly correlated. For example West Virginia has highest diabetes rate but it is not one of the top states having higher death rate.

Obesity and High Blood Pressure are high correlated with each other and with heart disease death rate as well. This stresses the point that adults who are obese are more prone to high blood pressure and heart diseases.

Smoking shows a good correlation but there are lot of mismatch in the data. For example, West Virginia shows uncorrelated relationship. The confidence interval is high for the states having high smoking prevalence.

***Surprising Facts:***

West Virginia is one of the states having highest diabetes prevalence, high blood pressure, obesity and smoking but it is not in the list of top five states having high death rate due to heart disease.

Another state that stands out is Arkansas. It has high prevalence of all the five risk factors but it is one of the states having very low death rate due to heart disease.

Yet another state is DC which shows similar kind of statistics.

It will be really useful to look at the health care policies and other factors that make the difference in these three states.

1. **County Wise Risk Factor Analysis**

With the previous analysis of state data we could drill down the top states having higher death rates – Mississippi, Alabama, Oklahoma, California and Louisiana. The analysis also showed better correlation factors as obesity and high blood pressure.

Scatter plot matrix could be used to further explore the county data for these top ranked states.

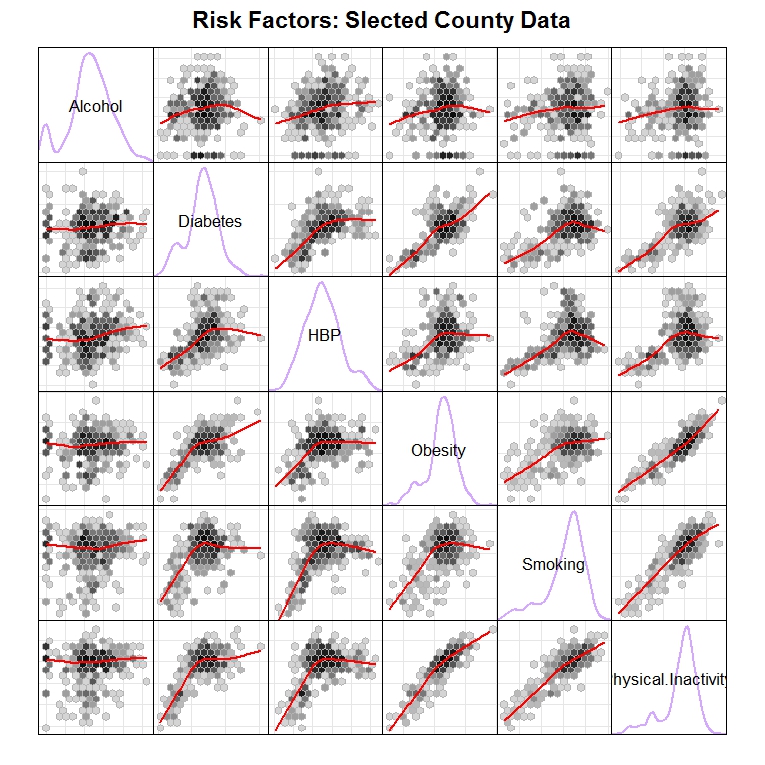


Fig 5.1 County Wise Risk Factors for states MI, AL, CA, LA and OK

Scatter plot matrix of risk factors shows another interesting factor that diabetes and obesity are proportional and have a linear fitting. We could figure out other relations such as smoking and physical inactivity has high correlation.

* 1. **Comparison of Obesity and Heart Disease for Mississippi**

Fitting a smooth over the ggplot provides a better understanding of the correlation. The graph in Fig 6.1 shows a ggplot with smooth for the county wise data for Mississippi. Because of lot of outliers the smooth is constant over the plot. Removing the outliers will give better fitting of data and it is depicted by simply removing one of the explicit outlier towards the left.

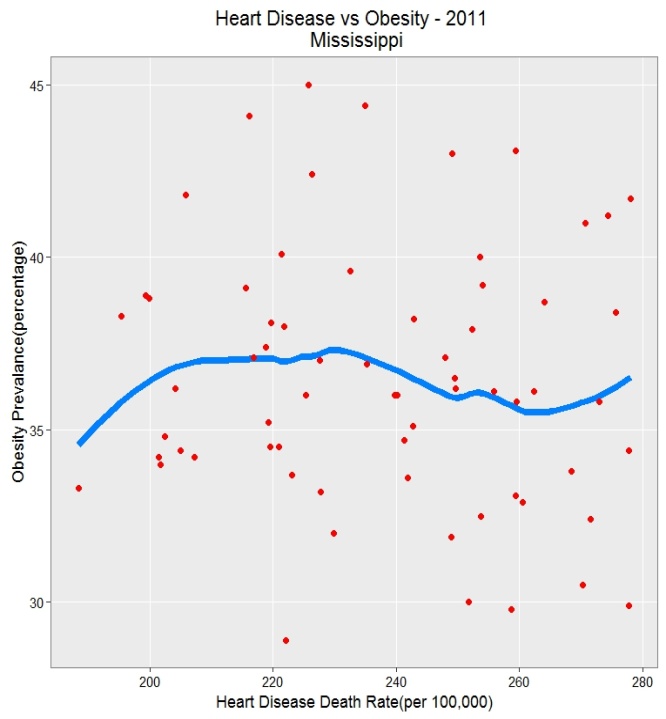
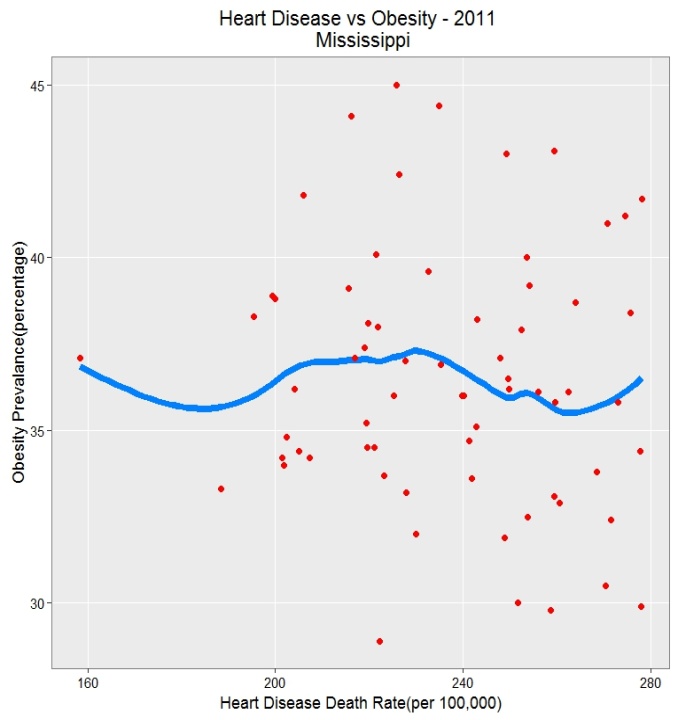


Fig 5.1.1 Heart Disease vs. Obesity with Outlier Fig 5.1.2 Heart Disease vs. Obesity without Outlier