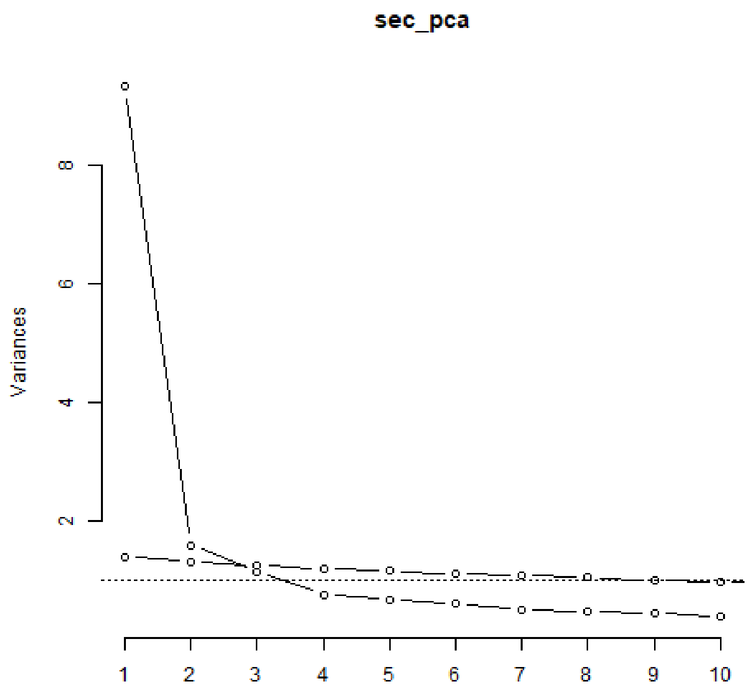
Due: 2022/05/22
Special thanks to 108071001

Question 1. (a) I make the scree plot by the following code:

```r
# Question 1 (a)
security <- read.csv("security_questions.csv") # read the dataset
sim_noise_ev <- function(n, p) # generate random noise
{
  noise <- data.frame(replicate(p, rnorm(n)))
  eigen(cor(noise))$values
}
evalues_noise <- replicate(100, sim_noise_ev(405, 18)) # The same size
evalues_mean <- apply(evalues_noise, 1, mean)
sec_pca <- prcomp(security, scale. = TRUE) # Apply PCA on dataset

png(filename = "1a.png")
screeplot(sec_pca, type="lines")
lines(evalues_mean, type="b")
abline(h=1, lty="dotted")
dev.off()
```



(b) By the plot, picking 2 PCs are enough.

■

Question 2. (a) The loadings of the first 3 principal components are written to a `.csv` file.

```r
# Question 2 (a)
sec_principal <- principal(security, nfactor=10, rotate="none", scores=TRUE)
sec_pc1 <- sec_principal$loadings[,"PC1"]
sec_pc2 <- sec_principal$loadings[,"PC2"]
sec_pc3 <- sec_principal$loadings[,"PC3"]
first_3_pc <- cbind(sec_pc1, sec_pc2, sec_pc3)
```

```r
names(first_3_pc) <- names(sec_principal)
write.table(first_3_pc, file="2a.csv", sep = ",", col.names=NA)
```

Then use some online LaTeXtool, we have

	sec_pc1	sec_pc2	sec_pc3
Q1	0.817	-0.139	-0.002
Q2	0.673	-0.014	0.089
Q3	0.766	-0.033	0.09
Q4	0.623	0.643	0.108
Q5	0.69	-0.031	-0.542
Q6	0.683	-0.105	0.207
Q7	0.657	-0.318	0.324
Q8	0.786	0.042	-0.343
Q9	0.723	-0.232	0.204
Q10	0.686	-0.099	-0.533
Q11	0.753	-0.261	0.173
Q12	0.63	0.638	0.122
Q13	0.712	-0.065	0.084
Q14	0.811	-0.1	0.157
Q15	0.704	0.011	-0.333
Q16	0.758	-0.203	0.183
Q17	0.618	0.664	0.11
Q18	0.807	-0.114	-0.065

- PC1: Q1, Q2, Q3, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q13, Q14, Q15, Q16, Q18.
- PC2: Q4, Q12, Q17.
- PC3: None.

(b) Compute each variance that the first 3 PCs capture by the following code:

```r
var1 <- sum(sec_principal$loadings[,"PC1"]^2)
var2 <- sum(sec_principal$loadings[,"PC2"]^2)
var3 <- sum(sec_principal$loadings[,"PC3"]^2)
total_var <- var1 + var2 + var3
```

We have PC1 captures 9.311, PC2 1.596, and PC3 1.150. Hence, They add up to 12.057.

