UNIVERSITY OF NOTRE DAME | MENDOZA COLLEGE OF BUSINESS

Why do police shootings keep happening?

ITAO - 40420 Spring 2021 | Final Project

Colton R. Crum

Introduction

The chaos of 2020 brought about many challenges in the United States and abroad, but argubly no more emotionally challenging than the tragic death of George Floyd. In the wake of his death under police custody, protests incited all over Minneapolis and surrounding U.S. cities. More specifically, many were concerned with police shootings with often unarmed and African American victims. In light of these challenges, I wanted to sift through the data and figure out some of the reasons why these things happen. In doing so, I hoped to find insights for both law enforcement and members of the community in an effort to mitigate these difficult circumstances.

Related work

The challenging relationship between law enforcement and minorities have been studied throughout the last fifty years. After the Civil Rights movment in the 1950’s and 1960’s, many acadmeics have looked at the complex and often messy relationship between police and minorities. A paper from Marshall W. Meyer in 1980 showed that African Americans in Los Angeles differed significantly in number, circumstances, and outcome of the shooting review process from Hispanics and whites. An empirical study by Chicago Law Enforcement Study Group in 1981 revealed showed racially linked patterns among population and arrest statistics. However, they found that blacks and whites were about equally likely to be shot by police given exposure to forcible felony arrests. A 2015 study by Cody T. Ross at the University of California Davis used geographically-resolved, mult-level Bayesian model to analyze racial bias in police shootings in the United States. He found that “police shooting data as a function of county-level predictors suggests that racial bias in police shootings is most likely to emerge in police departments in larger metropolitan counties with low median incomes and a sizable portion of black residents, especially when there is high financial inequality in that county. There is no relationship between county-level racial bias in police shootings and crime rates (even race-specific crime rates), meaning that the racial bias observed in police shootings in this data set is not explainable as a response to local-level crime rates.” As the awareness and activism for police brutality and targeted police shootings increased, I sought to use updated data to analyze some of the trends and variables at stake. More contemporary rhetoric involving police shootings has been around defunding the police and creating more police training to handle high stakes situations. Through our analysis, we found insights that bridges some of Ross's pervious work together with manageable insights.

Data Description

The Washington Post published an open-sourced dataset involving all police shootings from January 2, 2015 to May 4, 2021. The dataset was also published on Kaggle by Ahsen Nazir. The dataset included fourteen variables including key whether the victim was armed, their age, race, location, signs of mental illness, threat level, whether or not they fled, and whether a body camera was in use. City and state variables were converted into latitude and longitude for clustering. The data consisted of 6269 rows, each row a person who was fatally killed by police. Summary statistics showed a wide range of objects that victims were armed with from bricks to chainsaws. The majority of victims were white (2385), followed by black (2354), and Hispanic (878). Fleeing statistics showed most victims were not fleeing (3356), followed by car (889), and by foot (684).

Methods

Since the data only consists of those who have been killed by police and not injured or not shoot at all, I performed a cluster analysis to discover trends and important variables within the data. One of the challenges with cluster analysis is to find the appropriate number of clusters. Before running these, first I had to simplify the dataset. Specifically, I had to remove the unique armed weapons (flagpole, chainsaw) into more broad categories. Shards of glass were converted into blades, and fireworks were classified as explosives. After cleansing, I ran a few preliminary clusters looked like somewhere between k = 4 and k = 8 clusters would be the optimal amount (Figure 1). Next, Figure 2 explored the cluster magnitude of k = 4. While cluster k = 4 looks feasible, Figure 3 showed that k = 8 was a better way to go. Figure 4 showed the cluster magnitude vs. cardinality of k = 4, which shows each cluster falling off the line. Figure 5 shows a tighter spread of clusters on the line, further confirming this is the number of clusters we want moving forward. Figure 9 confirmed the best number of clusters were 7, so we used k = 7 for the remainder of our analysis. We ran three different clusters and made minor heat map modifications to show each parts of the cluster more clearly.

Results

The clustering analysis produced three heat maps, each consisting of 8 individual clusters. These heat maps showed unsupervised groupings that were significant among the data. For the scope of the project, I will only talk about the three most significant and interesting clusters from each heat map. These will include topics and variables that most outlets frequently miss about these tragic events.

Figure 6 showed a link between threat levels and women. Cluster 2 showed that women are more threatening in armed motor vehicles. Naturally, men are more intimating and threatening for law enforcement officers which result in far more police shootings. However, women significantly raise their threat level when fleeing in vehicles, and officers are challenged in discovering whether they are armed or not. Cluster 7 was dedicated entirely to whether the police officer had body cam in use or not. Body cams were only used in 30% of all fatalities. The cluster indicated that these were a significantly different shootings from the others. As with George Floyd's death, many police shootings captured on video provide contextual and visual analysis for the events leading up to the death. Finally, cluster 6 showed a group of armed undetermined, who were feeling with an undetermined threat level. When a victim flees it is significantly harder for law enforcement to tell whether the person is armed or not. Additionally, when the victim flees the event happens much more quickly, and an office likely makes a split-second decision as to what level of force is necessary.

The second heat map (Figure 7) showed more clustering around fleeing, age, and weapons of the victim. Cluster 2 indicated that the younger a person is the more likely they are to flee. Cluster 6 shows that those who show signs of mental illness are more likely to be armed with blades or blunt objects. Cluster 1 showed that young African Americans who flee around a specific latitude and longitude. These areas are often metropolis and have many communal challenges of violence and excessive use of force.

The last heat map (Figure 8) showed a relationship between threat level, age, and type of use of force. Cluster6 showed armed, underdetermined victims who flee have an undetermined threat level. Cluster 3 shows older victims armed with medieval weapons with body cameras not being used. Cluster 5 shows victims armed with blades, no body camera, who were shot and tased with another level of threat.

Discussion

The results are compatible with some of the issues police face in the United States. Mental illness is often not cited in many shootings and refusals to comply with police orders. The data is not entirely comprehensive of the work that needs to be analyzed. For example, variables like time of day, experience of the police officer, and age of the officer. Officer data could be merged to locate more detailed clusters and see if any new patterns emerge. The work performed indicates that fleeing and armed are a common theme with police shootings, where force is either justified or unjustified. Data involving the status of the case, including the discretionary conclusion of the court or police department could separate the type of shooting that occurred. Though the dataset is grim, many cases the officers were under attack and had to use necessary force. The lack of differentiation between these two types of shootings further conflicts the definitive conclusions of the study. Unjustified uses of force are the types of shootings that we should try and mitigate, and a de-escalation of those who are attacking police. Further variables could radically change the clusters and their interpretations of this study.

Conclusion and Future Work

Police brutality is some of the darkest topics in the United States. These difficult topics show a high level of complexity and cannot be easily quantified. The cluster analysis we performed showed that victims have to reduce their threat level. This may be women in vehicles, or armed with a weapon. Regardless, the police make split second decisions that are fatal under many of these unfortunate situations. Our analysis also found mental illness prevalent in the dataset. We found that mental illness increased the likelihood that victims were armed with blunt or blades. Finally, we showed that younger males are more likely to flee from police. Fleeing results in an increased likelihood of fatal shootings from increased threat levels.

Future work could use more comprehensive data in analyzing clusters. Additionally, data should be recorded when non-fatal shootings happen in order to analyze the differences that occur and alter the course of a person’s life. Finally, we suggest looking at new variables such as time of day to see if any new trends emerge.

Contributions

I worked with the generous help of Professor Martin Barron who advised me throughout the project’s data cleansing, cluster analysis, and graphing. He is literally the best.

