HW15_106023021

6/6/2021

Assistance: 108078518

Q1)

```
library(gdata)
```

```
## Warning: package 'gdata' was built under R version 4.0.5
```

```
## gdata: Unable to locate valid perl interpreter
## gdata:
## gdata: read.xls() will be unable to read Excel XLS and XLSX files
## gdata: unless the 'perl=' argument is used to specify the location of a
## gdata: valid perl intrpreter.
## gdata:
## gdata: (To avoid display of this message in the future, please ensure
## gdata: perl is installed and available on the executable search path.)
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLX' (Excel 97-2004) files.
```

```
##
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLSX' (Excel 2007+) files.
```

```
##
```

```
## gdata: Run the function 'installXLSXsupport()'
## gdata: to automatically download and install the perl
## gdata: libaries needed to support Excel XLS and XLSX formats.
```

```
##
## Attaching package: 'gdata'
```

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

```
## The following object is masked from 'package:utils':
##
##     object.size
```

```
## The following object is masked from 'package:base':
##
##     startsWith
```

```
sec = read.csv("security_questions.csv")
```

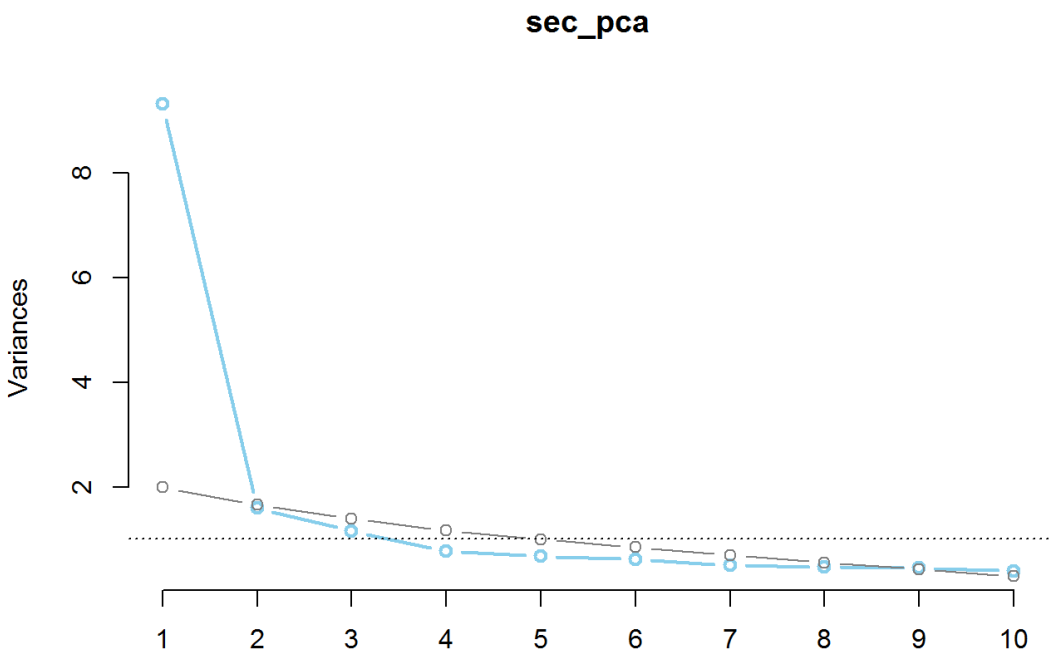
a.

```
sec_pca <- prcomp(sec, scale. = TRUE)

sim_noise_ev <- function(n, p) {
noise <- data.frame(replicate(p, rnorm(n)))
return( eigen(cor(noise))$values )
}

set.seed(114)
evalues_noise <- replicate(100, sim_noise_ev(33, 10))
evalues_mean <- apply(evalues_noise, 1, mean)
```

```
screeplot(sec_pca, type="lines", col="skyblue", lwd=2)
lines(evalues_mean, type="b", col="gray50")
abline(h=1, lty="dotted")
```



b.

```
# Use top 3 components above the line.
```

Q2)

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.0.5
```

a.

