Dimensionality Reduction

Ethan Huynh, Meinhard Capucao, Ryan Gagliardi, Andrew Gerungan

Run PCA on the unpopular songs data

```
library(caret)
```

```
## Loading required package: ggplot2
```

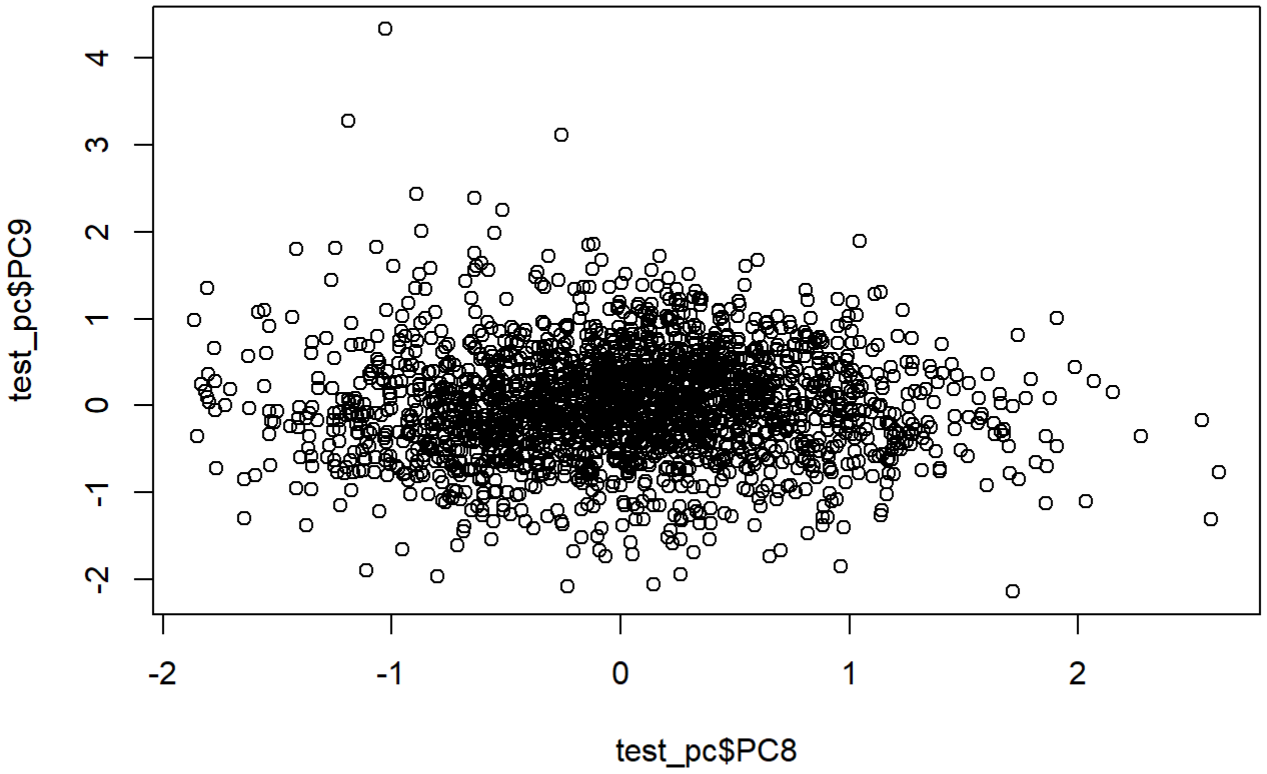
```
## Loading required package: lattice
```

```
df <- read.csv('unpopular_songs.csv')
set.seed(1526)
i <- sample(1:nrow(df), nrow(df)*0.8, replace=FALSE)
train <- df[i,]
test <- df[-i,]
pca_out <- preProcess(train[, c(1,2,4,6:12)], method=c("center", "scale", "pca"))
pca_out
```

```
## Created from 8701 samples and 10 variables
##
## Pre-processing:
##   - centered (10)
##   - ignored (0)
##   - principal component signal extraction (10)
##   - scaled (10)
##
## PCA needed 9 components to capture 95 percent of the variance
```

