QUANTIFY THE INFLUENCE OF UNEMPLOYMENT ON ADULT OBESITY

Dee Muralidharan

July 28, 2017

Capstone Project

```
This project is intended to study the causal relationship between unemployment and obe
sity. Two data sets will be utilized to perform this study.  The project is intended f
or law-makers , health advocates, public policy planners and anyone interested in thi
s causal relationship. The project will study the data to answer the questions below
```

- Impact of Geography and hence the weather on obesity (warm vs. cold)-
- Does areas with lower unemployment fare better in obesity prevention (i.e. economic demographics)
- Does political inclinations change skew the relationship (Democratic vs. Republican)
- Does occupation i.e. service based vs. labor alter the obesity
- Which sex is more affected by unemployment?
- What race is more affected by unemployment?

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(ggplot2)
library(ggmap)
```

```
## Warning: package 'ggmap' was built under R version 3.4.1
```

```
library(mapproj)
```

```
## Warning: package 'mapproj' was built under R version 3.4.1
```

```
## Loading required package: maps
```

```
## Warning: package 'maps' was built under R version 3.4.1
```

Step 1- Cleanup the 0besity_file

Coldest vs Warmest

Here the Climate effects on obesity will be studied to ensure a fair comparison, States will be picked that are similar in other aspects BUT weather. After looking at the data from the 48 states that are available the states that have the most similartities except for climatic difference in Nebraska and North Dakota

The factors that require to be similar are Unemployment, Racial Demographics, and political leanings

1. Unemployment rate Nebraska (Avg u-employ rate) = 3.10 and North Dakota - 3.84-Very similar
2. Demographics - Racial mixture should be similar-Nebraska -86% white , N. Dakota - 90% white
3. Political leanings- Both states have conservative views and have consistently picked republican candidates in the previous five presidential elections. This is important as it helps to illustrate that the lifestyle and dietary preference is very similar

This would be great for comparison as the significant factor that separates is their respective latitudes and thus the overall climatic condition.

```
setwd("D:/springboard Data science/Capstone Project/Final Pick/Rmarkdown Home Drive")
obesity_file <- read.csv ("obesity.csv", header=TRUE, stringsAsFactors = FALSE)
```

High Level Summary of the actions

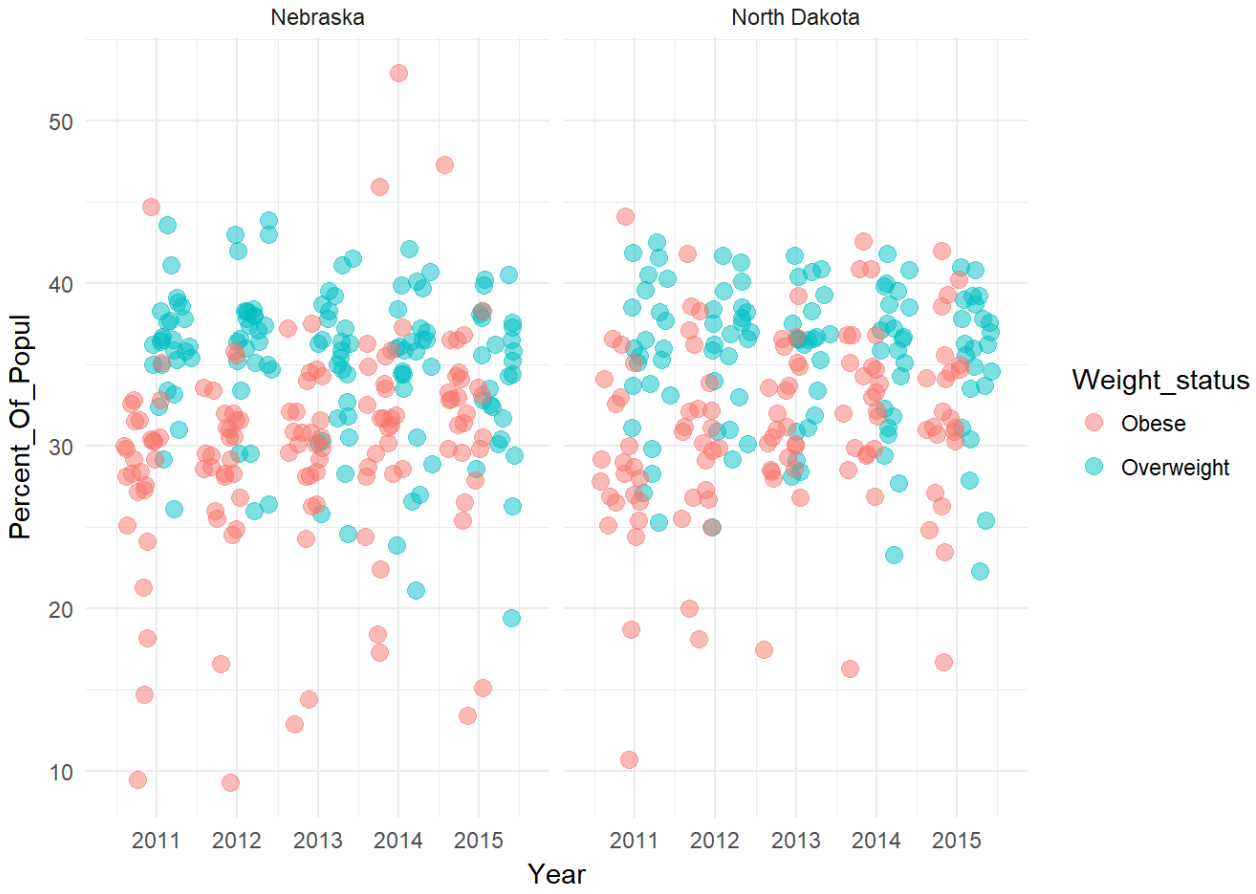
- Remove all obesity values that have a blank in them and replace them with a NA
- Filter Data based on just the obese and overweight by each state -Filter Data to reflect Nebraska and North Dakota
- Change names of Columns - Data_Value and YearEnd to Percent_Of_Popul and Year respectively.

