Assignment 2

1) This problem analyzes the Daily Hang Seng Index data

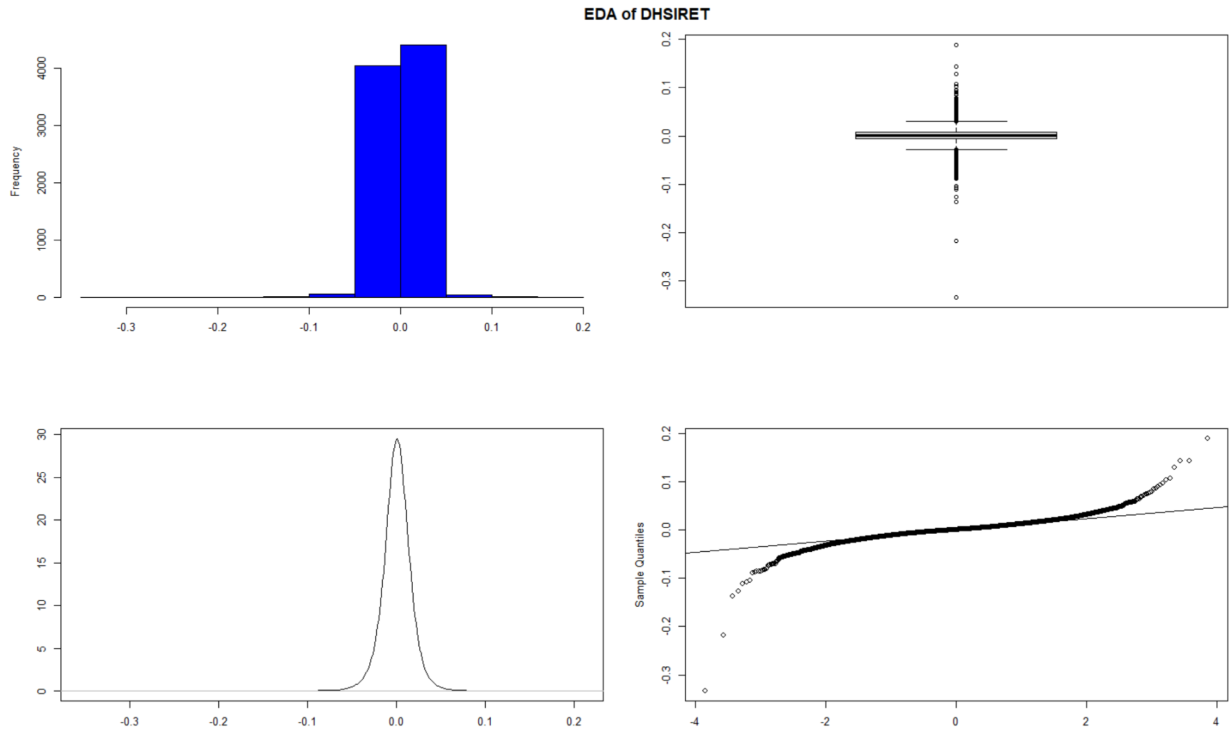
1.1)

```
HSI <- read.table("DHSI.csv",header = T, sep=",")

DHSI<-(HSI$Close)
n<-length(DHSI)
DHSIRET <- diff(DHSI)/DHSI[1:(n-1)] # compute raw returns (rate of return)

mu_DHSIRET <- mean(DHSIRET) # mean
sd_DHSIRET <- sd(DHSIRET) # standard deviation

round(c(mu_DHSIRET, sd_DHSIRET),4)
```

Mean: 0.0004
Standard Deviation: 0.0161

1.2)

