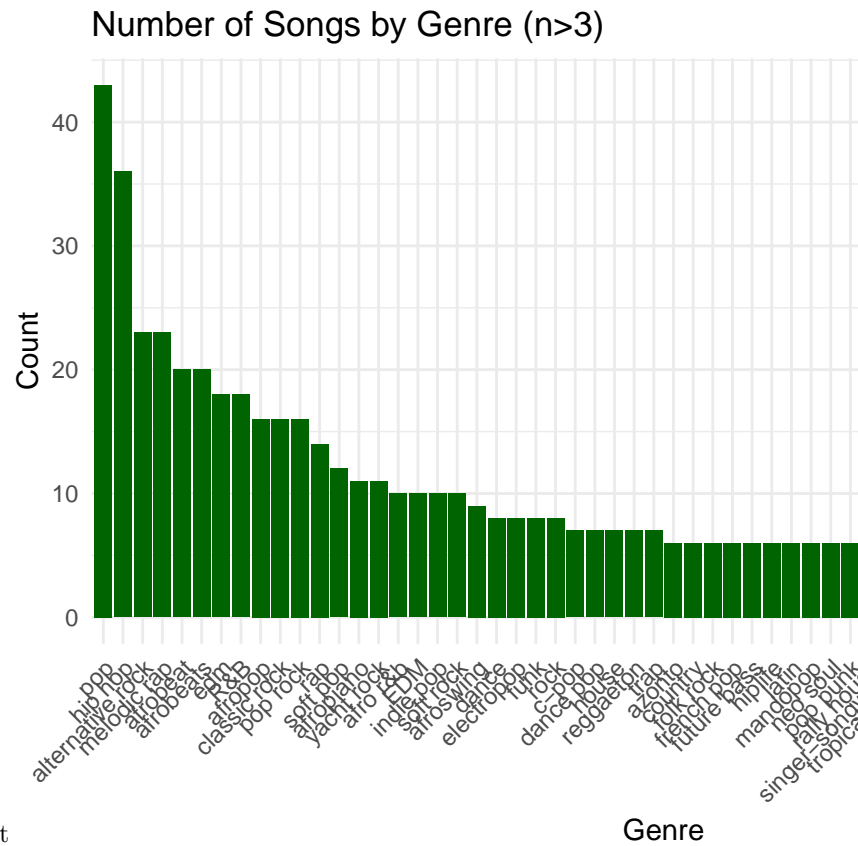


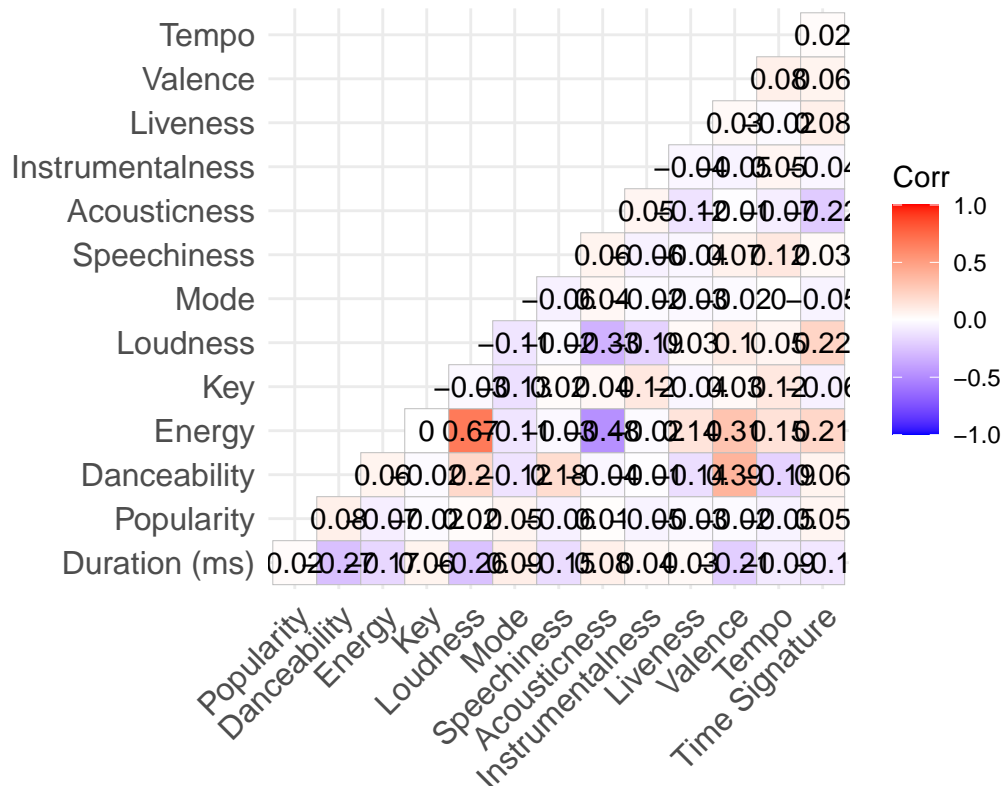
19\_adam

2025-06-30



here we have a lovely lil bar plot of all genres>3 count  
lets ermhhh do a lil correlation heatmap. also standardized data in here

### Correlation Heatmap of Song Features



clus-turd

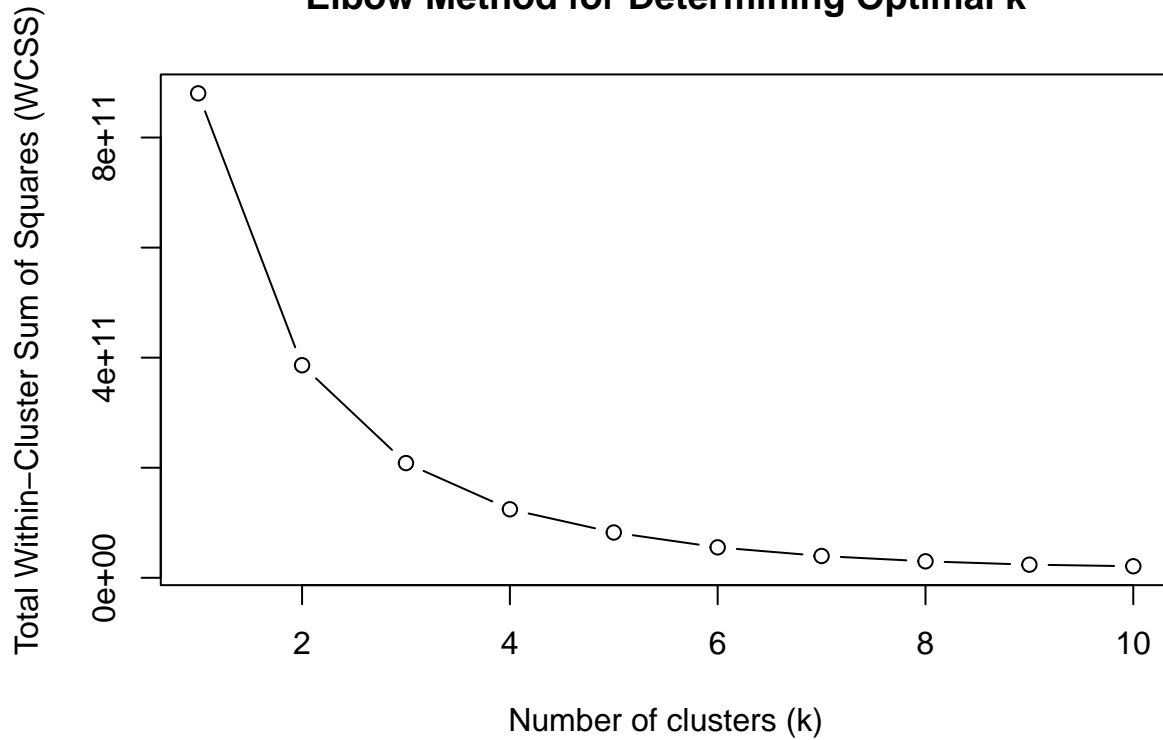
```
# Maximum number of clusters to try
max_k <- 10

# Vector to store total within-cluster sum of squares for each k
wcss <- numeric(max_k)

for (k in 1:max_k) {
  km <- kmeans(spotify_numeric, centers = k, nstart = 25) # Forgy + 10 random starts
  wcss[k] <- km$tot.withinss
}

# Plot the Elbow graph
plot(1:max_k, wcss, type = "b",
     xlab = "Number of clusters (k)",
     ylab = "Total Within-Cluster Sum of Squares (WCSS)",
     main = "Elbow Method for Determining Optimal k")
```