```
setwd("D:/springboard Data science/Capstone Project/Final Pick/Rmarkdown Home Drive")

obesity_file <- obesity_file  %>% mutate ( #Change Blanks to NA in the whole file
                                          Data_Value = replace( Data_Value , is.na(Data_
Value), NA ))


cld_vs_wrm <- obesity_file%>% filter(!is.na(Data_Value)) %>%

  filter (grepl("Percent of adults aged 18 years and older who have obesity",Question)

          | grepl("Percent of adults aged 18 years and older who have an overweight cl
assification",Question))

cld_vs_wrm <- cld_vs_wrm %>% filter(grepl("NE", LocationAbbr)| grepl("ND", LocationAbb
r))
cld_vs_wrm <- cld_vs_wrm %>% mutate(Weight_status = ifelse (grepl("Percent of adults a
ged 18 years and older who have obesity", Question), "Obese", "Overweight"))
colnames(cld_vs_wrm)[names(cld_vs_wrm) == 'Data_Value'] <- "Percent_Of_Popul"
colnames(cld_vs_wrm)[names(cld_vs_wrm) == 'YearEnd'] <- "Year"
write.csv(cld_vs_wrm,"obesity1.csv")
```

Exploratory Plots

This portion we will look at trends with the Data filtered above, the purpose of the plot is exploratory and is intended to gain the big picture.

```
ggplot(cld_vs_wrm, aes(x =Year, y = Percent_Of_Popul, col = Weight_status
 )) + geom_point(size = 3, alpha = 0.5, position = position_jitterdodge (jitter.width
= 0.5,dodge.width = 0.75)) + facet_grid(.~LocationDesc)+  theme_minimal()
```



It can be Visualized that the data spread is evenly (between 25% to 45%) in both Nebraska and North Dakota which would be sufficient to conclude that Climatic conditions and thus geographical locations do not effect obesity significantly. However the outliers at both at the 45 and above and 25 and below are presenting an interesting story. It would be benefetial to check the unemployment in these Counties.