ANLY 512: Factors influencing deaths due to cardiovascular diseases: A county-level analysis

Code

Raghu Sanugommula, Subhash Pemmaraju, Karan Parekh, Rohan Mashru

27th April 2019

Introduction

According to Blood Pressure Control article posted on CDC(Centers for Disease Control and Prevention), how uncontrolled blood pressure can lead to heart disease, stroke, kidney disease, and the possibility of death. In the same article, they mentioned about 70% of US adults; ages 65 or older have high blood pressure. Nearly 50% of adults ages 65 or older with high blood pressure do not have it under control. Five million adults, ages 65 or older, are not taking their blood pressure medicine as directed by the health care providers. Not only blood pressure is the reason, but there are also various factors that can influence to cardiovascular disease, For example, diabetes, intake of high cholesterol foods, ethnic background, smoking, physical inactivity, etc.

The Centers for Disease Control and Prevention (CDC) is the leading national public health institute of the United States. The CDC is a United States Federal Agency under the Department of Health and Human Services division. Its objective is to protect public health and safety through the control and prevention of disease, injury, and disability in the US and internationally. It is specific focuses is on infectious disease, foodborne pathogens, environmental health, occupational safety, and health promotion, injury prevention and education activities designed to improve the health of United States citizens. Also, the CDC researches and provided information on non-infectious diseases such as obesity and diabetes and is a founding member of the International Association of National Public Health Institute. CDC believes that public health and scientific advancement are best served when data are released to, or shared with, other public health agencies, academic researchers, and appropriate private researchers in an open, timely, and proper way.

In this research, we will try to study various factors influencing heart-related diseases, and how it is affected our existence with time using census and publically made available heart diseases related data on CDC. The intention behind giving a brief about our data source is to increase public outreach to the vital information about disease control and prevention measures.

We explored the following questions in this project and found some interesting trends:

1. What parts of the United States(US) have more prevelance deaths due to heart problems? How does the values look like by County Level?
2. How does the obesity and diabetes prevalance correlate?
3. How does the deaths due to heartdiseases looks like per state?
4. How does different factors correlate with deaths/1000? -diabetes prevelance Vs deaths/1000 correlate -obesity prevelance Vs deaths/1000 correlate -Inactivity Vs Deaths/1000 -Percentage 65 and Up vs Deaths/1000
5. How does the inactivity and Obesity are related?
6. How does the different factors by enthinicity looks like?
7. How does the normalized obesity by state looks like?
8. How does the normalized death by state looks like?

Project Work Breakdown

We as a team followed the typical industry standard project approach to attain our desired results within permissible periods. Below are the list of the tasks we have completed utilizing individual teammate area of strengths.

Project Introduction and Preliminary Discussions Team Members: RS, SP, KP and RM Timeline: February 2019 to March 2019 Status: Completed

- Project Kick off meeting where we discussed the various interesting topics
- Finalize the topic
- Research and post ideas related to the topic on teams site (Everyone on the team had done this exercise)
- List of all possible objectives we want to achieve
- Discuss various R methods could be used to achieve the objectives
- Write up the project material like Introduction summaries, backgrounds, objectives, etc.

Development and Testing Team Members: RS, SP, KP and RM Timeline: March 2019 to April 2019 Status:Completed

- Each team member worked in concurrent addressing different objectives using appropriate R coding techniques and Visualizations
- Data Preparation - Data Cleansing
- R Code addressing each objective
- Test the code for accuracy and completeness
- Fit visuals into the report document

Conclusion summary and Submit files on Moodle for Grade Resources: RS, SP, KP, and RM Timeline: End of the Semester (on or before the deadline) Status: Completed

- Create final HTML document using R markdown template provided by Professor
- Concluding text based on results
- Complete the project materials
- Results Evaluation
- Submit final HTML file on to Moodle

Data and Methods

The Census Bureau is the federal government's largest statistical agency. It has expertise in providing current facts about America's people, places, and the economy. Federal law protects the confidentiality of all the information the Census Bureau collects. It conducts a variety of surveys like American Community Surveys (ACS), Demographic Surveys, Economic Surveys, and Sponsored Surveys that shows numbers of what the U.S Population, income, poverty, education and many other subjects. These surveys help the government to allocate over 400$ billion federal funds every year. The United States Census Bureau is committed to confidentiality and guarantees non-disclosure of any addresses or personal information related to individuals or establishments. Only after 72 years does the data collected become available to other agencies or the public. We have choosen to use US Census Demographic Data from Kaggle, the data is extracted from the DP03 and DP05 tables of the 2015 American community survey 5- year estimates.The American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about our nation and its people[1],[2],[3],[4].

Information from the survey generates data that help determine how more than $675 billion in federal and state funds are distributed each year. Through the ACS, we know more about jobs and occupations, educational attainment, veterans, whether people own or rent their homes, and other topics. Public officials, planners, and entrepreneurs use this information to assess the past and plan the future.

When you respond to the ACS, you are doing your part to help your community plan for hospitals and schools, support school lunch programs, improve emergency services, build bridges, and inform businesses looking to add jobs and expand to new markets, and more.

The master source was cited for the references purposes.

More than 30 million people in the United States have diabetes. 84 million adults have prediabetes. That's nearly a third of Americans who have diabetes or are at high risk for type 2 diabetes. This is one of the most serious public health problems our nation has ever faced, and it has enormous and far-reaching consequences. CDC's Division of Diabetes Translation believes in the power of science to turn the tide in the diabetes epidemic. We are dedicated to putting that science into action through programs and policies that help people prevent type 2 diabetes and improve the health of everyone living with diabetes. We continue to make important strides with the understanding that much more needs to be done.

CDC is an open platform for data scientists to explore data, we have extracted variables like heart diseases, obesity distribution by counties, inactivity percentage distribution etc. we have merged this data set to the census for matching counties poplution, ethinicity etc.

Data Analysis Methods

Data Analysis techniques we leveraged to use for exploring our reserach questions are Mapping, box plots, correlations,regression models, barcharts. The data from CDC and Kaaggle for Census data has to be merged and create various categories for supporting our analysis. There are empty values has to be fixed in the data set to avoid the discrepancies. We primarily concentrated on identifying how the various factors are related to one another so more of correlations may be found in this project. The data was segregated by county and state level to support plotting for factors on the map. After merging the data sets and created a final master dataset the total variables are 72 and the observations are 3225.The final data set was merged with other supporting datset by countyID.The details of analysis and interpretations of visualizations are explained in the later sections of the document.

