Exploring Potential Environmental Racism in Cancer Alley, Louisiana

Ashley Weaver

May 2020

Table of Contents

[Introduction 1](#_Toc39830527)

[Methods 2](#_Toc39830528)

[Results and Discussion 3](#_Toc39830529)

[Cancer Alley Race Data 3](#_Toc39830530)

[Comparing Race and Health Data for the Entire State of Louisiana 5](#_Toc39830531)

[Analylizing Data Distributions 6](#_Toc39830532)

[Louisiana Health Compared to Health in the United States 8](#_Toc39830533)

[Cancer Alley Health Data 11](#_Toc39830534)

[Other Factors that could be Impacting Health 12](#_Toc39830535)

[Insurance 12](#_Toc39830536)

[Food and Exercise Availability 15](#_Toc39830537)

[Conclusions and Recommendations 16](#_Toc39830538)

[References 17](#_Toc39830539)

## Introduction

Cancer Alley, Louisiana is an area along the Mississippi River between Baton Rouge and New Orleans. It derived its notorious nickname, Cancer Alley, from the abundance of oil refineries and petrochemical companies along the river. It is already known that this area has an unusually high rate of cancers, which is attributed to the air and water contamination by industry. Environmental Racism in Cancer Alley, Louisiana has previously been explored but mostly in terms of cancer. This study will be exploring the other potential health risks that people living in Cancer Alley experience compared to the rest of Louisiana. This study will also look at other factors that could explain differences in quality of health by county.

The state of Louisiana is known for having very limited environmental regulations and therefore very high levels of pollution. Louisiana was ranked by US News as the worst pollution state in terms of exposure to pollution and related health risks. Additionally, there are numerous articles documenting the lack of enforcement of federal clean air, clean water, and hazardous waste laws. The lack of environmental regulation is part of the reason for the abundance of industrial sites in the state. High levels of industry and low levels of environmental protection have led to adverse health impacts throughout the state. Previous studies have documented the effects of these industrial sites on the number of cancer cases in the state.

There are measures of increased cancer risk and racial data for some of the counties along what is known as Cancer Alley. For instance, the lifetime risk of cancer in St. John Baptist is 800 times higher than an average american (Pasley 2020). Nationally, there has been a 16% decrease in toxic releases by plants over the last 30 year, but in Cancer Alley reported toxic releases grew by 25%. It is also noted that these oil refinery and chemical plants are located next to African American communities along the Mississippi River (Pasley 2020).

This study will be focused on exploring race and health data in the state of Louisiana to assess whether there is a link between health problems and race, particularly in Cancer Alley. It is important to note that this study is limited to looking at data by counties, though Cancer Alley itself is not entire counties. Populations that are closer to the industries and the Mississippi River (where these industries dump waste) are more likely to be impacted in terms of health. Since the study looks at health in Louisiana at a county level, it is possible that the adverse health impacts to communities on the river will not be reflected by the county-wide data.

## Methods

Three difference R packages were utilized to create the following analysis. The dplyr package was used extensively to create new descriptive variables and reorganize and restructure data frames to facilitate the analysis. The ggplot2 package was used in order to visualize the analysis through the creation of a variety of different plot types. Finally, the tidyr package was used to combine data from the race and health datasets so that they could be analyzed on the same plot.

The first portion of my analysis will focus on race data for counties that are considered to be part of Cancer Alley. These counties include: East Baton Rouge, West Baton Rouge, Ascension, St. John the Baptist, Iberville, St. Charles, Jefferson, and Orleans. My analysis will explore the differences in diversity of these areas compared to one another and the state as a whole. This will be followed by a discussion of the limitation of my racial analysis given the structure of the data.

The second part of my analysis will look at race and health data in the entire state of Louisiana. This portion of the analysis will focus on finding correlation between poor health and percentage of minorities in a county. This will include a scatter plot that examines health as a function of percentage of minorities in the community. This will give a more holistic view of Louisiana’s health by race. This will be followed by theoretical qq-plots to assess the normality of the data and isolate particular counties in Louisiana that appear to be in worse health. The health data will then be compared to other states in the United States to assess Louisiana’s statewide health compared to national health.