## Elbow Method for Determining Optimal k



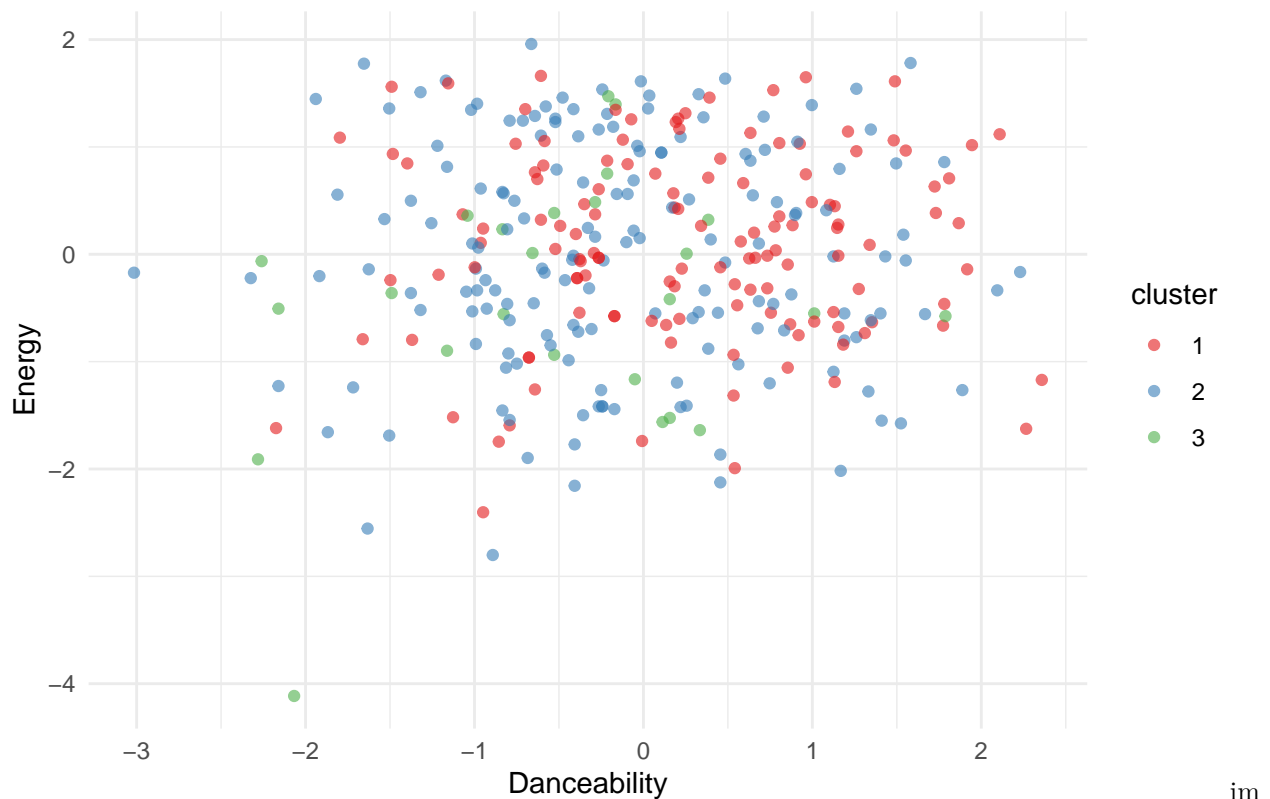
ima make k = 3 since 2 seems super small and not interesting though that is where the elbow is

```
k = 3
```

```
clusters = kmeans(spotify_numeric, centers = k, nstart = 25)

scaled_df$cluster <- factor(clusters$cluster)
# Plot the clusters
ggplot(scaled_df, aes(x = `Danceability`, y = `Energy`, color = cluster)) +
  geom_point(alpha = 0.6) +
  labs(title = "K-Means Clustering of Spotify Songs",
       x = "Danceability", y = "Energy") +
  theme_minimal() +
  scale_color_brewer(palette = "Set1")
```

