BRFSS 2013-2017 Survey EDA - Health indicators

Amanda Kimball

11/22/2020

The Smart: City and county survey data put together by the Center for Disease Control and Prevention: Behavioral risk factor surveillance system (Smart data, 2011-2017). Contains a number of survey response data that is equally distributed across a range of areas referred to as Metropolitan and Micropolitan statistical areas. The survey data covers health indicators, diet choices, health insurance availability, demographics, and economic factors. This file will focus on the diabetes heath indicator and will focused on individuals age 18-64.

library(foreign)  
BRFSS2017 <- read.xport("MMSA2017.XPT")  
BRFSS2016 <- read.xport("MMSA2016.XPT")  
BRFSS2015 <- read.xport("MMSA2015.XPT")  
BRFSS2014 <- read.xport("MMSA2014.XPT")  
BRFSS2013 <- read.xport("MMSA2013.XPT")

I will bind all 5 data sets together and then separate out the variables of interest. These variables where inline with the risk factors listed on the cdc website: The most important factors listed by the CDC as conditions that increase a risk for COVID19 are: cancer, chronic kidney disease, chronic obstructive pulmonary disease, heart disease, immunocompromised organ transplant, obesity, sickle cell disease, smoking, and type 2 diabetes mellitus (CDC, 2020b).

CVDCRHD4 - Ever Diagnosed with Angina or Coronary Heart Disease? Yes - 1, No - 2, not sure - 7, Refused - 9, blanks CHCOCNCR - Ever diagnosed with cancer (non-skin)? Yes - 1, No - 2, Not sure - 7, Refused - 9, blanks CHCKIDNY - Ever diagnosed with kidney disease? Yes - 1, No - 2, Not sure - 7, Refused - 9, blanks DIABETE3 - Ever diagnosed with diabetes? Yes - 1, Yes pregnant - 2, No - 3, No prediabetes - 4, not sure - 7, Refused - 9, blanks X\_RFSMOK3 - Adults who are current smokers? No - 1, Yes - 2, not sure/Refused - 9, blanks

I added the following due to it’s potential to impact covid prevalence: i.e. large household size, lack of healthcare,

GENHLTH - Would you say that in general your health is: Excellent - 1, Very Good - 2, Good - 3, Fair - 4, Poor - 5, not sure -7, Refused - 9, blanks HLTHPLN1 - Do you have any kind of health care coverage? Yes - 1, No - 2, Not sure - 7, Refused - 9, blanks

These were only available in the 2017 dataset and needed to be removed for an analysis over 5 years.

BPHIGH4 - Ever been told you have high blood pressure? Yes - 1, Yes pregnant - 2, No - 3, No borderline - 4, not sure - 7, Refused - 9, blanks TOLDHI2 - Ever been told you have high cholestrol? Yes - 1, No - 2, not sure - 7, Refused - 9, blanks

vars = c("GENHLTH", "HLTHPLN1",   
 #"BPHIGH4","TOLDHI2",   
 "CVDCRHD4", "CHCOCNCR", "CHCKIDNY",   
 "DIABETE3", "X\_AGE80", "X\_RFSMOK3", "X\_MMSA", "X\_MMSAWT", "MMSANAME")  
  
data17 <- subset(BRFSS2017, select = vars)  
data16 <- subset(BRFSS2016, select = vars)  
data15 <- subset(BRFSS2015, select = vars)  
data14 <- subset(BRFSS2014, select = vars)  
data13 <- subset(BRFSS2013, select = vars)  
  
data <- rbind(data13, data14, data15, data16, data17)  
  
data$MMSANAME <- lapply(data$MMSANAME, as.character)  
  
data$MMSANAME <- gsub(", Metropolitan Statistical Area", "", data$MMSANAME)  
data$MMSANAME <- gsub(", Micropolitan Statistical Area", "", data$MMSANAME)  
data$MMSANAME <- gsub(", Metropolitan Division", "", data$MMSANAME)  
names(data) <- c("General\_health", "Health\_coverage",   
 #"Blood\_pressure\_high", "Cholesterol\_high",  
 "Coronary\_heart\_disease", "cancer", "Kidney\_disease", "Diabetes", "Age", "Smoke",  
 "MMSA\_Number", "WT\_Factor", "MMSA\_NAME")  
str(data)

## 'data.frame': 1208648 obs. of 11 variables:  
## $ General\_health : num 1 4 3 4 3 2 2 1 4 4 ...  
## $ Health\_coverage : num 2 1 1 1 1 1 1 1 1 1 ...  
## $ Coronary\_heart\_disease: num 2 2 2 2 2 2 2 2 2 2 ...  
## $ cancer : num 2 2 2 2 2 2 2 2 2 2 ...  
## $ Kidney\_disease : num 2 2 2 1 2 2 2 2 2 2 ...  
## $ Diabetes : num 2 3 3 1 3 3 3 3 1 3 ...  
## $ Age : num 40 75 63 52 70 38 53 49 74 66 ...  
## $ Smoke : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ MMSA\_Number : num 10380 10380 10380 10380 10380 ...  
## $ WT\_Factor : num 294.3 89.2 251.1 448.6 60.2 ...  
## $ MMSA\_NAME : chr "Aguadilla-Isabela, PR" "Aguadilla-Isabela, PR" "Aguadilla-Isabela, PR" "Aguadilla-Isabela, PR" ...

