Analysis

```
setwd("~/Desktop/Thesis/analysis")
NAMCS.work <- read.csv("NAMCS08-11.csv", header = TRUE)
```

A quick summary of the data shows that by and large most of the observations continue to choose traditional means of obcuring primary care. Great quote: For our purposes, we do not presume to go quite this far, but we are applying this method as a form of retroduction, using observed evidence to create a research hypothesis that accounts for the observed facts *Sayer, 1992*.

Identifying Typologies with Cluster Analysis

While clustering methods are incresingly being used to generate scientific hypotheses, this analysis rather aims to apply clustering as a method of retroduction on a phenomenon that is currently vastly under studied. The results of this study are divided into three areas: (1) the determination of the number of clusters, (2) the examination of how the clusters differ in terms of multiple parameters, and (3) examining the clusters in light of the theoretical hypotheses posed in the introduction.

Because I am interested in two somewhat distinct aspects of the patients of Urgent Care– both their demographics and their patterns of use – The clustering was performed in two batches of parameters which aimed to help us understand the two different side to a visit to urgent care. Variables for *Age, Sex, Race, Urban Type, % Neighborhood Poverty, % Neighborhood college degree attainment, and Pay Type*, what I will refer to as the *demographic variables* from this point forward were first analyzed for subgroups. Secondly, some of the same variables were again analyzed with the behavior parameters of *Injury related, Primary Caregiver, Seen Before?, Past Visits, Major Reason, and the day of the week*, what I will refer to as the *behavior parameters.*

Determining the number of Clusters

soon to be in the appendix

Because urgent care centers have been greatly ignored by sociologists studying medical practices, there was little theoretical guidance in selecting a likely number of subgroups for the analysis. Similarly, because I am interested in understanding how an unsupervised analysis of patient data will reveal trends, it was particuarly important to the analysis that the number of clusters were both methematically acheivable and substantivly small enough for analysis.

To accomplish this, I began with hierarchal clustering of the random samples chosen from the data. Using a mixed-methods tool to calculate the distance matrix between the various observations, I placed each point into an althgorithm which subsequently minimized the variance between clusters. banner fig shows the banner for the initial agglomerative cluster methods for the behavioral variables. .

The white lines extending to the right represent clusters which differ from each other. After running the hierarchal, bottom up method for a number of trials, the agglomerative coefficient was always between .73 and .8, indicating that as the height for which the clusters should stop combining. Again, banner fig demonstrates that at that height, 8 clusters are have clearly seperated.

Differences in Demographic and Behavioral Factors Between Clusters

With cluster assignments established, I then analyzed how the identified groups differed from each other in terms of critical demographic chracteristics and behaviors that are described in the literature. Most importantly—in order to better understand the characteristics of individuals choosing to go to urgent care we must understand which chracteristics or bahaviors most distinguished the clusters in the anlaysis.

Beginning with demographic trends, it is immediatly clear that there are large homogenous groups within the selection of patients who utilize urgent care. Of the 8 clusters, 7 are predominantly or exclusively white.

Cluster	Age	Sex	Race	Urban	Pay Type	Injury
1	15-24 yr	Male	White	Large cent metro	Worker's comp	Yes
2	65-74 yr	Male	White	Fringe metro	Medicare	No
3	45-64 yr	Male	White	Large cent metro	Private insurance	No
4	25-44 yr	Female	White	Med metro	Private insurance	No
5	65-74 yr	Female	White	Rural	Medicare	No
6	45-64 yr	Female	White	Large cent metro	Private insurance	No

Cluster	Age	Sex	Race	Urban	Pay Type	Injury
7	45-64 yr	Male	White	Med metro	Private insurance	No
8	25-44 yr	Female	White	Large cent metro	Private insurance	No

Table 2: Correlation of Inheritance Factors for Parents and Child

Clusters	Primary	Seen Before	No. of vis	Major	Day of Wk
1	No	Yes, established patient	2	New problem	Tuesday
2	No	Yes, established patient	1	Chronic, routine	Tuesday
3	No	Yes, established patient	1	New problem	Friday
4	No	No, new patient	NA	New problem	Wednesday
5	Yes	Yes, established patient	0	New problem	Thursday
6	No	Yes, established patient	8	New problem	Tuesday
7	No	Yes, established patient	3	Chronic, routine	Monday
8	Yes	Yes, established patient	3	New problem	Monday

Table 2 shows that I don't know how to make tables.

With commentary

The exploratory analysis to follow was motivated by the current lack of statistical analysis and social theory surround urgent care centers, despite their rapid progression as an acceptable replacement for primary care (erhm). The exploratory nature of the findings will allow those intent on locating urgent care centers within the larger contextual framework of the American health care system to draw on the revealed typologies of urgent care seekers in order to better understand the industry's rapid growth. [UNCECESSARY!]