The third part of my analysis will focus primarily on the health in the counties that are part of Cancer Alley. This portion of the analysis uses a box plot to compare the percentage poor health in the counties of Cancer Valley, Louisiana to percentage poor health in the state and country as a whole.

The final portion of my analysis will review other factors that could be contributing to adverse health effects. This will involve an evaluation of the potential impact of lack of health insurance on Louisiana. This will be demonstrated through box plots that will compare the lack of insurance by age in the US and lack of insurance by age in Louisiana. Comparing the difference between National lack of insurance to just state lack of insurance in Louisiana will give us a better idea of if insurance (or lack thereof) can be used to explain the health disparity between most U.S. States and Louisiana. Next, this section will explore potential causes of obesity in Louisiana, which is shown in section 2 (“Comparing Race and Health Data for the Entire State of Louisiana”) to have the second highest rate of obesity in the US. This analysis will involve the use of scatterplots to compare obesity in Louisiana to access to healthy foods and physical inactivity.

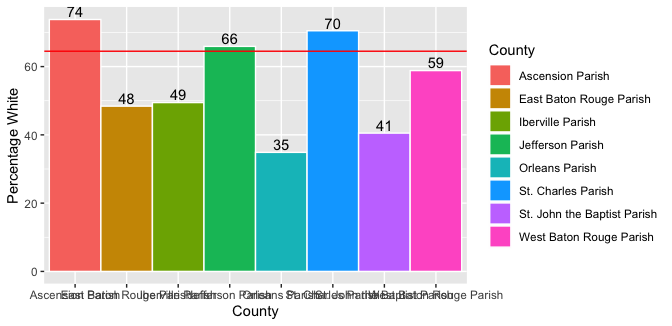
## Results and Discussion

### Cancer Alley Race Data

# Loading Race and Heatlh Datasets  
hdat <- read.csv("../Data/R11371317\_SL050.csv")  
racedat <- read.csv("../Data/cc-est2018-alldata-22.csv")  
  
# Organizing Race Dataset for Louisiana  
racedat1 <- racedat %>%   
 filter(YEAR == 9,  
 AGEGRP == 0) %>%   
 group\_by(COUNTY) %>%   
 mutate(White = sum(WA\_MALE, WA\_FEMALE),  
 Minority = TOT\_POP - White,  
 Perc\_White = (White / TOT\_POP) \* 100,  
 Perc\_Minority = (Minority / TOT\_POP) \* 100) %>%   
 ungroup() %>%   
 select(STNAME, CTYNAME, TOT\_POP, White,   
 Minority, Perc\_White, Perc\_Minority) %>%   
 mutate(LA\_POP = sum(TOT\_POP),  
 LA\_White\_Perc = mean(Perc\_White))  
  
# Creating Race Dataset for only Cancer Alley  
CAracedat1 <- racedat1 %>%   
 filter(CTYNAME %in% c("East Baton Rouge Parish",   
 "West Baton Rouge Parish",  
 "Ascension Parish",  
 "St. John the Baptist Parish",  
 "Iberville Parish",  
 "St. Charles Parish",  
 "Jefferson Parish",  
 "Orleans Parish"))

The first plot is a bar plot comparing the percentage white population in all the Cancer Alley counties. The red line on this plot represents the average percentage of white residents in each county, which is 65%. Of the counties that are part of Cancer Alley, five have a lower percentage of white population than the rest of the state of Louisiana. These five counties include East Baton Rouge, Iberville, Orleans, St. John the Baptist, and West Baton Rouge. These data suggest that much of the population that lives in Cancer Alley is not white. As previous studies have implicated Cancer Alley as a location with high levels of environmental racism, it should be expected that these communities are largely composed of minorities. The fact that three of these counties (Ascension, Jefferson, and St. Charles) have a higher percentage white population than the rest of the state is more than I expected.

