Assignment

Group 6

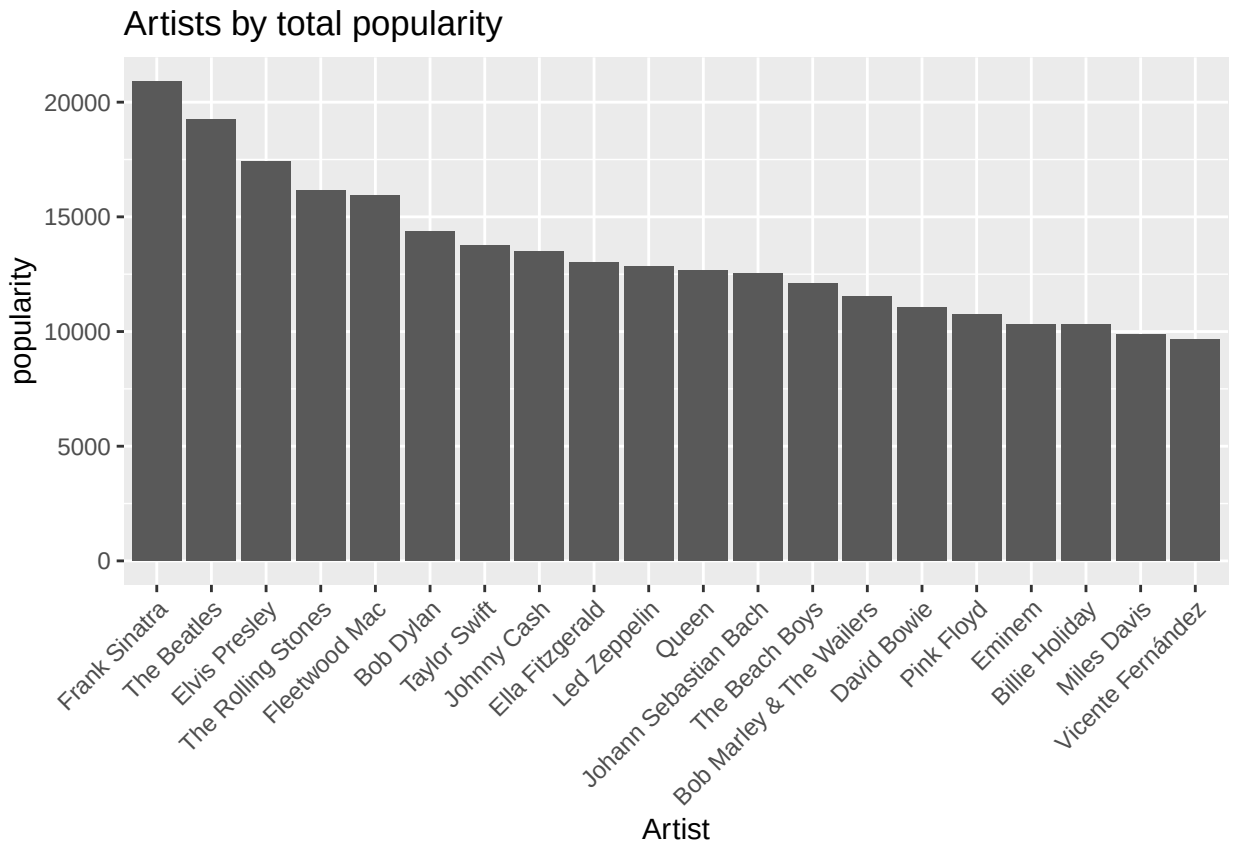
2/16/2021

```
data <- read_csv("data/data.csv")
```

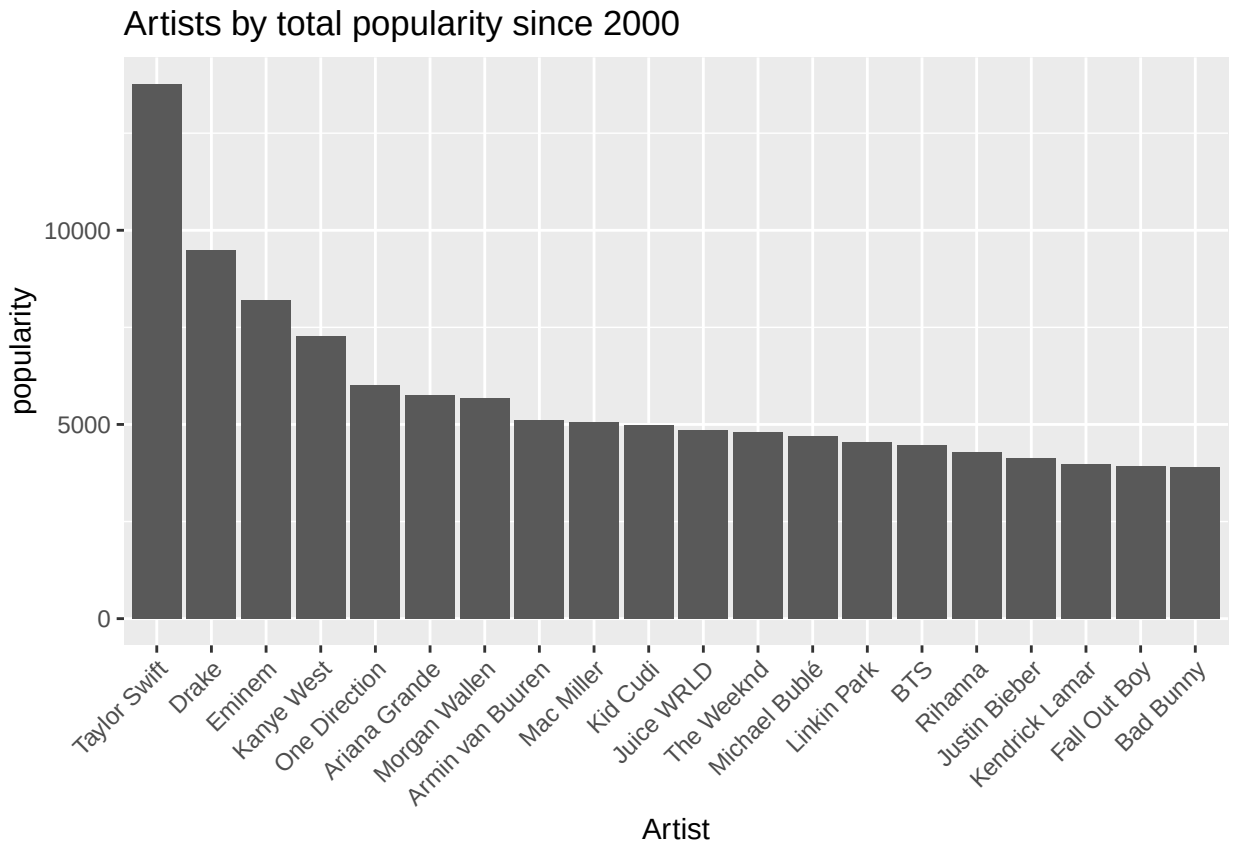
```
##
## -- Column specification --------------------------------------------------------
## cols(
##   acousticness = col_double(),
##   artists = col_character(),
##   danceability = col_double(),
##   duration_ms = col_double(),
##   energy = col_double(),
##   explicit = col_double(),
##   id = col_character(),
##   instrumentalness = col_double(),
##   key = col_double(),
##   liveness = col_double(),
##   loudness = col_double(),
##   mode = col_double(),
##   name = col_character(),
##   popularity = col_double(),
##   release_date = col_character(),
##   speechiness = col_double(),
##   tempo = col_double(),
##   valence = col_double(),
##   year = col_double()
## )
```

