PCA Codes

```r
library(readr)
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 4.1.3
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 4.1.3
```

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

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following objects are masked from 'package:plyr':
##
##      arrange, mutate, rename, summarise
```

```
## The following object is masked from 'package:stats':
##
##      filter
```

```
## The following object is masked from 'package:graphics':
##
##      layout
```

```
library(readr)
library(ggplot2)
library(ggbiplot)
```

```
## Loading required package: scales
```

```
## Warning: package 'scales' was built under R version 4.1.3
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:readr':
##
##      col_factor
```

```
## Loading required package: grid
```

```
dataset = read_csv("Customer Data.csv")
```

```
## Rows: 8950 Columns: 18
## -- Column specification ------------------------------------------------
## Delimiter: ","
## chr  (1): CUST_ID
## dbl (17): BALANCE, BALANCE_FREQUENCY, PURCHASES, ONEOFF_PURCHASES, INSTALLME...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
dataset_omitted = na.omit(dataset)
dataset_reduced = dataset_omitted[,-1]


#PCA Model of the data set
pca_model <- prcomp(dataset_reduced, scale=TRUE)

#Scree Plot
pve= 100 * pca_model$sdev^2 / sum(pca_model$sdev^2)

par(mfrow = c(1, 2))
plot(pve,  type = "o", ylab = "PVE",
     xlab = "Principal Component", col = "blue")