# Generating bar plot comparing percentage white populations in Cancer Alley  
ggplot(CAracedat1, aes(x = CTYNAME, y = Perc\_White, fill = CTYNAME)) +  
 geom\_bar(stat="identity", width=1, color="white") +  
 labs(x = "County", y = "Percentage White", fill = "County") +  
 geom\_abline(slope = 0, intercept = 64.5, col = "red") +  
 geom\_text(aes(label= round(Perc\_White)),   
 position=position\_dodge(width=0.9), vjust=-0.25)



Due to the high percentage white population in Ascension, Jefferson, and St. Charles, I found a more precise map of race along Cancer Alley which is shown below (Stepnick).

The above image shows the percentage of the population along the Mississippi River (Cancer Alley) that is black, while also indicating the locations of major industrial sites along the river. It appears that many of these industrial sites are located in areas with high percentages of black populations from 60-100%. The proximity of these majority black communities to industrial sites is an indication of environmental racism.

It is important to note the difference in the discoveries of my own graph and those in the above image. While the image clearly points to environmental racism, my graph suggests that only some of the counties that comprise Cancer Alley are majority non-White. The reason for this discrepancy is that the image looks at a smaller than county level. The health data is limited to the county level and can not be broken down into smaller increments. This highlights one of the limitations of this study, which is that among a county there can be a very high variability among the data. While the image clearly shows high percentages of black populations near industrial sites, my data looks at the diversity of the entire counties which extend miles from the Mississippi River and these industrial sites.

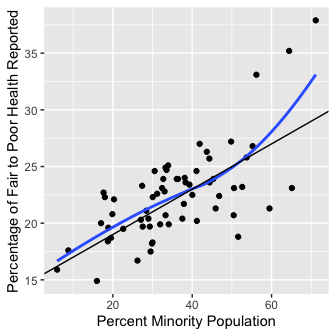
The limitation of this dataset of being broken down by county, instead of even smaller regions, will be important to the rest of the analysis. While people along the Mississippi River may experience adverse health impacts, this may not be reflected in the health data as it will be diluted by the rest of the county’s health.

### Comparing Race and Health Data for the Entire State of Louisiana

# Creating table of Poor Health Data in Louisiana by county  
hdat4 <- hdat %>%   
 filter(Geo\_STATE == 22) %>%   
 select(PoorHealth = "SE\_T002\_001", CTYNAME = "Geo\_NAME") %>%   
 arrange(PoorHealth)  
  
# Generating table of percentage minority by county   
racedat4 <- racedat1 %>%   
 select(CTYNAME, Perc\_Minority)  
  
# Generating a combined table of race and poor health data  
rhcombined <- left\_join(hdat4, racedat4, by = "CTYNAME")

It is important to understand the overall health of minorities in Louisiana to assess whether there is environmental racism. The below chart plots the percentage of minorities against the percentage of poor health by county.

# Plotting the Percentage of Minorities against Percentage of Poor Health Reported in Louisiana  
ggplot(rhcombined, aes(x= `Perc\_Minority`, y = `PoorHealth`)) +   
 geom\_point() +   
 stat\_smooth(method = "loess", se = FALSE, span = 1) + geom\_abline(intercept=15, slope=0.20) +   
 labs(x = "Percent Minority Population",   
 y = "Percentage of Fair to Poor Health Reported")

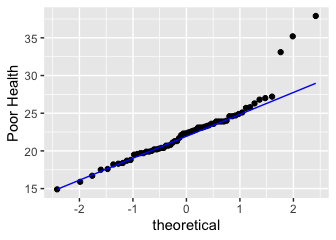


The above plot demonstrates that there is a correlation between percentage minorities and poor health. The plot has both a linear slope and a loess fit to determine if a linear slope best represents the data. The loess fit follows the linear slope pretty well until the last portion of the data where it increases because of the three outliers. The plot suggests that as the percentage of minorities in the population increases the percentage of fair to poor health reported also increases. This suggests that overall, in the state of Louisiana, there could be a connection between poor health and the counties with more minorities. Environmental racism is a potential explanation for this trend, but the potential causes for the fair to poor health of the minorities would need to be evaluated. The “Other Factors that could be Impacting Health” section will further discuss potential reasons for this trend.