Here are the calculated body mass indicator variables. As shown there are several missing values.

summary(data)

## General\_health Health\_coverage Coronary\_heart\_disease cancer   
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.00   
## 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:2.00   
## Median :2.000 Median :1.000 Median :2.000 Median :2.00   
## Mean :2.534 Mean :1.106 Mean :1.985 Mean :1.92   
## 3rd Qu.:3.000 3rd Qu.:1.000 3rd Qu.:2.000 3rd Qu.:2.00   
## Max. :9.000 Max. :9.000 Max. :9.000 Max. :9.00   
## NA's :10 NA's :1 NA's :3   
## Kidney\_disease Diabetes Age Smoke MMSA\_Number   
## Min. :1.000 Min. :1.000 Min. :18.00 Min. :1.00 Min. :10100   
## 1st Qu.:2.000 1st Qu.:3.000 1st Qu.:42.00 1st Qu.:1.00 1st Qu.:19740   
## Median :2.000 Median :3.000 Median :57.00 Median :1.00 Median :34980   
## Mean :1.984 Mean :2.767 Mean :54.45 Mean :1.47 Mean :31606   
## 3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:68.00 3rd Qu.:1.00 3rd Qu.:40380   
## Max. :9.000 Max. :9.000 Max. :80.00 Max. :9.00 Max. :49660   
## NA's :4 NA's :7   
## WT\_Factor MMSA\_NAME   
## Min. : 0.15 Length:1208648   
## 1st Qu.: 111.65 Class :character   
## Median : 274.50 Mode :character   
## Mean : 648.26   
## 3rd Qu.: 690.78   
## Max. :43227.53   
##

The 9 value’s are essential N/A’s. Everything else is presents absence characteristics. Even the high cholesterol indication is proceeded by a question asking if the participant had cholesterol in the last 5 years. THe N/As here are mostly participants indicating that they have not had a cholesterol check.

All N/As will be coded as 9.

index <- is.na(data[1])  
data[index, 1] <- 9  
index <- is.na(data[4])  
data[index, 4] <- 9  
index <- is.na(data[7])  
data[index, 7] <- 9  
index <- is.na(data[8])  
data[index, 8] <- 9

I am going to create dummy variables for each of the categories individually based off their conditions from above. I keep an unknown category and both the yes and no incase I need to debug in my subsequent analysis.

For general health, I will have good, bad and unknown categories. I keep an unknown value so that I can remove those individuals from the regional estimate for percentage of the region with a condition.

General Health - Would you say that in general your health is: Excellent - 1, Very Good - 2, Good - 3, Fair - 4, Poor - 5, not sure -7, Refused - 9, blanks

data$Gen\_Hlth\_good <- ifelse(data$General\_health == 1, 1,   
 ifelse(data$General\_health == 2, 1,   
 ifelse(data$General\_health == 3, 1, 0)))  
data$Gen\_Hlth\_bad <- ifelse(data$General\_health == 4, 1,   
 ifelse(data$General\_health == 5, 1, 0))  
data$Gen\_Hlth\_unkn <- ifelse(data$General\_health == 7, 1,   
 ifelse(data$General\_health == 9, 1, 0))  
  
summary(data[12:14])

## Gen\_Hlth\_good Gen\_Hlth\_bad Gen\_Hlth\_unkn   
## Min. :0.0000 Min. :0.0000 Min. :0.000000   
## 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.:0.000000   
## Median :1.0000 Median :0.0000 Median :0.000000   
## Mean :0.8226 Mean :0.1742 Mean :0.003166   
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:0.000000   
## Max. :1.0000 Max. :1.0000 Max. :1.000000

Health\_coverage - Do you have any kind of health care coverage? Yes - 1, No - 2, Not sure - 7, Refused - 9, blanks

data$Insured <- ifelse(data$Health\_coverage == 1, 1, 0)   
data$Uninsured <- ifelse(data$Health\_coverage == 2, 1, 0)   
data$Unkn\_insurance <- ifelse(data$Health\_coverage == 7, 1,   
 ifelse(data$Health\_coverage == 9, 1, 0))  
  
summary(data[15:17])

## Insured Uninsured Unkn\_insurance   
## Min. :0.0000 Min. :0.00000 Min. :0.000000   
## 1st Qu.:1.0000 1st Qu.:0.00000 1st Qu.:0.000000   
## Median :1.0000 Median :0.00000 Median :0.000000   
## Mean :0.9168 Mean :0.07926 Mean :0.003938   
## 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:0.000000   
## Max. :1.0000 Max. :1.00000 Max. :1.000000