Data Analysis and Visulization

Baseline Statistics

Hide

```
knitr::kable(d1,align=c('c','c','c'),caption="Number of States and Counties")
```

Number of States and Counties

State	Number of Counties
AK	27
AL	67
AR	75
AS	1
AZ	15
CA	58
CO	64
CT	8
DC	1
DE	3
FL	67
GA	159
GU	1
HI	5
IA	99
ID	44
IL	102
IN	92
KS	105
KY	120
LA	64
MA	14
MD	24
ME	16
MI	83

State	Number of Counties
MN	87
MO	115
MP	1
MS	82
MT	56
NC	100
ND	53
NE	93
NH	10
NJ	21
NM	33
NV	17
NY	62
OH	88
OK	77
OR	36
PA	67
PR	78
RI	5
SC	46
SD	66
TN	95
TX	254
UT	29
VA	134
VI	3
VT	14
WA	39
WI	72
WV	55
WY	23

Table 1

Hide

```
knitr::kable(d2,align=c('c','c','c'),caption="Population vs County")
```

Population vs County

County	Total Population
“Los Angeles, (CA)”	10057155
“Cook, (IL)”	5227575
“Harris, (TX)”	4434257
“Maricopa, (AZ)”	4088549
“San Diego, (CA)”	3253356

Table 2

Hide

```
knitr::kable(d3,align=c('c','c','c'),caption="Population vs State")
```

Population vs State

State	Total Population
CA	38654206
TX	26956435
FL	19934451
NY	19697457
IL	12851684

Table 3

Hide

```
knitr::kable(d4,align=c('c','c','c'),caption="Deaths/100,000 resulting from cardiovascular disease")
```

Deaths/100,000 resulting from cardiovascular disease

state	avgDeath
PR	14.90128
MS	13.50976
AL	13.44179
SC	12.58913

state	avgDeath
WV	12.19455

Table 4

Hide

```
knitr::kable(d5,align=c('c','c','c'),caption="Median Household Income ($1000s)")
```

Median Household Income ($1000s)

state	income
DC	73.10000
NJ	72.66190
CT	72.45000
MD	68.85417
MA	67.45714

Table 5

Hide

```
label(heart$Value)<-"Deaths/10,000"
label(heart$dm_prev_adj) <- "% with diabetes"
label(heart$ob_prev_adj) <- "% with obesity"
label(heart$ltpia_prev_adj) <- "% physically inactive"
label(heart$perc_white) <- "% that is white"
label(heart$perc_65up) <- "% over 65"
label(heart$pctui) <- "% Uninsured"
label(heart$medicaid) <- "% Medicaid eligible"
table1(~Value+dm_prev_adj+ob_prev_adj+ltpia_prev_adj+perc_white+perc_65up+pctui+medicaid, data=heart)
```

	Overall (n=3225)
Deaths/10,000	
Mean (SD)	239 (57.0)
Median [Min, Max]	233 [-1.00, 621]
% with diabetes	
Mean (SD)	9.77 (2.41)
Median [Min, Max]	9.50 [-1.00, 18.4]
% with obesity	
Mean (SD)	30.4 (7.03)
Median [Min, Max]	31.4 [-1.00, 48.0]
% physically inactive	
Mean (SD)	24.9 (6.55)
Median [Min, Max]	25.5 [-1.00, 42.1]
% that is white	
Mean (SD)	74.9 (23.4)
Median [Min, Max]	83.8 [-1.00, 100]
% over 65	
Mean (SD)	17.5 (4.58)
Median [Min, Max]	17.2 [-1.00, 53.1]
% Uninsured	
Mean (SD)	11.7 (5.45)
Median [Min, Max]	11.3 [-1.00, 37.4]
% Medicaid eligible	
Mean (SD)	20.8 (10.7)
Median [Min, Max]	21.3 [-1.00, 62.0]

Table 6

Summary of Baseline statistics

Table 1 shows the number of counties by state. The dataset comprises of 56 states and 3225 counties. The highest number of counties are 254 in the state of TX and lowest is 1 in the states of AS, DC, GU and MP.

Table 2 shows the highest populated counties and **Table 3** shows the most populated states. Los Angeles has the highest total population. Almost double of Cook. CA state has the hgihest population followed by Texas. CA is almost double of FL which comes third.

Table 4 shows the states with the highest average deaths/10000 due to cardiovascular disease. PR has the highest.

Table 5 shows the median household income by state. DC has the highest median salary followed by NJ and then CT

Table 6 shows baseline statistics of all the key variables we use in our analysis. This gives a good understanding of the data we are dealing with.

Hide

```
# Color Palette
pal <- colorQuantile("YlOrRd", NULL, n = 20)


# Plot Leaflet map 1
plot1  = leaflet(data = leafmap) %>% addTiles() %>%
  addPolygons(fillColor = ~pal(Value),
              fillOpacity = 0.8,
              color = "#BDBDC3",
              weight = 1) %>%
  setView(lat=33.5, lng=-90,zoom=5.2)%>%
  addRectangles(    lng1=-82, lat1=31,
                    lng2=-98, lat2=36,
                    fillColor = "transparent")


# Plot Leaflet map 2
plot2 = leaflet(data = leafmap) %>% addTiles() %>%
  addPolygons(fillColor = ~pal(Value),
              fillOpacity = 0.8,
              color = "#BDBDC3",
              weight = 1) %>%
  setView(lat=37.5, lng=-119,zoom=5.2)%>%
  addRectangles(    lng1=-115, lat1=33,
                    lng2=-123, lat2=42,
                    fillColor = "transparent")

# Plot the maps
sync(plot1, plot2)
```

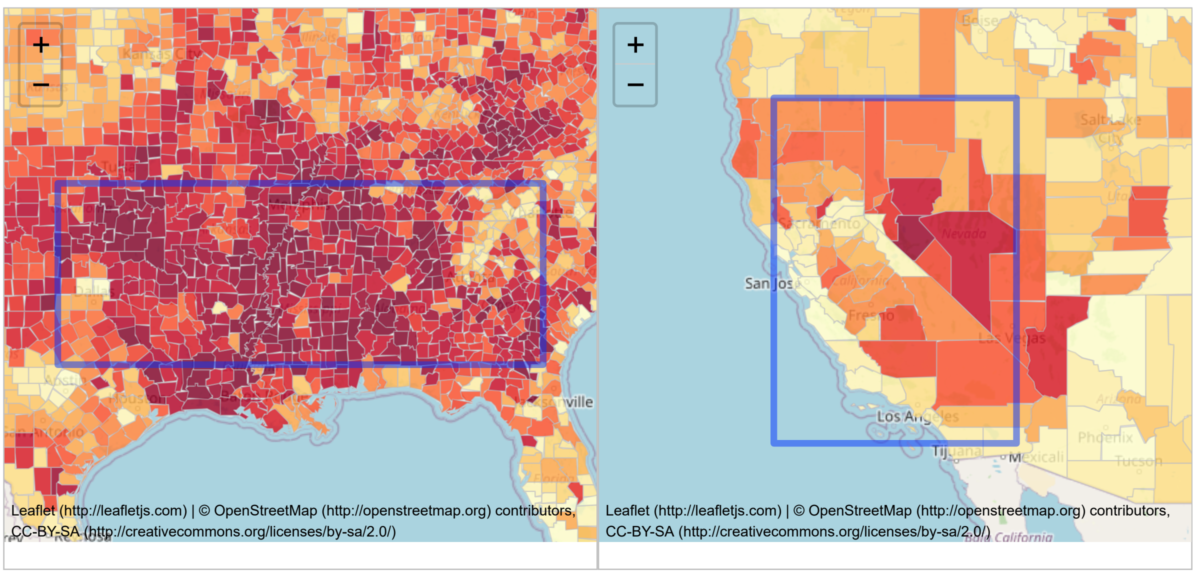

Figure 1: Plot of distribution of deaths due to heart disease across the US

VISUAL DESCRIPTION: COUNTIES HAVE MORE DEATHS DUE TO HEART PROBLEMS

As we stated earlier the data is segregated by county so the spread can be interpreated at micro level. The US map along with a boundaries of counties was displayed in this visualization. The color indication explains what counties have the deaths due to heart problems. counties are more darker have the more deaths due heart problems[5]. The regions highlighted blue have more deaths compare to other counties. it is evident that more deaths are within the southeast region. It seems, the reason for these regions having more deaths by heart diseases could be weekfood choices, poverty, physical inactivity, obesity, climate conditions etc. Awareness of healthy diets and living practices should be promoted in these regions. To further examine the results, we have created a box plot which shows the values comparision between the states.

