continuous.vars <- c("PCTPOVR", "PBAMORER", "URBANRUR", "AGER", "RACER", "SEX", "PAYTYPER")  
continuous.varset <- urgentCareOnly.df[continuous.vars]  
continuous.varset <- na.omit(continuous.varset)  
dim(continuous.varset)

## [1] 3618 7

behavioral.vars <- c("URBANRUR","PAYTYPER","INJURY", "PRIMCARE", "SENBEFOR", "PASTVIS", "MAJOR", "WEEKEND")  
behavioral.varset <- urgentCareOnly.df[behavioral.vars]  
bahavioral.varset <- na.omit(behavioral.vars)  
dim(behavioral.varset)

## [1] 3863 8

demographic.vars <- c("PCTPOVR", "HINCOMER", "PBAMORER", "URBANRUR", "RACEETH", "WEEKEND", "AGER", "SEX")  
demographic.varset <- urgentCareOnly.df[demographic.vars]  
demographic.varset <- na.omit(demographic.varset)  
dim(demographic.varset)

## [1] 3618 8

full.vars <- c("PCTPOVR", "HINCOMER", "PBAMORER", "URBANRUR", "RACEETH", "WEEKEND", "AGER", "SEX", "URBANRUR","PAYTYPER","INJURY", "PRIMCARE", "SENBEFOR", "PASTVIS", "MAJOR", "WEEKEND")  
full.varset <- urgentCareOnly.df[full.vars]  
full.varset <- na.omit(full.varset)  
dim(full.varset)

## [1] 3488 16

sample <- create\_sample.func(behavioral.varset, 250)  
sample <- na.omit(sample)  
dim(sample)

## [1] 233 8

cluster.func <- function (sample) {  
 #distance matrix  
 summary(sample.dist <- daisy(sample, metric = "gower"))  
 sample.dist <- sample.dist^2  
   
 # Investigate further z vs h differences  
 sample.clust <- hclust(sample.dist, method = "ward.D")  
 sample.agnes <- agnes(sample.dist, method = "ward")  
 par(mfrow=c(1,1))  
 temp <- sample.clust  
 temp$height <- sqrt(sample.clust$height)  
 plot(temp, labels = FALSE, main="X.hclust -- sqrt(height)")  
 plot(sample.agnes, which.plots=2, labels = FALSE, main="X.agnes.wardD2")  
   
  
 # Find number of clusters, need better way  
 # sample.heights <- sample.clust$height # height values  
 # sample.heights2 <- c( 0, sample.heights[ -length(sample.heights) ]) # finds height values at merge for n obs  
 # sample.max <- max(round(sample.heights - sample.heights2, 4)) # finds largest height value  
 # optimal <- which.max(round(sample.heights - sample.heights2, 4)) # Determines the location of first   
  
 sample.clust  
   
}  
  
sample.clust <- cluster.func(sample)

plot(sample.clust, labels = FALSE)

#cutree(sample.clust, k = 6:18) #k = 1 is trivial  
## Compare the 2 and 4 grouping:  
#g24 <- cutree(sample.clust, k = c(16,17))  
#table(grp2 = g24[,"16"], grp4 = g24[,"17"])  
  
k <- 10  
table(cutree(sample.clust, 10))

##   
## 1 2 3 4 5 6 7 8 9 10   
## 33 11 17 57 26 16 14 27 16 16

sample.kmode <- kmodes(sample, modes = k, iter.max = 5)  
  
Kmodes.func <- function (cluster, k){  
 table(cutree(cluster, k))  
 sample.kmode <- kmodes(sample, modes = k, iter.max = 10, weighted = FALSE)  
}  
  
#table(sample.kmode$modes)  
cluster.tbl <- as.data.frame(sample.kmode$modes)

cluster.tbl <- cluster.tbl[order(cluster.tbl$PAYTYPER, cluster.tbl$URBANRUR),]  
  
#sample.kmode$modes

#sample.dist2 <- sample.dist^2  
#sk2 <- silhouette(sample.kmode$cluster, sample.dist2)  
#plot(sk2)

#sample1 <- data.matrix(sample)  
#sample1 <- model.matrix(sample1)  
library(mclust)

