Megan Crawford
March 1, 2022
Machine Learning

Clustering Analysis of Concussion Data in Teenage Athletes Final Report

1. Introduction

On November 21, 2021, I hit my head very hard in a flag football game. Even though I knew something was wrong because my head hurt and I physically felt weak and awful, because I was responsive, coherent, and articulate in the conversation I had with the medic on the sideline, he told me that he didn't even think I had a concussion but to check in with UHS the next morning to be sure. I woke up the next morning and could barely even get out of bed and walk over to my UHS appointment where I had 18 out of the 22 concussion symptoms and failed every single balance, memory, and strength test. The nurse practitioner told me just to rest for a couple of days and that after Thanksgiving break I should feel better and be ready to go back to school. However, the injury ended up being so bad that I was out of classes for the rest of the semester, could barely travel home to California for winter break, was unable to participate in my study abroad program in Rome this semester, was only able to look at a computer screen for more than a couple of minutes at a time less than a week before the semester started, and at the beginning of the Spring Semester was only physically able to go to class and the dining hall. The sports medicine representative at the game clearly thought that incoherence, slurred speech, memory loss, and delayed responses in conversation are the symptoms that give the strongest indication of the existence of and the severity of the concussion, which is why he originally told me he did not think that I had one. However, the fact that the injury turned out to be very severe shows that those symptoms are not always the strongest predictors of the seriousness of a concussion and how long it will last in every situation. Since every person's brain is different and reacts differently to experiencing a hard hit, it is difficult to definitively predict how long a concussion will last and what the symptoms indicate for each individual person.

I examined a dataset from Harvard University that contains healthy control data for 13-18 year old athletes who played recreational sports. 106 of the 155 patients had previously experienced a concussion. However since this is a normative dataset, the symptoms it measures are in absence of a recent injury so therefore it provides insight into the lingering symptoms of concussions rather than their symptoms immediately following an injury or while they were still recovering. The objective of my report was to apply a clustering algorithm on the concussion dataset to identify and segment subgroups of patients with similar medical histories who experienced consistent symptoms. Through hierarchical and three rounds of k-means clustering, I identified a cluster characterized by patients who have previously experienced 1-2 concussions. In addition, many of the patients in this cluster ranked symptoms related to lack of sleep and increasing sadness highly, indicating that these could be lingering symptoms and/or long-term effects of concussions.

2. Related Work

There have been attempts to predict the recovery length of a concussion patient, to predict an athlete's risk of getting a concussion, and to detect whether or not an individual has experienced a concussion in the past utilizing machine learning models. For example, a group from the School of Biomedical Informatics at the University of Texas, Austin built a machine learning model to predict the concussion recovery time of high school athletes based on pre-injury risk factors, the severity of the initial injury, and the symptoms shown after the injury. In another study, lead researchers from Cornell University and the University of San Diego built a predictive model for sports-related concussions in college athletes and military cadets. The vast majority of models that have been built use supervised learning methods to predict or classify specific outcomes such as the length of recovery or whether or not somebody will get a concussion. However, from my personal experience, I know that every injury and every brain are different

so it is nearly impossible to predict which symptoms will have the greatest impact on recovery time for each individual. Therefore, the goal of my project is to utilize clustering, an unsupervised learning method, to identify patterns in the data and to segment subgroups of patients based on similar symptoms, severity of symptoms, and demographics. I will describe the strengths and limitations of clustering and why it is well-suited for this project in section 4.