Hide

```
# Identify the quartiles of the death data
##quantile(leafmap$Value[is.na(leafmap$Value)==0], c(0.25, 0.50, 0.75))

# Bucket the data approximately based on quartile
leafmap$heartdeath[leafmap$Value<200]<-'0-200'

leafmap$heartdeath[leafmap$Value>=200 & leafmap$Value < 250]<-'200-250'

leafmap$heartdeath[leafmap$Value>=250 & leafmap$Value < 300]<-'250-300'

leafmap$heartdeath[leafmap$Value>=300]<-'300 & higher'


# Plot a map of the data based on the buckets
factpal <- colorFactor("YlOrRd", leafmap$heartdeath)

leaflet(data = leafmap) %>% addTiles() %>%
  addPolygons(stroke=FALSE, color = ~factpal(heartdeath), fillOpacity=1) %>%
  addLegend("bottomright", pal = factpal, values = ~heartdeath,
            title = "Deaths from heart disease/10,000",
            opacity = 1) %>%
  setView(lat=39.8, lng=-98.6,zoom=3.5) %>%
  addRectangles(    lng1=-82, lat1=31,
                    lng2=-98, lat2=36,
                    fillColor = "transparent") %>%
  addRectangles(    lng1=-115, lat1=33,
                    lng2=-123, lat2=42,
                    fillColor = "transparent")
```

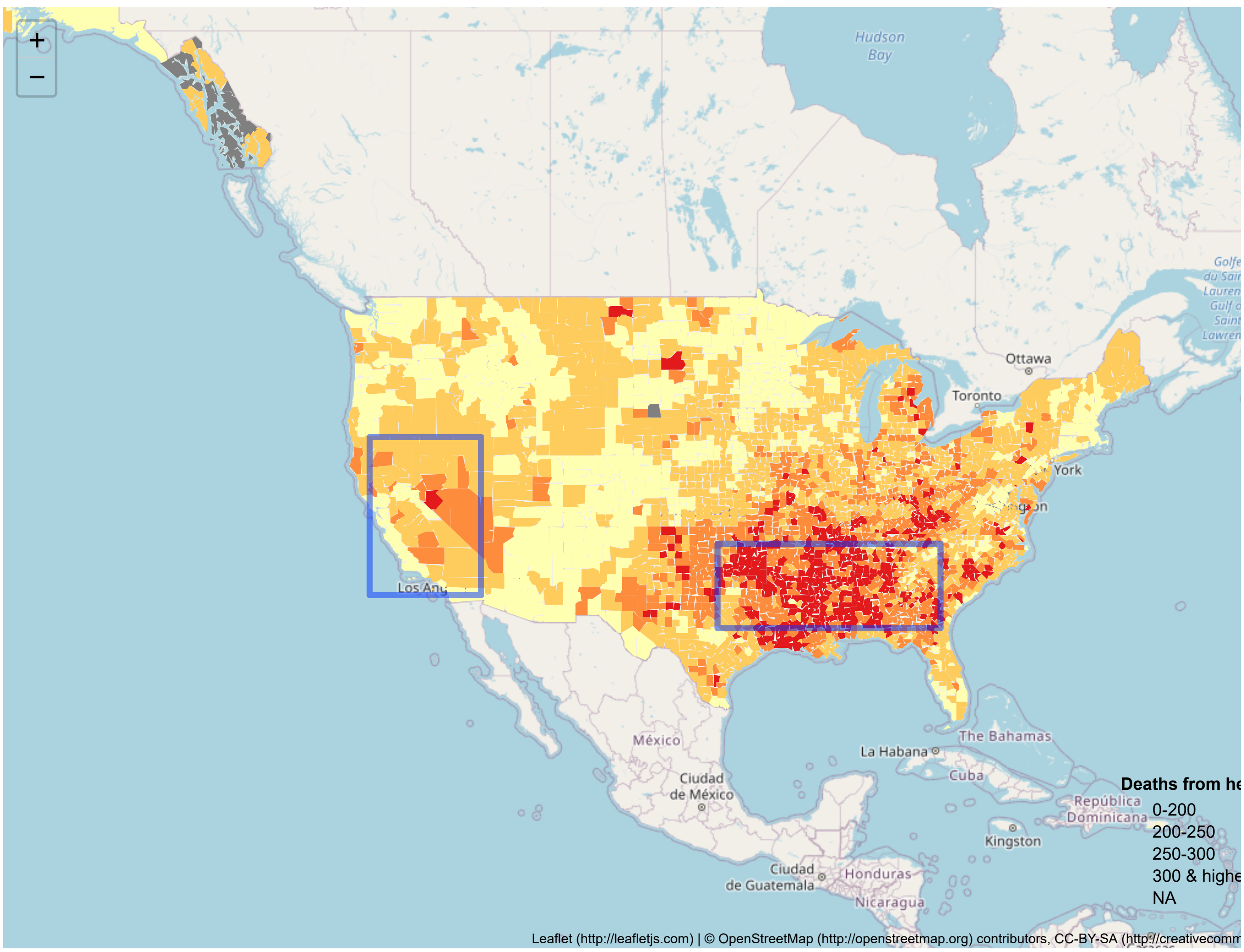

Figure 2: Deaths due to cardiovascular disease in the US

VISUAL DESCRIPTION: COUNTIES HAVE MORE DEATHS DUE TO HEART PROBLEMS; 250 Deaths per every 10,000 population

The visual map shows how the deaths due to heart poblems are distributed by counties. The color level indicates the scale of the deaths in easch county. It seems, south east regions have more deaths compare to others. In contrast to south east, west and central regions are relatively have low deaths due to heart problems. If we would interpret this in general, the climate could also be a strong contributor to living style diseases like, obesity, thus leads to the high cholestrol, which is a major contributor of heart disease.Physical inactivity is also a leading problem of lifetsyle diseases, and climate is a major cotributor of being active. However, these areas have to get more facilites and awareness for increasing the quality of lifestyle and avoid deaths due to heart problems.

Hide

```
### Extract the colums you need for plotting
data1<-data.frame(leafmap$Value, leafmap$dm_prev_adj, leafmap$ob_prev_adj, leafmap$ltpia_prev_adj, leafmap$perc_65up, leafma
p$heartdeath)

data1<-drop_na(data1)

#Plot the impact of diabetes and obesity on deaths due to heart disease
ggplot(data=data1, aes(x=leafmap.ob_prev_adj, y=leafmap.dm_prev_adj, color=leafmap.heartdeath))+geom_point()+labs(title = "O
besity Prevalence vs Diabetes Prevalence", x="Obesity %", y="Diabetes %", color="Deaths/10,000")
```