Figures

Figure 1

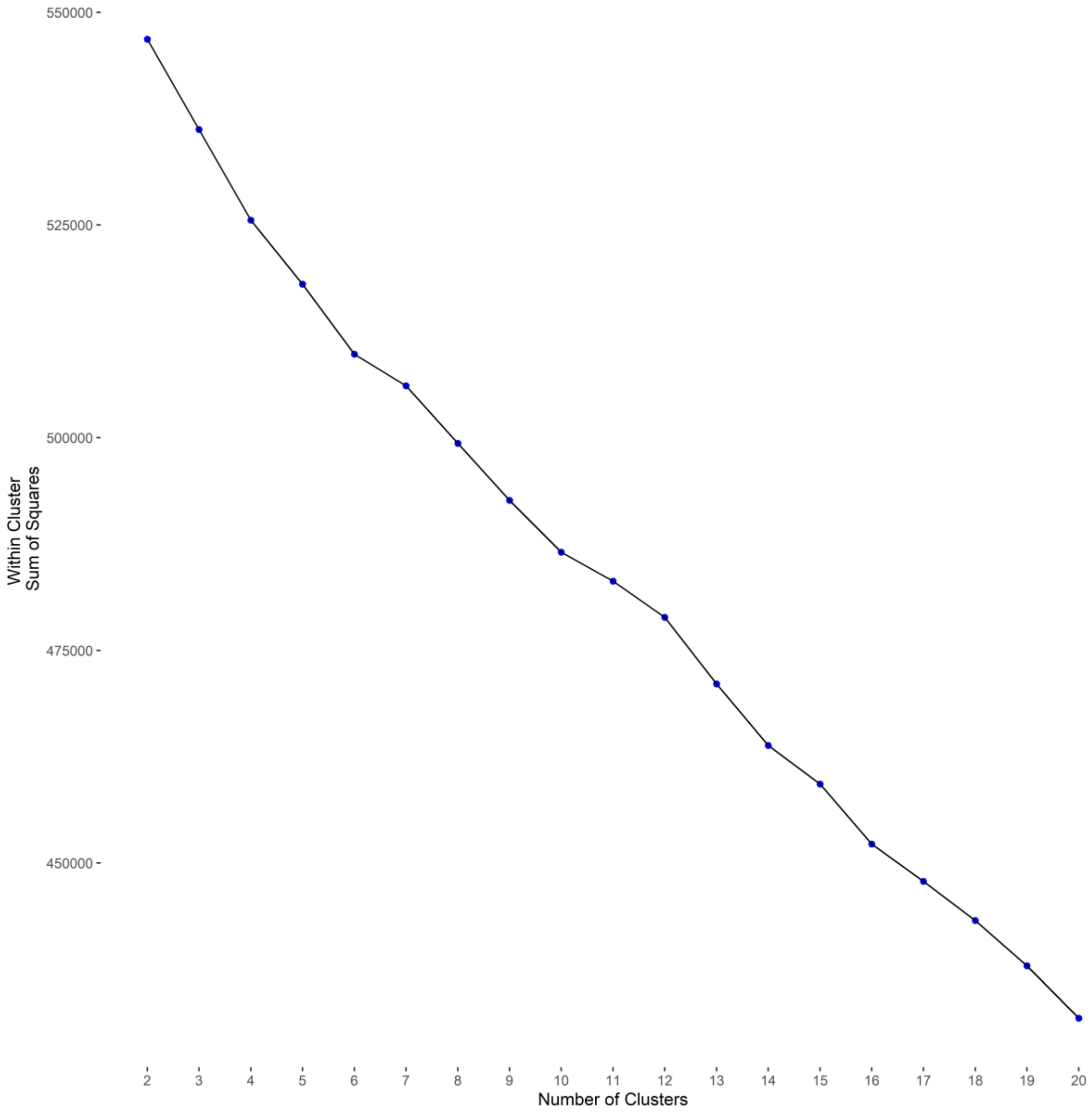


Figure 2

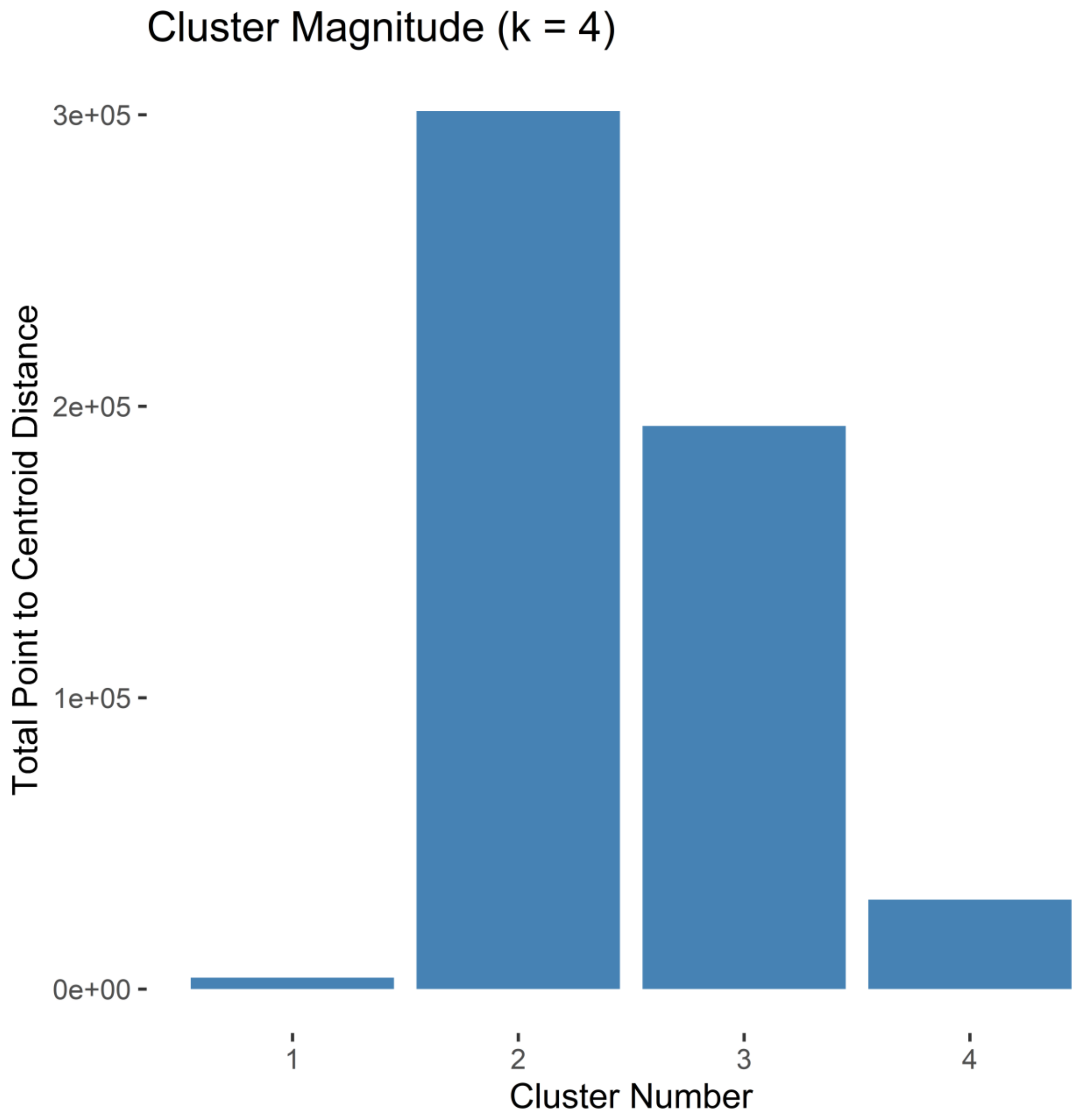

Cluster Magnitude (k = 4)
Total Point to Centroid Distance
3e+05
2e+05
1e+05
0e+00
1
2
3
4
Cluster Number