```
principal(sec, nfactor=3, rotate="none", scores=TRUE)
```

```
## Principal Components Analysis
## Call: principal(r = sec, nfactors = 3, rotate = "none", scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##      PC1   PC2   PC3   h2   u2 com
## Q1  0.82 -0.14  0.00 0.69 0.31 1.1
## Q2  0.67 -0.01  0.09 0.46 0.54 1.0
## Q3  0.77 -0.03  0.09 0.60 0.40 1.0
## Q4  0.62  0.64  0.11 0.81 0.19 2.1
## Q5  0.69 -0.03 -0.54 0.77 0.23 1.9
## Q6  0.68 -0.10  0.21 0.52 0.48 1.2
## Q7  0.66 -0.32  0.32 0.64 0.36 2.0
## Q8  0.79  0.04 -0.34 0.74 0.26 1.4
## Q9  0.72 -0.23  0.20 0.62 0.38 1.4
## Q10 0.69 -0.10 -0.53 0.76 0.24 1.9
## Q11 0.75 -0.26  0.17 0.66 0.34 1.4
## Q12 0.63  0.64  0.12 0.82 0.18 2.1
## Q13 0.71 -0.06  0.08 0.52 0.48 1.0
## Q14 0.81 -0.10  0.16 0.69 0.31 1.1
## Q15 0.70  0.01 -0.33 0.61 0.39 1.4
## Q16 0.76 -0.20  0.18 0.65 0.35 1.3
## Q17 0.62  0.66  0.11 0.83 0.17 2.0
## Q18 0.81 -0.11 -0.07 0.67 0.33 1.1
##
##                        PC1  PC2  PC3
## SS loadings           9.31 1.60 1.15
## Proportion Var        0.52 0.09 0.06
## Cumulative Var        0.52 0.61 0.67
## Proportion Explained  0.77 0.13 0.10
## Cumulative Proportion 0.77 0.90 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.05
##  with the empirical chi square  258.65  with prob <  1.4e-15
##
## Fit based upon off diagonal values = 0.99
```

```
# belong to component1.
```

b.

```
# Based on the chart there almost 67% variance has been captured by the top3 components.
```

c.

```
# Q2 has commonality=0.46 and uniqueness=0.53 which is less explained by the 3 components.
```

d.

```
# Q4, Q12, Q17 seem to share similar loadings in the 3 components.
```

e.

```
# With the high loading from Q1 & Q18, I guess the 1st component has a lot to do with "data privacy", because the
questions have some related keywords such as "confidentiality", "security".
```

Q3)

```
principal(sec, nfactor=3, rotate="varimax", scores=TRUE)
```

```
## Principal Components Analysis
## Call: principal(r = sec, 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
##
##                        RC1  RC3  RC2
## SS loadings           5.61 3.49 2.95
## Proportion Var        0.31 0.19 0.16
## Cumulative Var        0.31 0.51 0.67
## Proportion Explained  0.47 0.29 0.24
## Cumulative Proportion 0.47 0.76 1.00
##
## Mean item complexity =  1.6
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.05
##  with the empirical chi square  258.65  with prob <  1.4e-15
##
## Fit based upon off diagonal values = 0.99
```

a.

```
# Each component has different amount of variance than the corresponding principal components.
```

b.

```
# Even variance explained by each is different, the cumulative variance explained by the 3 are the same.
```

c.

```
# Yes, loadings of Q4 & Q12 & Q17 seem to have more clearly differentiated loadings among rotated components.
```

d.

```
# Q4, Q12 & Q17 have higher loadings. Form the wording of quesitons, they are with regard to protection against d
enial during the transaction process.
```

e.

```
principal(sec, nfactor=2, rotate="varimax", scores=TRUE)
```

```
## Principal Components Analysis
## Call: principal(r = sec, nfactors = 2, rotate = "varimax", scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##      RC1  RC2   h2   u2 com
## Q1  0.78 0.27 0.69 0.31 1.2
## Q2  0.60 0.31 0.45 0.55 1.5
## Q3  0.69 0.34 0.59 0.41 1.5
## Q4  0.24 0.86 0.80 0.20 1.1
## Q5  0.62 0.31 0.48 0.52 1.5
## Q6  0.65 0.24 0.48 0.52 1.3
## Q7  0.73 0.04 0.53 0.47 1.0
## Q8  0.67 0.42 0.62 0.38 1.7
## Q9  0.75 0.15 0.58 0.42 1.1
## Q10 0.65 0.24 0.48 0.52 1.3
## Q11 0.79 0.13 0.64 0.36 1.1
## Q12 0.25 0.86 0.80 0.20 1.2
## Q13 0.65 0.29 0.51 0.49 1.4
## Q14 0.76 0.30 0.67 0.33 1.3
## Q15 0.61 0.35 0.50 0.50 1.6
## Q16 0.76 0.19 0.62 0.38 1.1
## Q17 0.22 0.88 0.82 0.18 1.1
## Q18 0.76 0.29 0.66 0.34 1.3
##
##                        RC1  RC2
## SS loadings           7.52 3.39
## Proportion Var        0.42 0.19
## Cumulative Var        0.42 0.61
## Proportion Explained  0.69 0.31
## Cumulative Proportion 0.69 1.00
##
## Mean item complexity =  1.3
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##  with the empirical chi square  439.68  with prob <  1.3e-38
##
## Fit based upon off diagonal values = 0.99
```

```
# With only 2 component, the 1st one contains more information, not just the "data privacy" anymore. And the 2nd component seems to became similar to the 3rd component in the previous model that have high loading in Q4, Q12, Q17.

# So we can have conclusion that 1st component may contain information about "data privacy" &"personal identity", which is separated by the previous model into 2 component. And the 2nd component is similar to the 3rd component in the previous model with information about "protection of denial".
```

Looking back at all our results and analyses of this dataset (from this week and previous), how many components (1-3) do you believe we should extract and analyze to understand the security dataset?

```
# I think 3 components offers more information to better explain and understand the data.
```