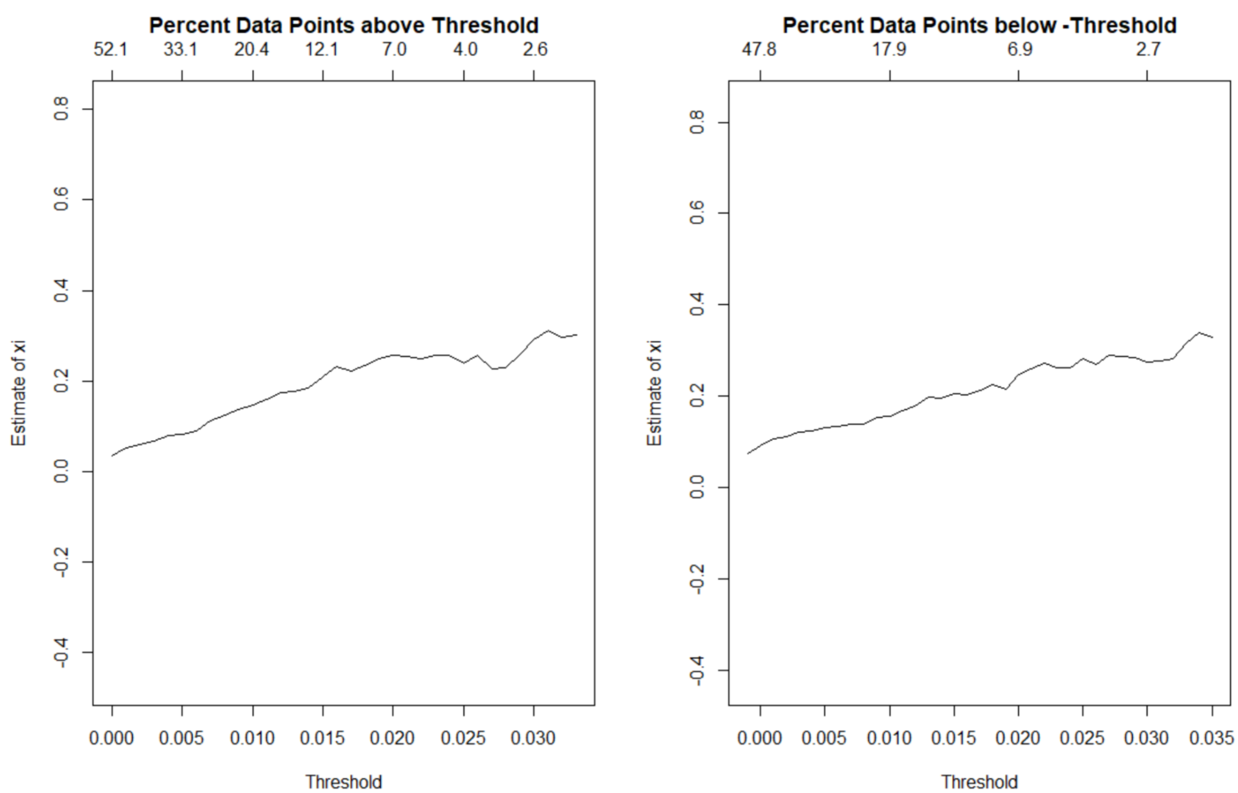
```
par(mfrow = c(1, 1))
eda.shape(DHSIRET)
```



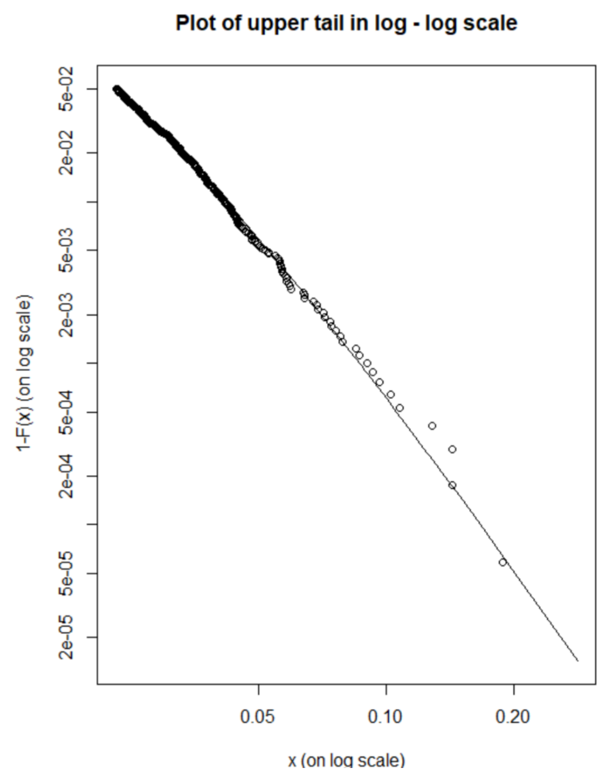
The graphs above shows DHSIRET summary: histogram, box-plot, kernel density estimate, and normal QQ plot. It can be deduced from the first 3 plots that DHSIRET returns are centered around 0, with minimal variance. From the boxplot, you can tell that there are some outliers, but they are negligible because they rarely show up. The shape of the distribution is quite tall and different from a normal distribution. It is confirmed by the normal QQ plot. The curved shape indicated on the left and and right tails of the sample distribution is heavier than the normal distribution.