3. Data Description

The dataset contains 155 samples of 13-18 year old athletes from Canada who participated in recreational sports at the time the data was collected. One drawback to the data is the fact that it is imbalanced because 106 out of the 155 patients have previously experienced a concussion. This is a control dataset, meaning that the data was collected when the athletes were healthy and not in response to a recent injury, although there is a variable in the dataset that indicates whether an athlete has experienced a concussion in the past or not. Using a control dataset allows us to compare the symptoms and demographics of teenagers who have and have not experienced concussions so we can examine what characteristics most strongly distinguish the patients who have experienced concussions. If the cluster contains mixed samples of patients who have and have not experienced concussions, the common symptoms of patients in that group are likely not as strong of an indicator as to whether a patient has a concussion or not. However, if a cluster is distinctly separate and only contains samples of teenagers who previously had concussions, it is likely that the common symptoms in that cluster are strong indicators of concussions. For the dataset, each row is an athlete. The explanatory variables are demographic statistics, the statistics about the medical history of the patient, the responses to the Post-Concussion Symptom scale questions (rank your symptoms such as headache and nausea on a scale of 0 to 6), and the responses to the Mood and Feelings Questionnaire (indicate whether the following statements such as ‘I felt I was no good anymore’ or ‘I felt miserable or unhappy’ were not true at all (0), sometimes true (1) or true (2) within the last 2 weeks). The inputs are available from the medical records of the hospital where the data was collected. As for pre-processing, the main issue is that the null values were encoded in the data as -1, which is a common practice in computer science. In order to address this, I replaced all of the -1 values with NA values and then used imputation to replace the null values with the mean of the values in the corresponding column. I have included images of the summary statistics of the dataset and of some summary graphs in the appendix.

4. Methods

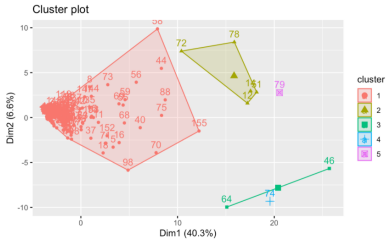
In this project, I used a clustering algorithm, which allowed me to find patterns in the samples and segment similar subgroups of patients based on common symptoms and medical histories. The goal of clustering is to group patients with similar symptoms and to detect trends and anomalies within and between those groups. Because every injury and its impact on the brain is so different, it is important to take a holistic approach when choosing a treatment strategy. Rather than relying on a single numerical value from a prediction decision or a label from a classification to make treatment decisions, it is important to look at a range of symptoms and how they interact with one another. Identifying the characteristics that define the subgroups can reveal a pattern rather than the doctor making a choice that is heavily based on a single number or factor. Also, this dataset does not contain any potential response variables, making it very difficult to use a supervised learning method to analyze.

I initially used hierarchical clustering to perform exploratory analysis, visualize the samples in the data, and to detect outliers. One weakness of hierarchical clustering is that the results become difficult to interpret when there are a high number of samples in the dataset. I ran into this issue in my project because my dataset contained 155 patient samples, which made the dendrograms very cluttered and the trends difficult to define. Also, because the data did not have a natural hierarchy, there was no clear height to cut the tree that would provide the optimal number of clusters to analyze. However, visualizing the dataset with the cluster plot was helpful in detecting outliers and getting a better perspective of where the samples fell in relation to one another. Then, I used k-means clustering to

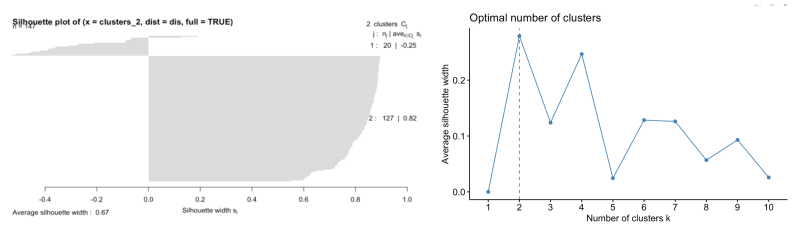
group the data points and discover trends and subgroups within the data to discover which symptoms were correlated with patients with prior concussions. One weakness of clustering is that there is no way to test the accuracy of the results using the test and training sets. As a result, I examined the quality of the k-means clusters by plotting the cluster cardinality (the number of samples in each cluster) against the cluster magnitude (within cluster sum of squares) and looking for the clusters that exhibited a positive correlation between cardinality and magnitude. There should be a linear relationship between how many samples there are and how much variation there is within each cluster, and it is better if a cluster falls close to the positive trend line. A cluster can be identified as anomalous when its cardinality does not correlate with magnitude relative to the other clusters.