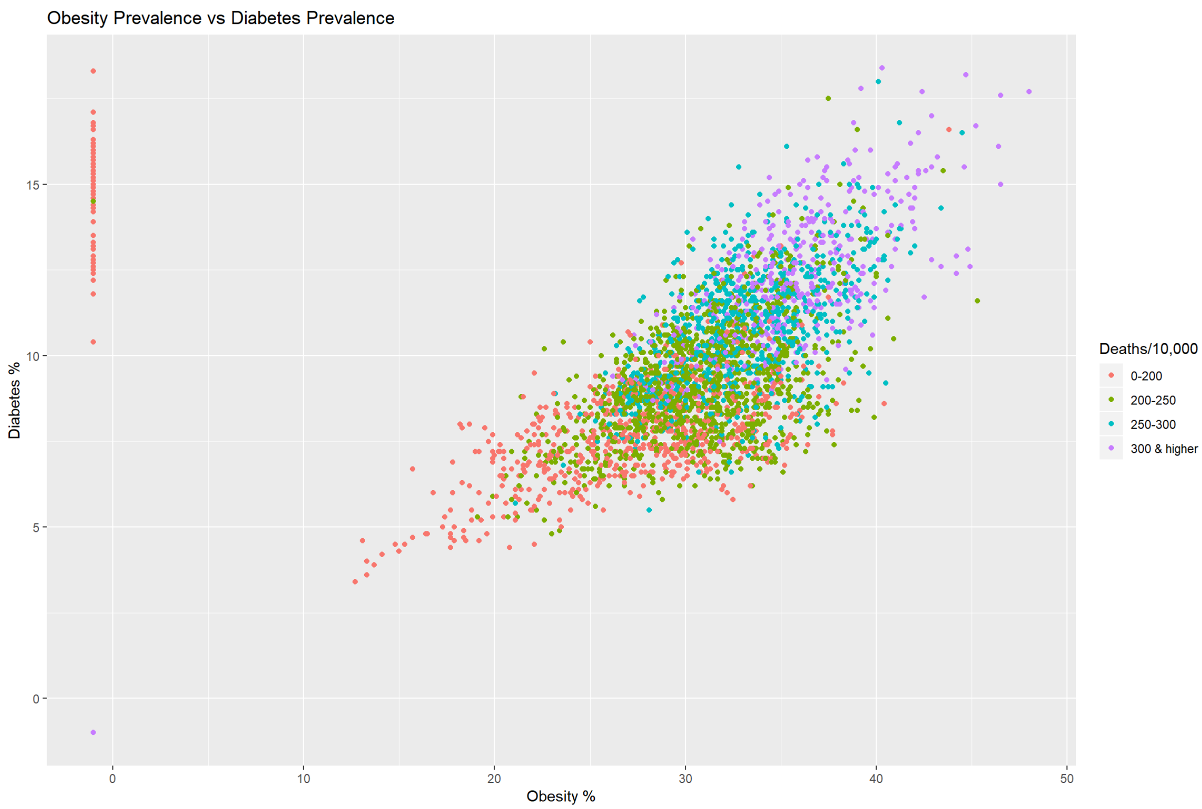

Figure 3: Obesity Prevalence and Diabetes Prevalence and its impact on deaths due to cardiovascular disease in the US

VISUAL DESCRIPTION: Correlation Between Obesity and Diabetes

As can be seen from the graph, there is a strong positive correlation between obesity and diabetes and these both in turn are correlated with deaths due to heart diseaseAt higher levels of obesity, there is more likelihood of diabetes and higher deaths due to heart disease.

This is a general knowledge that atleast everyother person has in todays time. However, there seems no preventive measures are taken to cotrol obesity, which is a major contributor of all kinds of diseases. Though the data was categorized into deaths per every 10000, still it has very strong positive correlation. Hence, it clearly indicates, controling the obesity may help prevent diseases and deaths at early ages.

Hide

```
# Extract abbreviated state names from the dataset's detailed county names
leafmap$state<-substr(leafmap$display_name, start=str_length(leafmap$display_name)-3, stop=str_length(leafmap$display_name)-
2)



data2<-data.frame(leafmap$Value, leafmap$dm_prev_adj, leafmap$ob_prev_adj, leafmap$ltpia_prev_adj, leafmap$perc_65up, leafma
p$heartdeath, leafmap$state, leafmap$medicaid, leafmap$pctui)

# Look at distribution of deaths due to heart disease by state
ggplot(data2, aes(x=leafmap.state, y=leafmap.Value, fill=leafmap$heartdeath))+geom_boxplot()+labs(title="Deaths due to heart
disease", x="State", y="Deaths/10,000", fill="Deaths/10,000")
```

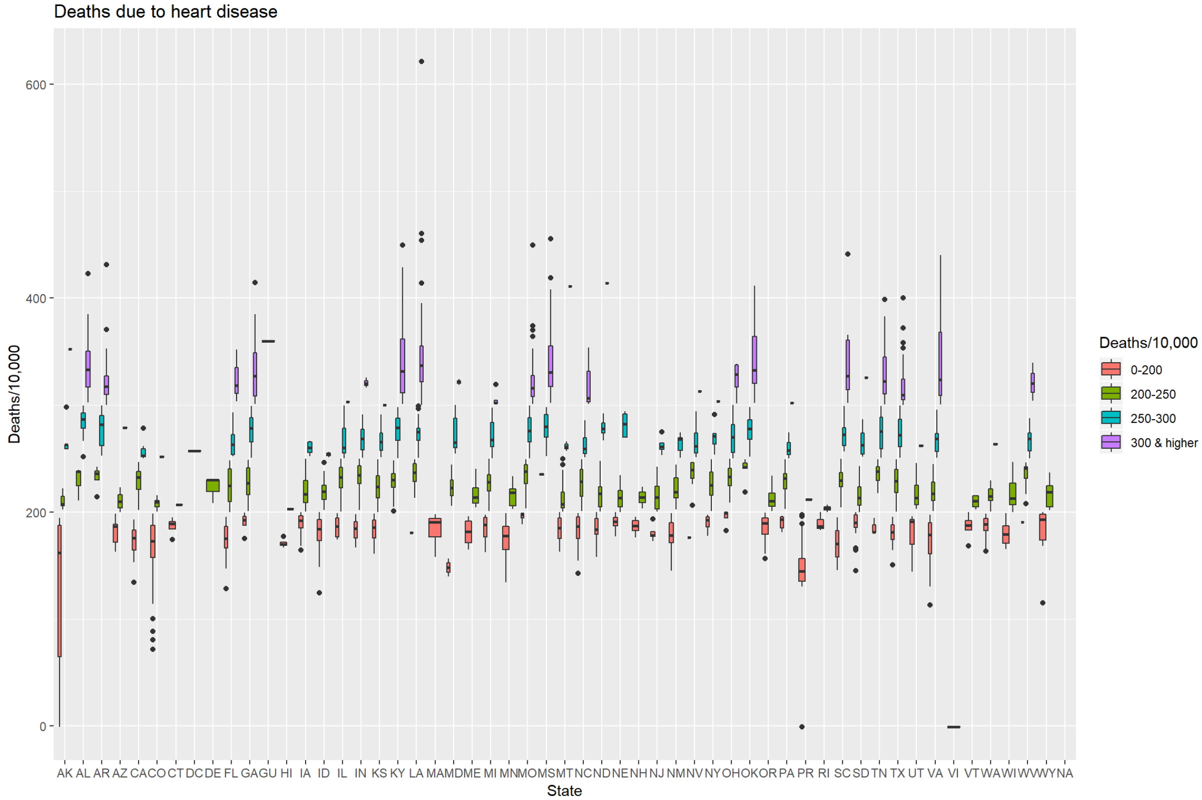

Figure 4: Deaths from heart disease by state

VISUAL DESCRIPTION As can be seen from the boxplot, there is almost a clear separation in the incidence of deaths due to diabetes in terms of the four segments. In addition, the data is presented for comparing the values between the different states in the US along with the four segments. From a quick peek, readers may understand that virginia has more deaths due to diabetes compare to other states.