Figure 3

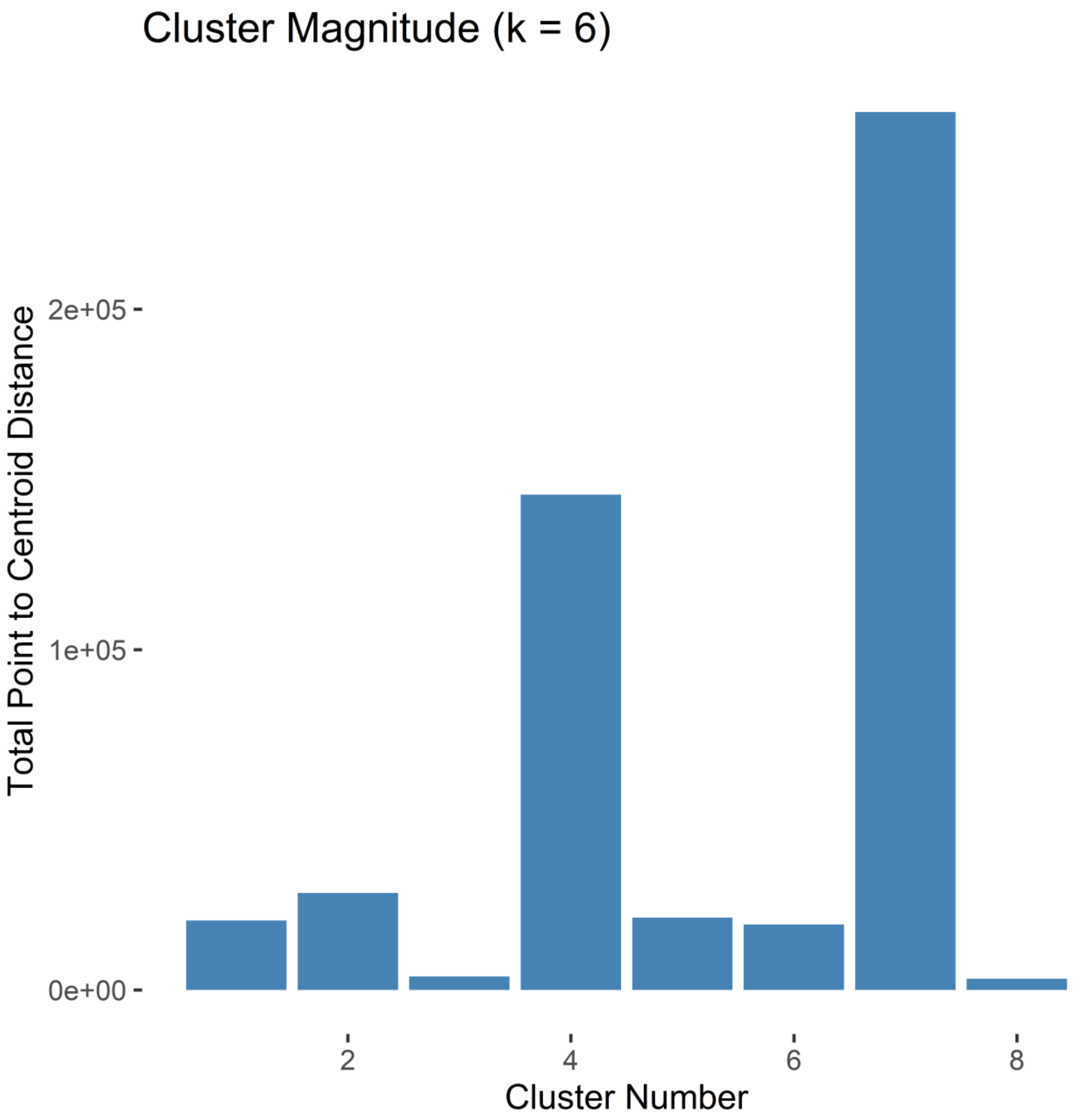

Cluster Magnitude (k = 6)
Total Point to Centroid Distance
2e+05
1e+05
0e+00
2
4
6
8
Cluster Number

Figure 4

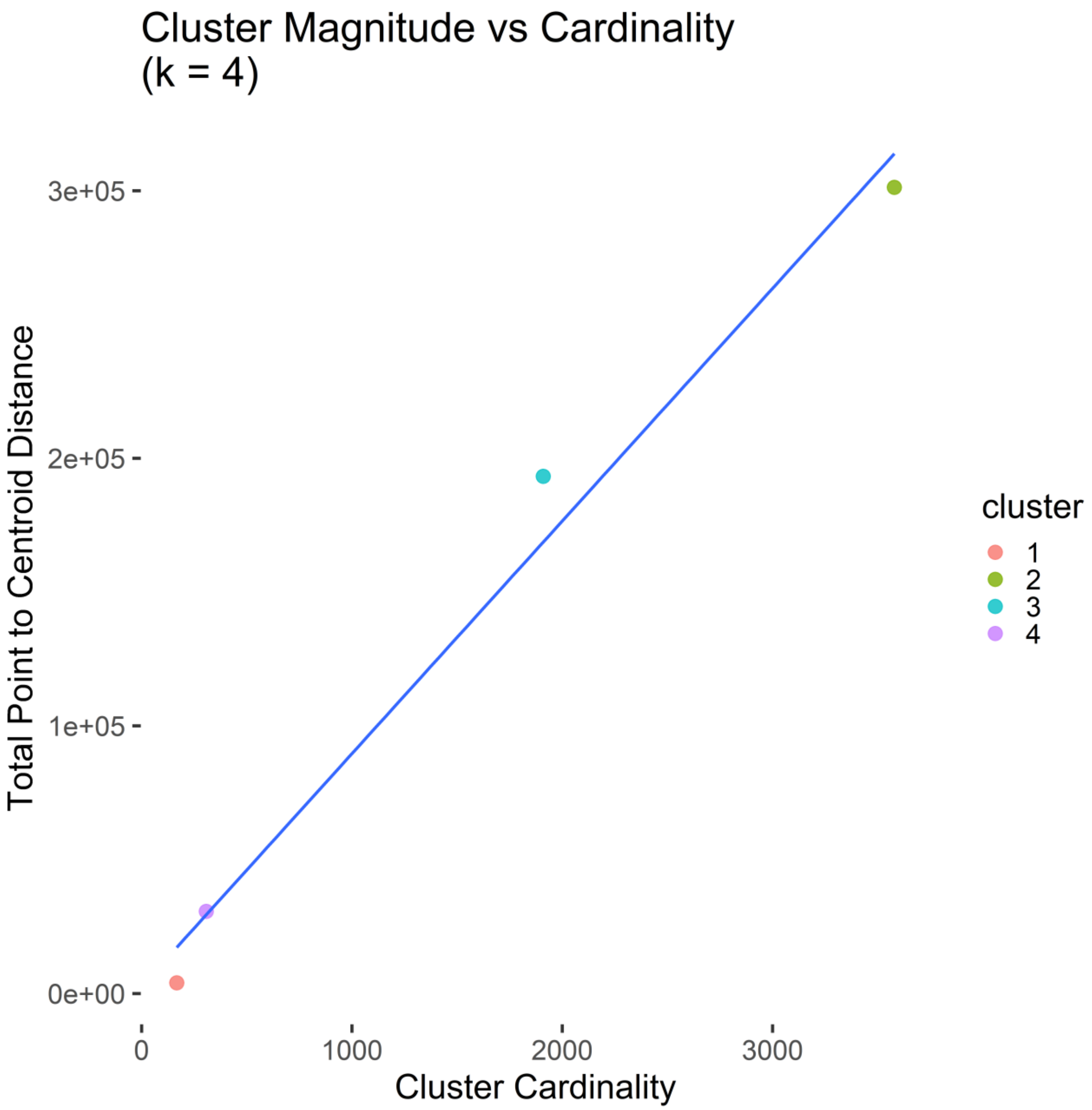

Cluster Magnitude vs Cardinality
(k = 4)
Total Point to Centroid Distance
3e+05
2e+05
1e+05
0e+00
0
1000
2000
3000
Cluster Cardinality
cluster
● 1
● 2
● 3
● 4