5. Results

I began by using hierarchical clustering to visualize the structure of my data and to detect outliers. First, I scaled the data, calculated the distance between the points using the euclidean measure, and then ran hierarchical clustering. Because there are 155 samples, the resulting dendrogram was very cluttered and it was difficult to find the optimal height to split the clusters at, so therefore I arbitrarily chose 5 as the number of clusters. The cluster plot revealed two key findings: that samples 74 and 79 were clear outliers and could be removed from the dataset and that the majority of the samples were very concentrated in a single part of cluster 1, indicating that this is the area that I would need to dive deeper into through k-means clustering to find the samples that are the most closely related. Below is an image of my hierarchical clustering plot:



Next, I ran my initial k-means clustering using 5 clusters because that was the number of clusters I had used previously in hierarchical clustering and I thought it would be a good place to start from. 5 ended up not being an optimal number because the resulting clusters were very imbalanced with cluster 1 containing 113 samples while clusters 2 and 5 only contained one sample each. I made a heat map to visualize how the center values for each cluster compare to each other. As a result of clusters 2 and 5 only having one sample each, their values were so distinct from the rest that they dominated the heatmap and it was difficult to see the values for the other clusters, which provided justification for me to remove them for the next round of sampling. For the next round of k-means clustering, I broke the samples up into 2 clusters because that was the number specified as the optimal number by the silhouette plot. However, once I broke them up this turned out to not be the best choice because the samples were really imbalanced with cluster 1 having 127 samples and cluster 2 containing 20. As a result, in the heat map the clusters for sample 1 were so distinct that they dominated the heat map and made it difficult to see the distinction between the different explanatory variables. After running another silhouette plot, pictured below on the left, it revealed that the one cluster with 127 samples would be ideal to do further analysis on:

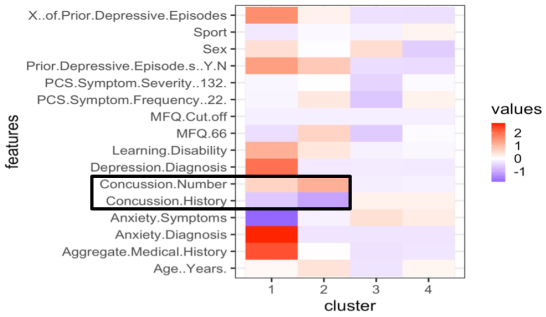


The image on the right shows the plot determining the optimal number of clusters to break this group of 127 samples into. I decided to use 4 clusters for this next round of k-means clustering because even though it is not the highest

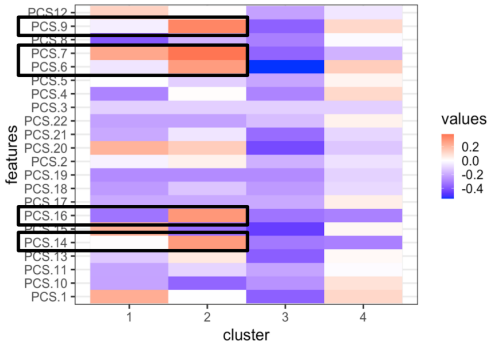
point it is another significant inflection point on the plot. This ended up being the optimal number of clusters to perform analysis on, and I will expand upon this part of the procedure and my findings in the discussion section. One challenge of clustering that I encountered in this project was determining the optimal number of clusters to use in both hierarchical and k-means clustering. Even when I used the silhouette plots, the results of those did not always guide me toward the optimal number of clusters to use. It took trial and error to remove outliers and to break the samples up into balanced clusters, but I ultimately was successful in breaking down the data.

6. Discussion

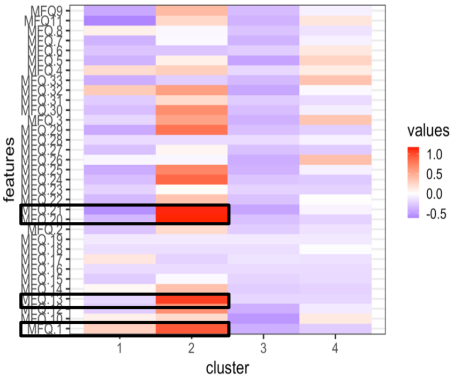
I decided to use k-means again to break the cluster of 127 samples that the silhouette plot of the previous k-means clusters revealed as optimal to perform further analysis on into 4 sub-clusters. I chose to use 4 sub-clusters based on the results of the silhouette plot summary chart. Although it was not perfectly balanced, this time the number of samples in each cluster was more even with 8 samples in cluster 1, 12 samples in cluster 2, 67 samples in cluster 3, and 40 samples in cluster 4. The heat maps of the data that I created revealed that cluster 2 was characterized by a high number of previous concussions and as samples with a concussion history, which highlighted that this was the cluster that would reveal the common symptoms among samples who had previously experienced concussions. The number of concussions was rated on a scale of 0 to 3 with 3 representing a patient who had 3 or more concussions. The fact that this box is shaded in as a lighter red shows that cluster 2 is characterized by patients who have either had 1 or 2 concussions. For concussion history, a binary response of 0 indicates that a patient has a concussion history, so the fact that the box is light blue shows that cluster 2 is characterized by patients who have a concussion history. Below is an image of the heatmap that shows the explanatory variables of the demographic and medical history portions of the questionnaire:



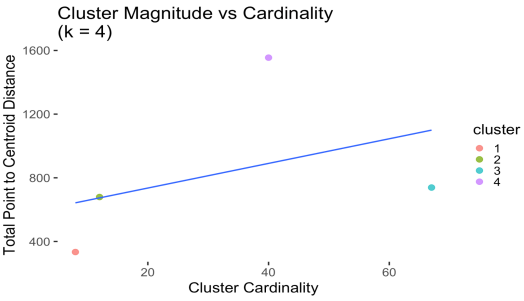
For the responses to the Post Concussion Symptom Scale (PCS) questions, participants ranked their symptoms on a scale of 0 to 6 with 6 being the highest. Cluster 2 is characterized by patients who ranked questions 6 (fatigue), 7 (trouble falling asleep), and 9 (loss of sleep), very highly as symptoms they were still experiencing. Cluster 2 is also characterized by patients who ranked questions 14 (more emotional) and 16 (sadness) highly on the symptom scale. The fact that cluster 2 is also characterized by patients who ranked symptoms related to lack of sleep and experiencing sadness highly shows that there may be a correlation between these symptoms and concussion history. Since this is a control dataset, the fact that patients who have been deemed healthy before taking this survey are still ranking these symptoms highly shows that these may be lingering symptoms of concussions or long-term effects rather than immediate symptoms. Below is an image of the heatmap showing the responses to the PCS questions:



For the mood and feelings questionnaire, participants indicated whether the statements about their emotional state were not true at all (0), sometimes true (1), or always true (2) within the last 2 weeks. Cluster 2 can possibly be characterized by patients who experienced extended sadness because they ranked Question 1 ("I felt miserable or unhappy") and Question 20 "I didn't want to see my friends") highly. Participants also ranked Question 13 ("I was talking more slowly than usual") and Question 21 ("I found it hard to think or concentrate") highly. While interesting, the high responses to these 2 questions do not seem to fit into a larger trend, at least within this cluster. Below is an image of the heatmap visualizing the responses to the MFQ questions:



As I talked about in section 4, a downside to the clustering method is that there is no definitive way to determine the accuracy of my predictions, so I examined the quality of my clusters by plotting the cluster cardinality against the cluster magnitude. Demonstrated in the image below, cluster 2 fits on the trend line perfectly while the other clusters are further away from the line, which shows that within the cluster 2 there is a linear relationship between how many samples there are and how much variation there is in the cluster, making the samples in cluster 2 the most closely related and therefore a verifiable source to examine trends within.



7. Conclusion and Future Work

The objective of my report was to utilize a clustering algorithm to distinguish subgroups of patients based on similar symptoms, severity of symptoms, and medical histories. To summarize, cluster 2 of the third round of k-means

clustering is characterized by patients who have a concussion history and have experienced 1-2 concussions. Further, many of the patients in this cluster ranked symptoms related to lack of sleep and increasing sadness highly, indicating that these could be lingering symptoms and long-term effects of concussions.

One downfall to this dataset is that all of the patients are teenagers, which makes it difficult to generalize the findings to people outside of this age bracket. However, when I was looking for a dataset to use it was extremely difficult to even find a reliable dataset on concussions. As a result, a potential next step could be to invest money in collecting concussion data from patients of all ages that would be made publicly available so any person looking to conduct research to help concussion patients could use it. I could then apply the machine learning clustering algorithm that I have already built to this new dataset in order to find trends in concussion symptoms for patients of all ages.

