

# Clustering Based on Zonal Statistics

```
setwd("/Users/jamesjensen/Documents/harris/q1_20/UML/Project/mapping-disaster-risk/data")  
  
roof_matrix <- read_csv("300_roof_matrix.csv")
```

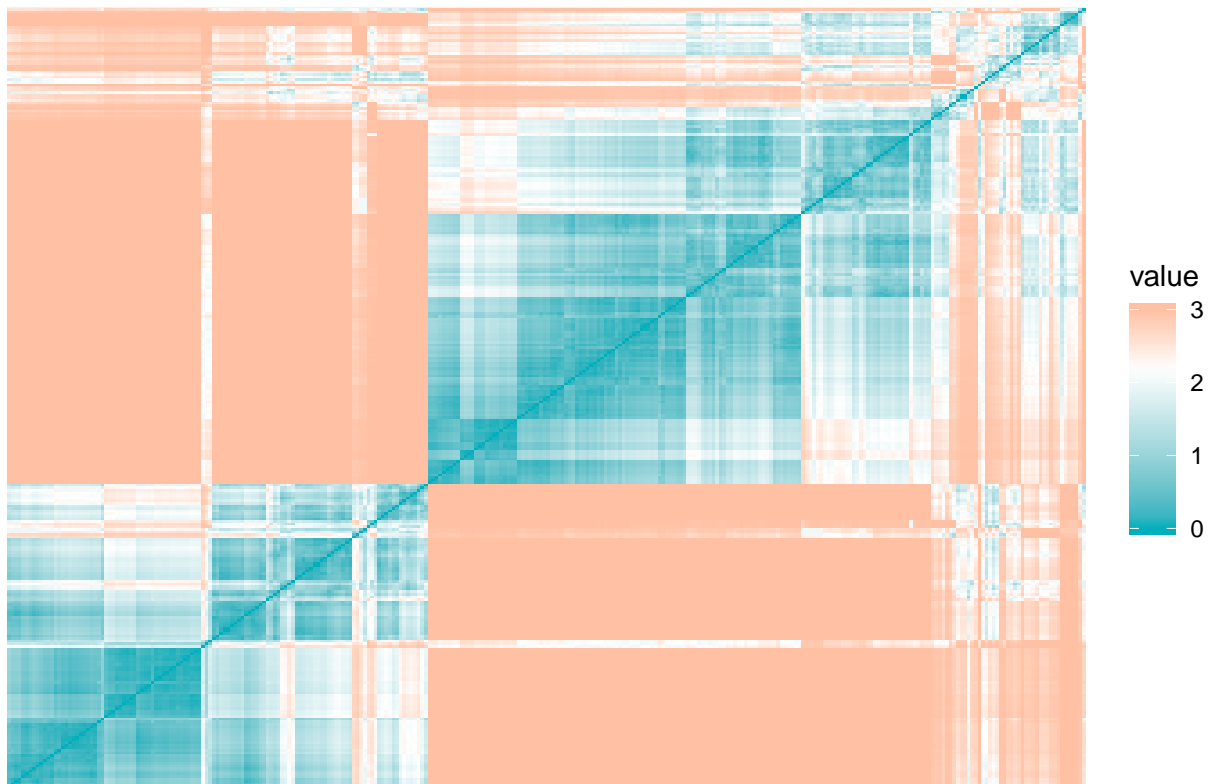
```
## Parsed with column specification:  
## cols(  
##   roof = col_character(),  
##   `1` = col_double(),  
##   `2` = col_double(),  
##   `3` = col_double()  
## )
```

```
# remove all zeroes  
# only necessary for matrices made without the zonal stats function  
#roof_matrix <- roof_matrix %>%  
#   select(-`3`) %>%  
#   filter(!(`0`==0 & `1`==0 & `2`==0))
```

## Ordered Dissimilarity Matrix

```
roof_scaled <- roof_matrix %>%  
  select( -roof) %>%  
  scale()  
  
dist_man <- roof_scaled %>%  
  dist(method="manhattan")  
  
dist_euc <- roof_scaled %>%  
  dist(method="euclidean")  
  
dist_can <- roof_scaled %>%  
  dist(method="canberra")  
  
hs <- hopkins(roof_scaled, n=50)  
hs <- round(as.numeric(as.character(unlist(hs))), digits=3)
```