Figure 5

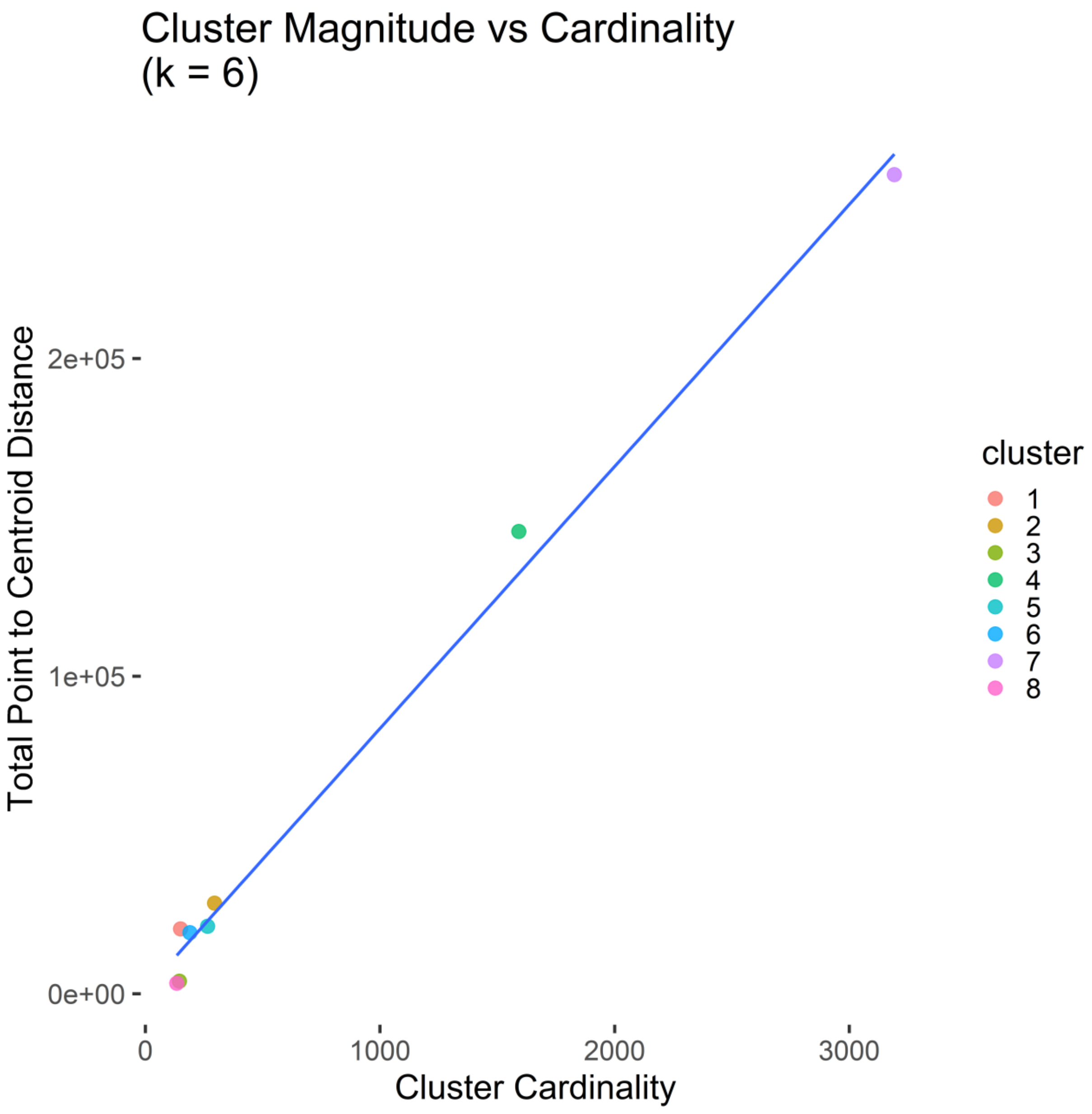

Cluster Magnitude vs Cardinality
(k = 6)
Total Point to Centroid Distance
2e+05
1e+05
0e+00
0
1000
2000
3000
Cluster Cardinality
cluster
1
2
3
4
5
6
7
8

