```
##############################################
#####    PRINCIPAL COMPONENTS ANALYSIS    #####
#####        EXERCISE AND SOLUTION        #####
##############################################

# Find the principal components of the following
# correlation matrix given by MacDonnell (1902)
# from measurements of seven physical charac-
# teristics in each of 3000 convicted criminals.

# How would you interpret the derived components?

# Head length           1.000
# Head breadth          0.402 1.000
# Face breadth          0.396 0.618 1.000
# Left finger length    0.301 0.150 0.321 1.000
# Left forearm length   0.305 0.135 0.289 0.846 1.000
# Left foot length      0.339 0.206 0.363 0.759 0.797 1.000
# Height                0.340 0.183 0.345 0.661 0.800 0.736 1.000

# read names of variables into vector
crim.var.labels <- c("Head-L","Head-B","Face-B",
                     "L-Fing","L-Fore","L-Foot",
                     "Height")

# read in the correlation data as a vector
cc <- c(
  1.000, 0.402, 0.396, 0.301, 0.305, 0.339, 0.340,
  0.402, 1.000, 0.618, 0.150, 0.135, 0.206, 0.183,
  0.396, 0.618, 1.000, 0.321, 0.289, 0.363, 0.345,
  0.301, 0.150, 0.321, 1.000, 0.846, 0.759, 0.661,
  0.305, 0.135, 0.289, 0.846, 1.000, 0.797, 0.800,
  0.339, 0.206, 0.363, 0.759, 0.797, 1.000, 0.736,
  0.340, 0.183, 0.345, 0.661, 0.800, 0.736, 1.000)

# put it in a matrix form
crimcorr <- matrix(cc, nrow = 7, byrow = TRUE)
crimcorr

# set row, col names to be variable names
rownames(crimcorr) <- crim.var.labels
colnames(crimcorr) <- crim.var.labels

# look at crimcorr again
crimcorr

# run the PCA on the correlation matrix w/ princomp()
# note that scores are true by default anyway
crm_pca <- princomp(crimcorr,
                    cor=TRUE,
```

```
                    scores=TRUE)

print(crm_pca)

# empty cells are close to zero
# in value
summary(crm_pca, loadings=TRUE)

# use prcomp() function
# get same results
crim_pca <- prcomp(crimcorr, scale = TRUE)
# look at results
print(crim_pca)
summary(crim_pca)

# linear combination for first PC is
a1 <- crim_pca$rotation[,1]
a1

# we extract the first PC from all pre-
# computed principal components:
predict(crim_pca)[,1]

# Seems clear that Head Breadth and Face
# Breadth have the most weight in computing
# criminal's individual scores on the first
# principal component which accounts for
# 75% of the variance in the correlation
# matrix

# The first two components account for 90%
# of the variance. A barplot of each compo-
# nent's variance shows how the first, and,
# to a lesser extent, the first two
# components dominate:
plot(crim_pca, main = "")

# seems that head size dimensions have
# the strongest relationships "explaining"
# criminal behavior (convictions).

# we call a biplot
biplot(crim_pca,
       col = c("gray", "black"))


# Biplot confirms conclusions: Head-Breadth
# and Face-Breadth dominate the variables in
# explaining criminal behavior (convictions)
```