Blood\_pressure\_high - Ever been told you have high blood pressure? Yes - 1, Yes pregnant - 2, No - 3, No borderline - 4, not sure - 7, Refused - 9, blanks

#data$BP\_High <- ifelse(data$Blood\_pressure\_high == 1, 1,   
# ifelse(data$Blood\_pressure\_high == 2, 1,   
# ifelse(data$Blood\_pressure\_high == 4, 1, 0)))  
#data$BP\_normal <- ifelse(data$Blood\_pressure\_high == 3, 1, 0)  
#data$BP\_unkn <- ifelse(data$Blood\_pressure\_high == 7, 1,   
# ifelse(data$Blood\_pressure\_high == 9, 1, 0))  
  
#summary(data[20:22])

Cholesterol\_high - Ever been told you have high cholestrol? Yes - 1, No - 2, not sure - 7, Refused - 9, blanks

#data$Chol\_High <- ifelse(data$Cholesterol\_high == 1, 1, 0)   
#data$Chol\_normal <- ifelse(data$Cholesterol\_high == 2, 1, 0)   
#data$Unkn\_chol <- ifelse(data$Cholesterol\_high == 7, 1,   
# ifelse(data$Cholesterol\_high == 9, 1, 0))  
  
#summary(data[23:25])

Coronary\_heart\_disease - Ever Diagnosed with Angina or Coronary Heart Disease? Yes - 1, No - 2, not sure - 7, Refused - 9, blanks

data$HD\_yes <- ifelse(data$Coronary\_heart\_disease == 1, 1, 0)   
data$HD\_no <- ifelse(data$Coronary\_heart\_disease == 2, 1, 0)   
data$Unkn\_HD <- ifelse(data$Coronary\_heart\_disease == 7, 1,   
 ifelse(data$Coronary\_heart\_disease == 9, 1, 0))  
  
summary(data[18:20])

## HD\_yes HD\_no Unkn\_HD   
## Min. :0.00000 Min. :0.0000 Min. :0.000000   
## 1st Qu.:0.00000 1st Qu.:1.0000 1st Qu.:0.000000   
## Median :0.00000 Median :1.0000 Median :0.000000   
## Mean :0.05492 Mean :0.9374 Mean :0.007699   
## 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:0.000000   
## Max. :1.00000 Max. :1.0000 Max. :1.000000   
## NA's :1 NA's :1 NA's :1

cancer - Ever diagnosed with cancer (non-skin)? Yes - 1, No - 2, Not sure - 7, Refused - 9, blanks

data$cancer\_yes <- ifelse(data$cancer == 1, 1, 0)   
data$cancer\_no <- ifelse(data$cancer == 2, 1, 0)   
data$Unkn\_cancer <- ifelse(data$cancer == 7, 1,   
 ifelse(data$cancer == 9, 1, 0))  
  
summary(data[21:23])

## cancer\_yes cancer\_no Unkn\_cancer   
## Min. :0.00000 Min. :0.0000 Min. :0.000000   
## 1st Qu.:0.00000 1st Qu.:1.0000 1st Qu.:0.000000   
## Median :0.00000 Median :1.0000 Median :0.000000   
## Mean :0.09378 Mean :0.9037 Mean :0.002483   
## 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:0.000000   
## Max. :1.00000 Max. :1.0000 Max. :1.000000

Kidney\_disease - Ever diagnosed with kidney disease? Yes - 1, No - 2, Not sure - 7, Refused - 9, blanks

data$KD\_yes <- ifelse(data$Kidney\_disease == 1, 1, 0)   
data$KD\_no <- ifelse(data$Kidney\_disease == 2, 1, 0)   
data$Unkn\_KD <- ifelse(data$Kidney\_disease == 7, 1,   
 ifelse(data$Kidney\_disease == 9, 1, 0))  
  
summary(data[24:26])

## KD\_yes KD\_no Unkn\_KD   
## Min. :0.00000 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.00000 1st Qu.:1.0000 1st Qu.:0.00000   
## Median :0.00000 Median :1.0000 Median :0.00000   
## Mean :0.03427 Mean :0.9623 Mean :0.00341   
## 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1.00000 Max. :1.0000 Max. :1.00000   
## NA's :4 NA's :4 NA's :4

Diabetes - Ever diagnosed with diabetes? Yes - 1, Yes pregnant - 2, No - 3, No prediabetes - 4, not sure - 7, Refused - 9, blanks

data$Diabetes\_yes <- ifelse(data$Diabetes == 1, 1,   
 ifelse(data$Diabetes == 2, 1,   
 ifelse(data$Diabetes == 4, 1, 0)))   
data$Diabetes\_no <- ifelse(data$Diabetes == 3, 1, 0)   
data$Unkn\_Diabetes <- ifelse(data$Diabetes == 7, 1,   
 ifelse(data$Diabetes == 9, 1, 0))  
  
summary(data[27:29])