#### Analylizing Data Distributions

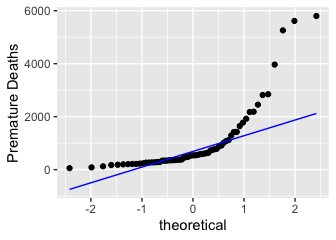
To determine if there were any regions in Louisiana that were particular outliers in terms of poor health or premature deaths, I created qq-plots of these data. Considering previous findings in Cancer Alley of negative health effects, I expected to find outliers in these data for the counties in Cancer Alley. Though there were some outliers in both the poor health and premature death data, neither indicated any of the counties in Cancer Alley as outliers.

# Theoretical QQ Plot of Fair to Poor Health Data   
ggplot(rhcombined) + aes(sample = PoorHealth) + geom\_qq(distribution = qnorm) +  
 geom\_qq\_line(line.p = c(0.25, 0.75), col = "blue") + ylab("Poor Health")



The plot above shows that the Fair to Poor Health data has a relatively normal distribution with the exception of 3 counties. These three counties are Tensas, Madison, and East Carroll, none of which are part of or nearby Cancer Alley. Looking at the next Theoretical QQ plot of premature deaths helps direct attention to a possible explanation for these outliers.

# Creating Dataset of Premature death by County   
hdatdeath <- hdat %>%   
 filter(Geo\_STATE == 22) %>%   
 select(PrematureDeath = "SE\_T007\_001", CTYNAME = "Geo\_NAME") %>%   
 arrange(PrematureDeath)  
  
# Theoretical QQ Plot of Premature Death Data   
ggplot(hdatdeath) + aes(sample = PrematureDeath) +   
 geom\_qq(distribution = qnorm) +  
 geom\_qq\_line(line.p = c(0.25, 0.75), col = "blue") + ylab("Premature Deaths")



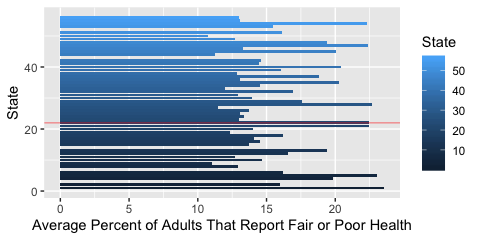
The theoretical QQ plot of premature deaths in Louisiana shows that these data do not have a normal distribution. Deaths were considered to be Premature if they were before the age of 75. Interestingly, Tensas has the second lowest, East Carroll has the third lowest, and Madison has the eighth lowest number of premature deaths. Considering the last QQ plot, where these counties were outliers for having higher reported percentages of fair to poor health, it would be expected for these counties to likely have higher levels of premature deaths. The reason these counties have such a lower premature death count is because they have much smaller total populations than many of the other counties in Louisiana. This could also be a potential explanation for why they had higher percentages of fair to poor health, because there just were not as many data points for people as the counties are smaller. This means that each person’s response in these counties impacts the overall county health evaluation more than people’s responses in larger counties. The upper three outliers on the theoretical QQ plot of premature deaths were Jefferson, East Baton Rouge, and Orleans. All of these counties are a part of Cancer Alley, but they also all have very large population sizes compared to the other counties in Louisiana, which is likely the explanation of their much higher number of premature deaths.