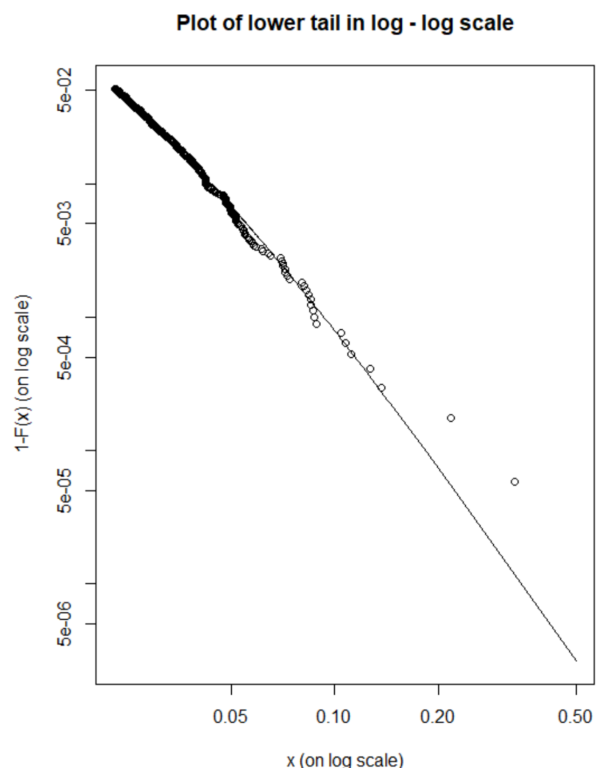
1.3)

```
SHAPE.XI=TRUE
par(mfrow=c(1,2))
shape.plot(DHSIRET,tail="upper")
shape.plot(DHSIRET,tail="lower")

  # fit

DHSIRET.est <- gpd.tail(DHSIRET, upper=0.023,lower=-0.023)

  # xi

DHSIRET.est[[1]]$par.ests[1] # upper
DHSIRET.est[[2]]$par.ests[1] # lower

  # goodness of fit

par(mfrow=c(1,2))

tailplot(DHSIRET.est[[1]]) # upper tail
title("Plot of upper tail in log - log scale")
tailplot(DHSIRET.est[[2]]) # lower tail
title("Plot of lower tail in log - log scale")
```



To fit the tail of the GPD, determine the xi parameter for the upper and lower tails. To do so, we would have to use a xi and threshold graph of the upper and lower tails. To deduce the xi parameter, we would have to determine the corresponding threshold first based on a couple requirements. Because we don't want the tail to contain too many observations, the left side of the threshold graph (near 0) is eliminated. Similarly, we also want the tail to contain enough observations to ensure the quality of the fitting; therefore, the right side of the threshold graph (near 0.03) is eliminated. Since the left and right sides of the graph are quite arbitrary, it means we would only focus on the center region. Lastly, we don't want the specification of the tail to affect the fitting too much, which means we want to pick a threshold where it has a relatively stable xi parameter. Based on these criteria, for the upper tail and lower tail, xi of 0.023 and -0.023 is used, respectively.