## Diabetes\_yes Diabetes\_no Unkn\_Diabetes   
## Min. :0.0000 Min. :0.0000 Min. :0.000000   
## 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:0.000000   
## Median :0.0000 Median :1.0000 Median :0.000000   
## Mean :0.1508 Mean :0.8474 Mean :0.001833   
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0.000000   
## Max. :1.0000 Max. :1.0000 Max. :1.000000   
## NA's :7 NA's :7 NA's :7

Smoke - Adults who are current smokers? No - 1, Yes - 2, not sure/Refused - 9, blanks

data$Smoke\_yes <- ifelse(data$Smoke == 1, 1, 0)   
data$Smoke\_no <- ifelse(data$Smoke == 2, 1, 0)   
data$Unkn\_Smoke <- ifelse(data$Smoke == 9, 1, 0)  
  
summary(data[30:32])

## Smoke\_yes Smoke\_no Unkn\_Smoke   
## Min. :0.000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :1.000 Median :0.0000 Median :0.0000   
## Mean :0.825 Mean :0.1329 Mean :0.0421   
## 3rd Qu.:1.000 3rd Qu.:0.0000 3rd Qu.:0.0000   
## Max. :1.000 Max. :1.0000 Max. :1.0000

Age - 18-80 with all values over 80 collapsed to 80.

data$age\_30minus <- ifelse(data$Age < 30, 1, 0)   
data$age\_30\_40 <- ifelse(data$Age < 30, 0, ifelse(data$Age >= 40, 0, 1))   
data$age\_40\_50 <- ifelse(data$Age < 40, 0, ifelse(data$Age >= 50, 0, 1))   
data$age\_50\_60 <- ifelse(data$Age < 50, 0, ifelse(data$Age >= 60, 0, 1))   
data$age\_60\_70 <- ifelse(data$Age < 60, 0, ifelse(data$Age >= 70, 0, 1))   
data$age\_70plus <- ifelse(data$Age < 70, 0, 1)   
summary(data[33:38])

## age\_30minus age\_30\_40 age\_40\_50 age\_50\_60   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
## Mean :0.1084 Mean :0.1169 Mean :0.1392 Mean :0.1957   
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## age\_60\_70 age\_70plus   
## Min. :0.0000 Min. :0.000   
## 1st Qu.:0.0000 1st Qu.:0.000   
## Median :0.0000 Median :0.000   
## Mean :0.2189 Mean :0.221   
## 3rd Qu.:0.0000 3rd Qu.:0.000   
## Max. :1.0000 Max. :1.000

I will aggregate the data as a proportion of the population for each value by county (MMSA Region).

#Each observation dummy value times it's weight factor.  
MMSA\_Health <- data[12:38]\*data$WT\_Factor  
#Keep MMSA by each observation  
MMSA\_Health$MMSA\_Name <- data$MMSA\_Number  
#Create total wt\_factor for each MMSA  
MMSAWT <- aggregate(data$WT\_Factor~data$MMSA\_Number, FUN = sum)  
names(MMSAWT) <- c("MMSA\_NAME", "WT\_Factor")  
#Aggregate the health indicators by MMSA  
MMSAHealth <- aggregate(.~MMSA\_Name, data = MMSA\_Health, FUN = sum)  
#Divide out each of the total wt factors.  
MMSAHealth[-1] <- MMSAHealth[-1]/MMSAWT$WT\_Factor  
summary(MMSAHealth)