8. Contributions

I worked alone so therefore I did all of the work

9. Bibliography

Castellanos, J., Phoo, C.P., Eckner, J.T. *et al.* Predicting Risk of Sport-Related Concussion in Collegiate Athletes and Military Cadets: A Machine Learning Approach Using Baseline Data from the CARE Consortium Study. *Sports Med* 51, 567–579 (2021). https://doi-org.proxy.library.nd.edu/10.1007/s40279-020-01390-w

Chu, Yan, et al. “Machine Learning to Predict Sports-Related Concussion Recovery Using Clinical Data.” *Annals of Physical and Rehabilitation Medicine*, U.S. National Library of Medicine, 2 Jan. 2022, https://pubmed.ncbi.nlm.nih.gov/34986402/#affiliation-1.

10. Appendix

Summary statistics of initial dataset:

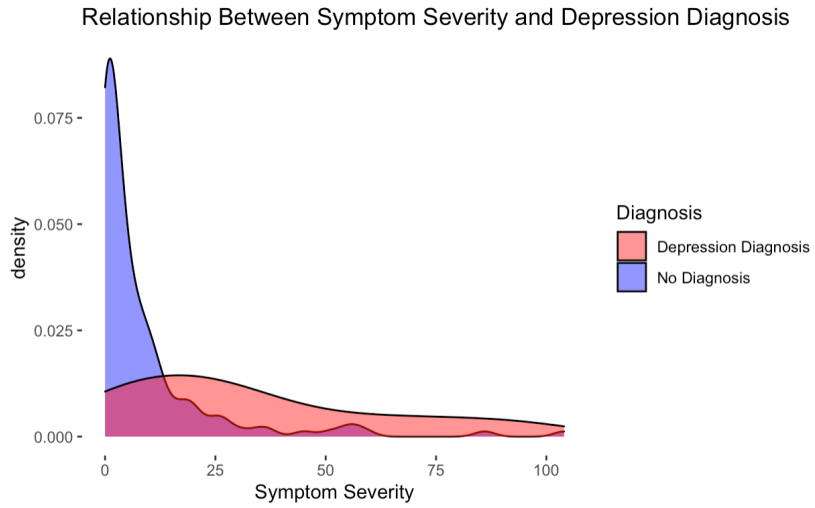
```
 Participant.ID    Age..Years.          Sex              Sport
 Min.   :  1.00   Min.   :-1.00   Min.   :0.0000   Min.   :1.00
 1st Qu.: 46.50   1st Qu.:13.00   1st Qu.:0.0000   1st Qu.:2.00
 Median : 86.00   Median :14.00   Median :0.0000   Median :2.00
 Mean   : 85.68   Mean   :14.33   Mean   :0.4903   Mean   :2.89
 3rd Qu.:126.50   3rd Qu.:16.00   3rd Qu.:1.0000   3rd Qu.:5.00
 Max.   :165.00   Max.   :18.00   Max.   :1.0000   Max.   :6.00
 Concussion.History Concussion.Number Learning.Disability Anxiety.Diagnosis
 Min.   :0.0000     Min.   :0.0000    Min.   :-1.00000    Min.   :-1.00000
 1st Qu.:0.0000     1st Qu.:0.0000    1st Qu.: 0.00000    1st Qu.: 0.00000
 Median :1.0000     Median :0.0000    Median : 0.00000    Median : 0.00000
 Mean   :0.6839     Mean   :0.4258    Mean   : 0.08387    Mean   : 0.09677
 3rd Qu.:1.0000     3rd Qu.:1.0000    3rd Qu.: 0.00000    3rd Qu.: 0.00000
 Max.   :1.0000     Max.   :3.0000    Max.   : 1.00000    Max.   : 1.00000
 Anxiety.Symptoms  Depression.Diagnosis X..of.Prior.Depressive.Episodes
 Min.   :-1.0000   Min.   :-1.00000     Min.   :-1.000
 1st Qu.: 1.0000   1st Qu.: 0.00000     1st Qu.: 0.000
 Median : 1.0000   Median : 0.00000     Median : 0.000
 Mean   : 0.7613   Mean   : 0.05806     Mean   : 0.471
 3rd Qu.: 1.0000   3rd Qu.: 0.00000     3rd Qu.: 0.000
 Max.   : 1.0000   Max.   : 1.00000     Max.   : 4.000
 Prior.Depressive.Episode.s..Y.N Aggregate.Medical.History
 Min.   :-1.0000                 Min.   :-1.0000
 1st Qu.: 0.0000                 1st Qu.: 0.0000
 Median : 0.0000                 Median : 0.0000
 Mean   : 0.1677                 Mean   : 0.1613
 3rd Qu.: 0.0000                 3rd Qu.: 0.0000
 Max.   : 1.0000                 Max.   : 1.0000
 PCS.Symptom.Frequency..22. PCS.Symptom.Severity..132.     MFQ.66
 Min.   : 0.000             Min.   :  0.00             Min.   : 0.000
 1st Qu.: 0.000             1st Qu.:  0.00             1st Qu.: 1.000
 Median : 2.000             Median :  3.00             Median : 3.000
 Mean   : 4.219             Mean   : 10.14             Mean   : 7.787
 3rd Qu.: 6.000             3rd Qu.: 11.00             3rd Qu.: 9.000
 Max.   :22.000             Max.   :104.00             Max.   :53.000
```