plot(cumsum(pve), type = "o", ylab = "Cumulative PVE",
     xlab = "Principal Component", col = "brown3")
```
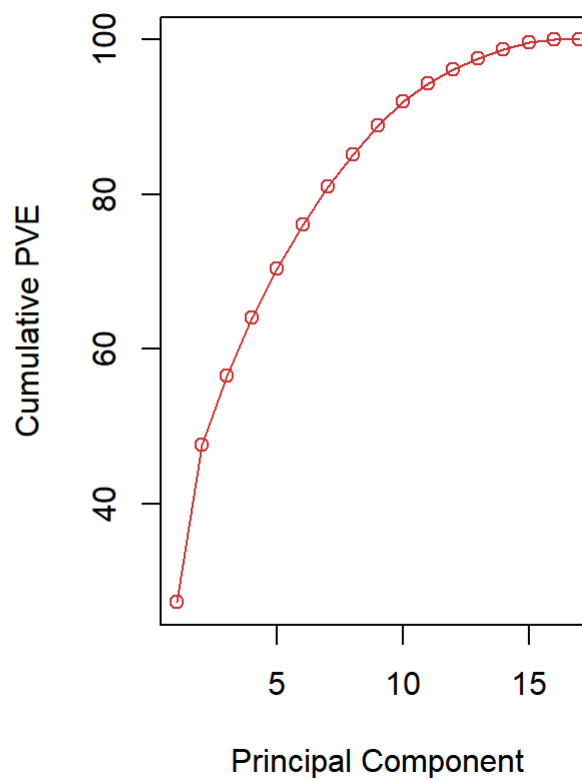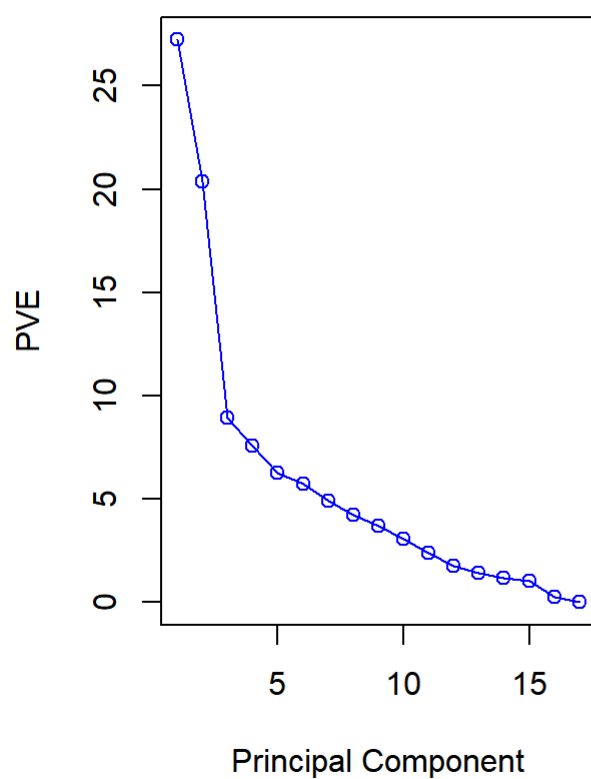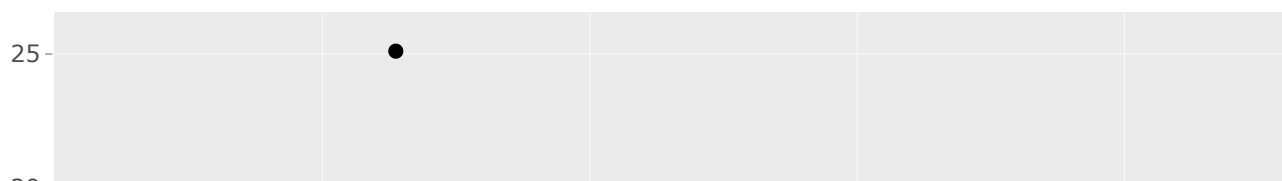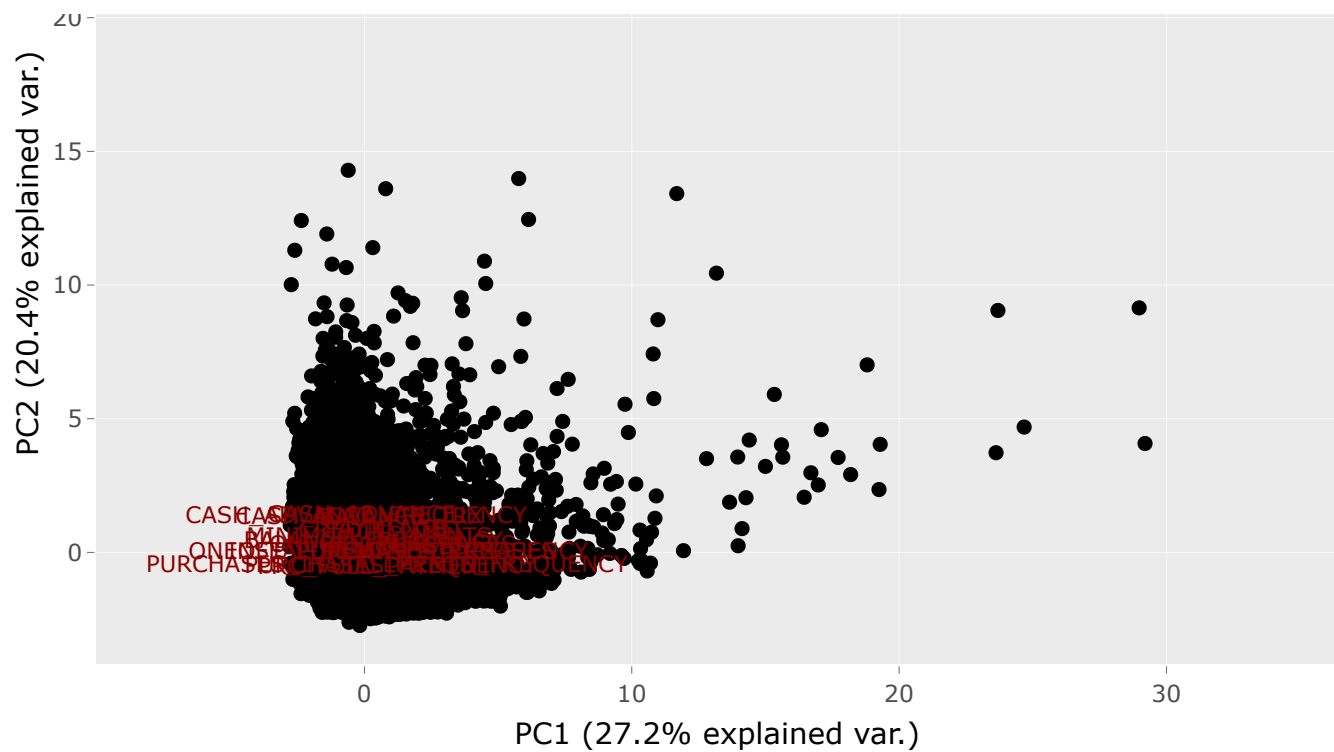


