Project 5 Part 1

Kaushal Khatiwada

2024-05-13

Use NCI60 data of ISLR2 package

```
library(ISLR2)
```

```
## Warning: package 'ISLR2' was built under R version 4.3.3
```

A) Define nci labels (NCI$labs$) as nci.labs and ncidata (NCIdata) and nic.data

```
nci.labs <- NCI60$labs
nci.data <- NCI60$data
```

B) Check dimension of nci.data object and interpret it carefully

```
dim(nci.data)   # 64 rows 6830 columns
```

```
## [1]   64 6830
```

C) Check first four cancer types using nci.labs object

```
head(nci.labs,4)
```

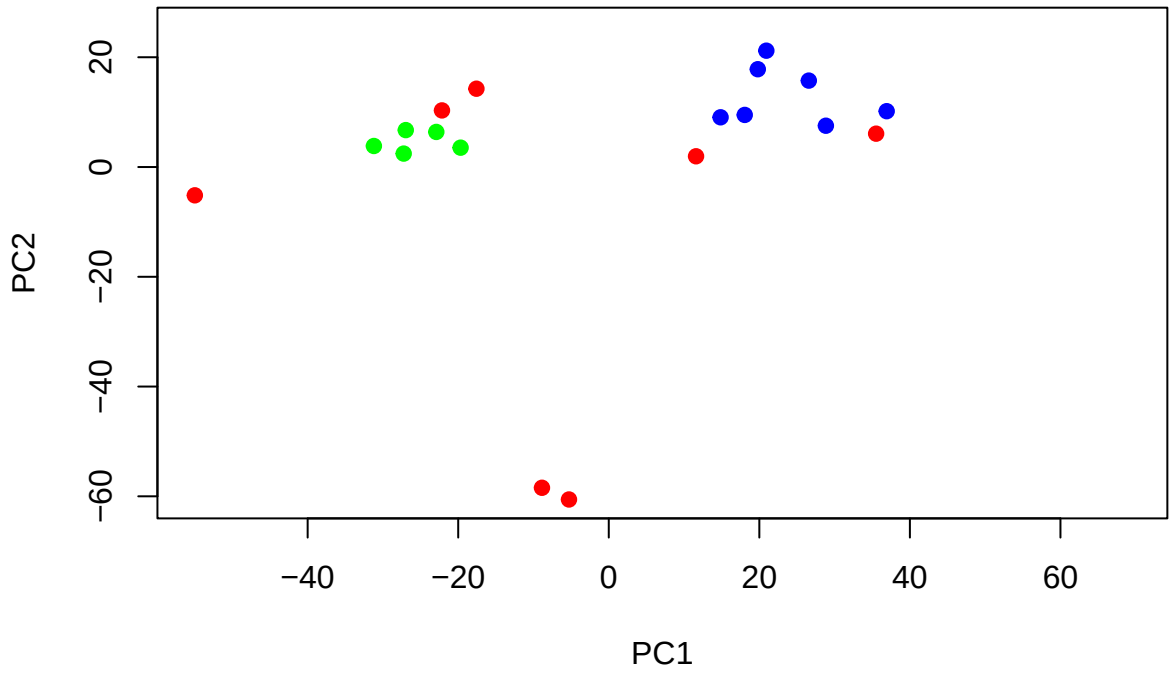
```
## [1] "CNS"   "CNS"   "CNS"   "RENAL"
```

D) Fit principal component analysis (PCA) on nci.data with scale = TRUE argument as pr.out object

```
pr.out <- prcomp(nci.data,scale=T)
```

E) Create a plot showing first three PCA components with three different colors

```r
plot(pr.out$x[, 1:3],
     col = c("red", "green", "blue")[factor(nci.labs)],
     pch = 19,
     xlab = "PC1",
     ylab = "PC2")
```