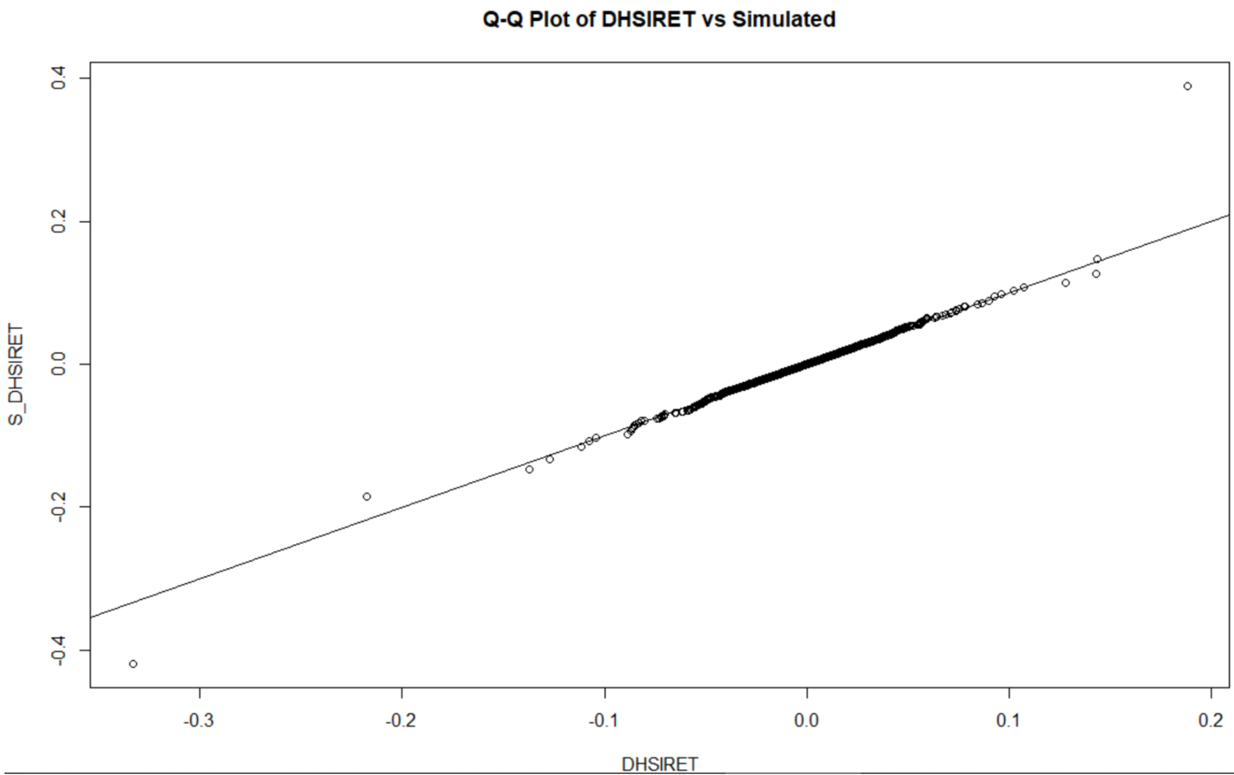




The fitting result is quite spectacular. Both the upper and lower tail seem to fit very well. When benchmarked against the empirical data (line) is theoretical data (points) seem to line up quite well. The only thing notable is that the lower tail (right graph) has points that are higher than the line, which indicates that there is a slight underestimation of risk.

1.4)

```
par(mfrow=c(1,1))

S_DHSIRET=gpd.2q(runif(100000), DHSIRET.est)
qqplot(DHSIRET,S_DHSIRET,main = "Q-Q Plot of DHSIRET vs Simulated")
abline(a=0,b=1)
```



Based on the QQ plot, it shows that the simulated sample S_DHSIRET and the empirical sample DHSIRET has more or less the same distribution. This is indicated by how straight the line of fit is, which shows that the distribution are essentially the same.

1.5)

```r
q<-0.005

  # a. empiracle quantile

VaR_emp <- - quantile(DHSIRET,q)

  # b. normally distributed

VaR_N <- - qnorm(q,mean(DHSIRET),sd(DHSIRET))

  # c. fitted GPD distribution

xi_upper = DHSIRET.est[[1]]$par.ests[1]
threshold_upper = DHSIRET.est[[1]]$threshold
lambda_upper = DHSIRET.est[[1]]$par.ests[2]

xi_lower = DHSIRET.est[[2]]$par.ests[1]
threshold_lower = DHSIRET.est[[2]]$threshold
lambda_lower = DHSIRET.est[[2]]$par.ests[2]

Cond_p=length(DHSIRET[-DHSIRET>threshold_lower])/length(DHSIRET)

VaR_GPD <- qgpd(1-0.005/Cond_p,xi_lower,threshold_lower,lambda_lower)

  # d. Monte Carlo S_DHSIRET

VaR_Monte <- - quantile(S_DHSIRET,q)

  # e. comparison

VaR_Ratio <-  (VaR_N - VaR_GPD)/VaR_GPD

round(c(VaR_emp,VaR_N,VaR_GPD,VaR_Monte,VaR_Ratio),3)
```

