ES218 Final Project (Health Data)

Peter Brown

5/4/2020

library(dplyr)  
library(tidyr)  
library(ggplot2)  
library(tibble)  
library(USAboundaries)  
library(tmap)  
library(tibble)  
  
# Set working directory  
# Load the data (found in the repository, Health/data/...)  
health.dat <- read.csv(url("https://raw.githubusercontent.com/Peternbrown/es218\_project/master/Health/data/R11371317\_SL050.csv"))  
  
# Load county income data  
income.dat <- read.csv(url("https://raw.githubusercontent.com/Peternbrown/es218\_project/master/Health/data/County\_GDP\_percapita.csv"))  
  
# ---- CLEAN AND PREPARE DATA ----  
  
# Create state only income data  
income.state <- income.dat %>%   
 select(-Rank, -County.or.county.equivalent) %>%   
 add\_column("Geo\_STATE" = c(1, 2, 4:6, 8:10, 12, 13, 15:42, 44:51, 53:56))  
  
# Remove uneeded columns from health data  
health.dat2 <- health.dat %>%   
 select(Geo\_STATE, SE\_T001\_001, SE\_T001\_002, SE\_T004\_001, SE\_T004\_002,  
 SE\_T008\_004, SE\_T005\_001, SE\_NV003\_001, SE\_NV003\_002, SE\_NV006\_004) %>%   
 rename("Physically Unhealthy Days per Month" = SE\_T001\_001,  
 "Mentally Unhealthy Days per Month" = SE\_T001\_002,  
 "Primary Care Physicians (PCP)" = SE\_T004\_001,  
 "Mental Health Providers (MHP)" = SE\_T004\_002,  
 "Health Care Costs Price-adjusted Medicare Reimbursements" = SE\_T005\_001,  
 "Drug Poisoning Mortality Count" = SE\_T008\_004,  
 "Drug Poisoning Mortality Rate" = SE\_NV006\_004,  
 "PCP per 100,000 people" = SE\_NV003\_001,  
 "MHP per 100,000 people" = SE\_NV003\_002) %>%   
 group\_by(Geo\_STATE) %>%   
 summarise\_all(mean, na.rm = TRUE) %>%   
 mutate\_all(funs(round(., digits = 2)))  
  
# Join the income and health tables  
all.data <- inner\_join(health.dat2, income.state, by = "Geo\_STATE") %>%   
 rename("State" = State..federal.district.or.territory)  
  
# Create New-England only data  
ne.dat <- all.data %>%   
 filter(State == "Maine" | State == "New Hampshire" | State == "Vermont" |  
 State == "Massachusetts" | State == "Connecticut" |State == "Rhode Island") %>%   
 rename("state\_name" = State)  
  
# Create New England Counties  
attach(health.dat)  
ne.counties <- health.dat[order(Geo\_FIPS), ]  
detach(health.dat)  
  
all.dat.rm <- all.data %>%   
 select(Geo\_STATE, State)  
  