F) Get summary of pr.out object and interpret it carefully

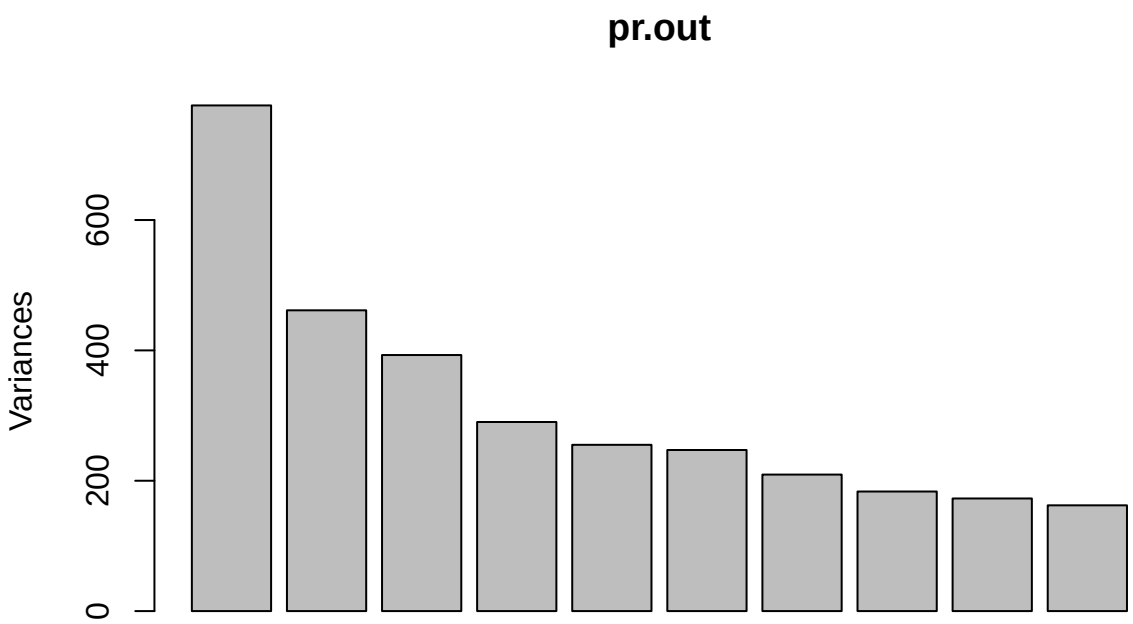
```r
summary(pr.out)
```

```
## Importance of components:
##                             PC1      PC2      PC3      PC4      PC5      PC6
## Standard deviation     27.8535 21.48136 19.82046 17.03256 15.97181 15.72108
## Proportion of Variance  0.1136  0.06756  0.05752  0.04248  0.03735  0.03619
## Cumulative Proportion   0.1136  0.18115  0.23867  0.28115  0.31850  0.35468
##                             PC7      PC8      PC9     PC10     PC11     PC12
## Standard deviation     14.47145 13.54427 13.14400 12.73860 12.68672 12.15769
## Proportion of Variance  0.03066  0.02686  0.02529  0.02376  0.02357  0.02164
## Cumulative Proportion   0.38534  0.41220  0.43750  0.46126  0.48482  0.50646
##                            PC13     PC14     PC15     PC16     PC17     PC18
## Standard deviation     11.83019 11.62554 11.43779 11.00051 10.65666 10.48880
## Proportion of Variance  0.02049  0.01979  0.01915  0.01772  0.01663  0.01611
```

```
## Cumulative Proportion   0.52695  0.54674  0.56590  0.58361  0.60024  0.61635
##                            PC19    PC20     PC21    PC22    PC23    PC24
## Standard deviation     10.43518 10.3219 10.14608 10.0544 9.90265 9.64766
## Proportion of Variance  0.01594  0.0156  0.01507  0.0148 0.01436 0.01363
## Cumulative Proportion   0.63229  0.6479  0.66296  0.6778 0.69212 0.70575
##                            PC25    PC26    PC27   PC28    PC29    PC30    PC31
## Standard deviation     9.50764 9.33253 9.27320 9.0900 8.98117 8.75003 8.59962
## Proportion of Variance 0.01324 0.01275 0.01259 0.0121 0.01181 0.01121 0.01083
## Cumulative Proportion  0.71899 0.73174 0.74433 0.7564 0.76824 0.77945 0.79027
##                            PC32    PC33    PC34    PC35    PC36    PC37    PC38
## Standard deviation     8.44738 8.37305 8.21579 8.15731 7.97465 7.90446 7.82127
## Proportion of Variance 0.01045 0.01026 0.00988 0.00974 0.00931 0.00915 0.00896
## Cumulative Proportion  0.80072 0.81099 0.82087 0.83061 0.83992 0.84907 0.85803
##                            PC39    PC40    PC41   PC42    PC43   PC44    PC45
## Standard deviation     7.72156 7.58603 7.45619 7.3444 7.10449 7.0131 6.95839
## Proportion of Variance 0.00873 0.00843 0.00814 0.0079 0.00739 0.0072 0.00709
## Cumulative Proportion  0.86676 0.87518 0.88332 0.8912 0.89861 0.9058 0.91290
##                          PC46    PC47    PC48    PC49    PC50    PC51    PC52
## Standard deviation     6.8663 6.80744 6.64763 6.61607 6.40793 6.21984 6.20326
## Proportion of Variance 0.0069 0.00678 0.00647 0.00641 0.00601 0.00566 0.00563
## Cumulative Proportion  0.9198 0.92659 0.93306 0.93947 0.94548 0.95114 0.95678
##                            PC53    PC54    PC55    PC56    PC57   PC58    PC59
## Standard deviation     6.06706 5.91805 5.91233 5.73539 5.47261 5.2921 5.02117
## Proportion of Variance 0.00539 0.00513 0.00512 0.00482 0.00438 0.0041 0.00369
## Cumulative Proportion  0.96216 0.96729 0.97241 0.97723 0.98161 0.9857 0.98940
##                            PC60    PC61    PC62    PC63      PC64
## Standard deviation     4.68398 4.17567 4.08212 4.04124 1.951e-14
## Proportion of Variance 0.00321 0.00255 0.00244 0.00239 0.000e+00
## Cumulative Proportion  0.99262 0.99517 0.99761 1.00000 1.000e+00
```