PCA plot

```
train_pc <- predict(pca_out, train[, c(1,2,4,6:12)])
test_pc <- predict(pca_out, test[, c(1,2,4,6:12)])
plot(test_pc$PC8, test_pc$PC9, pch=c(23,21,22)[unclass(test_pc$popularity)], bg=c("red","green",
"blue")[unclass(test$popularity)])
```



###

Scale the data

Now let’s see if our nine principal components can predict popularity

```
train_scaled <- train_pc[, 1:9]  # don't scale popularity
means <- sapply(train_scaled, mean)
stdvs <- sapply(train_scaled, sd)
train_scaled <- scale(train_scaled, center=means, scale=stdvs)
test_scaled <- scale(test_pc[, 1:9], center=means, scale=stdvs)
```

PCA data in knn

```
library(class)
set.seed(1526)
# fit the model
fit <- knnreg(train_scaled, train$popularity, k=50)
# evaluate
pred <- predict(fit, test_scaled)
cor_knn <- cor(pred, test$popularity)
mse_knn <- mean((pred - test$popularity)^2)
print(paste("cor=", cor_knn))
```

```
## [1] "cor= 0.286903075341191"
```

```
print(paste("mse=", mse_knn))
```

```
## [1] "mse= 14.6802245814408"
```

```
print(paste("rmse=", sqrt(mse_knn)))
```

```
## [1] "rmse= 3.83147811966097"
```

The correlation is higher than if we used all 10 predictors (0.265). Perhaps the 10th predictor somehow reduced the accuracy.

```
train_df <- data.frame(train_pc$PC1, train_pc$PC2, train_pc$PC3, train_pc$PC4, train_pc$PC5, train_pc$PC6, train_pc$PC7, train_pc$PC8, train_pc$PC9, train$popularity)
test_df <- data.frame(test_pc$PC1, test_pc$PC2, test_pc$PC3, test_pc$PC4, test_pc$PC5, test_pc$PC6, test_pc$PC7, test_pc$PC8, test_pc$PC9, test$popularity)
```

LDA

```
library(MASS)
lda1 <- lda(popularity~danceability+energy+loudness+speechiness+acousticness+instrumentalness+liveness+valence+tempo+duration_ms, data=train)
lda1$means
```

```
##    danceability    energy   loudness speechiness acousticness instrumentalness
## 0     0.5376149 0.5647934 -12.003471   0.1325391    0.3786337       0.24211663
## 1     0.5485832 0.5296254 -12.924138   0.1514061    0.4010062       0.26909728
## 2     0.5640532 0.5141037 -11.886859   0.1272897    0.3752843       0.26188599
## 3     0.5922725 0.5404041 -10.477344   0.1156091    0.3369195       0.24546405
## 4     0.6114753 0.5930456  -9.449223   0.1367113    0.2511059       0.17052066
## 5     0.6026176 0.5844542  -9.916807   0.1265592    0.2997897       0.22919177
## 6     0.6035030 0.6193928  -9.018042   0.1756515    0.2717404       0.12123045
## 7     0.6158426 0.5805222  -9.675148   0.1343583    0.2947714       0.12389855
## 8     0.6156383 0.5963085  -8.901362   0.1497957    0.2781445       0.09477849
## 9     0.6137271 0.6129687  -9.355625   0.1628292    0.2822854       0.16735240
## 10    0.5744000 0.6464000  -8.292092   0.1256985    0.2248053       0.10474583
## 11    0.5757895 0.5864447  -9.385526   0.1012921    0.3390473       0.11698398
## 12    0.6323958 0.5856875  -9.188542   0.1245000    0.2647504       0.12134127
## 13    0.6547783 0.5725542  -9.625670   0.1840335    0.2688670       0.12385511
## 14    0.6562659 0.5858688  -9.134251   0.1655195    0.2667917       0.21416937
## 15    0.6564500 0.5753325  -9.409308   0.1814700    0.2671372       0.18241168
## 16    0.5686000 0.6175333  -8.797400   0.1220733    0.3937238       0.12521260
## 17    0.7736667 0.6146667  -9.424333   0.2193333    0.1489367       0.00003900
## 18    0.7770000 0.6310000  -7.489500   0.1168500    0.1445000       0.00017276
##     liveness   valence    tempo duration_ms
## 0  0.2368147 0.4484339 116.2031    205295.3
## 1  0.2266787 0.4405126 116.3751    199334.9
## 2  0.2071915 0.4616358 117.9493    205230.9
## 3  0.1852842 0.5037757 117.2805    215876.4
## 4  0.1957052 0.4951926 120.8683    221605.4
## 5  0.2082643 0.4634088 120.9351    214849.3
## 6  0.2108251 0.4920802 124.3874    213216.3
## 7  0.2127713 0.5064500 117.2473    218807.8
## 8  0.2130702 0.4368948 124.1773    217488.1
## 9  0.1906229 0.5048438 121.4240    198132.6
## 10 0.2063554 0.4455940 128.1820    228663.5
## 11 0.2011289 0.4847158 130.8187    222485.8
## 12 0.1711125 0.5495625 119.6255    206596.8
## 13 0.1829537 0.4612507 123.3421    192594.8
## 14 0.1808241 0.4884569 120.4090    189865.8
## 15 0.2095608 0.4897842 120.5655    174495.2
## 16 0.2148933 0.3472827 102.9085    185938.0
## 17 0.1796667 0.4454667  94.8550    176170.7
## 18 0.1596750 0.4805000 112.6128    167738.5
```