#### Louisiana Health Compared to Health in the United States

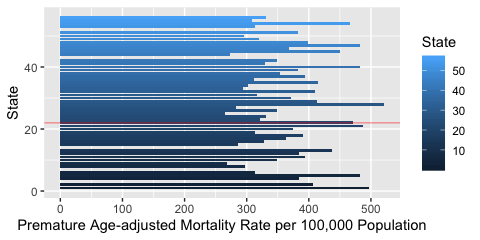
# Health Dataset with mean values for all of the United States  
hdatUS <- hdat %>%   
 select(PoorHealth = "SE\_T002\_001", State = "Geo\_STATE",   
 County = "Geo\_QNAME", PrematureMortality = "SE\_NV006\_003",   
 Perc\_Obese = "SE\_T012\_003") %>%   
 arrange(State) %>%   
 group\_by(State) %>%   
 na.omit() %>%   
 mutate(State\_PoorHealth = mean(PoorHealth), State\_Obese = mean(Perc\_Obese),  
 State\_PrematureMortality = mean(PrematureMortality)) %>%   
 select(State, State\_PoorHealth, State\_Obese, State\_PrematureMortality) %>%   
 distinct()

The next three plots will compare the health of Louisiana to other U.S. states. The red line on all the plots indicates the value that corresponds with the state of Louisiana. The states are numbered but do not labelled, as the focus is solely on the state of Louisiana compared to the other states. The spaces where there are bars missing indicate a lack of health data for these 5 states.

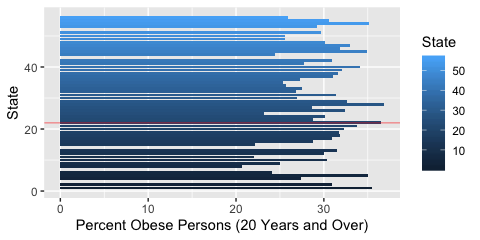
# Generates plot of Average Percent of Adults That Report Fair or Poor Health by State  
ggplot(hdatUS, aes(x = State, y = State\_PoorHealth, fill = State)) +  
 geom\_bar(stat = "identity") +  
 coord\_flip() +   
 geom\_vline(xintercept = 22, col = "red", alpha = 0.4) +  
 scale\_y\_continuous(  
 name = "Average Percent of Adults That Report Fair or Poor Health")



# Generates plot of Premature Age-adjusted Mortality Rate per 100,000 Population by State  
ggplot(hdatUS, aes(x = State, y = State\_PrematureMortality, fill = State)) +  
 geom\_bar(stat = "identity") +  
 coord\_flip() +   
 geom\_vline(xintercept = 22, col = "red", alpha = 0.4) +  
 scale\_y\_continuous(  
 name =   
 "Premature Age-adjusted Mortality Rate per 100,000 Population")



# Generates plot of Percent Obese Persons (20 Years and Over) by State  
ggplot(hdatUS, aes(x = State, y = State\_Obese, fill = State)) +  
 geom\_bar(stat = "identity") +  
 coord\_flip() +   
 geom\_vline(xintercept = 22, col = "red", alpha = 0.4) +  
 scale\_y\_continuous(  
 name =   
 "Percent Obese Persons (20 Years and Over)")



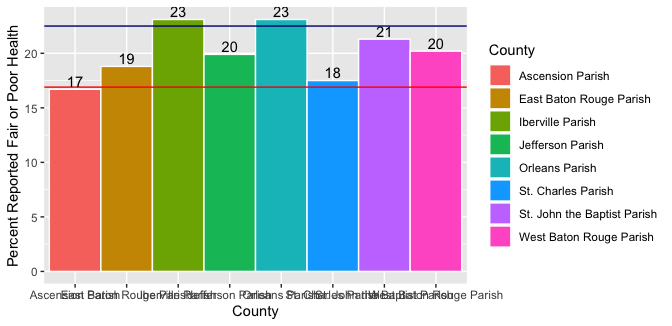
Looking at these plots it is very clear that, compared to other U.S. states, the population in Louisiana overall has some of the worst health. The first plot indicates that Louisiana is in the top 5 states that had the highest reported fair or poor health. The second plot indicates that Louisiana has the seventh highest Premature Mortality Rate per 100,000 population of all the US States. The third plot indicates that Louisiana has the second highest percentage obese persons of all of the US States.

The high rates of poor health, premature deaths, and obesity suggest that Louisiana is in much worse health by comparison to other U.S. states. The poor overall health of Louisiana could possibly have connections to their weak environmental regulations, which enable industrial waste to contaminate their air and water. That said, further research would be needed to fully demonstrate a link between their lack of environmental regulation and adverse health impacts.