VaR Empirical: 0.053
VaR Normal: 0.041
VaR GPD Fitted: 0.056
VaR Monte Carlo: 0.056

Ratio (VaR Normal: VaR GPD Fitted): -0.262

Comparison

VaR Normal < VaR Empirical < VaR GPD Fitted = VaR Monte Carlo

Based on the results above, the normal VaR underestimates the risk compared to the empirical VaR by -1.2%, which is a huge difference when the portfolio is large. Also the GPD fitted VaR and the Monte Carlo VaR are essentially the same thing, and they have a VaR that overestimates the risk by +0.3%. The VaR under normal assumption is smaller than the VaR under GPD assumption by about 26.2%. In risk management, it is better to use a more conversative approach; therefore, the GPD fitting / Monte Carlo simulation VaR is more suitable.

1.6)

```
  # a. empiracle conditional mean

ES_emp <- mean(- DHSIRET[- DHSIRET > VaR_emp])

  # b. normally distributed

N<-100000
X<-rnorm(N,mu_DHSIRET,sd_DHSIRET)
ES_N <- mean( - X[- X > VaR_N])

  # c. fitted GPD distribution

ES_GPD <- mean(- S_DHSIRET[- S_DHSIRET > VaR_GPD])

  # e. comparison

round(c(ES_emp, ES_N, ES_GPD),3)
```

Expected Shortfall Empirical: 0.083
Expected Shortfall Normal: 0.046
Expected Shortfall Fitted GPD: 0.081

Comparison
Expected Shortfall Normal < Expected Shortfall Empirical < Expected Shortfall Fitted GPD

The expected shortfall for normal is smaller than empirical and then followed by fitted GPD. The normal model greatly underestimates the risk, while the fitted GPD is approximately equal to the empirical model.

2) Problem 3.7 Parts 1 & 2 on page 187 of the book by Carmona

2.1)

```
# 1. means, standard deviations, correlation coef of X and Y

mu_X <- mean(X)
sd_X <- sd(X)
var_X <- var(X)

mu_Y <- mean(Y)
sd_Y <- sd(Y)
var_Y <- var(Y)

cor_XY <- cor(X,Y)

round(c(mu_X,sd_X,var_X,mu_Y,sd_Y,var_Y,cor_XY),4)
```

Mean X: 0.0024
Mean Y: 0.0005
Standard Deviation X: 0.0270
Standard Deviation Y: 0.0364
Variance X: 0.0007
Variance Y: 0.0013
Correlation X + Y: 0.5446

2.2)

```
# 2. 2-percentiles of variables X + Y and X - Y, empirical estimate of percentiles

# estimated

qnorm(0.02, mean=(mu_X+mu_Y), sd=sqrt(var_X+var_Y+2*cor_XY*sd_X*sd_Y)) # X + Y
qnorm(0.02, mean=(mu_X-mu_Y), sd=sqrt(var_X+var_Y-2*cor_XY*sd_X*sd_Y)) # X - Y

# empiracle

quantile(X+Y,0.02) # X + Y
quantile(X-Y,0.02) # X - Y
```