predict on test

```
lda_pred <- predict(lda1, newdata=test)
lda_pred$class
```

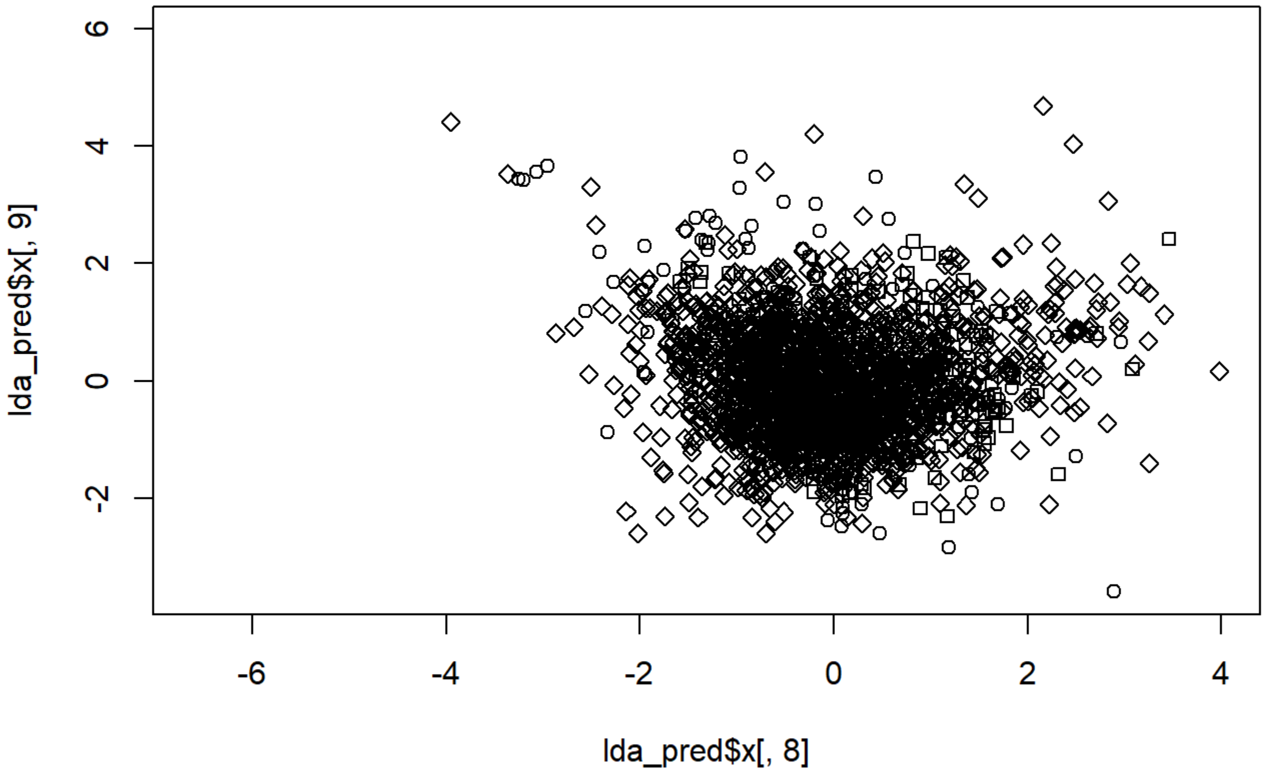
```
##    [1] 0  2  0  3  0  0  0  0  0  0  0  0  0  0  0  3  2  2  0  2  0  0  0  0
##   [25] 0  0  0  0  2  0  2  0  0  0  0  0  2  0  0  2  0  0  2  3  2  2  2  2
##   [49] 2  0  0  0  0  2  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  2  3  0
##   [73] 0  0  1  1  0  0  0  2  2  2  2  2  1  2  2  1  0  0  2  2  0  0  0  0
##   [97] 0  0  2  1  2  2  2  2  1  2  2  2  1  1  2  2  2  2  1  1  1  1  0  3
##  [121] 0  0  2  2  0  2  2  0  2  0  0  0  0  0  0  0  0  0  2  2  2  2  2  3
##  [145] 1  1  1  1  1  2  0  2  2  0  0  0  3  0  1  0  0  0  0  0  0  0  0  0
##  [169] 0  0  0  0  0  0  0  0  0  0  0  0  1  0  2  0  0  0  0  1  0  0  0  0
##  [193] 0  0  0  2  2  2  0  0  0  0  0  0  2  0  0  0  0  0  1  1  1  0  2  0
##  [217] 0  0  0  0  2  2  0  0  0  0  0  0  0  2  2  1  0  0  0  0  0  0  0  2
##  [241] 0  3  0  3  0  1  0  0  0  0  0  0  0  0  1  0  1  1  0  2  3  1  2  1
##  [265] 1  2  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  1
##  [289] 1  1  0  2  0  0  0  0  0  0  3  0  1  1  2  1  1  0  0  0  0  0  0  0
##  [313] 0  0  0  0  0  0  1  0  0  0  1  0  0  2  0  0  1  0  1  0  0  0  2  0
##  [337] 0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  2  0  0  0  0  2  0  0  0
##  [361] 0  3  14 2  1  0  0  0  2  1  0  0  0  2  0  0  2  0  3  0  0  0  0  0
##  [385] 0  0  2  0  1  0  0  0  0  0  0  2  2  2  2  2  2  2  0  3  0  0  0  2
##  [409] 1  1  1  1  2  2  2  0  0  0  0  0  0  0  0  2  0  0  2  0  0  0  0  2
##  [433] 0  0  0  1  1  1  1  0  2  1  0  1  0  0  0  2  2  2  2  2  2  1  0  0
##  [457] 0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  1  2  0  0  0
##  [481] 2  0  0  2  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [505] 0  0  1  2  2  3  0  2  2  0  1  1  0  2  0  0  2  2  0  0  0  0  0  0
##  [529] 0  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2  