Los angeles has the max outlier and similarly few other states have the outliers but not identical compare to LA. It is clearly visible in this box plot that south east states are having the significant deaths due to heart problems supporting our map in earlier sectios of the document.

Hide

```
# Plot all the variables that impact death due to heart disease and study their relationships
p1<-ggplot(data=data2, aes(leafmap.dm_prev_adj, leafmap.Value))+geom_point()+geom_smooth()+labs(title = "Diabetes Prevalence vs Deaths/10000", x="Diabetes %", y="Deaths")
p2<-ggplot(data=data2, aes(leafmap.ob_prev_adj, leafmap.Value))+geom_point()+geom_smooth()+labs(title = "Obesity Prevalence vs Deaths/10000", x="Obesity %", y="Deaths")
p3<-ggplot(data=data2, aes(leafmap.ltpia_prev_adj, leafmap.Value))+geom_point()+geom_smooth()+labs(title = "Inactivity vs Deaths/10000", x="Inactivity %", y="Deaths")
p4<-ggplot(data=data2, aes(leafmap.perc_65up, leafmap.Value))+geom_point()+geom_smooth()+labs(title = "Percentage 65 and Up vs Deaths/10000", x="65 & Up %", y="Deaths")

grid.arrange(p1, p2, p3, p4)
```

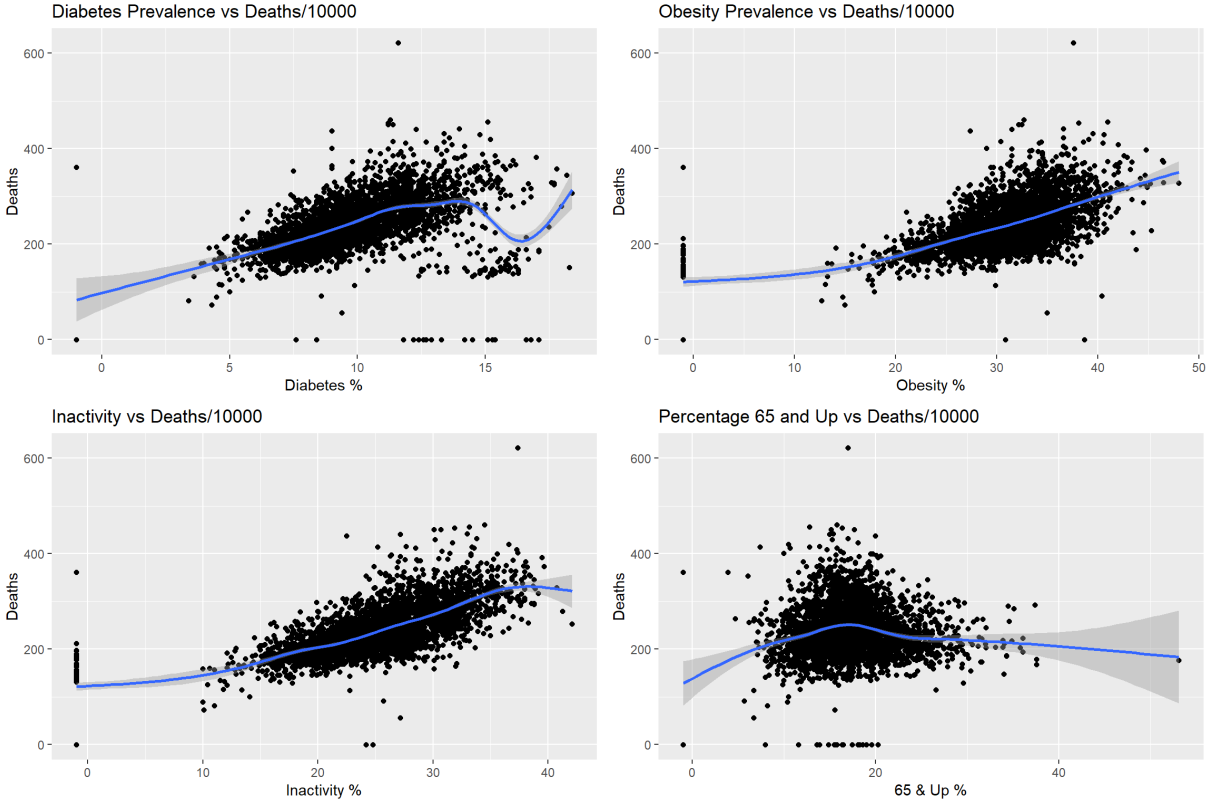

Figure 4: Impact of Diabetes, Obesity and Inactivity on deaths in the US

VISUAL DESCRIPTION

As can be seen from the plots, the strongest influence on deaths due to heart disease appears to be from inactivity and obesity, followed by diabetes and the % of population over 65 has the least influence.

Hide

```
ggplot(data=data2, aes(leafmap.ltpia_prev_adj, leafmap.ob_prev_adj, color=leafmap.heartdeath))+geom_point()+labs(title = "In
activity % vs Obesity %", x="Inactivity %", y="Obesity %", color="Deaths")
```

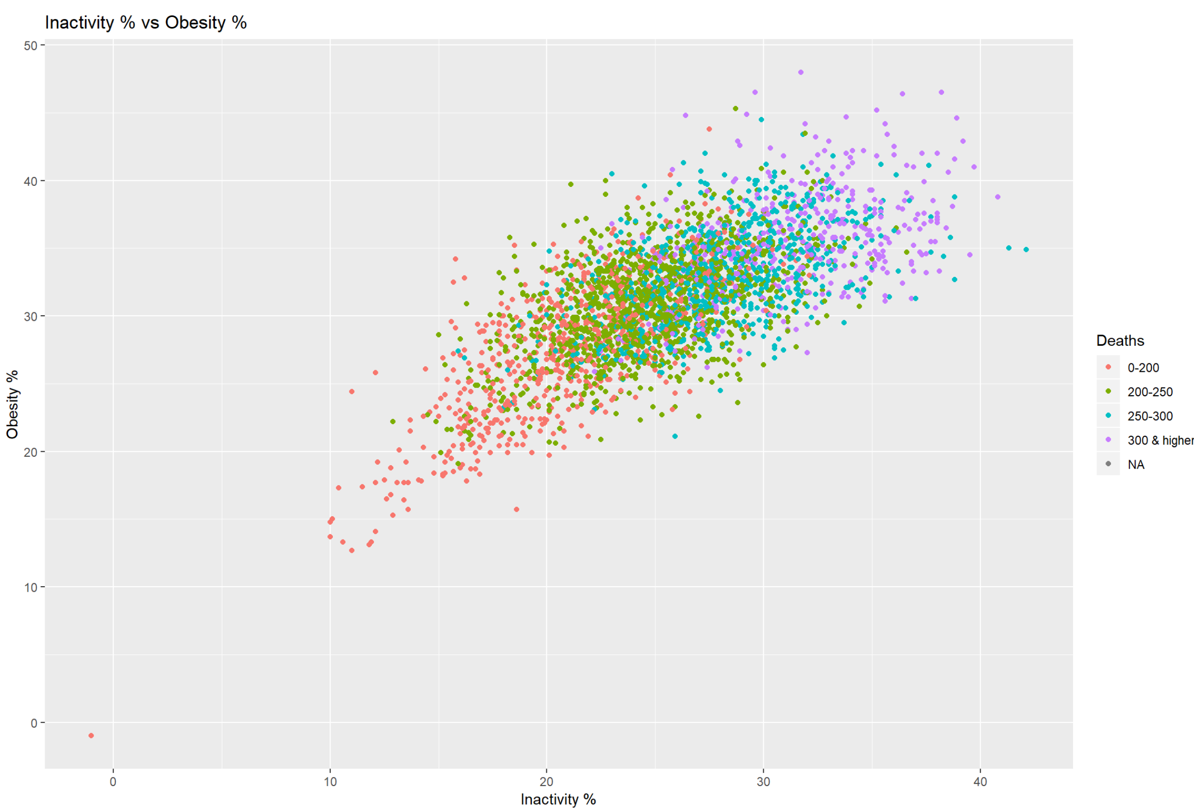

Figure 5: Inactivity vs Obesity and deaths in the US

VISUAL DESCRIPTION

It is evident from the graph that the High inactivity and high obesity leads to more deathhs since it shows a strong postitve correlation.It is a known fact and our data solidifies it. More effort spent on being active, the more less of the deaths due to obesity.