<u>Estimated</u>

X+Y: -0.112
X–Y: -0.063

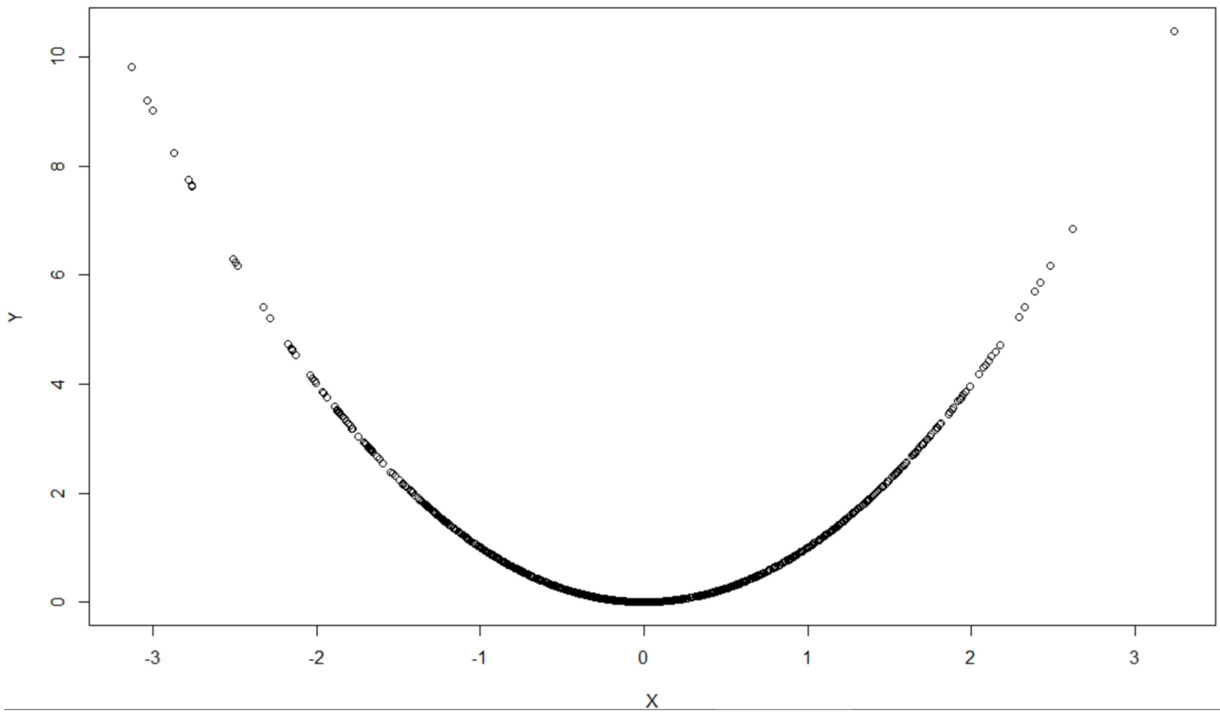
<u>Empirical</u>

X+Y: -0.101
X–Y: -0.058

3) Lecture 6 pp. 29-48 Analysis

3.1)

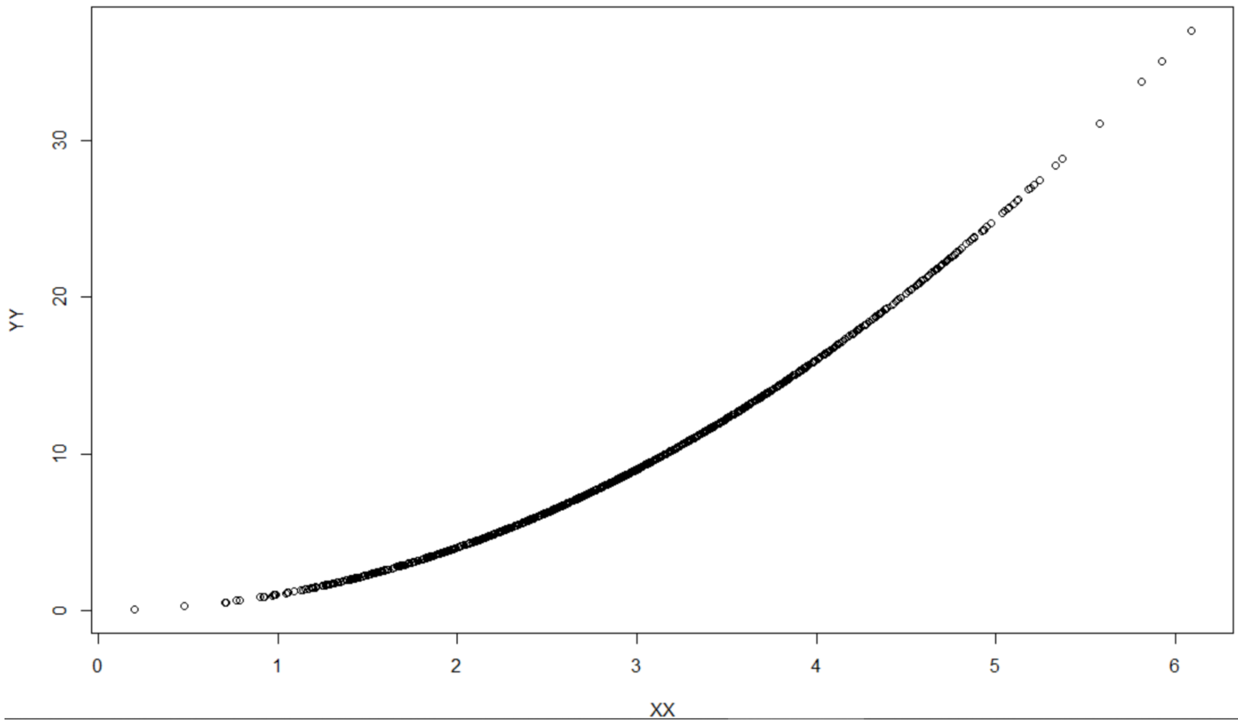
```
X <- rnorm(1024)
Y <- X^2

par(mfrow=c(1,1))
plot(X,Y)
cor(X,Y)
```