Figure 6

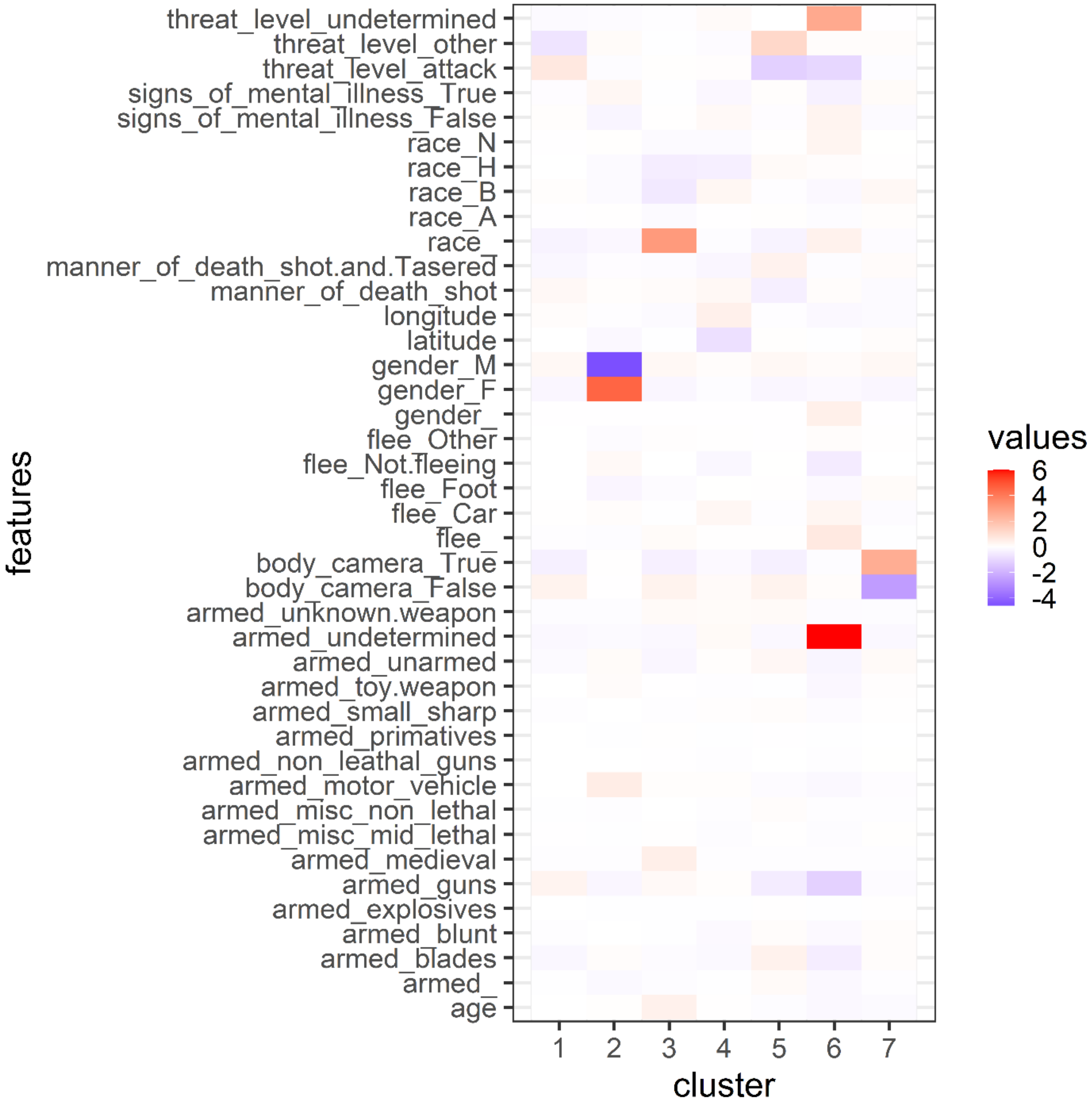

threat_level_undetermined
threat_level_other
threat_level_attack
signs_of_mental_illness_True
signs_of_mental_illness_False
race_N
race_H
race_B
race_A
race_
manner_of_death_shot.and.Tasered
manner_of_death_shot
longitude
latitude
gender_M
gender_F
gender_
flee_Other
flee_Not.fleeing
flee_Foot
flee_Car
flee_
body_camera_True
body_camera_False
armed_unknown.weapon
armed_undetermined
armed_unarmed
armed_toy.weapon
armed_small_sharp
armed_primatives
armed_non_leathal_guns
armed_motor_vehicle
armed_misc_non_lethal
armed_misc_mid_lethal
armed_medieval
armed_guns
armed_explosives
armed_blunt
armed_blades
armed_
age
features
1
2
3
4
5
6
7
cluster
values
6
4
2
0
-2
-4

Figure 7

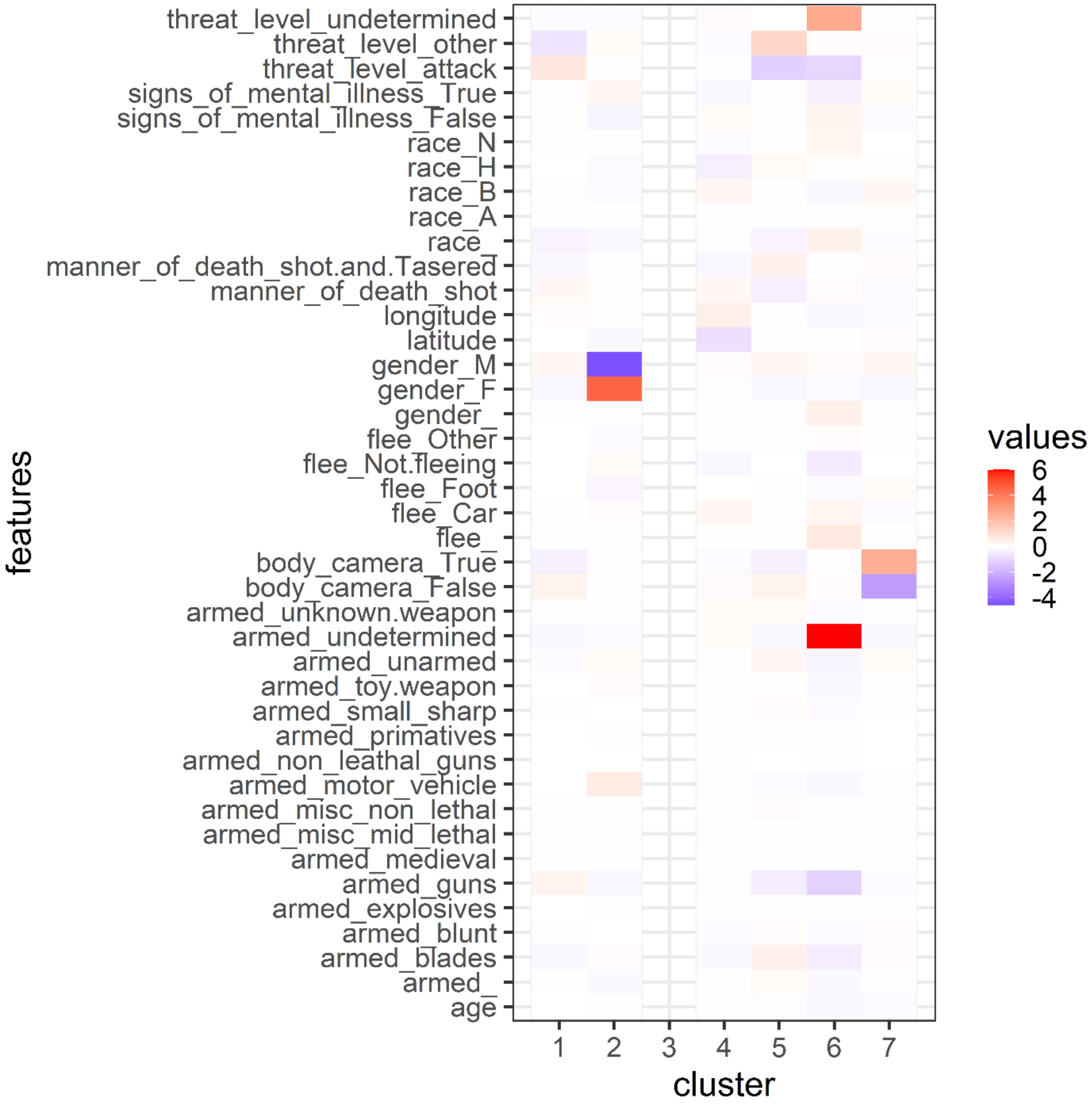

threat_level_undetermined
threat_level_other
threat_level_attack
signs_of_mental_illness_True
signs_of_mental_illness_False
race_N
race_H
race_B
race_A
race_
manner_of_death_shot.and.Tasered
manner_of_death_shot
longitude
latitude
gender_M
gender_F
gender_
flee_Other
flee_Not.fleeing
flee_Foot
flee_Car
flee_
body_camera_True
body_camera_False
armed_unknown.weapon
armed_undetermined
armed_unarmed
armed_toy.weapon
armed_small_sharp
armed_primatives
armed_non_leathal_guns
armed_motor_vehicle
armed_misc_non_lethal
armed_misc_mid_lethal
armed_medieval
armed_guns
armed_explosives
armed_blunt
armed_blades
armed_
age
features
1
2
3
4
5
6
7
cluster
values
6
4
2
0
-2
-4