## Package 'mclust' version 5.1

## Type 'citation("mclust")' for citing this R package in publications.

#View(sample)  
#d\_clust <- Mclust(as.matrix(sample), G=1:20)

kable(cluster.tbl, col.names = c("Urban/Rural", "Pay Type", "Injury", "Primary Care", "Seen Before", "Past Visits", "Major Reason", "Weekend" ),  
 caption = "Clustering Table \\label{tab:clust}", label(scale = .75))

##   
## \begin{figure}[htbp]  
## \centering  
## \includegraphics[scale = 0.75,angle = 0]{}  
## \caption[Default caption]{\normalsize{Default caption}}  
## \label{fig:def}  
## \end{figure}

Clustering Tables

Behavioral clusters: Listed are the responses that occurred *most*  in each cluster found.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Urban/Rural | Pay Type | Injury | Primary Care | Seen Before? | Past Visits | Major Reason | Weekend |
| Large central metro | Medicaid | No | Yes | Established | 3 | Preventive care | No |
| Small metro | Medicaid | No | No | Established | 1 | New Problem | No |
| Large central metro | Medicare | No | Unknown | Established | 1 | Chronic problem, routine | No |
| Medium metro | Medicare | No | No | Established | 1 | Chronic problem, routine | No |
| Rural | Medicare | Yes | No | New | NA | New Problem | No |
| Large central metro | Private | No | Yes | Established | 3 | Chronic problem, routine | No |
| Large central metro | Private | No | No | New | NA | New Problem | No |
| Large fringe metro | Private | No | Yes | Established | 6 | New Problem | No |
| Rural | Private | No | No | Established | 2 | Pre-/Post-surgery | No |
| Large central metro | Worker's compensation | Yes | No | Established | 2 | New Problem | No |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Primary Care** | **Payment** | **Urban/ Rural** | **HH Income** | **% Pov** | **% Bac Degree** | **Race** | **Age** | **Injury Related** | **Seen Before** | **Past Visits** | **Major Reason** |
| Yes | Medicare | Large central metro | Q1 ($32,793 or less) | Q4 (>20%) | Q1 (<12%) | White | >75 | No | Est. Patient | 10 | New problem |
| Yes | Private | Large fringe metro | Q3 ($40,627-$52,387) | Q2 (5-10%) | Q3 (20-32%) | White | 45-64 years | No | Est. Patient | 2 | Chronic problem, routine |
| Yes | Private | Medium metro | Q3 ($40,627-$52,387) | Q2 (5-10%) | Q2 (12-20%) | White | 45-64 years | No | Est. Patient | 3 | New Problem |
| No | Medicare | Rural | Q2 ($32,794-$40,626) | Q3 (10-20%) | Q2 (12-20%) | White | 45-64 years | No | Est. Patient | 3 | Chronic problem, routine |
| No | Private | Large fringe metro | Q4 ($52,388 or more) | Q1 (<5%) | Q4 (>32%) | White | Under 15 years | No | Est. Patient | 1 | Chronic problem, routine |
| No | Private | Medium metro | Q2 ($32,794-$40,626) | Q3 (10-20%) | Q1 (<12%) | White | 25-44 years | No | Est. Patient | 2 | Chronic problem, routine |
| No | Private | Large central metro | Q4 ($52,388 or more) | Q2 (5-10%) | Q4 (>32%) | White | 25-44 years | No | Est. Patient | 1 | New Problem |
| No | Private | Large central metro | Q1 ($32,793 or less) | Q4 (>20%) | Q1 (<12%) | White | 25-44 years | No | New Patient | NA | New Problem |
| No | Private | Large fringe metro | Q4 ($52,388 or more) | Q1 (<5%) | Q3 (20-32%) | White | 15-24 years | No | New Patient | NA | New Problem |
| No | Private | Medium metro | Q4 ($52,388 or more) | Q2 (5-10%) | Q4 (>32%) | White | 45-64 years | No | New Patient | NA | New Problem |

Logit decisions decisions:

2 options:

of those who use urgent care, what parameters predict whether or not it is their primary care physician

OR: of all recorded visits, what parameters predict the decision to *go* to urgent care?