## MMSA\_Name Gen\_Hlth\_good Gen\_Hlth\_bad Gen\_Hlth\_unkn   
## Min. :10100 Min. :0.6354 Min. :0.09126 Min. :0.000000   
## 1st Qu.:19450 1st Qu.:0.8056 1st Qu.:0.14215 1st Qu.:0.001396   
## Median :30970 Median :0.8295 Median :0.16634 Median :0.002207   
## Mean :30098 Mean :0.8250 Mean :0.17184 Mean :0.003129   
## 3rd Qu.:40200 3rd Qu.:0.8548 3rd Qu.:0.19117 3rd Qu.:0.003957   
## Max. :49660 Max. :0.9085 Max. :0.36271 Max. :0.016987   
## Insured Uninsured Unkn\_insurance HD\_yes   
## Min. :0.6865 Min. :0.03879 Min. :0.000000 Min. :0.01799   
## 1st Qu.:0.8428 1st Qu.:0.08560 1st Qu.:0.004244 1st Qu.:0.03457   
## Median :0.8738 Median :0.11923 Median :0.005831 Median :0.04078   
## Mean :0.8716 Mean :0.12246 Mean :0.005902 Mean :0.04285   
## 3rd Qu.:0.9094 3rd Qu.:0.15043 3rd Qu.:0.007006 3rd Qu.:0.04898   
## Max. :0.9612 Max. :0.30602 Max. :0.032258 Max. :0.10274   
## HD\_no Unkn\_HD cancer\_yes cancer\_no   
## Min. :0.8891 Min. :0.001627 Min. :0.03092 Min. :0.9007   
## 1st Qu.:0.9443 1st Qu.:0.005188 1st Qu.:0.06064 1st Qu.:0.9207   
## Median :0.9520 Median :0.006480 Median :0.06578 Median :0.9317   
## Mean :0.9501 Mean :0.007041 Mean :0.06804 Mean :0.9297   
## 3rd Qu.:0.9598 3rd Qu.:0.008192 3rd Qu.:0.07676 3rd Qu.:0.9370   
## Max. :0.9790 Max. :0.021922 Max. :0.09696 Max. :0.9672   
## Unkn\_cancer KD\_yes KD\_no Unkn\_KD   
## Min. :0.000000 Min. :0.01446 Min. :0.9466 Min. :0.000000   
## 1st Qu.:0.001324 1st Qu.:0.02398 1st Qu.:0.9658 1st Qu.:0.001831   
## Median :0.002079 Median :0.02712 Median :0.9700 Median :0.002629   
## Mean :0.002253 Mean :0.02799 Mean :0.9692 Mean :0.002789   
## 3rd Qu.:0.002942 3rd Qu.:0.03114 3rd Qu.:0.9739 3rd Qu.:0.003511   
## Max. :0.011902 Max. :0.04673 Max. :0.9834 Max. :0.014370   
## Diabetes\_yes Diabetes\_no Unkn\_Diabetes Smoke\_yes   
## Min. :0.07175 Min. :0.7988 Min. :0.0000000 Min. :0.7007   
## 1st Qu.:0.11186 1st Qu.:0.8580 1st Qu.:0.0009942 1st Qu.:0.7606   
## Median :0.12706 Median :0.8714 Median :0.0017772 Median :0.7856   
## Mean :0.12840 Mean :0.8697 Mean :0.0019134 Mean :0.7870   
## 3rd Qu.:0.14022 3rd Qu.:0.8855 3rd Qu.:0.0024190 3rd Qu.:0.8103   
## Max. :0.19898 Max. :0.9256 Max. :0.0103879 Max. :0.9169   
## Smoke\_no Unkn\_Smoke age\_30minus age\_30\_40   
## Min. :0.05118 Min. :0.006358 Min. :0.1377 Min. :0.1086   
## 1st Qu.:0.14126 1st Qu.:0.032717 1st Qu.:0.1920 1st Qu.:0.1560   
## Median :0.17345 Median :0.040532 Median :0.2067 Median :0.1703   
## Mean :0.17009 Mean :0.042928 Mean :0.2129 Mean :0.1705   
## 3rd Qu.:0.19575 3rd Qu.:0.050203 3rd Qu.:0.2243 3rd Qu.:0.1848   
## Max. :0.28748 Max. :0.117774 Max. :0.4350 Max. :0.2291   
## age\_40\_50 age\_50\_60 age\_60\_70 age\_70plus   
## Min. :0.09238 Min. :0.1170 Min. :0.0833 Min. :0.07676   
## 1st Qu.:0.14962 1st Qu.:0.1709 1st Qu.:0.1385 1st Qu.:0.11536   
## Median :0.16157 Median :0.1778 Median :0.1493 Median :0.12766   
## Mean :0.15858 Mean :0.1776 Mean :0.1494 Mean :0.13099   
## 3rd Qu.:0.17041 3rd Qu.:0.1864 3rd Qu.:0.1601 3rd Qu.:0.14276   
## Max. :0.19899 Max. :0.2161 Max. :0.2156 Max. :0.26090

Assuming that I did that all correctly, then the sum of the individual category columns for each MMSA should be 1 (i.e. 100%). This was a good review, because I found that I missed some values (3 and 4) in the Diabetes categories.

MMSA <- "10740"  
  
paste("General Health has", sum(MMSAHealth[2:4][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")

## [1] "General Health has 100 % of data."

paste("Health coverage has", sum(MMSAHealth[5:7][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")

## [1] "Health coverage has 100 % of data."

#paste("Blood Pressure has", sum(MMSAHealth[8:10][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")  
#paste("Cholestrol has", sum(MMSAHealth[11:13][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")  
paste("Heart Disease has", sum(MMSAHealth[8:10][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")

## [1] "Heart Disease has 100 % of data."

paste("Cancer has", sum(MMSAHealth[11:13][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")

## [1] "Cancer has 100 % of data."

paste("Kidney Disease has", sum(MMSAHealth[14:16][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")

## [1] "Kidney Disease has 100 % of data."

paste("Diabetes has", sum(MMSAHealth[17:19][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")

## [1] "Diabetes has 100 % of data."

paste("Smoke has", sum(MMSAHealth[20:22][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")

## [1] "Smoke has 100 % of data."

paste("Age has", sum(MMSAHealth[23:28][MMSAHealth$MMSA\_Name == MMSA,])\*100, "% of data.")