Figure 8

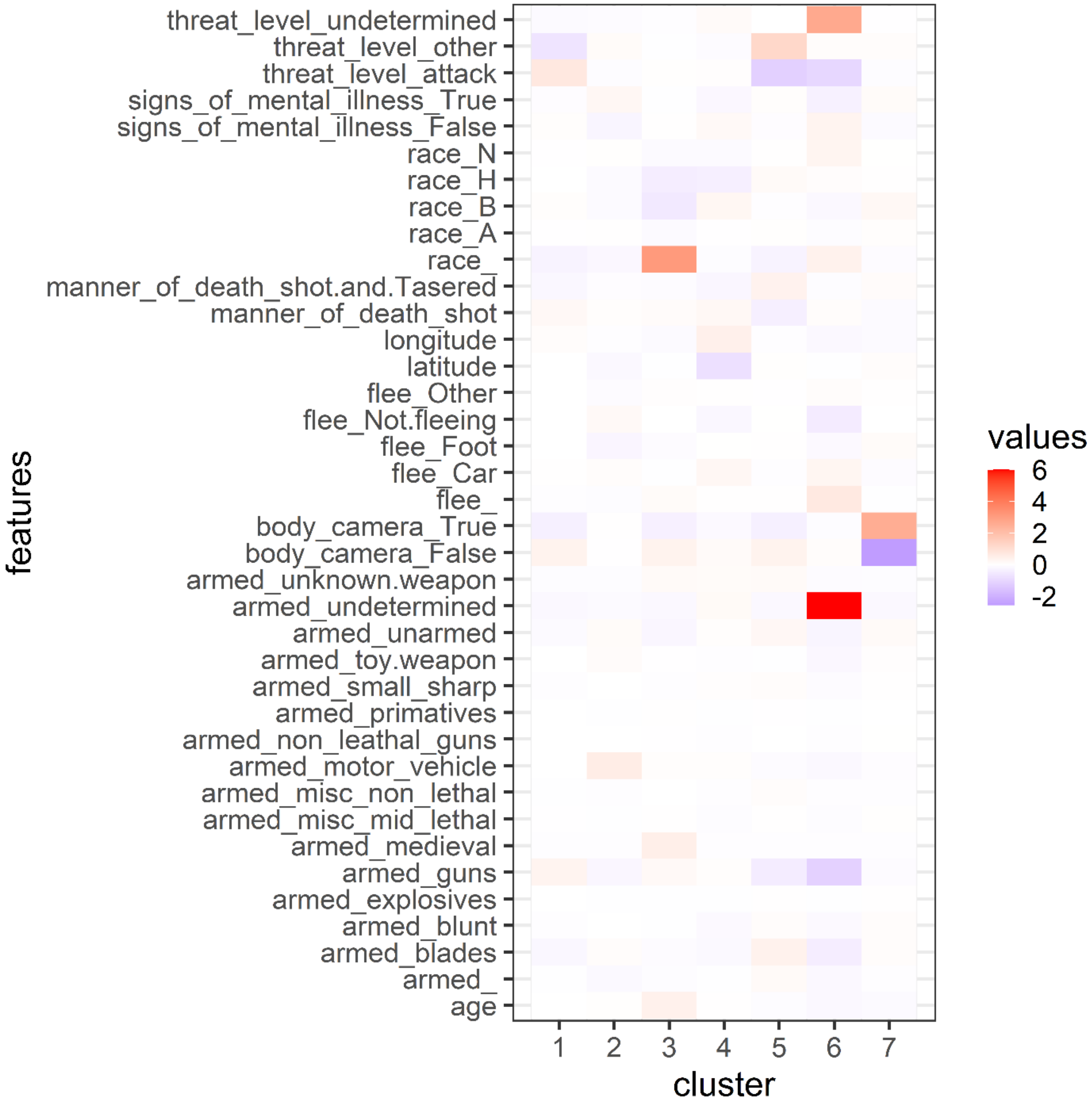
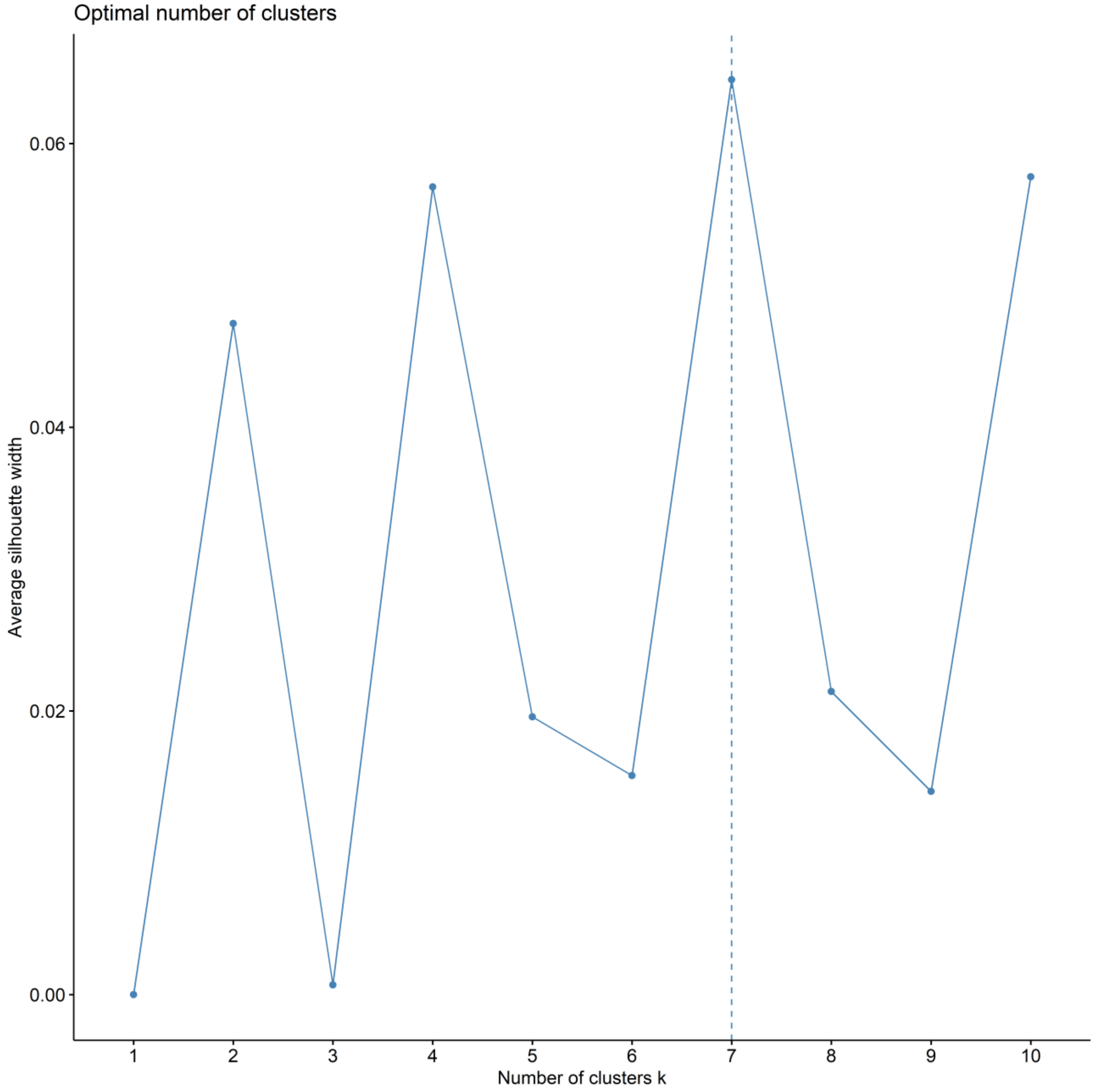

threat_level_undetermined
threat_level_other
threat_level_attack
signs_of_mental_illness_True
signs_of_mental_illness_False
race_N
race_H
race_B
race_A
race_
manner_of_death_shot.and.Tasered
manner_of_death_shot
longitude
latitude
flee_Other
flee_Not.fleeing
flee_Foot
flee_Car
flee_
body_camera_True
body_camera_False
armed_unknown.weapon
armed_undetermined
armed_unarmed
armed_toy.weapon
armed_small_sharp
armed_primatives
armed_non_leathal_guns
armed_motor_vehicle
armed_misc_non_lethal
armed_misc_mid_lethal
armed_medieval
armed_guns
armed_explosives
armed_blunt
armed_blades
armed_
age
features
1
2
3
4
5
6
7
cluster
values
6
4
2
0
-2

Figure 9


Optimal number of clusters
0.06
0.04
0.02
0.00
Average silhouette width
1
2
3
4
5
6
7
8
9
10
Number of clusters k

Bibliography

American Psychological Association. (n.d.). *What works to reduce police brutality*. Monitor on Psychology. https://www.apa.org/monitor/2020/10/cover-police-brutality.