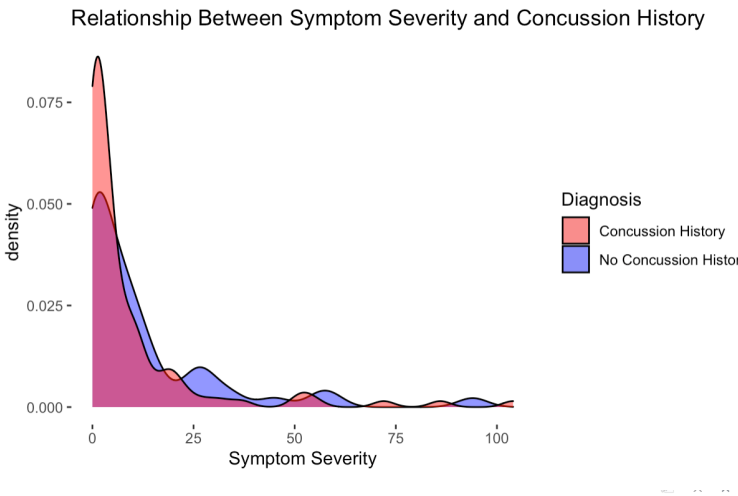
```
 MFQ.Cut.off        PCS.1            PCS.2            PCS.3
 Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
 1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
 Median :0.00000   Median :0.0000   Median :0.0000   Median :0.00000
 Mean   :0.08387   Mean   :0.6129   Mean   :0.1935   Mean   :0.08387
 3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
 Max.   :1.00000   Max.   :6.0000   Max.   :5.0000   Max.   :5.00000
      PCS.4            PCS.5            PCS.6            PCS.7
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0
 Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0
 Mean   :0.2645   Mean   :0.3419   Mean   :0.9806   Mean   :0.8
 3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:2.0000   3rd Qu.:1.0
 Max.   :5.0000   Max.   :6.0000   Max.   :6.0000   Max.   :6.0
      PCS.8            PCS.9            PCS.10           PCS.11
 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   :0.5806   Mean   :0.5548   Mean   :0.4258   Mean   :0.2581
 3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
 Max.   :6.0000   Max.   :6.0000   Max.   :5.0000   Max.   :6.0000
      PCS12            PCS.13           PCS.14           PCS.15
 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   :0.2258   Mean   :0.4774   Mean   :0.5419   Mean   :0.7419
 3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
 Max.   :6.0000   Max.   :6.0000   Max.   :6.0000   Max.   :5.0000
      PCS.16           PCS.17           PCS.18           PCS.19
 Min.   :0.0000   Min.   :0.0000   Min.   :0.000    Min.   :0.0000
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000    1st Qu.:0.0000
 Median :0.0000   Median :0.0000   Median :0.000    Median :0.0000
 Mean   :0.4839   Mean   :0.2581   Mean   :0.329    Mean   :0.4129
 3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000    3rd Qu.:0.0000
 Max.   :6.0000   Max.   :6.0000   Max.   :6.000    Max.   :6.0000
      PCS.20           PCS.21           PCS.22           MFQ.1
 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   :0.6903   Mean   :0.6065   Mean   :0.2774   Mean   :0.4065
 3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
 Max.   :6.0000   Max.   :6.0000   Max.   :5.0000   Max.   :7.0000
```