## K-Means Clustering of Spotify Songs



boutta pop off with this graph just watch

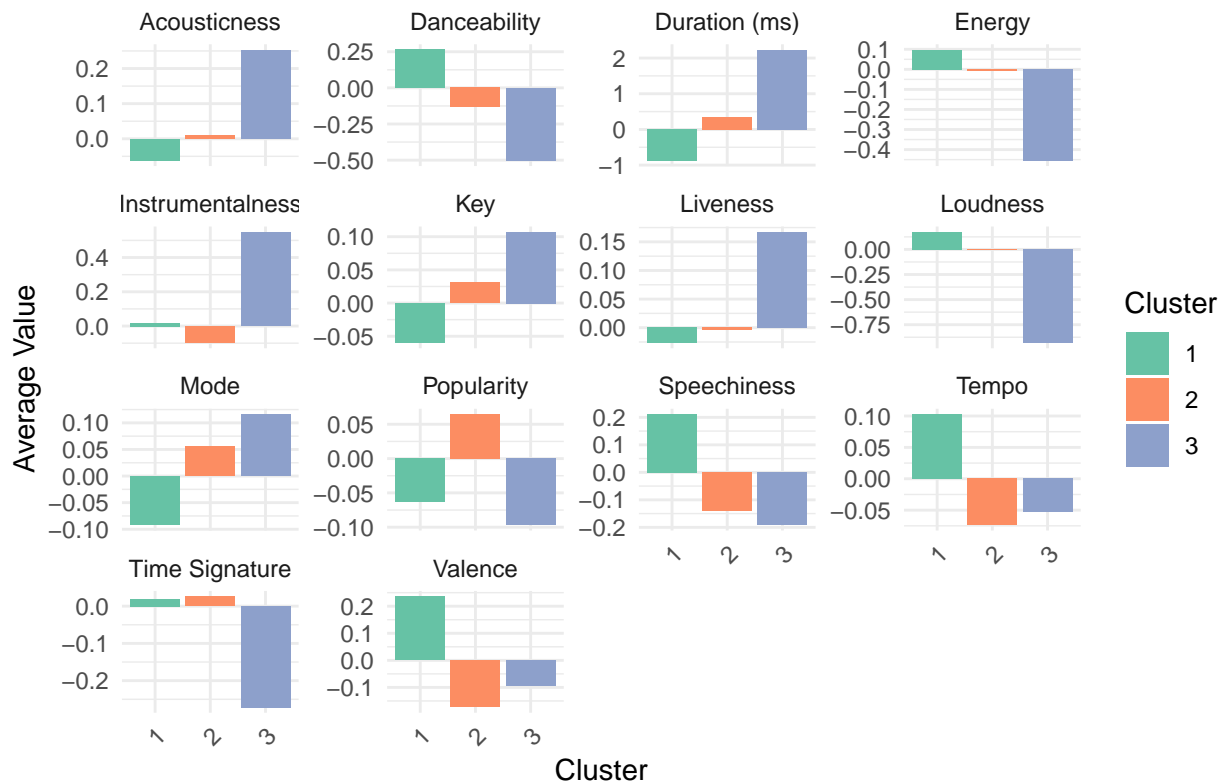
```
numeric_columns <- colnames(scaled_df)[sapply(scaled_df, is.numeric)]

# Calculate means by cluster for all numeric variables
cluster_means <- scaled_df %>%
  group_by(cluster) %>%
  summarise(across(all_of(numeric_columns), mean))

# Reshape to long format for plotting
cluster_means_long <- cluster_means %>%
  pivot_longer(cols = -cluster, names_to = "Variable", values_to = "Average")

ggplot(cluster_means_long, aes(x = cluster, y = Average, fill = cluster)) +
  geom_col(position = position_dodge()) +
  facet_wrap(~ Variable, scales = "free_y") +
  labs(title = "Average Values by Cluster for All Variables",
       x = "Cluster",
       y = "Average Value",
       fill = "Cluster") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_brewer(palette = "Set2")
```

## Average Values by Cluster for All Variables



ok so 3 is spivey?? slower, softer, less energy, longer, not rly danceable, more acoustic, more instrumental

1 is more upbeat faster songs, more electronic ones, but less popular than 2. 2 is slower and pretty middle on most categories except popularity. 2 is prob aiwen a lot. eh shes also probably in 1 a bunch but more 2. one i think has the rap/hiphop