36-315 Final Project PCA

Liz Chu

December 11, 2022

Contents

```
zipinfo <- read.csv(file = "zipout.csv")
zipinfo = na.omit(zipinfo)
zipinfo.subset = subset(zipinfo, select = c(n_rats, n_rest, score_avg, population, tax_rating))
zipinfo.subset = apply(zipinfo.subset, MARGIN = 2, FUN = function(x) x - mean(x))
zipinfo.subset = apply(zipinfo.subset, MARGIN = 2, FUN = function(x) x/sd(x))
zipinfo.pca = prcomp(zipinfo.subset)
summary(zipinfo.pca)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5
## Standard deviation     1.4621 1.1530 0.8984 0.67644 0.51781
## Proportion of Variance 0.4276 0.2659 0.1614 0.09152 0.05363
## Cumulative Proportion  0.4276 0.6934 0.8549 0.94637 1.00000
```

```
X = as.matrix(zipinfo.subset)
prin_comps = as.data.frame(X %*% zipinfo.pca$rotation)
prin_comps$borough = zipinfo$borough


library(devtools)
```

```
## Loading required package: usethis
```

```
library(ggbiplot)
```

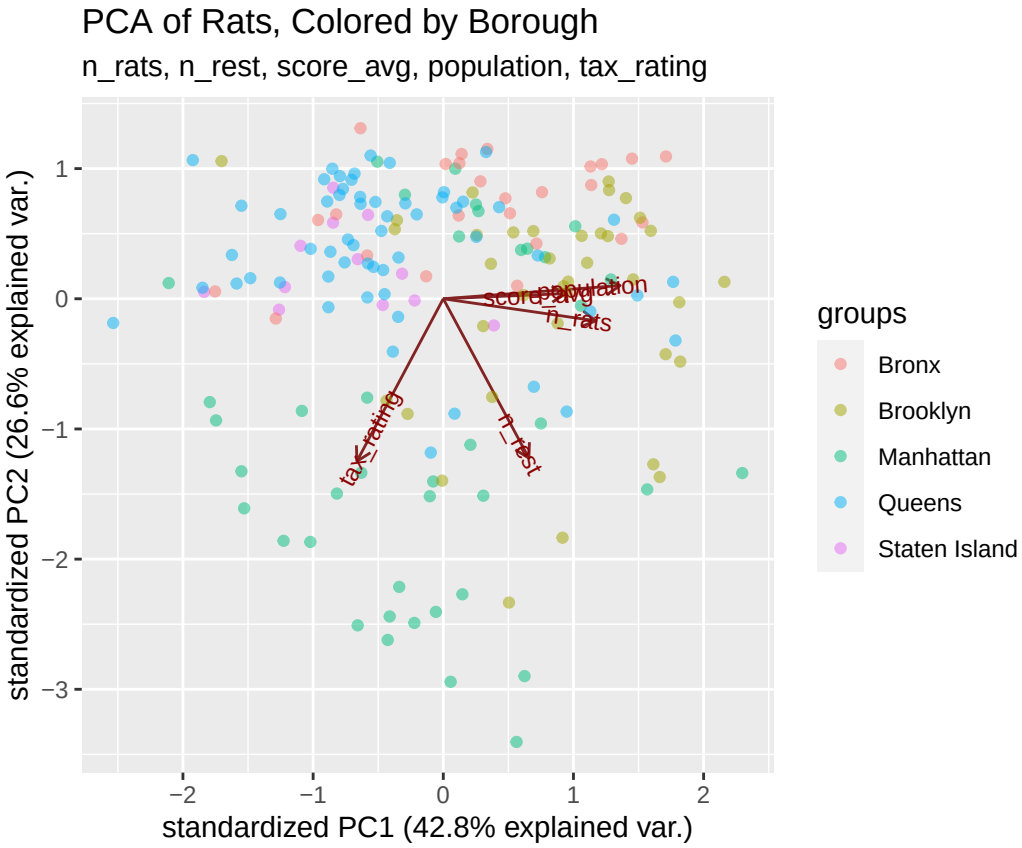
```
## Loading required package: ggplot2

## Loading required package: plyr

## Loading required package: scales

## Loading required package: grid
```

```
ggbiplot(zipinfo.pca, alpha = 0.5, groups = zipinfo$borough, varname.adjust=-0.5) + labs(
  title = "PCA of Rats, Colored by Borough",
  subtitle = "n_rats, n_rest, score_avg, population, tax_rating"
)
```


PCA of Rats, Colored by Borough
n_rats, n_rest, score_avg, population, tax_rating
standardized PC2 (26.6% explained var.)
standardized PC1 (42.8% explained var.)
1
0
−1
−2
−3
−2
−1
0
1
2
tax_rating
n_rest
score_avg
population
n_rats
groups
Bronx
Brooklyn
Manhattan
Queens
Staten Island