0  2  0  0  0
##  [553] 0  0  0  0  0  0  0  3  0  0  0  0  0  0  0  2  0  0  1  0  0  0  0  0
##  [577] 0  0  2  0  2  2  0  2  1  2  2  0  0  0  2  2  2  0  0  0  0  0  0  0
##  [601] 0  2  0  0  0  0  0  2  0  0  0  0  0  0  0  2  1  0  0  0  0  0  0  0
##  [625] 0  1  0  0  0  0  0  2  0  0  0  0  0  0  0  2  0  2  2  2  0  0  0  0
##  [649] 0  0  2  0  0  0  0  0  2  0  0  0  0  0  1  0  2  0  3  0  1  0  0  0
##  [673] 0  0  2  1  2  2  2  0  2  0  0  0  0  1  2  0  2  1  0  0  0  0  0  2
##  [697] 0  0  2  0  0  1  2  2  0  2  0  0  1  0  0  0  0  0  0  0  0  0  2  2
##  [721] 2  1  2  2  3  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0
##  [745] 1  1  0  2  1  2  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  0
##  [769] 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  1  2  1  1  0  0  0
##  [793] 0  2  0  0  0  0  0  0  2  2  2  2  2  0  2  0  0  0  0  0  0  2  0  0
##  [817] 0  0  0  3  1  1  0  0  0  0  0  0  0  2  0  0  0  0  2  1  0  0  0  0
##  [841] 0  0  0  0  0  2  0  1  0  0  2  1  1  1  0  0  0  0  2  0  0  0  0  0
##  [865] 0  1  2  0  0  2  2  1  2  2  0  0  1  0  0  0  2  2  0  0  0  0  1  1
##  [889] 2  2  2  0  0  2  0  0  0  0  0  0  0  2  0  0  0  0  0  0  0  0  0  2
##  [913] 0  0  0  0  0  0  0  0  0  0  0  2  2  0  0  2  0  0  1  1  0  0  0  0
##  [937] 2  0  2  2  2  0  0  0  0  0  0  1  0  1  0  0  0  2  2  2  2  0  0  0
##  [961] 0  0  0  0  2  0  0  0  0  0  0  0  0  0  0  0  2  2  2  0  0  1  1  2
##  [985] 0  0  0  0  0  2  0  2  2  1  0  0  0  2  0  1  1  0  0  2  2  0  0  0
## [1009] 0  0  0  2  0  2  0  0  2  0  2  2  2  2  0  0  0  0  0  0  2  0  0  0
## [1033] 2  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2  0  0  2  0  0  1
## [1057] 0  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [1081] 0  0  0  0  2  2  2  2  2  1  0  0  2  0  0  0  0  1  0  0  0  2  2  0
## [1105] 1  2  1  1  1  1  1  2  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0