## Canberra



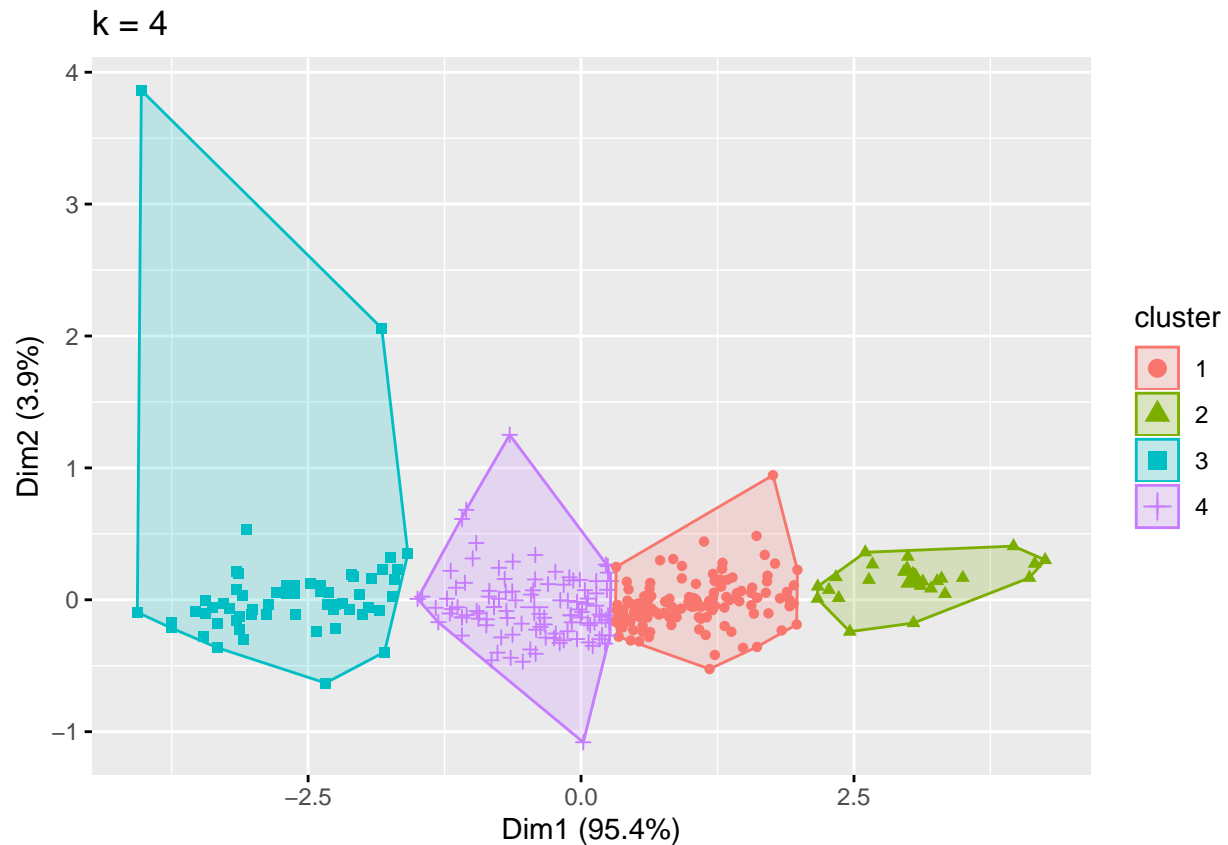
```
k2 <- kmeans(roof_scaled, centers = 4, nstart = 15)

output <- roof_matrix %>%
  dplyr::select(roof)

output$K_Cluster <- as.factor(k2$cluster)

p1 <- fviz_cluster(k2, geom = "point", data = roof_scaled) + ggtitle("k = 4")

p1
```



### Compare Groupings to Actual Labels

```
grouped <- aggregate(output, by=list(output$roof, output$K_Cluster),
                      FUN=length)
colnames(grouped)[colnames(grouped)=="Group.2"] <- "Cluster"
table(output$roof[output$K_Cluster == 1])
```

```
##
## concrete_cement    healthy_metal    incomplete irregular_metal
##                2                56                3                54
```

```
table(output$roof[output$K_Cluster == 2])
```

```
##
## healthy_metal irregular_metal
##            20                5
```

```
actual_label_count <- table(roof_matrix$roof)

ggplot(grouped, aes(fill=Cluster, y=K_Cluster, x=Group.1)) +
  geom_bar(position="dodge", stat="identity") +
```

```
xlab("Roof Material") +  
ylab("Count")
```