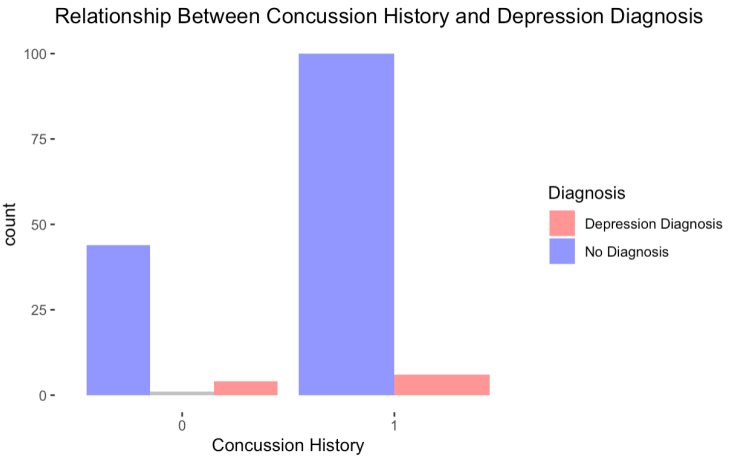
```
    MFQ.2           MFQ.3           MFQ.4           MFQ.5
Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
Median :0.0000   Median :0.000   Median :0.0000   Median :0.0000
Mean   :0.1419   Mean   :0.271   Mean   :0.2903   Mean   :0.5677
3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:1.0000
Max.   :2.0000   Max.   :2.000   Max.   :2.0000   Max.   :2.0000
    MFQ.6           MFQ.7           MFQ.8           MFQ9
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   :0.1677   Mean   :0.2387   Mean   :0.1871   Mean   :0.2903
3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
Max.   :2.0000   Max.   :2.0000   Max.   :2.0000   Max.   :2.0000
    MFQ.10          MFQ11           MFQ.12          MFQ.13
Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
Median :0.0000   Median :0.0000   Median :0.0000   Median :0.00000
Mean   :0.5032   Mean   :0.4258   Mean   :0.3097   Mean   :0.09677
3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.5000   3rd Qu.:0.00000
Max.   :2.0000   Max.   :2.0000   Max.   :2.0000   Max.   :2.00000
    MFQ.14          MFQ.15          MFQ.16          MFQ.17
Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000
Mean   :0.1677   Mean   :0.1419   Mean   :0.09032   Mean   :0.1161
3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
Max.   :2.0000   Max.   :2.0000   Max.   :2.00000   Max.   :2.0000
    MFQ.18          MFQ.19          MFQ.20          MFQ.21
Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
Median :0.00000   Median :0.00000   Median :0.0000   Median :0.0000
Mean   :0.09032   Mean   :0.06452   Mean   :0.1806   Mean   :0.3806
3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:1.0000
Max.   :2.00000   Max.   :2.00000   Max.   :2.0000   Max.   :2.0000
    MFQ.22          MFQ.23          MFQ.24          MFQ.25
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   :0.1806   Mean   :0.1548   Mean   :0.1935   Mean   :0.2581
3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
Max.   :2.0000   Max.   :2.0000   Max.   :2.0000   Max.   :2.0000
```

```
    MFQ.26          MFQ.27          MFQ.28          MFQ.29
Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
Median :0.000   Median :0.0000   Median :0.0000   Median :0.0000
Mean   :0.329   Mean   :0.2323   Mean   :0.1226   Mean   :0.3097
3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
Max.   :2.000   Max.   :2.0000   Max.   :2.0000   Max.   :2.0000
    MFQ.30          MFQ.31          MFQ.32          MFQ.33
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   :0.2065   Mean   :0.1484   Mean   :0.2839   Mean   :0.2387
3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
Max.   :2.0000   Max.   :2.0000   Max.   :2.0000   Max.   :2.0000
```

Summary graphs:


Relationship Between Symptom Severity and Depression Diagnosis
density
0.075
0.050
0.025
0.000
0
25
50
75
100
Symptom Severity
Diagnosis
Depression Diagnosis
No Diagnosis


Relationship Between Symptom Severity and Concussion History
density
0.075
0.050
0.025
0.000
0
25
50
75
100
Symptom Severity
Diagnosis
Concussion History
No Concussion History


Relationship Between Concussion History and Depression Diagnosis
count
100
75
50
25
0
0
1
Concussion History
Diagnosis
Depression Diagnosis
No Diagnosis