(c) Use the command

```
> sec_pca_rot
```

to inspect the communality. It looks like most of the items are less than adequately explained. say Q1, Q2, Q3, Q6, Q7, Q9, Q11, Q13, Q14, Q15, Q16, Q18.

```
> sec_pca_rot
Principal Components Analysis
Call: principal(r = security, nfactors = 3, rotate = "varimax",
    scores = TRUE)
Standardized loadings (pattern matrix) based upon correlation matrix
     RC1  RC3  RC2   h2   u2 com
Q1  0.66 0.45 0.22 0.69 0.31 2.0
Q2  0.54 0.29 0.29 0.46 0.54 2.1
Q3  0.62 0.34 0.31 0.60 0.40 2.1
Q4  0.22 0.19 0.85 0.81 0.19 1.2
Q5  0.24 0.83 0.16 0.77 0.23 1.3
Q6  0.65 0.20 0.23 0.52 0.48 1.5
Q7  0.79 0.10 0.06 0.64 0.36 1.0
Q8  0.38 0.71 0.30 0.74 0.26 2.0
Q9  0.74 0.23 0.14 0.62 0.38 1.3
Q10 0.28 0.82 0.10 0.76 0.24 1.3
Q11 0.76 0.28 0.12 0.66 0.34 1.3
Q12 0.23 0.19 0.85 0.82 0.18 1.2
Q13 0.59 0.32 0.26 0.52 0.48 1.9
Q14 0.72 0.31 0.28 0.69 0.31 1.7
Q15 0.34 0.66 0.24 0.61 0.39 1.8
Q16 0.74 0.27 0.17 0.65 0.35 1.4
Q17 0.21 0.19 0.87 0.83 0.17 1.2
Q18 0.61 0.50 0.23 0.67 0.33 2.2
```

(d) Based on the table in (a), among the components whose loading of the PC it belongs to is less than 0.7, given that the loading of other PCs are not closed to 0, I found that Q4, Q7, Q12 and Q17 share similar loadings between 2 or more components.

(e) We may sum up the questions in PC1 as "the site protects my information security".

■

Question 3. (a) Let's inspect the variance directly.

```r
# Question 3 (a)
sec_pca_org <- principal(security, # The original PCs
                         nfactor=3,
                         rotate="none",
                         scores=TRUE)
sec_pca_rot <- principal(security, # rotate the pcs
                         nfactor=3,
                         rotate="varimax",
                         scores=TRUE) # just call these two model

var1_org <- sum(sec_pca_org$loadings[,"PC1"]^2)
var2_org <- sum(sec_pca_org$loadings[,"PC2"]^2)
var3_org <- sum(sec_pca_org$loadings[,"PC3"]^2)
total_var_org <- var1_org + var2_org + var3_org

var1_rot <- sum(sec_pca_rot$loadings[,"RC1"]^2)
var2_rot <- sum(sec_pca_rot$loadings[,"RC2"]^2)
var3_rot <- sum(sec_pca_rot$loadings[,"RC3"]^2)
total_var_rot <- var1_rot + var2_rot + var3_rot
```

It looks like even if the variance each PCs/RCs explain are different, the total variance remains the same.

	PC	RC
1	9.311	5.613
2	1.597	2.954
3	1.150	3.490
total	12.057	12.057

(b) Similar to 2(c), just call `sec_pca_org` and `sec_pca_rot` to inspect the `Cumulative Var`. RC1 and RC2 explain less cumulative variance as PC1 and PC2. However it turns out that PC3 and RC3 share the same.