Because I am particularly interested in two somewhat individual facets of the patients of urgent care—both their demographics and their patterns of use—the clustering was performed in two batches of parameters which aimed to help develop an understanding of the two components which are involved in a visit to urgent care[WHOA TORTURNED SENTENCE] . Variables for Age, Sex, Race, Urban Type, % Neighborhood

Poverty, % Neighborhood college degree attainment, and Pay Type, which I will refer to as the demographic variables from this point forward, were first analyzed for subgroups.

Subsequently, some of the same variables were again analyzed with the parameters of Injury related, Primary Caregiver, Seen Before, Past Visits, Major Reason, and the day of the week, which I will refer to as the behavior parameters. *[FYI, you should have a section of your methods discussion dedicated to explaining each of these variables.]*

The analysis of this study is divided into three areas: (1) the determination of the number of clusters, (2) the examination of how the clusters differ in terms of multiple parameters, and (3) examining the clusters in light of the theoretical hypotheses posed in the introduction.

Determining the number of Clusters

[**Most of this can go in an appendix**]

Because urgent care centers have been significantly overlooked by sociologists studying medical practices, there was little theoretical guidance in selecting a likely number of subgroups for the cluster analysis, often the first step in such examinations. Similarly, because I am interested in understanding how an unsupervised analysis of patient data will reveal trends, rather than imposing them onto the data, it was particularly important to the investigation that the number of clusters be both mathematically achievable and substantively small enough for analysis while still remaining unsupervised. The general rule of thumb for cluster analysis suggested by Mardia, Kent and Bibbby (1997) stipulates that the number of clusters k is approximately the square root of n / 2. However, in the case of our comparably small sample of urgent care visitors recorded in the NAMCS data set, such a rule would indicate at least 30 clusters—clearly not a workable number to make sense of patterns in urgent care patients and their use.

Instead, to estimate a k which would produce heterogeneous groups while still remaining theoretically enlightening, I began with non-guided hierarchal clustering of samples of 200 visits to urgent care centers randomly chosen from the data. Using the Gower methodology of finding the distance between dissimilar variable types, I calculated the distance matrix between the various observations for each sample. These distance measures were then used as the input to a hierarchal clustering algorithm which attempted to minimize the distance between observations within the same groups while maximizing the distance between them. Figure 1 shows the 'banner' for the initial agglomerative cluster methods for the behavioral variables, and can be understood as a graphical representation of the points at which a cluster breaks away from the pack of observations. The white lines extending to the right represent clusters at their separation point, where larger red areas between white lines indicate stronger outside group variance.

Figure 1. Agglomerative Cluster Banner

After running the hierarchal method for a number of trials (on-going), the agglomerative coefficient for the behavioral parameters remained between .73 and .8 (so-far), indicating that level as the most-likely optimal height for which the clusters should stop combining in order to maximize their heterogeneity and avoid overfitting the groups. For the demographic variables, the coefficient ranged between .8 and .85, with an average of 12 clusters at those heights. Again, Figure 1 demonstrates graphically that at roughly height .75, 8 clusters have clearly separated and to go further left would dramatically decrease the differences between groups. For the demographic variables, the coefficient ranged between .8 and .85, with an average of 12 clusters at those heights. Accordingly, the demographic set of parameters and the behavioral set of parameters were assigned 8 and 12 clusters respectively, which were then used to sort the samples into homogenous groups.

Differences in Demographic and Behavioral Factors Between Clusters

In order to better understand the characteristics of individuals choosing to go to urgent care, we must first understand which characteristics or behaviors most distinguished the clusters in the cluster analysis, and to do this I used the k-modes clustering algorithm to situation observations into similar groups of patients. With cluster assignments established by the hierarchal agglomerative methods above, I then analyzed how the identified groups or 8 and 12 differed from each other in terms of critical demographic characteristics and visit behaviors that were described in the literature review.

Beginning with demographic trends, the cluster analysis makes immediately clear that there are distinct homogenous groups within the selection of patients who utilize urgent care. Race in particular is immediately noticeable for its lack of variety: of the 12 clusters, 9 are predominantly or exclusively white, an unsurprising trend when one looks at the summary statistics (of the 3863 urgent care visits included in the data, 87% (3352) of those were white individuals). Table 1 shows the modes of each cluster for an illustrative sample of patients, and it is immediately clear that the majority of urgent care goers are white, but that within clusters of white patients there is further segmentation.

For an example of such segmentation, we can examine the socioeconomic indicators for the patients' ZIP codes which were included as demographic parameters. Unsurprisingly, these appeared to cluster and correlate with both race and income: for both males and females, there were consistent clusters of White, 25-44-year-old patients with private insurance, and both of these clusters were in the highest quartile of percent population with Bachelor's degrees, the second lowest percent population under the poverty line (5-10%), and both visited clinics coded as being located in a “Large central metro” (Table 1.).