Hide

```
### High inactivity and high obesity leads to more deaths as can be seen from this chart

# Calculate % of non-white
leafmap$perc_nonwhite<-100-leafmap$perc_white

leafmap$nonwhite[leafmap$perc_nonwhite>=50]<-'>=50% Non-white'
leafmap$nonwhite[leafmap$perc_nonwhite<50]<-'<50% Non-white'

data3<-data.frame(leafmap$Value, leafmap$nonwhite, leafmap$dm_prev_adj, leafmap$ob_prev_adj, leafmap$ltpia_prev_adj, leafmap
$perc_65up)


# Difference by racial distribution
p1<-ggplot(data=subset(data3, !is.na(leafmap.nonwhite)), aes(x=leafmap.nonwhite, y=leafmap.dm_prev_adj, fill=leafmap.nonwhit
e))+geom_boxplot()+labs(y = "Diabetes Prevalence", x="% Non-White", fill="% Non-white")
p2<-ggplot(data=subset(data3, !is.na(leafmap.nonwhite)), aes(x=leafmap.nonwhite, y=leafmap.ob_prev_adj, fill=leafmap.nonwhit
e))+geom_boxplot()+labs(y = "Obesity Prevalence", x="% Non-White", fill="% Non-white")
p3<-ggplot(data=subset(data3, !is.na(leafmap.nonwhite)), aes(x=leafmap.nonwhite, y=leafmap.ltpia_prev_adj, fill=leafmap.nonw
hite))+geom_boxplot()+labs(y = "Inactivity Prevalence", x="% Non-White", fill="% Non-white")
p4<-ggplot(data=subset(data3, !is.na(leafmap.nonwhite)), aes(x=leafmap.nonwhite, y=leafmap.perc_65up, fill=leafmap.nonwhit
e))+geom_boxplot()+labs(y = "65 and Up %", x="% Non-White", fill="% Non-white")
grid.arrange(p1, p2, p3, p4)
```

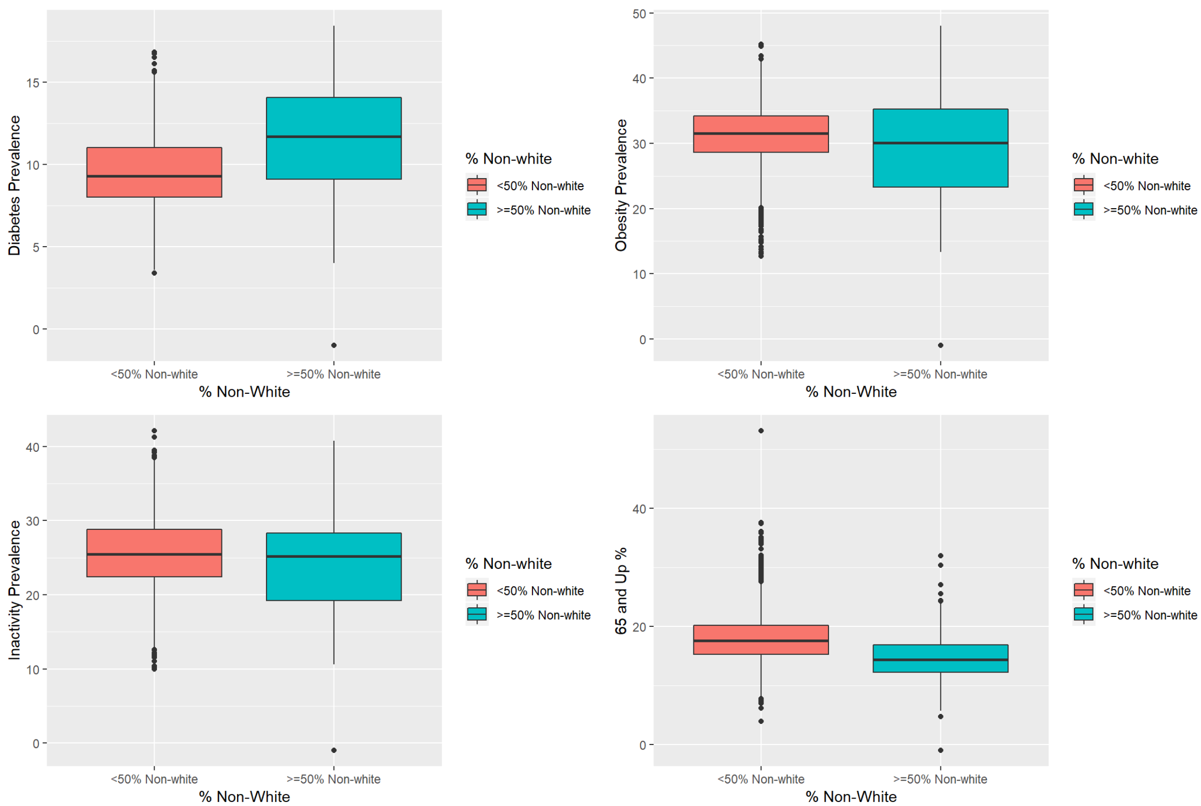

Figure 6: Obesity, Diabetes, Inactivity and 65 & above population by race

VISUAL DESCRIPTION

It is intresting to find that the non white population are more compare to the white population have diabetes prevalence, obesity prevalance and inactivity prevalence.In contrast, non white 65 and Up % are low compare to the other age groups though the data was minimal[6,7].

