BAILA_clustering

This file walks you through BAILA analysis

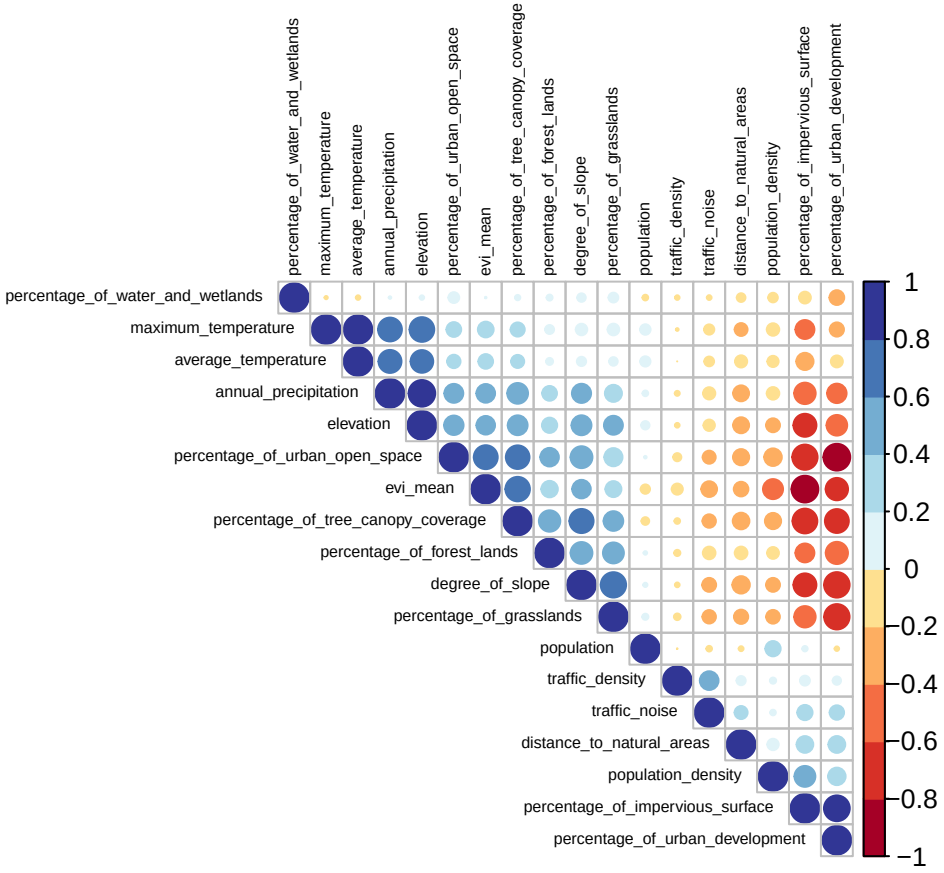
```
load("generated_data/clean_baila_data.rda")
```

```
library(tidyverse)
library(corrplot)
library(RColorBrewer)
library(sf)
```

```
scaled.dat <-
  clean_baila_data %>%
  select(-"bg_id") %>%
  st_set_geometry(NULL) %>%
  scale()
```

Correlation Plot

```
M <-cor(scaled.dat)
corrplot(M, order="hclust",type="upper",
         col=brewer.pal(n=10, name="RdYlBu"), tl.col = 'black',tl.cex = 0.5)
```



Create dendrogram

```
library(dendextend)
```

```
dend <-
  scaled.dat %>% # Scale the data
  dist (method = "euclidean") %>% # calculate a distance matrix
  hclust(method = "ward.D2")  %>% # Hierarchical clustering
  as.dendrogram # Turn the object into a dendrogram
```

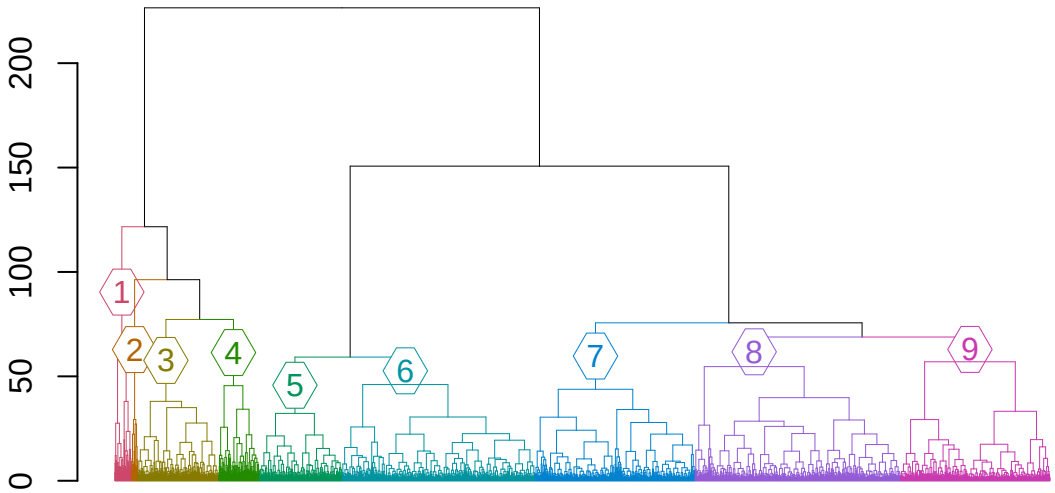
Gap Statistic deciding optimal number of clusters

```
library(cluster)
library(factoextra)
gap_stat <-
  clean_baila_data %>%
  select(-"bg_id") %>%
  st_set_geometry(NULL) %>%
  scale()%>%
  clusGap(FUN = hcut, K.max = 15, B = 30) # gap statistic to decide optimal number of clsuters

fviz_gap_stat(gap_stat) #Plot gap statistic
gap_stat$Tab #check gap statistic
```