Binder, A. (1983). Split-second decisions: Shootings of and by Chicago police. *Journal of Criminal Justice*, *11*(2), 181–185. https://doi.org/10.1016/0047-2352(83)90055-7

DEFUND THE POLICE. M4BL. (2020, August 20). https://m4bl.org/defund-the-police/.

Home. Black Lives Matter. (2021, April 28). https://blacklivesmatter.com/.

Meyer, M. W. (1980). Police Shootings at Minorities: The Case of Los Angeles. *The ANNALS of the American Academy of Political and Social Science*, *452*(1), 98–110. https://doi.org/10.1177/000271628045200110

Ross, C. (2014). Introducing the United States Police-Shooting Database: A Multi-Level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States, 2011-2014. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2534673

WP Company. (2020, January 22). *Fatal Force: Police shootings database*. The Washington Post. https://www.washingtonpost.com/graphics/investigations/police-shootings-database/.

R-code

~~~
---
title: "Why does this keep happening?"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## Loading in the Data
~~~

~~~
```{r}
library(ggplot2)
library(factoextra) # clustering algorithms & visualization
library(tidyr)
library(fastDummies)

```




```{r}
#load("C:/Users/ccrum/Desktop/fatal-police-shootings-data.csv")

```

## Taking a Peak at the Data

```{r}
data <- fatal.police.shootings.data
head(data)
tail(data)
summary(as.factor(data$flee))

```



```{r}
summary(as.factor(data$armed))
new_armed <- as.character(data$armed)
new_armed[data$armed %in% c("gun", "gun and vehicle", "vehicle and gun", "machete and
gun", "gun and machete", "hatchet and gun", "guns and explosive", "gun and knife", "gun and
sword", "gun and car", "guns and explosives")] <- "guns"

new_armed[data$armed %in% c("hatchet","knife","machete","samurai sword","sharp
object","ax", "pick-axe", "bayonet", "meat cleaver", "baseball bat and knife","sword", "vehicle and
machete", "knife and vehicle", "lawn mower blade", "car, knife and mace", "pole and knife")] <-
"blades"


new_armed[data$armed %in% c("baseball bat and bottle", "baton", "bottle", "brick", "garden
tool", "piece of wood", "pole","hammer","metal pipe","metal stick","pipe", "shovel","tire iron",
"wrench", "baseball bat", "crowbar", "oar", "pitchfork", "walking stick", "blunt object", "baseball
bat and fireplace poker", "ice pick", "metal pole", "barstool", "metal rake", "flashlight")] <- "blunt"
~~~

~~~~
new_armed[data$armed %in% c("Airsoft pistol", "BB gun and vehicle", "BB gun", "air pistol",
"bean-bag gun", "pellet gun")] <- "non_leathal_guns"

new_armed[data$armed %in% c("vehicle", "motorcycle")] <- "motor_vehicle"


new_armed[data$armed %in% c("flagpole", "contractor's level", "chain", "binoculars", "stapler",
"chair", "air conditioner", "cordless drill", "wasp spray", "pepper spray", "metal object", "pen",
"microphone", "claimed to be armed")] <- "misc_non_lethal"


new_armed[data$armed %in% c("scissors", "glass shard", "screwdriver","beer bottle","straight
edge razor", "box cutter", "metal hand tool", "railroad spikes")] <- "small_sharp"

new_armed[data$armed %in% c("bow and arrow","crossbow","spear")] <- "medieval"

new_armed[data$armed %in% c("rock", "hand torch")] <- "primatives"


new_armed[data$armed %in% c("grenade", "fireworks", "incendiary device")] <- "explosives"

new_armed[data$armed %in% c("chain saw", "carjack","Taser", "nail gun" , "chainsaw")] <-
"misc_mid_lethal"


summary(as.factor(new_armed))
```

```{r}
table(new_armed,data$race)
data$armed <- new_armed

```



## Scaling the Categorical Variables

```{r}
data$age[is.na(data$age)] <- mean(data$age,na.rm = TRUE)

data_2 <- dummy_cols(data[,c(1,4:8, 10:16)])
# Create x variables
#x_vars <- model.matrix(id~.,
~~~~

~~~
#                     data[,c(1,4:8, 10:16)])[,-1]
x_vars <- na.omit(data_2[,c(4,12, 13, 14:104)])

# Scale data
x_vars_s <- scale(x_vars)

```


## Initial Clustering

```{r}
set.seed(12345) # Set seed for reproducibility
fit_1 <- kmeans(x = x_vars_s, # Set data as explantory variables
                centers = 4,  # Set number of clusters
                nstart = 25, # Set number of starts
                iter.max = 100 ) # Set maximum number of iterations to use


```


```{r}
fit_1$centers
```


```{r}
# Create function to try different cluster numbers
kmean_withinss <- function(k) {
  cluster <- kmeans( x = x_vars_s,  # Set data to use
                    centers = k,  # Set number of clusters as k, changes with input into function
                    nstart = 25, # Set number of starts
                    iter.max = 100) # Set max number of iterations
  return (cluster$tot.withinss) # Return cluster error/within cluster sum of squares
}


# Set maximum cluster number
max_k <-20
# Run algorithm over a range of cluster numbers
wss <- sapply(2:max_k, kmean_withinss)


# Create a data frame to plot the graph
elbow <-data.frame(2:max_k, wss)
~~~

~~~
# Plot the graph with gglop
g_1 <- ggplot(elbow, aes(x = X2.max_k, y = wss)) +
  theme_set(theme_bw(base_size = 22) ) +
  geom_point(color = "blue") +
  geom_line() +
  scale_x_continuous(breaks = seq(1, 20, by = 1)) +
  labs(x = "Number of Clusters", y="Within Cluster \nSum of Squares") +
  theme(panel.grid.major = element_blank(), # Turn of the background grid
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank())
g_1

ggsave(g_1, file = "g_1.png", width = 10, height = 10, dpi = 600)

```


```{r}
set.seed(12345) # Set seed for reproducibility
fit_2 <- kmeans(x = x_vars_s, # Set data as explantory variables
                centers = 8,  # Set number of clusters
                nstart = 25, # Set number of starts
                iter.max = 100 ) # Set maximum number of iterations to use
clusters_2 <- fit_2$cluster
clusters_1 <- fit_1$cluster
```




```{r}

plot_clust_cardinality <- cbind.data.frame(clusters_1, clusters_2) # Join clusters with  k =4 and k=6
names(plot_clust_cardinality) <- c("k_4", "k_6") # Set names
# Create bar plots
g_2 <- ggplot(plot_clust_cardinality, aes( x = factor(k_4))) + # Set x as cluster values
  geom_bar(stat = "count", fill = "steelblue") + # Use geom_bar with stat = "count" to count observations
    labs(x = "Cluster Number", y="Points in Cluster", # Set labels
         title = "Cluster Cardinality (k = 4)") +
  theme(panel.grid.major = element_blank(), # Turn of the background grid
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank())
~~~