ne.counties2 <- ne.counties %>%  
 filter(Geo\_FIPS == "9001" | Geo\_FIPS == "9003" | Geo\_FIPS == "9005" | Geo\_FIPS == "9007" | Geo\_FIPS == "9009" |  
 Geo\_FIPS == "9011" | Geo\_FIPS == "9013" | Geo\_FIPS == "9015" | Geo\_FIPS == "23001" | Geo\_FIPS == "23003" |  
 Geo\_FIPS == "23005" | Geo\_FIPS == "23007" | Geo\_FIPS == "23009" | Geo\_FIPS == "23011" | Geo\_FIPS == "23013" |   
 Geo\_FIPS == "23015" | Geo\_FIPS == "23017" | Geo\_FIPS == "23019" | Geo\_FIPS == "23021" | Geo\_FIPS == "23023" |   
 Geo\_FIPS == "23025" | Geo\_FIPS == "23027" | Geo\_FIPS == "23029" | Geo\_FIPS == "23031" | Geo\_FIPS == "25001" |   
 Geo\_FIPS == "25003" | Geo\_FIPS == "25005" | Geo\_FIPS == "25007" | Geo\_FIPS == "25009" | Geo\_FIPS == "25011" |  
 Geo\_FIPS == "25013" | Geo\_FIPS == "25015" | Geo\_FIPS == "25017" | Geo\_FIPS == "25019" | Geo\_FIPS == "25021" |  
 Geo\_FIPS == "25023" | Geo\_FIPS == "25025" | Geo\_FIPS == "25027" | Geo\_FIPS == "33001" | Geo\_FIPS == "33003" |   
 Geo\_FIPS == "33005" | Geo\_FIPS == "33007" | Geo\_FIPS == "33009" | Geo\_FIPS == "33011" | Geo\_FIPS == "33013" |  
 Geo\_FIPS == "33015" | Geo\_FIPS == "33017" | Geo\_FIPS == "33019" | Geo\_FIPS == "44001" | Geo\_FIPS == "44003" |   
 Geo\_FIPS == "44005" | Geo\_FIPS == "44007" | Geo\_FIPS == "44009" | Geo\_FIPS == "50001" | Geo\_FIPS == "50003" |   
 Geo\_FIPS == "50005" | Geo\_FIPS == "50007" | Geo\_FIPS == "50009" | Geo\_FIPS == "50011" | Geo\_FIPS == "50013" |   
 Geo\_FIPS == "50015" | Geo\_FIPS == "50017" | Geo\_FIPS == "50019" | Geo\_FIPS == "50021" | Geo\_FIPS == "50023" |   
 Geo\_FIPS == "50025" | Geo\_FIPS == "50027") %>%   
 select(Geo\_STATE, Geo\_NAME, Geo\_QNAME, SE\_T001\_001, SE\_T001\_002, SE\_T004\_001, SE\_T004\_002,  
 SE\_T008\_004, SE\_T005\_001, SE\_NV003\_001, SE\_NV003\_002, SE\_NV006\_004) %>%   
 rename("Physically Unhealthy Days per Month" = SE\_T001\_001,  
 "Mentally Unhealthy Days per Month" = SE\_T001\_002,  
 "Primary Care Physicians (PCP)" = SE\_T004\_001,  
 "Mental Health Providers (MHP)" = SE\_T004\_002,  
 "Health Care Costs Price-adjusted Medicare Reimbursements" = SE\_T005\_001,  
 "Drug Poisoning Mortality Count" = SE\_T008\_004,  
 "Drug Poisoning Mortality Rate" = SE\_NV006\_004,  
 "PCP per 100,000 people" = SE\_NV003\_001,  
 "MHP per 100,000 people" = SE\_NV003\_002) %>%   
 add\_column(State = c("Connecticut", "Connecticut", "Connecticut", "Connecticut", "Connecticut", "Connecticut", "Connecticut",  
 "Connecticut", "Maine", "Maine", "Maine", "Maine", "Maine", "Maine", "Maine", "Maine", "Maine", "Maine",   
 "Maine", "Maine", "Maine", "Maine", "Maine", "Maine", "Massachusetts", "Massachusetts", "Massachusetts",   
 "Massachusetts", "Massachusetts", "Massachusetts", "Massachusetts", "Massachusetts", "Massachusetts",   
 "Massachusetts", "Massachusetts", "Massachusetts", "Massachusetts", "Massachusetts", "New Hampshire",   
 "New Hampshire", "New Hampshire", "New Hampshire", "New Hampshire", "New Hampshire", "New Hampshire",   
 "New Hampshire", "New Hampshire", "New Hampshire", "Rhode Island", "Rhode Island", "Rhode Island",  
 "Rhode Island", "Rhode Island", "Vermont", "Vermont", "Vermont", "Vermont", "Vermont", "Vermont", "Vermont",   
 "Vermont", "Vermont", "Vermont", "Vermont", "Vermont", "Vermont", "Vermont"))

## Introduction

As this project is being created during COVID-19 lockdown, the importance of mental health is a factor for many people around the globe. This inspired me to use this project to look into the relationship between mental health, mental healthcare access, and drug poisoning mortality in New England. Overall, what I find is that there is little to now relation between mental health and drug induced mortality.

## Methods

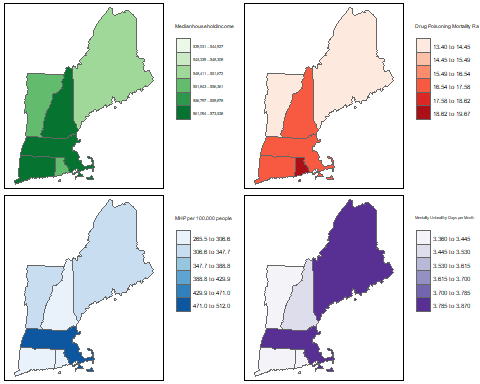
The data was sourced from my class, as well as one .csv file containing state wide income levels that I grabbed from Wikipedia. I analyzed the data using several R packages including tmap and ggplot2.

## Results and Discussion

ne.map <- us\_states(states = c("Massachusetts", "Vermont", "Maine",  
 "New Hampshire", "Rhode Island",  
 "Connecticut"))  
  
ne.shp <- full\_join(ne.map, ne.dat, by = "state\_name")  
  
tmap\_options(max.categories = 6)  
  
income.map <- tm\_shape(ne.shp, projection = 26919) +   
 tm\_polygons("Medianhouseholdincome", style = "quantile", n = 6, palette = "Greens") +  
 tm\_legend(outside = TRUE)  
  
# Map drug mortality  
drug.map <- tm\_shape(ne.shp, projection = 26919) +   
 tm\_polygons("Drug Poisoning Mortality Rate", style = "quantile", n = 6, palette = "Reds") +  
 tm\_legend(outside = TRUE)  
  
# Map mental health providers  
MHP.map <- tm\_shape(ne.shp, projection = 26919) +   
 tm\_polygons("MHP per 100,000 people",  
 style = "quantile", n = 6, palette = "Blues") +  
 tm\_legend(outside = TRUE)  
  
  
mentally\_unhealthy.map <- tm\_shape(ne.shp, projection = 26919) +   
 tm\_polygons("Mentally Unhealthy Days per Month",  
 style = "quantile", n = 6, palette = "Purples") +  
 tm\_legend(outside = TRUE)

