Injury Severity Analysis of Traffic Incidents

By: Doris Gao and Claer Jestin
Advisor: Joel Greenhouse Client: Dr. Leonard Weiss

Introduction

- We aim to assess the severity of injuries sustained in traffic incidents to improve preventative measures
- Accurate injury severity assessments ensure that patients promptly receive the appropriate level of care
- **Research Goal:** Develop an injury severity index that classifies patients' injuries from minor to fatal

Data

- Data pulled from Pittsburgh Emergency Services & 911 Dispatch Services medical records of traffic incidents over one-year period (N = 7,681)
- **Vitals Measurements Dataset**
  - 22 variables recording patients' vitals measurements over time for 4,178 observations (vitals are recorded only for non-cancelled calls)
  - Variables include quantitative (e.g. heart rate) and qualitative (e.g. respiratory effort) metrics of each patient's physical condition
- **Data Preprocessing:** Filter for each patient's earliest recorded vitals measurements
- **Feature Engineering:** Add variable to record the number of times the vitals measurements were taken per patient

Methods

About K-Means Cluster Analysis

- Partition observations into predetermined number of clusters, *k*
- Assign each observation to a cluster such that observations within a cluster are as similar to each other as possible and as different from observations in any other cluster as possible
- (Dis)similarity defined by squared Euclidean distance

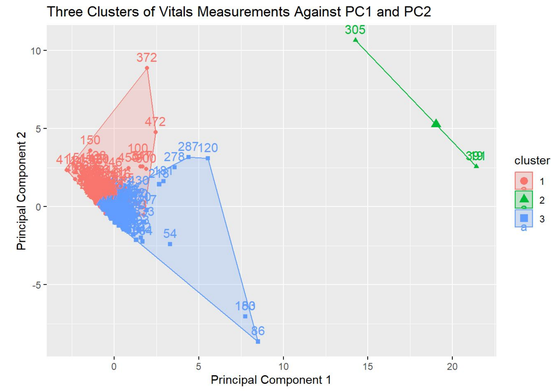
Implementing K-Means Cluster Analysis

- Variables Used: Glasgow coma scale component scores; heart rate; systolic and diastolic blood pressure; oxygen saturation; respiratory rate; and the number of times the vitals measurements were taken
- All observations with missing values are removed
- Scale variables so that those of different magnitudes are comparable
- Optimal number of clusters, *k*, chosen using scree plot of total within-cluster sum of squares

Results & Analysis

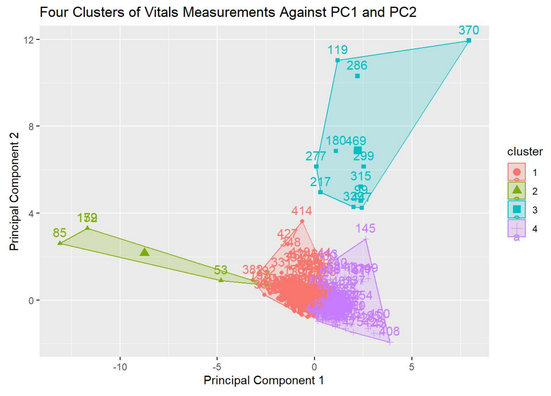
First Iteration

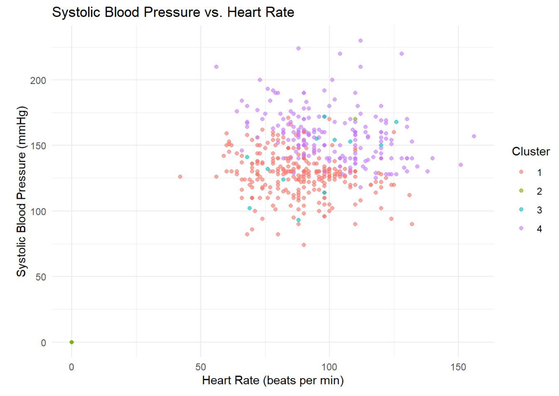
- Model with 3 clusters generates two large clusters of 224 and 252 patients, respectively, and third tiny cluster of only 3 patients
- 3 clusters separates patients in two larger clusters by low and high blood pressure and heart rate values
- Patients in tiny cluster have abnormally low Glasgow coma scale component scores, which measure consciousness
- Review of the paramedics' notes indicates that these patients suffered critical or fatal injuries



Second Iteration

- Re-ran K-means clustering without 3 extreme observations using 4 clusters (based on scree plot)
- Larger clusters remain nearly identical but model picks out two new clusters of 12 and 5 patients
- Cluster of 12 patients (Teal) distinguished by lower Glasgow coma scale component scores
- Cluster of 5 patients (Green) have extremely low oxygen saturation but higher values for other vitals measurements
- Review of paramedics' notes confirms these are **incorrect inputs**





Conclusions & Limitations

- K-means cluster analysis was not able to identify subgroups that align with injury severity
  - Clusters do not capture abnormal ranges since both too low or too high values are abnormal for blood pressure and heart rate
- **Analysis is limited by low number of patients with higher injury severity (N = 8)**
- Future research should re-run clustering with additional years of data to include more critical patients