## [1129] 0  0  0  0  0  2  0  0  0  2  0  0  2  0  1  0  1  0  0  0  2  0  0  0
## [1153] 0  0  0  0  0  0  2  0  0  0  0  3  3  2  0  2  2  0  0  0  2  0  0  0
## [1177] 0  0  2  0  0  0  0  0  0  0  0  0  0  0  0  2  2  0  2  0  0  0  0  0
## [1201] 1  2  0  0  0  0  0  2  0  0  0  0  0  2  1  2  0  0  0  2  0  2  2  2
## [1225] 2  2  2  2  2  0  2  2  0  0  0  0  0  0  1  2  0  2  2  0  0  3  3  0
## [1249] 0  2  2  0  0  0  2  0  2  0  0  0  0  3  0  0  0  2  0  0  0  0  2  2 
## [1273] 0  2  0  0  2  0  0  0  0  3  2  0  0  0  0  0  0  0  0  0  0  2  3  2 
## [1297] 0  2  2  0  0  2  2  2  0  0  0  2  2  0  0  0  0  2  0  1  0  0  0  0 
## [1321] 0  0  0  0  0  0  0  0  0  0  2  0  0  0  2  2  3  0  0  0  0  0  2  2 
## [1345] 0  1  0  0  0  0  0  0  0  2  2  0  1  1  2  2  0  0  0  0  0  0  0  0 
## [1369] 0  0  2  0  0  1  2  2  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0 
## [1393] 0  0  0  0  2  0  0  3  0  0  0  0  0  0  14 2  0  2  2  2  0  0  0  2 
## [1417] 0  0  0  0  0  0  0  0  0  0  0  0  0  1  1  1  0  0  0  0  0  0  0  2 
## [1441] 2  2  0  2  0  0  0  0  0  0  0  0  0  0  0  0  2  0  1  0  0  0  0  0 
## [1465] 0  0  2  0  0  0  0  0  2  2  0  0  0  1  0  0  0  0  0  0  0  0  2  0 
## [1489] 2  0  0  1  0  1  0  1  0  0  1  0  0  0  0  2  0  0  0  0  0  0  0  0 
## [1513] 0  1  1  1  1  1  2  2  0  2  14 2  2  3  2  0  2  2  3  2  2  0  0  0 
## [1537] 1  0  0  2  2  0  0  0  0  0  0  0  0  0  2  0  2  0  0  2  0  0  0  0 
## [1561] 0  0  0  0  2  2  0  0  0  0  1  1  0  0  2  0  1  1  0  1  2  2  0  0 
## [1585] 0  0  0  0  0  0  0  0  2  0  0  0  0  0  2  2  0  0  2  0  0  0  0  0 
## [1609] 0  0  0  2  1  2  2  1  0  0  0  0  2  2  0  0  0  0  1  0  0  0  0  2 
## [1633] 0  0  0  0  0  2  2  2  0  0  0  0  0  0  1  1  0  0  0  1  0  0  0  1 
## [1657] 2  2  0  0  0  2  2  3  2  2  1  1  1  1  1  1  0  0  0  0  1  0  0  2 
## [1681] 1  1  0  1  0  3  2  1  0  2  1  1  0  1  1  0  0  0  0  0  0  0  0  0 