```
ggplotly(ggbiplot(pca_model,scale = 0,labels=rownames(pca_model$x)))
```
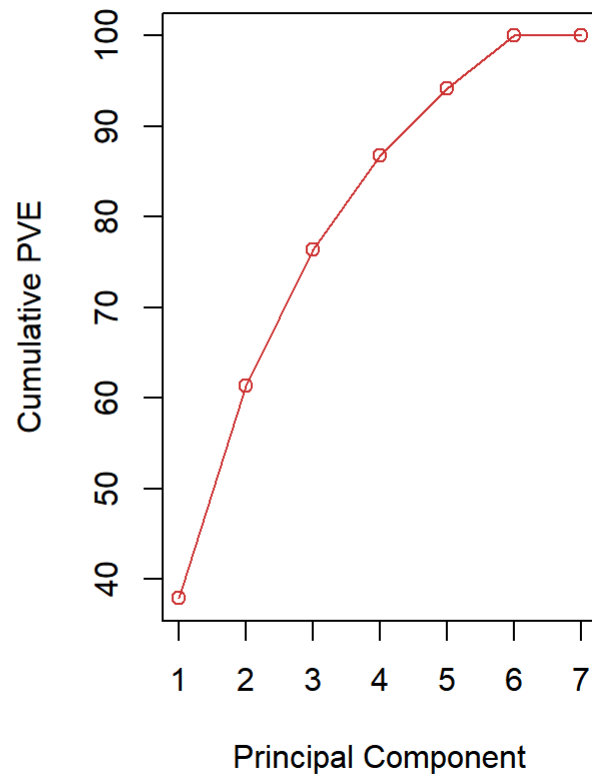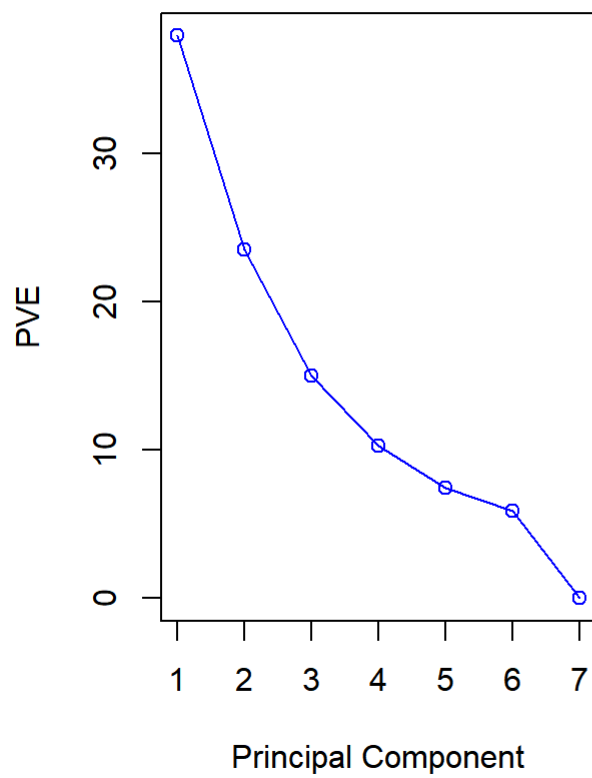
```
dataset_reduced_new = dataset_omitted[,c(2,3,4,5,6,7,8)]


pca_model_new <- prcomp(dataset_reduced_new, scale=TRUE)

#Scree Plot
pve_new = 100 * pca_model_new$sdev^2 / sum(pca_model_new$sdev^2)

par(mfrow = c(1, 2))
plot(pve_new,  type = "o", ylab = "PVE",
    xlab = "Principal Component", col = "blue")

plot(cumsum(pve_new), type = "o", ylab = "Cumulative PVE",
    xlab = "Principal Component", col = "brown3")
```

```
ggplotly(ggbiplot(pca_model_new,scale = 0,labels=rownames(pca_model_new$x)))
```

0   10   20   30

PC1 (37.9% explained var.)