	PC	RC
1	0.52	0.31
2	0.61	0.51
3	0.67	0.67

(c) Yes, all of them have more clearly differentiated loadings.

```
> sec_pca_rot
Principal Components Analysis
Call: principal(r = security, nfactors = 3, rotate = "varimax",
    scores = TRUE)
Standardized loadings (pattern matrix) based upon correlation matrix
     RC1  RC3  RC2   h2   u2 com
Q1  0.66 0.45 0.22 0.69 0.31 2.0
Q2  0.54 0.29 0.29 0.46 0.54 2.1
Q3  0.62 0.34 0.31 0.60 0.40 2.1
Q4  0.22 0.19 0.85 0.81 0.19 1.2
Q5  0.24 0.83 0.16 0.77 0.23 1.3
Q6  0.65 0.20 0.23 0.52 0.48 1.5
Q7  0.79 0.10 0.06 0.64 0.36 1.0
Q8  0.38 0.71 0.30 0.74 0.26 2.0
Q9  0.74 0.23 0.14 0.62 0.38 1.3
Q10 0.28 0.82 0.10 0.76 0.24 1.3
Q11 0.76 0.28 0.12 0.66 0.34 1.3
Q12 0.23 0.19 0.85 0.82 0.18 1.2
Q13 0.59 0.32 0.26 0.52 0.48 1.9
Q14 0.72 0.31 0.28 0.69 0.31 1.7
Q15 0.34 0.66 0.24 0.61 0.39 1.8
Q16 0.74 0.27 0.17 0.65 0.35 1.4
Q17 0.21 0.19 0.87 0.83 0.17 1.2
Q18 0.61 0.50 0.23 0.67 0.33 2.2
```

(d) The loadings of the first 3 rotated components are written to a `.csv` file.

```r
# Question 3 (d)
rot_pc1 <- sec_pca_rot$loadings[,"RC1"] # Rotated component's loading
```

```
rot_pc2 <- sec_pca_rot$loadings[,"RC2"]
rot_pc3 <- sec_pca_rot$loadings[,"RC3"]
loading_rot <- round(cbind(rot_pc1, rot_pc2, rot_pc3), digits=3)
names(loading_rot) <- c("RC1", "RC2", "RC3")
write.table(loading_rot, file="3d.csv", sep = ",", col.names=NA)
```

	RC1	RC2	RC3
Q1	0.66	0.221	0.45
Q2	0.544	0.288	0.286
Q3	0.621	0.311	0.337
Q4	0.218	0.854	0.193
Q5	0.244	0.162	0.828
Q6	0.652	0.234	0.199
Q7	0.79	0.056	0.103
Q8	0.382	0.305	0.706
Q9	0.738	0.138	0.234
Q10	0.277	0.102	0.823
Q11	0.757	0.118	0.278
Q12	0.233	0.854	0.186
Q13	0.593	0.259	0.315
Q14	0.719	0.283	0.31
Q15	0.342	0.244	0.656
Q16	0.74	0.174	0.267
Q17	0.205	0.87	0.187
Q18	0.609	0.227	0.495

Similar to 2(a),

- RC1: Q1, Q2, Q3, Q6, Q7, Q9, Q11, Q13, Q14, Q16, Q18.
- RC2: Q4, Q12, Q17.
- RC3: Q5, Q8, Q10, Q15.

To sum up in my own words,

- RC1: My information safety is protected.
- RC2: The transaction would not be easily denied after processing.
- RC3: The site check my identity and I would not be transmitted to a fake site.

(e) Now reduce the number of extracted and rotated components to 2.

```
# Question 3 (e)
sec_pca_rot2 <- principal(security, # rotate the pcs
                          nfactor=2,
                          rotate="varimax",
                          scores=TRUE) # just call these two model

rot2_pc1 <- sec_pca_rot2$loadings[,"RC1"] # Rotated component's loading
rot2_pc2 <- sec_pca_rot2$loadings[,"RC2"]
loading_compare_2comp <- round(cbind(rot2_pc1, rot2_pc2), digits=3)
names(loading_compare_2comp) <- c("RC1", "RC2")
write.table(loading_compare_2comp, file="3e.csv", sep = ",", col.names=NA)
```

	rot2_pc1	rot2_pc2
Q1	0.783	0.271
Q2	0.596	0.312
Q3	0.687	0.34
Q4	0.236	0.864
Q5	0.62	0.305
Q6	0.649	0.237
Q7	0.728	0.038
Q8	0.668	0.416
Q9	0.745	0.145
Q10	0.649	0.244
Q11	0.786	0.134
Q12	0.245	0.862
Q13	0.655	0.286
Q14	0.759	0.304
Q15	0.612	0.348
Q16	0.762	0.187
Q17	0.221	0.88
Q18	0.762	0.289

- RC1: Q1, Q2, Q3, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q13, Q14, Q15, Q16, Q18.
- RC2: Q4, Q12, Q17.

That means the RC3 in 3(d) is merged to RC1 here. Hence I sum up these two RC as,

- RC1: My would not suffer from information safety risk with my identity checked.
- RC2: The transaction would not be easily denied after processing.

■

The Last Question. After reading the questions, I believe picking 3 RCs as 3(d) are the best choice if we'd like to explain the dataset to the others in words. ■