Correlation XY: -0.04868983

3.2)

```
XX <- rnorm(1024, 3, 1)
YY <- XX^2

plot(XX,YY)
cor(XX,YY)
```



Correlation XXYY: 0.9794414

3.3)

```
  # n = 1
A <- rnorm(1)
B = A^2
cor(A,B)

  # n = 100000
A <- rnorm(100000)
B = A^2
cor(A,B)
```

Correlation AB (n = 1): N/A
Correlation AB (n = 100,000): -0.002703943 ≈ 0

Supplementary Proof

$$E(A) = E(u+\sigma Z) = u + \sigma E(Z) = u + \sigma(0) = u$$

$$E(A^2) = E[(u+\sigma Z)^2] = E(u^2 + 2u\sigma Z + \sigma^2 Z^2) = u^2 + 2u\sigma E(Z) + \sigma^2 E(Z^2) = u^2 + 2u\sigma(0) + \sigma^2(1) = u^2 + \sigma^2$$

$$E(A^3) = E[(u+\sigma Z)^3] = E(u^3 + 3u^2\sigma Z + 3u\sigma^2 Z^2 + \sigma^3 Z^3) = u^3 + 3u^2\sigma E(Z) + 3u\sigma^2 E(Z^2) + \sigma^3 E(Z^3) = u^3 + 3u^2\sigma(0) + 3u\sigma^2(1) + \sigma^3(0) = u^3 + 3u\sigma^2$$

$$E(A^4) = E((u+\sigma Z)^4) = E(u^4 + 4u^3\sigma Z + 6u^2\sigma^2 Z^2 + 4u\sigma^3 Z^3 + \sigma^4 Z^4)$$
$$= u^4 + 4u^3\sigma E(Z) + 6u^2\sigma^2 E(Z^2) + 4u\sigma^3 E(Z^3) + \sigma^4 E(Z^4)$$
$$= u^4 + 4u^3\sigma(0) + 6u^2\sigma^2(1) + 4u\sigma^3(0) + \sigma^4(3)$$
$$= u^4 + 6u^2\sigma^2 + 3\sigma^4$$