```
data <- data %>% mutate(artists= str_split(str_remove_all(artists, regex("\\[|\\]|'")),","))
data <- data %>% mutate(first_artist=map_chr(artists, ~ .[1]))
```

```
data %>% group_by(first_artist) %>% summarise(popularity=sum(popularity)) %>% slice_max(order_by=popula
  ggplot() +
    geom_col(aes(x=reorder(first_artist, -popularity), y=popularity))+
    theme(axis.text.x = element_text(angle = 45,hjust = 1)) +
    xlab("Artist") +
  ggtitle("Artists by total popularity")
```



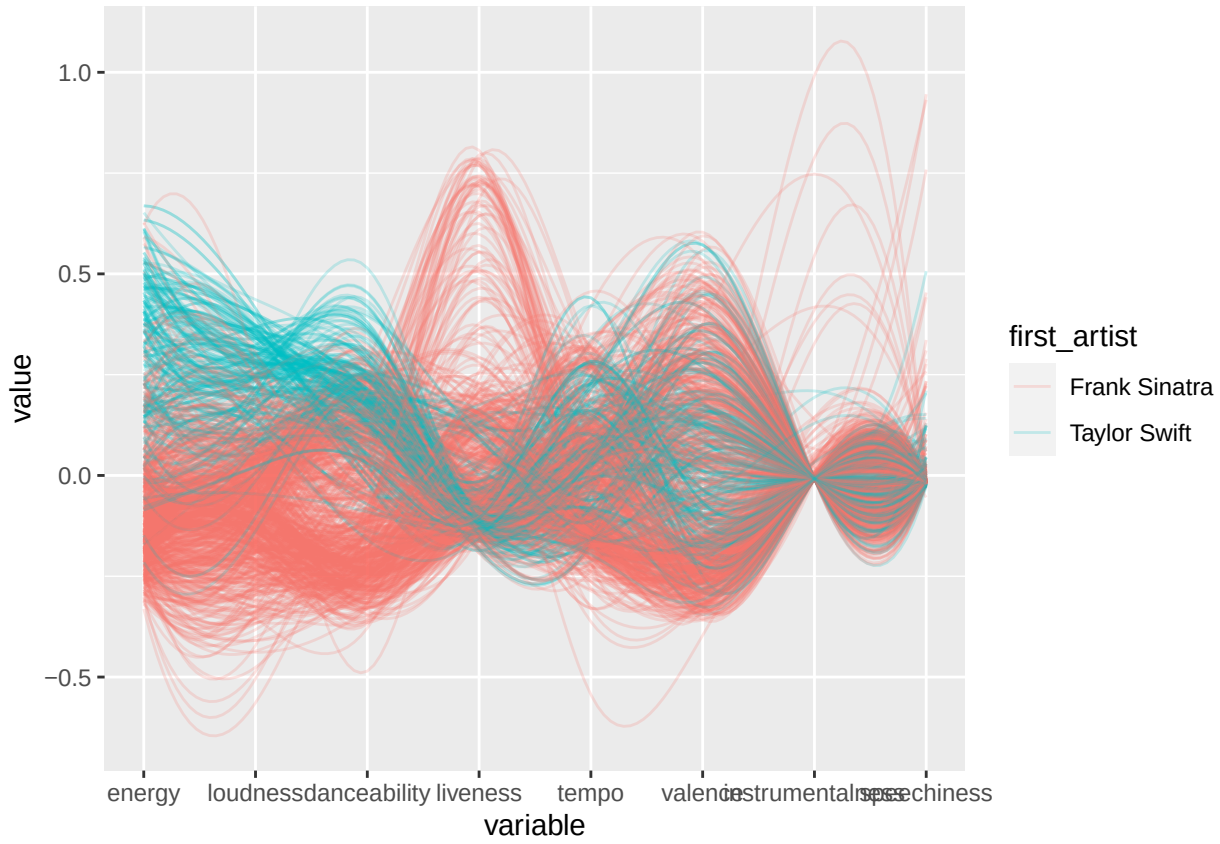

```
data %>%filter(year>=2000) %>%  group_by(first_artist) %>% summarise(popularity=sum(popularity)) %>% sl
  ggplot() + geom_col(aes(x=reorder(first_artist, -popularity), y=popularity))+ theme(axis.text.x = ele