### Cancer Alley Health Data

The next plot will hone in on health data specifically in Cancer Alley, Louisiana to determine if there is a correlation between poor health and these counties. The navy blue line represents the average percent reported fair or poor health in all of the United States. The red line represents the average percent reported fair or poor health in the state of Louisiana.

# Creating table of Cancer Alley Health Data  
CA\_Health <- hdat %>%   
 filter(Geo\_STATE == 22) %>%   
 select(PoorHealth = "SE\_T002\_001", Perc\_Obese = "SE\_T012\_003",  
 PrematureMortality = "SE\_NV006\_003", CTYNAME = "Geo\_NAME") %>%   
 filter(CTYNAME %in% c("East Baton Rouge Parish",   
 "West Baton Rouge Parish",  
 "Ascension Parish",  
 "St. John the Baptist Parish",  
 "Iberville Parish",  
 "St. Charles Parish",  
 "Jefferson Parish",  
 "Orleans Parish"))  
  
# Calculating the mean poor health in the U.S. and Louisiana  
USmeanPoorHealth <- mean(hdat$SE\_T002\_001, na.rm = TRUE)  
LouisianaPoorHealth <- mean(hdat4$PoorHealth)  
  
# Generating bar blot of Cancer Alley Poor Health with Louisiana and U.S. Poor Health Averages  
ggplot(CA\_Health, aes(x = CTYNAME, y = PoorHealth, fill = CTYNAME)) +  
 geom\_bar(stat="identity", width=1, color="white") +  
 labs(x = "County", y = "Percent Reported Fair or Poor Health",   
 fill = "County") +  
 geom\_abline(slope = 0, intercept = 16.9, col = "red") +  
 geom\_abline(slope = 0, intercept = 22.5, col = "navyblue") +  
 geom\_text(aes(label= round(PoorHealth)),   
 position=position\_dodge(width=0.9), vjust=-0.25)



The percentage of poor health in Cancer Alley, Louisiana counties is at or above the poor health reported in the United States. That said, most percentages of fair or poor health reported in these counties are below the average for Louisiana. This indicates that there tend to be higher levels of poor health in Louisiana than in the United States. Whether poor health is, in fact, affected by living in Cancer Alley can not be determined by this graph. The health impacts of living in Cancer Alley would need to be assessed at a level that is smaller than the county level to definitively convey adverse health impacts on the communities in and closest to the Cancer Alley region along the Mississippi River.

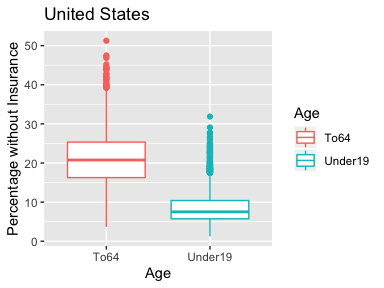
### Other Factors that could be Impacting Health

It was made clear by previous analysis that the state of Louisiana has significantly poorer health than most other U.S. States. This section covers some other factors that could be impacting health in Louisiana outside of potential environmental factors.

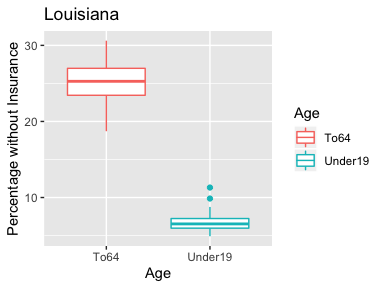
#### Insurance

The following two plots explore the differences in percentage of the Population without insurance by age in the U.S. as a whole and in only Louisiana.