## [1] "Age has 100 % of data."

I will reload the body mass index data and merge the 2 datasets by MMSA.

BMIdata <- read.csv('healthdata4.csv')  
MMSAHealth <- merge(x = BMIdata, y = MMSAHealth, by.x ="MMSA", by.y = "MMSA\_Name")  
summary(MMSAHealth)

## MMSA BMI\_.kg.m2. underweight normalweight   
## Min. :10100 Min. :25.94 Min. :0.007477 Min. :0.2342   
## 1st Qu.:19450 1st Qu.:27.49 1st Qu.:0.014803 1st Qu.:0.2892   
## Median :30970 Median :27.93 Median :0.016730 Median :0.3065   
## Mean :30098 Mean :27.88 Mean :0.017439 Mean :0.3095   
## 3rd Qu.:40200 3rd Qu.:28.24 3rd Qu.:0.019679 3rd Qu.:0.3285   
## Max. :49660 Max. :29.61 Max. :0.036068 Max. :0.4393   
## overweight obese extremeobesity Gen\_Hlth\_good   
## Min. :0.3384 Min. :0.1313 Min. :0.01793 Min. :0.6354   
## 1st Qu.:0.3844 1st Qu.:0.2132 1st Qu.:0.03537 1st Qu.:0.8056   
## Median :0.3964 Median :0.2342 Median :0.04298 Median :0.8295   
## Mean :0.3963 Mean :0.2336 Mean :0.04324 Mean :0.8250   
## 3rd Qu.:0.4094 3rd Qu.:0.2509 3rd Qu.:0.05107 3rd Qu.:0.8548   
## Max. :0.4556 Max. :0.3593 Max. :0.07540 Max. :0.9085   
## Gen\_Hlth\_bad Gen\_Hlth\_unkn Insured Uninsured   
## Min. :0.09126 Min. :0.000000 Min. :0.6865 Min. :0.03879   
## 1st Qu.:0.14215 1st Qu.:0.001396 1st Qu.:0.8428 1st Qu.:0.08560   
## Median :0.16634 Median :0.002207 Median :0.8738 Median :0.11923   
## Mean :0.17184 Mean :0.003129 Mean :0.8716 Mean :0.12246   
## 3rd Qu.:0.19117 3rd Qu.:0.003957 3rd Qu.:0.9094 3rd Qu.:0.15043   
## Max. :0.36271 Max. :0.016987 Max. :0.9612 Max. :0.30602   
## Unkn\_insurance HD\_yes HD\_no Unkn\_HD   
## Min. :0.000000 Min. :0.01799 Min. :0.8891 Min. :0.001627   
## 1st Qu.:0.004244 1st Qu.:0.03457 1st Qu.:0.9443 1st Qu.:0.005188   
## Median :0.005831 Median :0.04078 Median :0.9520 Median :0.006480   
## Mean :0.005902 Mean :0.04285 Mean :0.9501 Mean :0.007041   
## 3rd Qu.:0.007006 3rd Qu.:0.04898 3rd Qu.:0.9598 3rd Qu.:0.008192   
## Max. :0.032258 Max. :0.10274 Max. :0.9790 Max. :0.021922   
## cancer\_yes cancer\_no Unkn\_cancer KD\_yes   
## Min. :0.03092 Min. :0.9007 Min. :0.000000 Min. :0.01446   
## 1st Qu.:0.06064 1st Qu.:0.9207 1st Qu.:0.001324 1st Qu.:0.02398   
## Median :0.06578 Median :0.9317 Median :0.002079 Median :0.02712   
## Mean :0.06804 Mean :0.9297 Mean :0.002253 Mean :0.02799   
## 3rd Qu.:0.07676 3rd Qu.:0.9370 3rd Qu.:0.002942 3rd Qu.:0.03114   
## Max. :0.09696 Max. :0.9672 Max. :0.011902 Max. :0.04673   
## KD\_no Unkn\_KD Diabetes\_yes Diabetes\_no   
## Min. :0.9466 Min. :0.000000 Min. :0.07175 Min. :0.7988   
## 1st Qu.:0.9658 1st Qu.:0.001831 1st Qu.:0.11186 1st Qu.:0.8580   
## Median :0.9700 Median :0.002629 Median :0.12706 Median :0.8714   
## Mean :0.9692 Mean :0.002789 Mean :0.12840 Mean :0.8697   
## 3rd Qu.:0.9739 3rd Qu.:0.003511 3rd Qu.:0.14022 3rd Qu.:0.8855   
## Max. :0.9834 Max. :0.014370 Max. :0.19898 Max. :0.9256   
## Unkn\_Diabetes Smoke\_yes Smoke\_no Unkn\_Smoke   
## Min. :0.0000000 Min. :0.7007 Min. :0.05118 Min. :0.006358   
## 1st Qu.:0.0009942 1st Qu.:0.7606 1st Qu.:0.14126 1st Qu.:0.032717   
## Median :0.0017772 Median :0.7856 Median :0.17345 Median :0.040532   
## Mean :0.0019134 Mean :0.7870 Mean :0.17009 Mean :0.042928   
## 3rd Qu.:0.0024190 3rd Qu.:0.8103 3rd Qu.:0.19575 3rd Qu.:0.050203   
## Max. :0.0103879 Max. :0.9169 Max. :0.28748 Max. :0.117774   
## age\_30minus age\_30\_40 age\_40\_50 age\_50\_60   
## Min. :0.1377 Min. :0.1086 Min. :0.09238 Min. :0.1170   
## 1st Qu.:0.1920 1st Qu.:0.1560 1st Qu.:0.14962 1st Qu.:0.1709   
## Median :0.2067 Median :0.1703 Median :0.16157 Median :0.1778   
## Mean :0.2129 Mean :0.1705 Mean :0.15858 Mean :0.1776   
## 3rd Qu.:0.2243 3rd Qu.:0.1848 3rd Qu.:0.17041 3rd Qu.:0.1864   
## Max. :0.4350 Max. :0.2291 Max. :0.19899 Max. :0.2161   
## age\_60\_70 age\_70plus   
## Min. :0.0833 Min. :0.07676   
## 1st Qu.:0.1385 1st Qu.:0.11536   
## Median :0.1493 Median :0.12766   
## Mean :0.1494 Mean :0.13099   
## 3rd Qu.:0.1601 3rd Qu.:0.14276   
## Max. :0.2156 Max. :0.26090