## [1705] 0  0  2  0  0  0  1  1  1  2  0  0  0  0  0  0  0  0  0  3  0  0  0  2 
## [1729] 0  2  0  0  0  0  0  0  0  2  0  0  0  0  1  1  1  2  2  0  2  2  0  1 
## [1753] 0  0  0  0  0  1  1  1  1  1  1  1  1  1  1  1  1  0  2  2  0  0  0  0 
## [1777] 0  2  2  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2  0  2  0  0 
## [1801] 0  2  0  2  2  0  0  0  0  0  0  2  0  1  0  0  0  1  0  2  0  0  0  2 
## [1825] 0  2  0  0  0  0  0  0  0  0  0  0  1  3  2  2  2  0  0  0  0  0  0  0 
## [1849] 0  3  0  0  1  0  1  0  0  0  0  0  0  0  0  0  1  1  0  1  0  0  2  0 
## [1873] 0  0  0  0  0  0  0  0  2  0  0  0  2  0  1  0  2  0  0  0  0  0  0  1 
## [1897] 2  0  2  2  0  1  0  0  0  0  0  0  0  0  1  1  0  0  0  0  0  0  3  3 
## [1921] 2  2  0  0  0  2  0  2  0  2  2  2  2  0  0  0  0  0  1  0  0  0  0  0 
## [1945] 0  0  0  0  0  0  0  0  0  0  0  2  0  2  0  2  0  0  0  2  2  1  0  0 
## [1969] 2  0  2  2  2  2  0  0  0  0  0  0  0  0  0  0  0  0  0  1  2  0  0  0 
## [1993] 0  0  0  0  0  0  2  0  0  0  0  0  0  2  0  1  0  2  1  0  0  0  0  0 
## [2017] 0  0  2  2  0  0  0  0  0  2  0  2  0  0  0  0  0  0  0  0  0  0  2  0 
## [2041] 0  0  0  0  0  0  0  2  0  0  0  0  0  1  0  2  0  2  2  2  0  0  0  0 
## [2065] 0  0  0  0  0  0  0  0  0  0  1  1  0  0  0  0  0  2  0  0  1  2  0  0 
## [2089] 0  0  0  0  1  0  1  1  0  0  1  0  1  0  0  0  0  0  1  0  0  0  1  0 
## [2113] 1  0  0  0  0  2  2  2  2  0  0  0  0  0  0  0  0  0  2  0  2  1  0  0 
## [2137] 2  0  0  2  0  0  0  0  0  0  2  0  2  2  0  0  0  0  0  0  0  0  0  0 
## [2161] 0  2  2  1  0  1  1  0  1  0  0  0  0  0  0  0 
## Levels: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
```

```
mean(lda_pred$class==test$popularity)
```

```
## [1] 0.245864
```

plot

```
plot(lda_pred$x[,8], lda_pred$x[,9], pch=c(23,21,22)[unclass(lda_pred$class)], bg=c("red","gree
n","blue")[unclass(test_pc$popularity)])
```


6
4
2
0
-2
lda_pred$x[, 9]
-6
-4
-2
0
2
4
lda_pred$x[, 8]