In fact, if one predominant trend becomes clear from the demographic clustering, it is that urgent care centers are almost entirely an urban affair. Of the clusters found, all but one consisted of visits located in medium to large metros, indicating that while other variation between patients exists, most urgent care consumers are located in larger cities. And if its singularity wasn’t enough, the cluster of visits which did occur in a rural setting also differed from most other clusters in the other parameters as well. As can be seen in the first row of Table 1., the rural (“Micropolitan/noncore”) cluster of visits had one of the highest levels of localized poverty, one of the lowest levels of educational attainment, and was one of only two clusters where the majority of visits were paid with Medicaid.

Also informative are the clusters’ most occurring payment types, which complicate some of the theoretical notions of urgent care described in Chapter 1. While the largest clusters were consistently comprised of private insurance payers, Medicare without a doubt plays a role in getting people to urgent care centers. Medicaid had the lowest numbers of visits across all trials, averaging around 12 % of the observations. These clusters are crucial to understanding the decision process behind choosing a health care provider, while simultaneously raising the question of the relationship between the surprising number of elderly patients on Medicare seeking treatment at an urgent care centers.[THIS PARAGRAPH SHOULD OFFER A CLEARER STORY – AND CLARIFY THE DISTINCTION BETWEEN MEDICARE AND MEDICAID]

If the demographic clusters illuminate the subgroup characteristics of urgent care seekers with socioeconomic status in mind, the behavioral parameters included in the second wave of cluster analyses allow us to go further in examining how urgent care is utilized by different groups. [READ THAT SENTENCE AND THINK OF YOUR POOR MOTHER.] It should be noted that though the analyses were performed in two waves, the clusters should be examined with each other in mind, and I have included some of the key demographic variables in with the behavioral variables to that end. [OH YEAH, THIS IS GOING IN THE APPENDIX]

Figure 2. Color coded clusters.

Figure 1 above shows the typical spread of the behavioral parameters’ clusters, and is useful for keeping the proportions of clusters in mind. By far the largest and most homogenous clusters isolated consist of white, privately insured patients in medium to large urban environments. These trends match those found in the demographic clusters, and we can learn even more about this group by examining the cluster segmentation that occurred across their behavioral parameters in the second wave. Of the two clusters which consist almost entirely of white, 25-44-year-old men and women (yellow and light pink in Figure 1 above), both were almost entirely classed as "established patients", with at least one past visit. Also theoretically interesting, neither cluster had observations whose reason for visiting was "injury related" and neither had recorded visits on weekends. Cluster one in particular (Row 1, Table 2), consisted of 25-44-year-old white females with private insurance, an established history of visiting urgent care, and a reason for their current visit coded as a "chronic, routine problem". [SUGGESTION: LOW-TECH TABLE TO ILLUSTRATE YOUR CLUSTERS. COLUMNS FOR URBAN/RURAL, RACE, AND OTHER KEY VARIABLES.]

In fact, only 32 visits in the example model were coded as new patients (Figure 2, purple and orange), and examining these two clusters reveals what I referred to earlier as the traditional patient characteristics ascribed to urgent care visits. These two cluster show what many would expect, both consisted of white, privately insured, new patients, and both clusters contain the only observations which occurred on a weekend or were coded as injury related. These characteristics are in line with what many hold as the purpose of urgent care centers, but they remain a small fraction of the recorded number of visits. Completely contradicting this idea, the other seven of the nine clusters identified have high levels of established patients with chronic or routine problems, with ages ranging for 15 to over 75.

As for the Medicare question, if the clusters which consist of primarily elderly patients are examined—rows 5 and 9 in Table 2 (Figure 2, magenta and blue)—we can begin to understand a little better the relatively large percentage of what many would consider non-emergent care utilizers and their relationship to urgent care. Cluster 5 consists entirely of white patients between 65 and 74, whose visits were recorded as routine and who have a minimum of 3 past visits at the same center. Cluster nine on the other hand has a slightly older age range of above 75, and visits which were all considered new problems. Also distinguishing, cluster nine is the only cluster with a majority of rural observations. When compared, it appears that there are two typologies of elderly urgent care seekers, the first of which may be using urgent in much the same ways as the younger and larger clusters of privately insured urbanites, while the second seems to rely on urgent care much more as an actually resource for urgent problems.

What do these mean?

```
The question remains how the observed trends align with the literature that assess the ways Americans a
```

(working on making aggregated/pretty/readable versions of these) Table 1.

Demographic Parameters

Table 2. Behavior Parameters