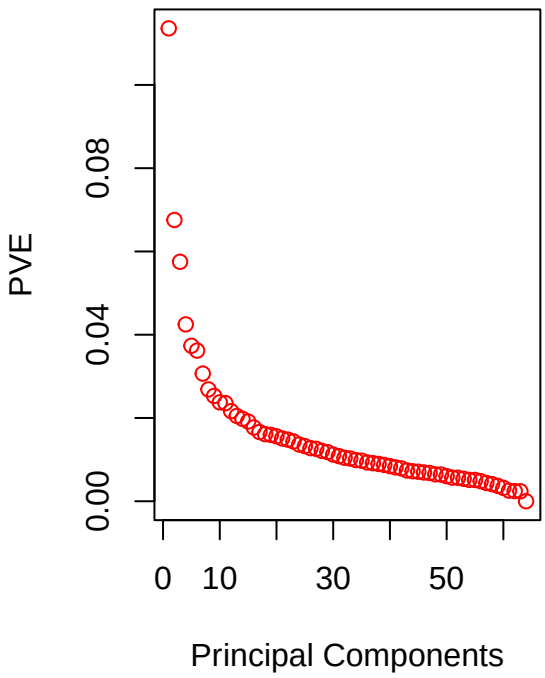
G Plot pro.out object and interpret it carefully

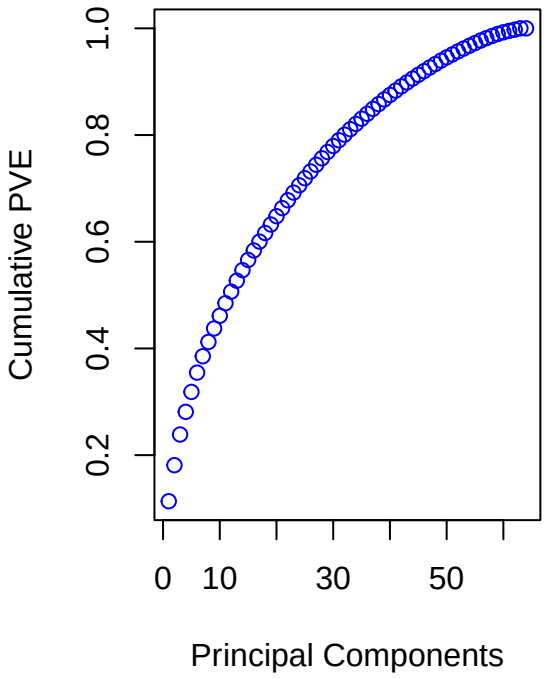
```
plot(pr.out)
```



H) Create custom scatterplots with principal components in x-axis and proportion variance explained (PVE) in y-axis for the first plot and cumulative PVE in the y-axis for the second plot and interpret them carefully

```
par(mfrow = c(1, 2))
var_expalined <- pr.out$sdev^2 / sum(pr.out$sdev^2)
plot(var_expalined,
     xlab = "Principal Components",
     ylab = "PVE",
     col="red")
plot(cumsum(var_expalined),
     xlab = "Principal Components",
     ylab = "Cumulative PVE",
     col="blue")
```