Cut and display dendrogram

```
dend %>%
  set("labels", NA) %>%
  color_branches(k=9, groupLabels =TRUE) %>%
  set("labels_colors", k = 9 ) %>%
  set("branches_lwd", 0.5) %>% # Branches line width
  set("branches_k_color", k = 9) %>%
  plot()
```



Link urban type info with BGs

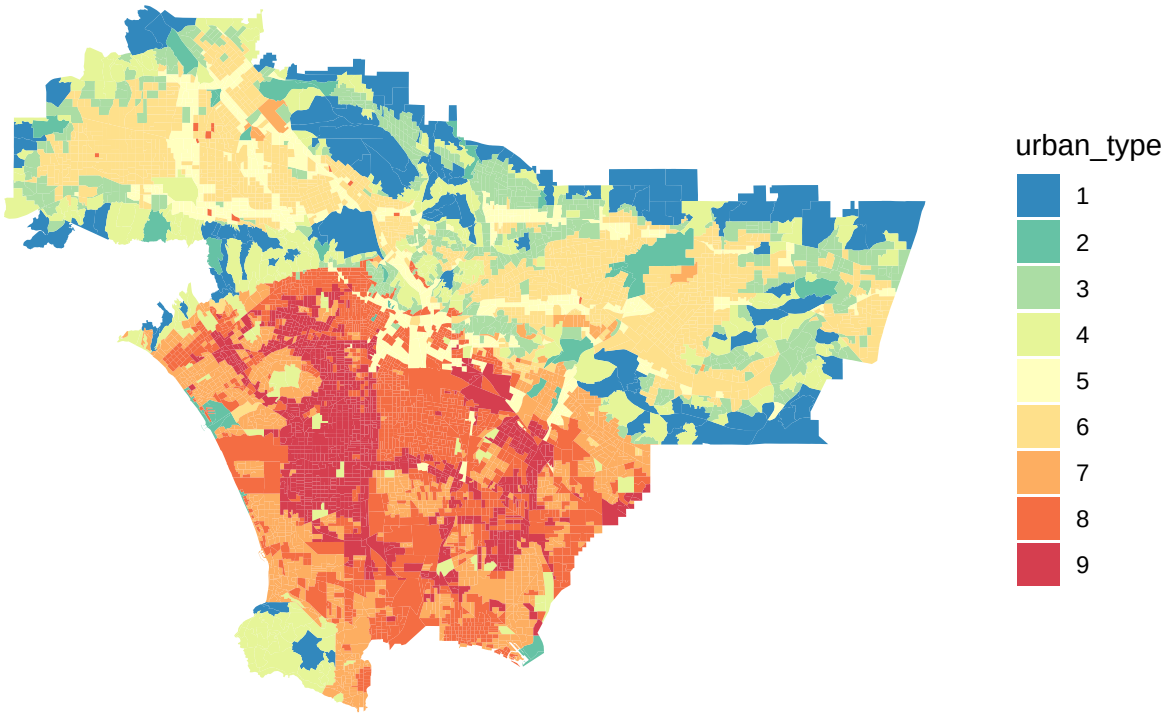
```
cluster <- cutree(dend,k=9, order_clusters_as_data=FALSE) ## cut tree into 9 types

urban_type <-
  cluster %>%
  as.data.frame() %>%
  rownames_to_column("id") %>%
  mutate(id = as.numeric(id)) %>%
  rename("urban_type" = ".") %>%
  arrange(id)

#add urban type info to original dataset
clean_baila_data<-
  clean_baila_data %>%
  mutate(urban_type  = urban_type$urban_type) %>%
  mutate(urban_type = as.factor(urban_type))
```

Plot typology map

```
clean_baila_data %>%
  ggplot()+
  geom_sf(aes(fill= urban_type), color = NA) +
  scale_fill_brewer(palette = "Spectral", direction = -1) +
  coord_sf(crs = st_crs(clean_baila_data), datum = NA) +
  theme_classic()
```



Summarise number of BG in each type

```
clean_baila_data %>%
  st_set_geometry(NULL) %>%
  group_by(urban_type) %>%
  summarise(n=n())
```

```
## # A tibble: 9 x 2
##   urban_type     n
##   <fct>      <int>
## 1 1            109
## 2 2             39
## 3 3            524
## 4 4            262
## 5 5            536
## 6 6           1245
## 7 7           1033
## 8 8           1328
## 9 9            964
```