Hide

```
#State level correlation of obesity and deaths due to heart disease
data4<-data.frame(leafmap$Value, leafmap$ob_prev_adj, leafmap$state)


data4$death_z<-round((data4$leafmap.Value-mean(data4$leafmap.Value, na.rm=TRUE))/sd(data4$leafmap.Value, na.rm=TRUE),2)
data4$death_type<-ifelse(data4$death_z<0, "Below", "Above")
data4<-data4[order(data4$death_z),]
data4$leafmap.state<-factor(data4$leafmap.state, levels=unique(data4$leafmap.state))


data4$ob_z<-round((data4$leafmap.ob_prev_adj-mean(data4$leafmap.ob_prev_adj, na.rm=TRUE))/sd(data4$leafmap.ob_prev_adj, na.r
m=TRUE),2)
data4$ob_type<-ifelse(data4$ob_z<0, "Below", "Above")
data4<-data4[order(data4$ob_z),]
data4$leafmap.state<-factor(data4$leafmap.state, levels=unique(data4$leafmap.state))

p1<-ggplot(data4, aes(x=reorder(leafmap.state, ob_z), y=ob_z, label=ob_z)) +
  geom_bar(stat='identity', aes(fill=ob_type), width=.5)  +
  labs(title="Normalised obesity by state",
       x="normalized obesity by state", y="state") +
  coord_flip()


p2<-ggplot(data4, aes(x=reorder(leafmap.state, death_z), y=death_z, label=death_z)) +
  geom_bar(stat='identity', aes(fill=death_type), width=.5)  +
  labs(title="Normalised death by state",
       x="normalized death by state", y="state") +
  coord_flip()
grid.arrange(p1,p2, ncol=2)
```

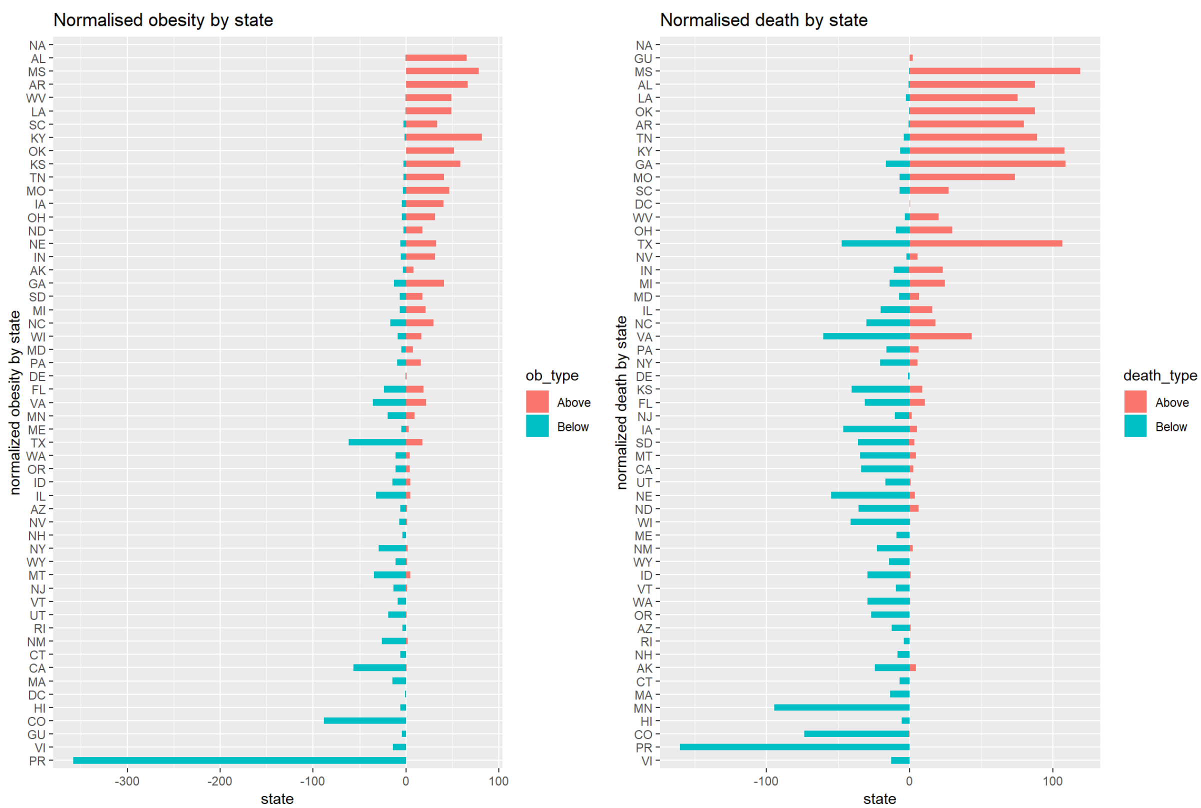

Figure 7: Normalized Obesity and Deaths by State

VISUAL DESCRIPTION

In our previous analysis it was proven that the obesity is a major cause of deaths. Here in this bar chart, it shows how the normalized deaths and obesity is distributed by state, and in many of the states with the highest incidence of obesity also have the highest incidence of deaths due to heart disease Eg: Alabama, Mississippi, Arkansas,Kentucky, Oklahoma etc[8],[9].

Hide

```
# Look at distribution of Percentage without Health Insurance, Under Age 65
ggplot(data2, aes(x=leafmap.state, y=leafmap.pctui))+geom_bar(stat = 'identity', aes(fill=leafmap$heartdeath), width =.5)+la
bs(title="Percentage without Health Insurance, Under Age 65", x="State", y="% without Health Insurance (Under 65)", fill="De
aths/10,000")
```

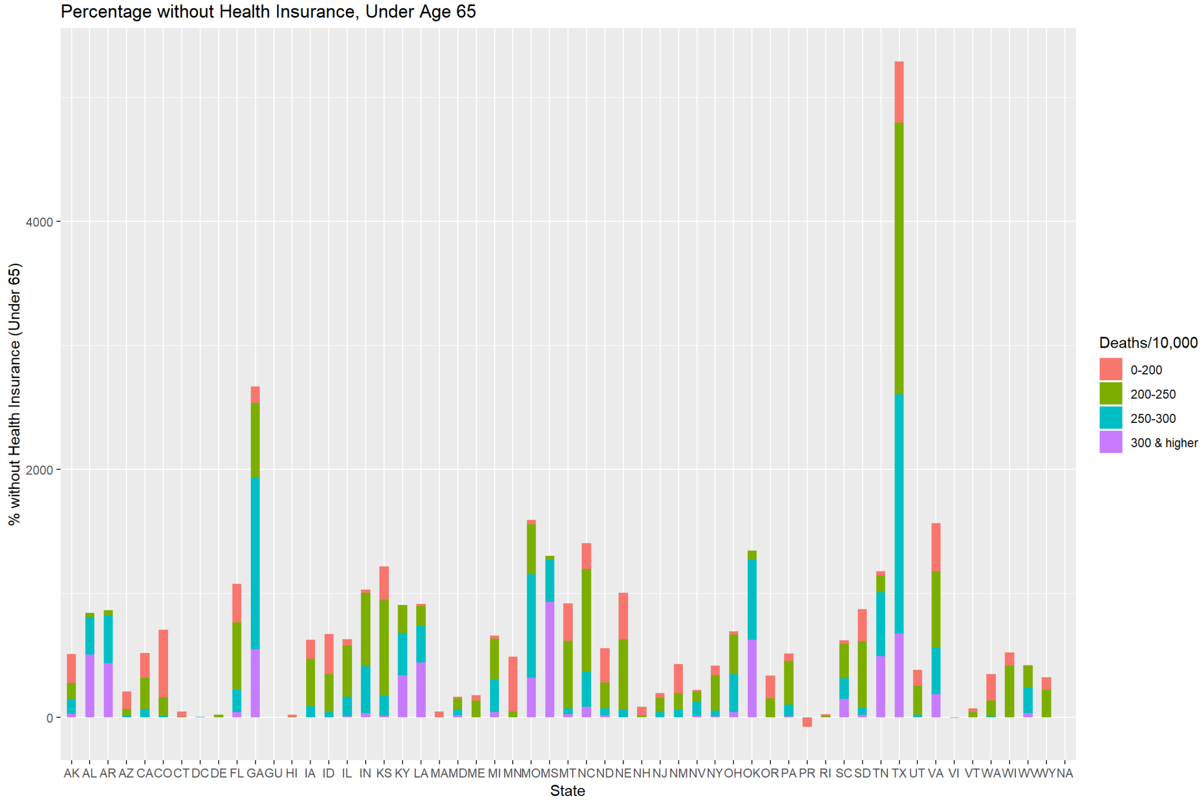

Figure 8: Distribution of population without health insurance by State

Hide

```
# Look at distribution of Percentage who are Medicaid Eligible
ggplot(data2, aes(x=leafmap.state, y=leafmap.medicaid))+geom_bar(stat = 'identity', aes(fill=leafmap$heartdeath), width =.5)
+labs(title="Percentage with Medicaid Eligible", x="State", y="% with Medicaid Eligible", fill="Deaths/10,000")
```