I will reload the covid data and merge that data by MMSA.

coviddata <- read.csv('covid\_MMSA.csv')  
# I created a listing of MMSA number vs MMSA name since each data set has a slightly different   
# method. The numbers are consistent across sets.Then merged it with the health data.  
MMSA <- subset(data, select = c("MMSA\_NAME","MMSA\_Number"))  
MMSA\_Health <- merge(x = MMSA[!duplicated(MMSA$MMSA\_Number),], y = MMSAHealth,   
 by.x ="MMSA\_Number", by.y = "MMSA")  
MMSA\_Health <- merge(x = coviddata, y = MMSA\_Health, by ="MMSA\_NAME")  
summary(MMSA\_Health)

## MMSA\_NAME CASES DEATHS   
## Aberdeen, SD : 1 Min. : 391 Min. : 2.00   
## Akron, OH : 1 1st Qu.: 5188 1st Qu.: 77.25   
## Albany-Schenectady-Troy, NY : 1 Median : 10346 Median : 207.00   
## Albuquerque, NM : 1 Mean : 22595 Mean : 522.81   
## Allentown-Bethlehem-Easton, PA-NJ: 1 3rd Qu.: 23600 3rd Qu.: 500.75   
## Anchorage, AK : 1 Max. :324050 Max. :7844.00   
## (Other) :124   
## death\_rate MMSA\_Number BMI\_.kg.m2. underweight   
## Min. :0.2046 Min. :10100 Min. :26.20 Min. :0.007477   
## 1st Qu.:1.3271 1st Qu.:19680 1st Qu.:27.64 1st Qu.:0.014334   
## Median :1.8956 Median :30970 Median :28.02 Median :0.016860   
## Mean :2.2524 Mean :30338 Mean :27.98 Mean :0.017376   
## 3rd Qu.:2.7432 3rd Qu.:40310 3rd Qu.:28.34 3rd Qu.:0.019652   
## Max. :9.5563 Max. :49660 Max. :29.61 Max. :0.036068   
##   
## normalweight overweight obese extremeobesity   
## Min. :0.2342 Min. :0.3384 Min. :0.1515 Min. :0.01793   
## 1st Qu.:0.2822 1st Qu.:0.3822 1st Qu.:0.2224 1st Qu.:0.03740   
## Median :0.3022 Median :0.3954 Median :0.2407 Median :0.04507   
## Mean :0.3047 Mean :0.3941 Mean :0.2389 Mean :0.04493   
## 3rd Qu.:0.3252 3rd Qu.:0.4078 3rd Qu.:0.2563 3rd Qu.:0.05231   
## Max. :0.4153 Max. :0.4556 Max. :0.3593 Max. :0.07540   
##   
## Gen\_Hlth\_good Gen\_Hlth\_bad Gen\_Hlth\_unkn Insured   
## Min. :0.6760 Min. :0.09126 Min. :0.000000 Min. :0.6865   
## 1st Qu.:0.8029 1st Qu.:0.14551 1st Qu.:0.001396 1st Qu.:0.8423   
## Median :0.8268 Median :0.17026 Median :0.002290 Median :0.8713   
## Mean :0.8254 Mean :0.17141 Mean :0.003155 Mean :0.8681   
## 3rd Qu.:0.8511 3rd Qu.:0.19247 3rd Qu.:0.003603 3rd Qu.:0.9016   
## Max. :0.9085 Max. :0.31400 Max. :0.016987 Max. :0.9558   
##   
## Uninsured Unkn\_insurance HD\_yes HD\_no   
## Min. :0.04226 Min. :0.000000 Min. :0.01799 Min. :0.8891   
## 1st Qu.:0.08977 1st Qu.:0.004265 1st Qu.:0.03457 1st Qu.:0.9437   
## Median :0.12240 Median :0.005927 Median :0.04139 Median :0.9513   
## Mean :0.12575 Mean :0.006124 Mean :0.04253 Mean :0.9500   
## 3rd Qu.:0.15095 3rd Qu.:0.007115 3rd Qu.:0.04886 3rd Qu.:0.9597   
## Max. :0.30602 Max. :0.032258 Max. :0.09526 Max. :0.9790   
