

AC4 - DVA - GID: mmendiola3

Data

I used a subset of the LAPD crime data available at the following URL, which includes the GPS coordinates and Crime Code for all LA crimes reported in 2016.

<https://data.lacity.org/A-Safe-City/LAPD-Crime-and-Collision-Raw-Data-for-2016/ttiz-7an8/data>

Analysis

I analyzed the distribution of Crimes Code Catagories between the two major Los Angeles population centers (San Fernando Vally and City of Los Angeles).

```
# Perform k-means cluster analysis
cl = kmeans(as.matrix(cbind(df$lat, df$lon), ncol=2), 2)
df$cluster = factor(cl$cluster)

# Reshape to show relative proportions of each crime ranking by cluster
c <- df %>% select(Crm_rank, cluster) %>% group_by(Crm_rank, cluster) %>%
  mutate(count = n()) %>% slice(1)
c_wide <- dcast(c, Crm_rank ~ cluster)
```

```
## [1] "Max difference in relative percentages of Crime Rankings: 0.060721"
```

Visualization

The following plot shows the geographical distribution of reported crimes across Los Angeles. Shapes show the k-means (k=2) cluster membership of each report. Color indicates the Crime Code value (100s indicating major crimes to 900s indicating minor crimes). The max difference in relative percentages of Crime rankings from above indicates that between the two population clusters, there is little difference in the distribution of the seriousness of reported crimes.