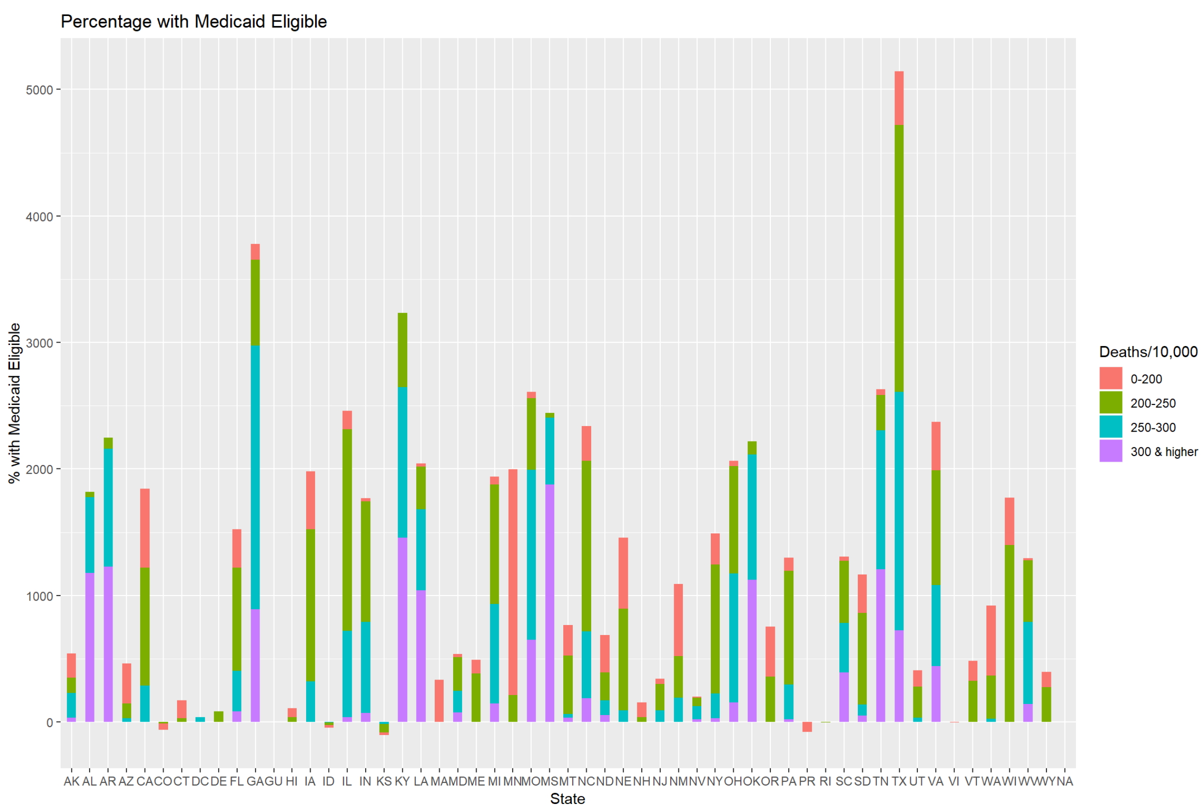

Figure 9: Distribution of population eligible for Medicaid by State

VISUAL DESCRIPTION

Figures 8 and 9 show the distribution of populations[10] across the state by two key variables: percentage of population without health insurance and percentage of population eligible for Medicaid. As can be seen, states like Mississippi, Kentucky, Tennessee, Oklahoma, Georgia andd Texas have a high level of people without health insurance and high level of people eligible for Medicaid and a number of people dying due to heart disease.

CONCLUSION AND DISCUSSION

We conclude our project with some intresting findings like obesity is an Vital contributor for cardio vascular diseases.Physical inactivity is also have strong correlation with obesity and heart problems. Intrestingly, non white popluation have more prevelance to obesity and diabetes. Another intresting finding was most of the states have significant deaths due to heart problems are in the south east region.

Therefore, we could interpret this has the climate is also a major contibutor for lifestyle diseases. Climates in general contributes to human phycological behaviour,and a good climate may keep human motivated for being physically active and allow spend more time towards physical activities. Since we cannot manipulate climate, atleast, an awareness shall be brought to these regions that your communities are most affected from these diseases could keep them physchologically active and put themselves from inactive to active with time.Though the researh hepled us finding a lot of intresting facts, it is a subject needed a deeper investigation which can aid decision making bodies take an effective measures preventing these deaths going forward.We will try to hold our pace and will continue research on this subject though this project is specific to this semester. Thank you for giving an oppurtunity to study this subject using effective R Data methodlogies. Data can save the planet earth if it's intensity is properly utilized.

REFERENCES

1. Center for Disease Control: US County level data on blood presssure related illnesses; https://www.cdc.gov/vitalsigns/blood-pressure/index.html (https://www.cdc.gov/vitalsigns/blood-pressure/index.html)↩
2. Center for Disease Control policy for sharing and releasing data; https://www.cdc.gov/maso/policy/releasingdata.pdf (https://www.cdc.gov/maso/policy/releasingdata.pdf)↩
3. Kaggle dataset on US Demographics; https://www.kaggle.com/muonneutrino/us-census-demographic-data (https://www.kaggle.com/muonneutrino/us-census-demographic-data)↩
4. US Census Data on American Community Survey; https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2015/5-year.html (https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2015/5-year.html)↩
5. Color palettes in R; https://stat.ethz.ch/R-manual/R-devel/library/grDevices/html/palettes.html (https://stat.ethz.ch/R-manual/R-devel/library/grDevices/html/palettes.html)↩
6. Getting rid of NAs in plots; https://stackoverflow.com/questions/17146213/nas-are-being-plotted-in-boxplot-ggplot2/40829195 (https://stackoverflow.com/questions/17146213/nas-are-being-plotted-in-boxplot-ggplot2/40829195)↩
7. Side-by-side plots with GGplot2; https://stackoverflow.com/questions/1249548/side-by-side-plots-with-ggplot2 (https://stackoverflow.com/questions/1249548/side-by-side-plots-with-ggplot2)↩
8. Reorder, Factor functions in R; https://stat.ethz.ch/R-manual/R-devel/library/stats/html/reorder.factor.html (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/reorder.factor.html)↩
9. Plotting data in descending order; https://stackoverflow.com/questions/16961921/plot-data-in-descending-order-as-appears-in-data-frame (https://stackoverflow.com/questions/16961921/plot-data-in-descending-order-as-appears-in-data-frame)↩
10. Plotting an overlapping histogram in R; https://www.r-bloggers.com/overlapping-histogram-in-r/ (https://www.r-bloggers.com/overlapping-histogram-in-r/)↩