I) Perform PCA with varimax rotation and compare it with the PCA result obtained above

```
library(FactoMineR)
```

```
## Warning: package 'FactoMineR' was built under R version 4.3.3
```

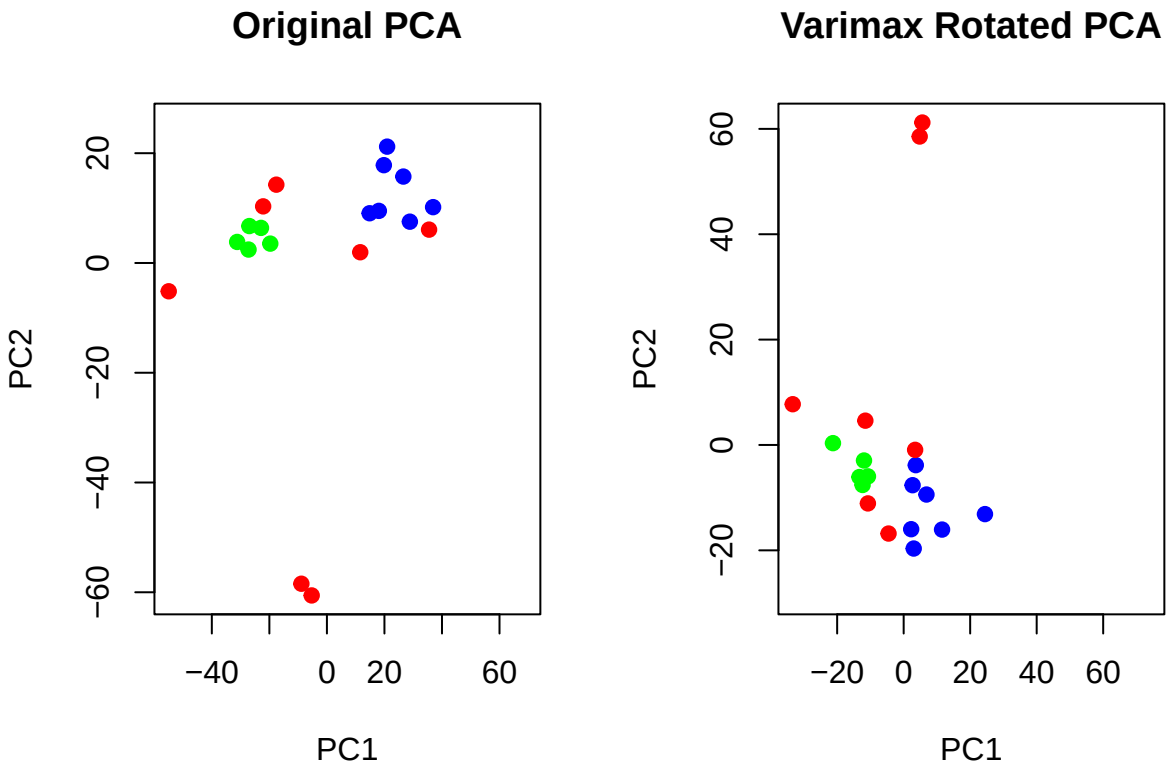
```
pca.1 <- PCA(nci.data, scale = TRUE, ncp = 3, graph = FALSE)
varimax_rot <- varimax(pca.1$ind$coord[, 1:3])
```

Compare with the PCA result obtained above

```
par(mfrow = c(1, 2))
plot(pr.out$x[, 1:3],
     col = c("red", "green", "blue")[factor(nci.labs)],
     pch = 19, xlab = "PC1",
     ylab = "PC2",
     main = "Original PCA")

plot(varimax_rot$loadings,
     col = c("red", "green", "blue")[factor(nci.labs)],
```

```
    pch = 19,
    xlab = "PC1",
    ylab = "PC2",
    main = "Varimax Rotated PCA")
```



J) Write summary of the results and conclusion based on your findings