  xlab("Artist") +
  ggtitle("Artists by total popularity since 2000")
```




```
temp <- data %>% filter(first_artist=="Taylor Swift"|first_artist=="Frank Sinatra") %>% mutate(first_ar
```

```
comp_cols <- c("danceability", "energy", "loudness", "speechiness", "liveness", "instrumentalness","tem
temp %>%   select("first_artist", comp_cols) %>%
ggparcoord(groupColumn= 1, columns=comp_cols, order = "anyClass", alphaLines = 0.2, scale="center", spl
```

```
## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(comp_cols)` instead of `comp_cols` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```



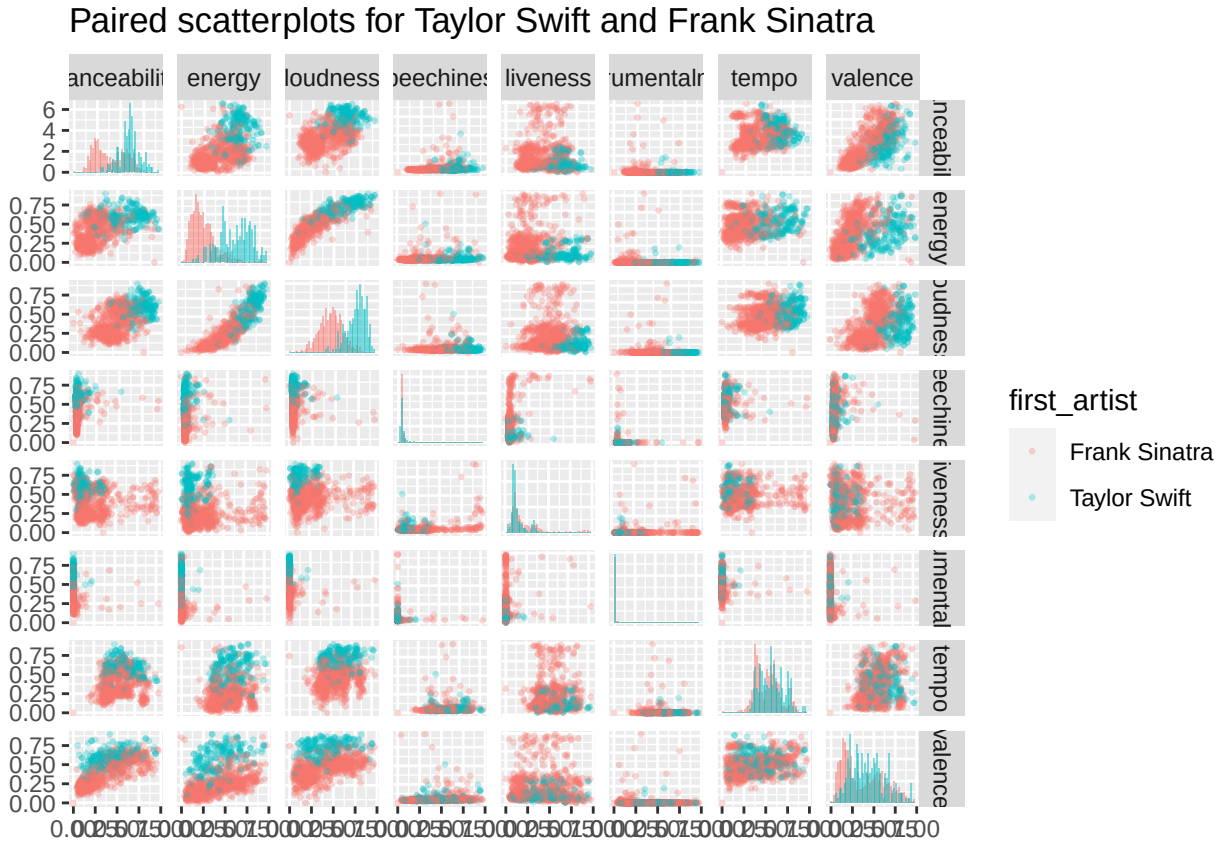

```
pairplot <- function(data, title=NULL, columns=NULL, geom_type=geom_point, include_dist=FALSE, legend=NU