~~~~
g_3 <- ggplot(plot_clust_cardinality, aes( x = factor(k_6))) + # Set x as cluster values
  geom_bar(stat = "count", fill = "steelblue") + # Use geom_bar with stat = "count" to count
observations
    labs(x = "Cluster Number", y="Points in Cluster", # Set labels
         title = "Cluster Cardinality (k = 6)") +
  theme(panel.grid.major = element_blank(), # Turn of the background grid
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank())
g_2

g_3

```


```{r}
k_4_mag <- cbind.data.frame(c(1:4), fit_1$withinss) # Extract within cluster sum of squares
names(k_4_mag) <- c("cluster", "withinss") # Fix names for plot data
g_4 <- ggplot(k_4_mag, aes(x = cluster, y = withinss)) + # Set x as cluster, y as withinss
  geom_bar(stat = "identity", fill = "steelblue") + # Use geom bar and stat = "identity" to plot
values directly
   labs(x = "Cluster Number", y="Total Point to Centroid Distance", # Set labels
         title = "Cluster Magnitude (k = 4)") +
  theme(panel.grid.major = element_blank(), # Turn of the background grid
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank())

k_6_mag <- cbind.data.frame(c(1:8), fit_2$withinss) # Extract within cluster sum of squares
names(k_6_mag) <- c("cluster", "withinss") # Fix names for plot data
g_5 <- ggplot(k_6_mag, aes(x = cluster, y = withinss)) +  # Set x as cluster, y as withinss
  geom_bar(stat = "identity", fill = "steelblue") + # Use geom bar and stat = "identity" to plot
values directly
   labs(x = "Cluster Number", y="Total Point to Centroid Distance", # Set labels
         title = "Cluster Magnitude (k = 6)") +
  theme(panel.grid.major = element_blank(), # Turn of the background grid
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank())


g_4
~~~~

~~~
g_5

ggsave(g_4, file = "g_4.png", width = 10, height = 10, dpi = 600)
ggsave(g_5, file = "g_5.png", width = 10, height = 10, dpi = 600)
```

```{r}
k_4_dat <- cbind.data.frame(table(clusters_1), k_4_mag[,2]) # Join magnitude and cardinality
names(k_4_dat) <- c("cluster", "cardinality", "magnitude") # Fix plot data names

g_6 <- ggplot(k_4_dat, aes(x = cardinality, y = magnitude, color = cluster)) + # Set aesthetics
  geom_point(alpha = 0.8, size  = 4) +  # Set geom point for scatter
 geom_smooth(aes(x = cardinality, y = magnitude), method = "lm",
              se = FALSE, inherit.aes = FALSE, alpha = 0.5) + # Set trend  line
  labs(x = "Cluster Cardinality", y="Total Point to Centroid Distance", # Set labels
        title = "Cluster Magnitude vs Cardinality \n(k = 4)") +
  theme(panel.grid.major = element_blank(), # Turn of the background grid
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank())


k_6_dat <- cbind.data.frame(table(clusters_2), k_6_mag[,2]) # Join magnitude and cardinality
names(k_6_dat) <- c("cluster", "cardinality", "magnitude") # Fix plot data names

g_7 <- ggplot(k_6_dat, aes(x = cardinality, y = magnitude, color = cluster)) + # Set aesthetics
  geom_point(alpha = 0.8, size = 4) +  # Set geom point for scatter
  geom_smooth(aes(x = cardinality, y = magnitude), method = "lm",
              se = FALSE, inherit.aes = FALSE, alpha = 0.5) + # Set trend  line
  labs(x = "Cluster Cardinality", y="Total Point to Centroid Distance", # Set labels
        title = "Cluster Magnitude vs Cardinality \n(k = 6)") +
  theme(panel.grid.major = element_blank(), # Turn of the background grid
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank())


g_6
g_7

ggsave(g_6, file = "g_6.png", width = 10, height = 10, dpi = 600)
ggsave(g_7, file = "g_7.png", width = 10, height = 10, dpi = 600)
~~~

~~~
```

```{r}
# Create silhouette plot
silhouette_plot <- fviz_nbclust(x_vars_s, kmeans, method = "silhouette")
ggsave(silhouette_plot, file = "silhouette_plot.png", width = 10, height = 10, dpi = 600)


```

```{r}
set.seed(12345) # Set seed for reproducibility
fit_3 <- kmeans(x = x_vars_s, # Set data as explantory variables
                centers = 7,  # Set number of clusters
                nstart = 25, # Set number of starts
                iter.max = 100 ) # Set maximum number of iterations to use


```

```{r}
centers_3 <- fit_3$centers
# Create vector of clusters
cluster <- c(1: 7)
# Extract centers
center_df <- data.frame(cluster, centers_3)

# Reshape the data
plot_data <- center_df[,c(1:30, 84:95)]
center_reshape <- gather(plot_data, features, values, age:body_camera_True)
head(center_reshape)


```

```{r}
g_heat_1 <- ggplot(data = center_reshape, aes(x = features, y = cluster, fill = values)) +
  scale_y_continuous(breaks = seq(1, 7, by = 1)) +
  geom_tile() +
  coord_equal() +
  theme_set(theme_bw(base_size = 22) ) +
  scale_fill_gradient2(low = "blue", # Choose low color
                       mid = "white", # Choose mid color
                       high = "red", # Choose high color
~~~

~~~~
                       midpoint =0, # Choose mid point
                       space = "Lab",
                       na.value ="grey", # Choose NA value
                       guide = "colourbar", # Set color bar
                       aesthetics = "fill") + # Select aesthetics to apply
  coord_flip()


g_heat_1
ggsave(g_heat_1, file = "heat_map_1.png", width = 10, height = 10, dpi = 600)

```


```{r}
summary(as.factor(fit_3$cluster))

```


```{r}
g_heat_2 <- ggplot(data = center_reshape[center_reshape$cluster!=3,], aes(x = features, y =
cluster, fill = values)) +
  scale_y_continuous(breaks = seq(1, 7, by = 1)) +
  geom_tile() +
  coord_equal() +
  theme_set(theme_bw(base_size = 22) ) +
  scale_fill_gradient2(low = "blue", # Choose low color
                       mid = "white", # Choose mid color
                       high = "red", # Choose high color
                       midpoint =0, # Choose mid point
                       space = "Lab",
                       na.value ="grey", # Choose NA value
                       guide = "colourbar", # Set color bar
                       aesthetics = "fill") + # Select aesthetics to apply
  coord_flip()


g_heat_2
ggsave(g_heat_2, file = "heat_map_2.png", width = 10, height = 10, dpi = 600)

```


```{r}
plot_data_2 <- center_df[,c(1:22,26:30,84:95)]
center_reshape_2 <- gather(plot_data_2, features, values, age:body_camera_True)
~~~~

~~~
head(center_reshape_2)

```


```{r}
g_heat_3 <- ggplot(data = center_reshape_2, aes(x = features, y = cluster, fill = values)) +
  scale_y_continuous(breaks = seq(1, 7, by = 1)) +
  geom_tile() +
  coord_equal() +
  theme_set(theme_bw(base_size = 22) ) +
  scale_fill_gradient2(low = "blue", # Choose low color
                       mid = "white", # Choose mid color
                       high = "red", # Choose high color
                       midpoint =0, # Choose mid point
                       space = "Lab",
                       na.value ="grey", # Choose NA value
                       guide = "colourbar", # Set color bar
                       aesthetics = "fill") + # Select aesthetics to apply
  coord_flip()


g_heat_3
ggsave(g_heat_3, file = "heat_map_3.png", width = 10, height = 10, dpi = 600)
```
~~~