# Creating Dataset for Insurance by Age in United States  
hdat\_in\_US <- hdat %>%   
 select(Under19 = "SE\_T006\_001",   
 To64 = "SE\_T006\_002") %>%   
 pivot\_longer(names\_to = "Age", values\_to = "Perc\_Insurance", cols = 1:2)  
  
# Plotting Percentage of Population without Insurance by Age in United States  
ggplot(hdat\_in\_US, aes(x = Age, y = Perc\_Insurance, col = Age)) +   
 geom\_boxplot() +  
 labs(y = "Percentage without Insurance") +  
 ggtitle("United States")



# Creating Dataset for Insurance by Age in Louisiana  
hdat\_in <- hdat %>%   
 filter(Geo\_STATE == 22) %>%  
 select(Under19 = "SE\_T006\_001",   
 To64 = "SE\_T006\_002") %>%   
 pivot\_longer(names\_to = "Age", values\_to = "Perc\_Insurance", cols = 1:2)  
  
# Plotting Percentage of Population without Insurance by Age  
ggplot(hdat\_in, aes(x = Age, y = Perc\_Insurance, col = Age)) +   
 geom\_boxplot() +  
 labs(y = "Percentage without Insurance") +  
 ggtitle("Louisiana")



In both the United States and Louisiana people between the ages of 18 to 64 have much higher percentages of lacking insurance than people below the age of 19. In the entire United States half of the counties in the US have populations with roughly 16-25% people between 18 and 64 who do not have insurance. Half of the counties in Louisiana, on the other hand, have populations where roughly 24.5-27% of people between 18 and 64 do not have insurance. This is a much higher percentage of counties that have a lack of insurance than the United States as a whole. Though, as expected, the spread of the percent of population without insurance is much higher for the entire U.S. than just the state of Louisiana because it accounts for many more counties.

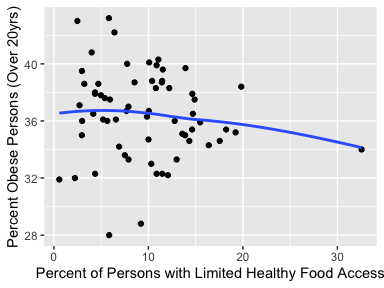
Just looking at the plot of Louisiana it is clear that the percentage of people without insurance by county is much greater for the age range of 18 to 64 than for people under 19. All quartiles of the boxplot for ages 18 to 64 are greater than all the quartiles of the boxplot for under 19. A total of 75%, the upper 3 quartiles, of the data in Louisiana for ages 18 to 64 is greater than the United States median of about 21%. In fact, the majority of Louisiana’s data for the age range of 18 to 64 falls above the United States median of 21%. This means that overall, the counties in Louisiana tend to have a higher percentage of the population without insurance than an average U.S. county.

A higher lack of insurance in Louisiana, particularly among the age range of 18 to 64, is another possible explanation for the poor health in the state. People without insurance are less likely to seek medical help when needed and get routine check-ups, which would contribute to the reported poor health of residents in the state. A lack of health insurance could also be an indication of lower incomes, another factor that could be contributing to the poor health of the state by comparison to the United States as a whole.

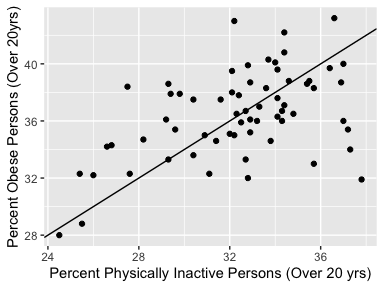
#### Food and Exercise Availability

The next two plots will look at the relationship between Louisiana’s high obesity and food and exercise availability.

# Creating Obesity Dataset for Louisiana  
hdatobese <- hdat %>%   
 filter(Geo\_STATE == 22) %>%   
 select(Perc\_Obese = "SE\_T012\_003", Perc\_LHF\_Access = "SE\_T012\_001",  
 Perc\_Inactive = "SE\_T012\_004", CTYNAME = "Geo\_NAME") %>%   
 arrange(Perc\_Obese)  
  
# Creating Scatterplot Comparing Limited Access to Healthy Food to Obesity   
ggplot(hdatobese, aes(x = Perc\_LHF\_Access, y = Perc\_Obese)) +  
 geom\_point() +   
 stat\_smooth(method = "loess", se = FALSE, span = 1) +   
 labs(x = "Percent of Persons with Limited Healthy Food Access",   
 y = "Percent Obese Persons (Over 20yrs)")