### Geographical Overview of Study Area

# side-by-side maps (change figure size when knitting)  
tmap\_arrange(income.map, drug.map, MHP.map, mentally\_unhealthy.map,   
 nrow = 2, ncol = 2)



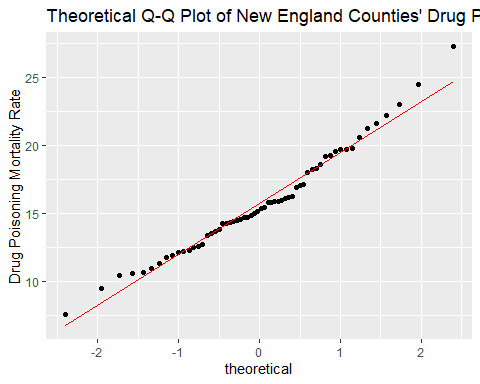
The above maps represent a graphical overview of New England states’ statistics of factors that may be intertwined. In the following analysis I take a look at the relation between median household income, drug poisoning mortalities, number of mental health providers and the number of mentally unhealthy days by a person per month.

At first glance we can see some trends. First, despite Massachussets having relatively high income and the most mental health care providers (MHP) per 100,000 people, the state has high number of mentally unhealthy days among citizens and high drug mortality rate. Maine on the other hand has the lowest median household income, a low amount of MHP per 100,000 people and low drug poisoning mortality rate. However, they have among the highest mentally unhelathy days per month. This prelimianry look at our data prompts a few questions that we can further explore.

The relation between access to mental health care and drug mortality rates is of particular interest.

### Drug Mortality Rate

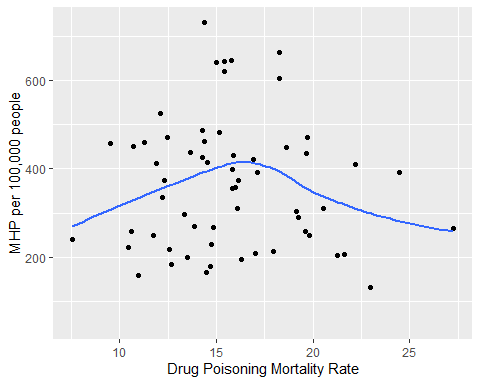
# ---- qq-plot of all new england counties' drug mortality rate ----  
  
ggplot() + aes(sample = ne.counties2$`Drug Poisoning Mortality Rate`) + geom\_qq(distribution = qnorm) +   
 geom\_qq\_line(line.p = c(0.25, 0.75), col = "red") + ylab("Drug Poisoning Mortality Rate") +  
 ggtitle("Theoretical Q-Q Plot of New England Counties' Drug Poisoning Mortality Rates")



The theoretical q-q plot above shows that drug posioning mortality rates is fairly normally distributed across New England counties. There is slight tailing at both ends of the graph and the middle points don’t follow the abline well, giving the points a slight “S” shape. This suggests that the data may be under-dispersed, although barely.

I am curious to see how drug poisoning mortality rate looks when plotted against access to mental health care. Or in this instance, the number of mental health care providers per 100,000 people.

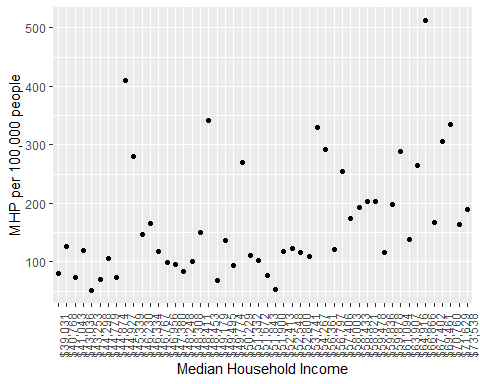
ggplot(ne.counties2, aes(x = `Drug Poisoning Mortality Rate`, y = `MHP per 100,000 people`)) +  
 geom\_point() + stat\_smooth(method = "loess", se = FALSE, span = 0.9)



There appears to be no relation between access to mental health providers and drug poisoning mortality rates between New England counties. However, using a loess curve with a span of 0.9, we can see there is a parabolic relation between the two variables. However, given the wide span of the loess curve and the resulting shape of the curve, this is a rather weak relationship.

At this point, lets take a step back and see if there is any relationshuip between income and access to mental health care providers nationwide.

ggplot(all.data, aes(x = `Medianhouseholdincome`, y = `MHP per 100,000 people`)) + geom\_point() + xlab("Median Household Income") +  
 theme(axis.text.x = element\_text(angle = 90))



There appearst to be slight positive correlation between the two variables.

I had trouble plotting the residuals, but this is where they would go.

## Conclusion

After several analyses of New England mental health variables, it appears there is no correlation between mental health and drug poisoning mortality rate. However, analysis does show that there might be a correlation between income in a state and access to mental health care providers per 100,000 people.

REFERENCE:

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.