  if (is.null(columns)){
    columns <- 1:ncol(data)
  }
  col_names <- names(data)[columns]
  plot_list <- list()
  k <- 1
  for (i in 1:length(columns)){
    for (j in 1:length(columns)){
      if(i==j){
        p <- ggplot(data,mapping=aes_string(x=col_names[i], fill="first_artist")) + geom_histogram(data=
          geom_histogram(data=subset(data, first_artist=="Taylor Swift"), mapping=aes(y=..density..), co
        if(include_dist){
          p <- p + geom_density(alpha=.2, fill="#FF6666")
        }

        plot_list[[k]] <-  p
      } else{
        plot_list[[k]] <- ggplot(data) + geom_type(mapping=aes_string(x=col_names[i], y=col_names[j], co
      }

      k <- k+1
    }
  }
  ggmatrix(plot_list, nrow = length(col_names), ncol = length(col_names), xAxisLabels
           =col_names, yAxisLabels=col_names, title=title, legend=legend)
```

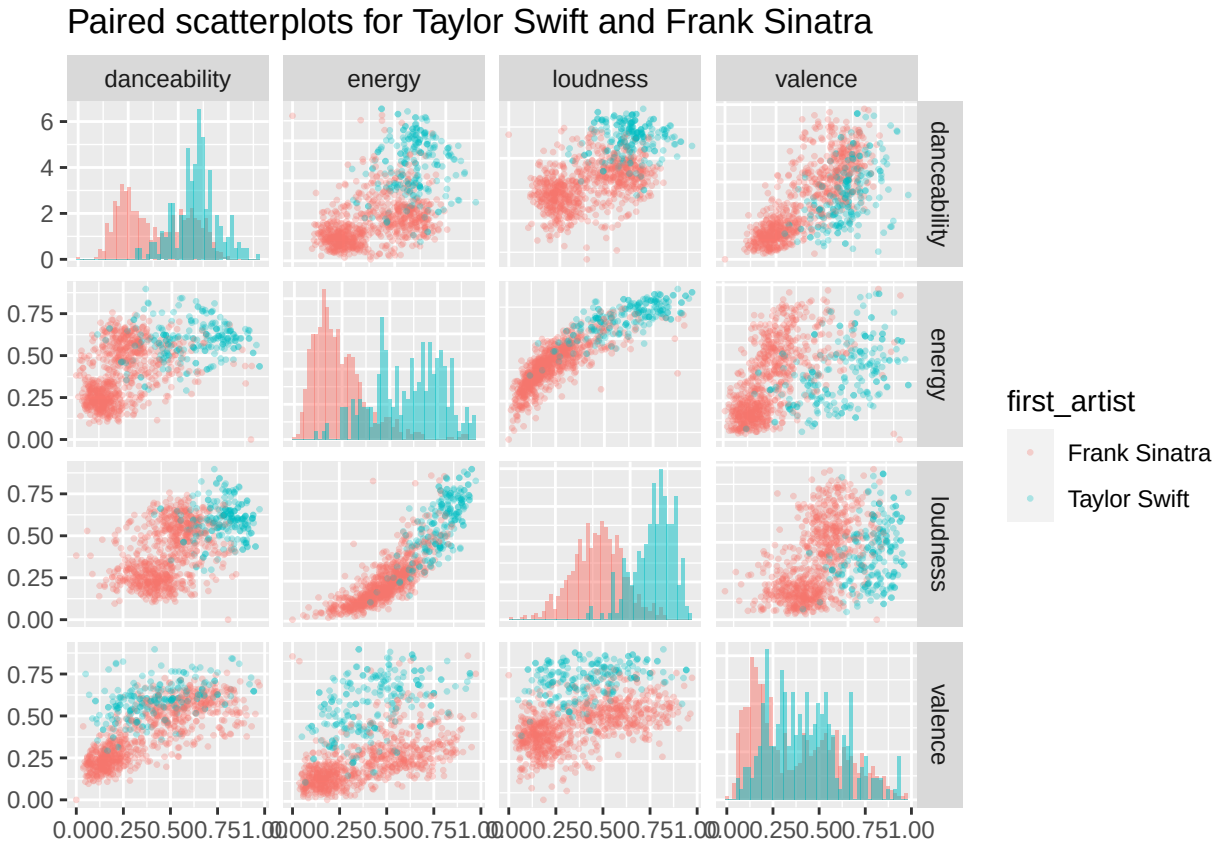
```
}
```

```
temp %>% select("first_artist", comp_cols) %>%
pairplot(columns=c(2:(length(comp_cols)+1)), geom_type=geom_point, include_dist = FALSE, size=0.5, alpha
```



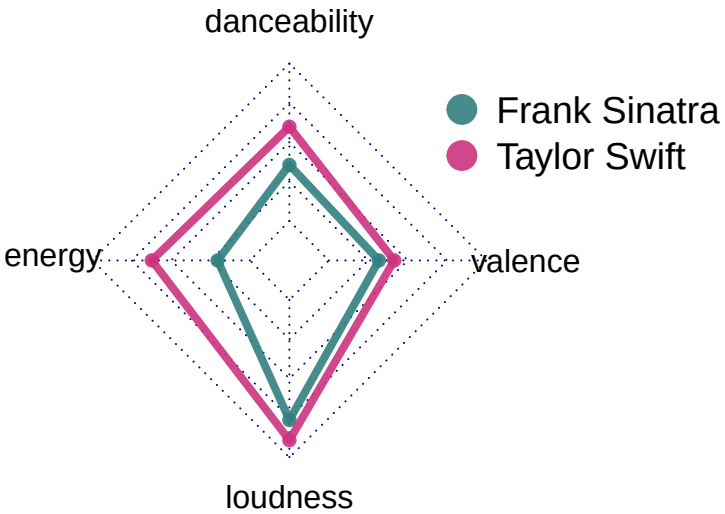

```
detailed_comp_cols <- c("danceability", "energy", "loudness","valence")
temp %>% select("first_artist", detailed_comp_cols) %>%
pairplot(columns=c(2:(length(detailed_comp_cols)+1)), geom_type=geom_point, include_dist = FALSE, size=
```

```
## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(detailed_comp_cols)` instead of `detailed_comp_cols` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```




```
temp %>% group_by(first_artist) %>% summarise_if(is.numeric, median) %>%  select("first_artist", detail
rownames(temp2) <- temp2$group
temp2 <- temp2 %>% select(-group)
temp2 <- rbind(rep(1, 4), rep(0, 4), as.data.frame(temp2))
colors_border=c( rgb(0.2,0.5,0.5,0.9), rgb(0.8,0.2,0.5,0.9))
colors_in=c( rgb(0.2,0.5,0.5,0.4), rgb(0.8,0.2,0.5,0.4) )

temp2 %>% radarchart(pcol=colors_border ,
                     #pfcol=colors_in ,
                     plwd=4 , plty=1,)
legend(x=0.7, y=1, legend = rownames(temp2[-c(1,2),]), bty = "n", pch=20 , col=colors_border , text.col
```



```
comp_cols <- c("danceability", "energy", "loudness", "valence")
temp %>%   select("first_artist", comp_cols) %>%
ggparcoord(groupColumn= 1, columns=comp_cols, order = "anyClass", alphaLines = 0.2, scale="center", spli
```