# Creating Scatterplot Comparing Physical Inactivity to Obesity  
ggplot(hdatobese, aes(x = Perc\_Inactive, y = Perc\_Obese)) +  
 geom\_point() +   
 geom\_abline(intercept = 4, slope = 1) +  
 labs(x = "Percent Physically Inactive Persons (Over 20 yrs)",   
 y = "Percent Obese Persons (Over 20yrs)")



The plot on the left shows percent obesity as a function of percent persons with limited healthy food access. There does not appear to be a correlation between the percent of persons with limited healthy food access to the percent obesity in Louisiana. The data is very spread out and does not point towards any trends. Though limited access to healthy food is often linked to obesity, this does not appear to be the case in Louisiana (at least at the county level).

The plot on the right shows percent of obese persons as a function of percent of physically inactive persons. There appears to be a correlation between percent physical inactivity and percent obesity in Louisiana. There is a small cluster of five points of counties with high percentages of inactive people but lower percentages of obesity than the rest of the trend. Overall, the counties in Louisiana that have higher percentages of inactivity, tend to also have higher percentages of obesity.

The slope of the data shown on the right plot indicates that a 1% increase in inactivity appears to be correlated with about a 1% increase in obesity of a county. Though, using the slope, we see that percent obesity appears to consistently be 4% higher than percent inactivity. This indicates the additive shift of 4 to the intercept of the trendline so that it will more accurately reflect this trend in the data.

## Conclusions and Recommendations

Though there has been previous research indicating a correlation between poor health and living in Cancer Alley, Louisiana, these data do not indicate this relationship. It appears that though minority communities do experience a higher degree of health issues, there is no demonstrated correlation by these data of minorities in Cancer Alley experiencing adverse health impacts greater than other counties. Other factors that could be impacting health in Louisiana could include a lack of insurance for those between the ages of 18 and 64, and high levels of physical inactivity.

Further research would need to be conducted to demonstrate negative health impacts, other than cancers, in these regions. This study was limited to county wide data, though Stepnicks plot showed that closer to the Mississippi River there were higher percentages of black populations. This means that black populations tend to live closer to these industrial sites, but the county as a whole could be much more diverse. Similarly, the data for poor health in Cancer Alley was limited to entire communities, though the people that are most directly impacted live right on the Mississippi River. I suspect that these black populations living in Cancer Alley right next to the industrial sites would have a high level of other non-cancer health issues, but further research would be needed to make such conclusions. Future studies would need to isolate smaller regions within each county for analysis. This could involve breaking down the health data by zip code. It would be very effective to analyze health and diversity as a function of distance from these industrial sites (though data collection for this might be challenging).

## References

Census. (2019). County Characteristics Resident Population Estimates. US Census Bureau, Population Division. Retrived from: <https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html>

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Hadley Wickham and Lionel Henry (2020). tidyr: Tidy Messy Data. R package version 1.0.2. <https://CRAN.R-project.org/package=tidyr>

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 0.8.4. <https://CRAN.R-project.org/package=dplyr>

Pasley, James. “Inside Louisiana’s Horrifying ‘Cancer Alley,’ an 85-Mile Stretch of Pollution and Environmental Racism That’s Now Dealing with Some of the Highest Coronavirus Death Rates in the Country.” Business Insider, Business Insider, 9 Apr. 2020, www.businessinsider.com/louisiana-cancer-alley-photos-oil-refineries-chemicals-pollution-2019-11.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Stepnick, Micheal. (2015). Industry & Infrastructure: Cancer Alley, LA and Detroit, MI. Retrived from:  
<https://due-parsons.github.io/methods3-fall2015/projects/industry-infrastructure-cancer-alley-la-and-detroit-mi/>

“These States Have the Least Amount of Pollution.” U.S. News & World Report, U.S. News & World Report, www.usnews.com/news/best-states/rankings/natural-environment/pollution.