

Question 8:

a) Before doing anything, I scaled the data using the scale function. I followed the book's code to produce the percent of variance explained for the scaled data.

```
data = scale(data)

# a)
pr = prcomp(data)
pr.var = pr$sdev^2
pve = pr.var / sum(pr.var)
```

b) According to Equation 12.10, our PVE can be expressed by using the component scores (which is the $\phi_i \cdot x_i$ term) squared, divided by the sum of all the data squared. I did this, and received the

```
> pve
[1] 0.62006039 0.24744129 0.08914080 0.04335752
```

same PVE as the book.

```
# b)
scores = pr$x
lis = rep(0,4)
for (i in 1:4) {
  lis[i] = sum(scores[, i]^2) / sum(data^2)
}
print(lis)
```

Question 10

a) I wrote a long for loop to create some clusters of data. They are very separated, which should bode well for PCA and K-means clustering when we get there on Wednesday.

b) I could not get the plot to color the different clusters, however here is the plot, with the 3 clear distinct clusters. You can't see the center cluster super well since the directions are covering it, but the data points are there.