##   
## Unkn\_HD cancer\_yes cancer\_no Unkn\_cancer   
## Min. :0.002738 Min. :0.04069 Min. :0.9007 Min. :0.000000   
## 1st Qu.:0.005446 1st Qu.:0.06133 1st Qu.:0.9207 1st Qu.:0.001255   
## Median :0.006759 Median :0.06697 Median :0.9310 Median :0.001910   
## Mean :0.007417 Mean :0.06889 Mean :0.9289 Mean :0.002188   
## 3rd Qu.:0.008593 3rd Qu.:0.07684 3rd Qu.:0.9365 3rd Qu.:0.002878   
## Max. :0.021922 Max. :0.09696 Max. :0.9568 Max. :0.011902   
##   
## KD\_yes KD\_no Unkn\_KD Diabetes\_yes   
## Min. :0.01446 Min. :0.9466 Min. :0.000000 Min. :0.07175   
## 1st Qu.:0.02410 1st Qu.:0.9653 1st Qu.:0.001713 1st Qu.:0.11227   
## Median :0.02728 Median :0.9698 Median :0.002538 Median :0.12770   
## Mean :0.02851 Mean :0.9687 Mean :0.002777 Mean :0.12863   
## 3rd Qu.:0.03178 3rd Qu.:0.9739 3rd Qu.:0.003499 3rd Qu.:0.14214   
## Max. :0.04673 Max. :0.9834 Max. :0.014370 Max. :0.19898   
##   
## Diabetes\_no Unkn\_Diabetes Smoke\_yes Smoke\_no   
## Min. :0.7988 Min. :0.000000 Min. :0.7007 Min. :0.05118   
## 1st Qu.:0.8569 1st Qu.:0.000954 1st Qu.:0.7563 1st Qu.:0.15358   
## Median :0.8705 Median :0.001657 Median :0.7815 Median :0.18131   
## Mean :0.8695 Mean :0.001825 Mean :0.7814 Mean :0.17641   
## 3rd Qu.:0.8853 3rd Qu.:0.002351 3rd Qu.:0.8020 3rd Qu.:0.19817   
## Max. :0.9256 Max. :0.009252 Max. :0.9169 Max. :0.28748   
##   
## Unkn\_Smoke age\_30minus age\_30\_40 age\_40\_50   
## Min. :0.008756 Min. :0.1377 Min. :0.1119 Min. :0.09238   
## 1st Qu.:0.032657 1st Qu.:0.1933 1st Qu.:0.1575 1st Qu.:0.14672   
## Median :0.040064 Median :0.2091 Median :0.1722 Median :0.15880   
## Mean :0.042201 Mean :0.2160 Mean :0.1714 Mean :0.15621   
## 3rd Qu.:0.048349 3rd Qu.:0.2290 3rd Qu.:0.1856 3rd Qu.:0.16778   
## Max. :0.117774 Max. :0.4350 Max. :0.2291 Max. :0.19318   
##   
## age\_50\_60 age\_60\_70 age\_70plus   
## Min. :0.1170 Min. :0.0833 Min. :0.07676   
## 1st Qu.:0.1707 1st Qu.:0.1389 1st Qu.:0.11484   
## Median :0.1776 Median :0.1503 Median :0.12736   
## Mean :0.1759 Mean :0.1496 Mean :0.13085   
## 3rd Qu.:0.1843 3rd Qu.:0.1622 3rd Qu.:0.14269   
## Max. :0.2013 Max. :0.2156 Max. :0.26090   
##

write.csv(MMSA\_Health, 'healthcoviddata2.csv', row.names = FALSE)

References:

Unknown. (2017) Smart: City and county survey data: 2017 Data. Center for Disease Control and Prevention: Behavioral risk factor surveillance system. Retrieved November 1, 2020 from: <https://www.cdc.gov/brfss/smart/Smart_data.htm>

Unknown. (2020b). Coronavirus disease 2019 (COVID-19): People with certain medical conditions. Centers for Disease Control and Prevention. Retrieved from: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fneed-extra-precautions%2Fgroups-at-higher-risk.html>