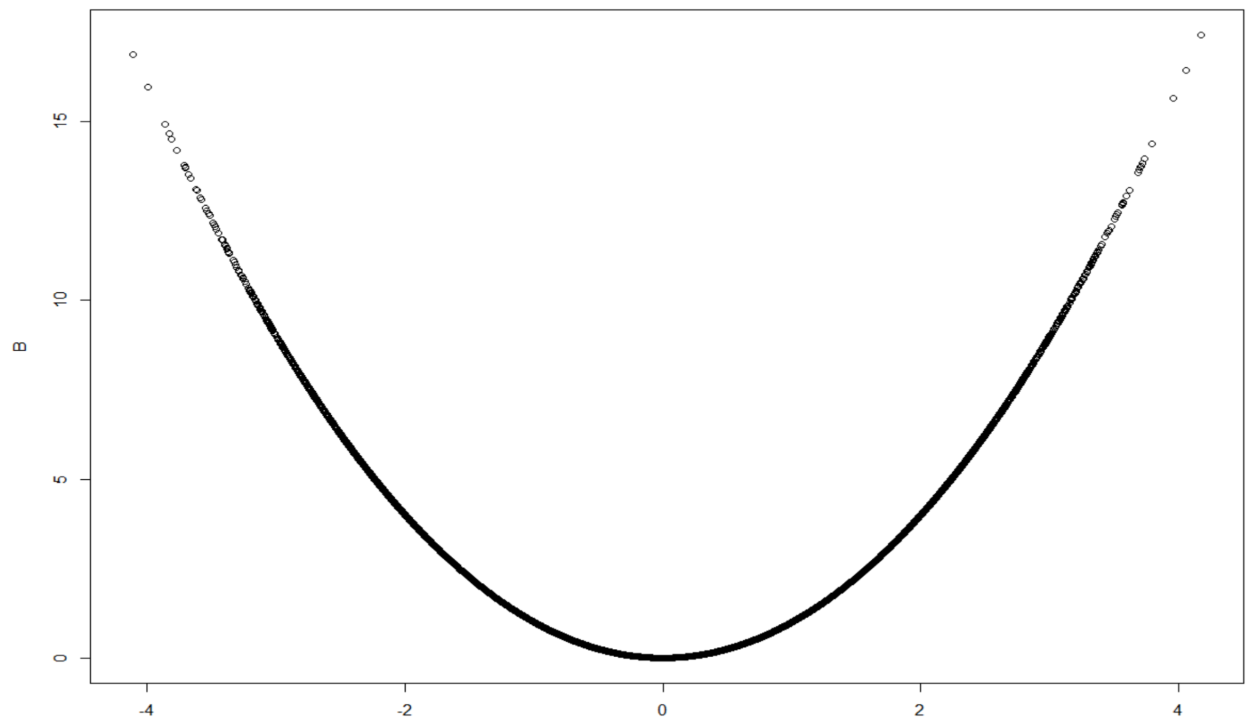
$$Cov(A, A^3) = EA^3 - EA + EA^2 = u^3 + 3u\sigma^2 - u^3 - u\sigma^2 = 2u\sigma^2$$

$$P(A, A^2) = \frac{Cor(A, A^2)}{\sqrt{Var(A)Va(A^2)}} = \frac{2u\sigma^2}{\sqrt{\sigma^2 \times (\sigma^2(4u^2+2\sigma^2)}} = \frac{2u}{\sqrt{4u^2+2\sigma^2}}$$

$$Var(A) = EA^2 - (EA)^2 = (u^2+\sigma^2) - (u)^2 = u^2 + \sigma^2 - u^2 = \sigma^2$$

$$Var(A^2) = E(A^4) - (E(A^2))^2 = (u^4 + 6u^2 + \sigma^2 + 3\sigma^4) - (u^2+\sigma^2)^3$$
$$= u^4 + 6u^2\sigma^2 + 3\sigma^4 - u^4 - 2u^2\sigma^2 - \sigma^4 = 4u^2\sigma^2 + 2\sigma^4 = \sigma^2(4u^2 + 2\sigma^2)$$

3.4)



According to the calculation above, it can be assumed that A and B are uncorrelated. As n→∞ (sample size increases), it can be deduced that the correlation between A and B approaches 0. Additionally, the correlation method of Kendall's Tau and Spearman's Rho has been used and similar behavior is also exhibited. The reason is because based on the graph, the negative correlation when A < 0 and the positive correlation when A > 0 causes the correlation to be 0 when it cancels each other out. However, this uncorrelation does not indicate that A and B are independent. To determine independence there are more constraints to satisfy and are more restrictive. But based on the graph, it can be determined that A and B are somewhat correlated. As A increases in magnitude (positive and negative) B also increases, which means that B depends on the magnitude of A.

Supplementary Proof

$\mu_{A,B} = 0$

$$\text{corr}(A,B) = \frac{\text{cov}(A,B)}{\sqrt{\text{var}(A)\text{var}(B)}} = \frac{E[(A-\mu_A)(B-\mu_B)]}{\sqrt{\text{var}(A)\text{var}(B)}} = \frac{E[(A)(B)]}{\sqrt{\text{var}(A)\text{var}(B)}}$$

$$= \frac{E(AB)}{\sqrt{\text{var}(A)\text{var}(B)}} = \frac{E(A \cdot A^2)}{\sqrt{\text{var}(A)\text{var}(B)}} = \frac{E(A^3)}{\sqrt{\text{var}(A)\text{var}(B)}}$$

$$E(A) = E(A^3) = 0$$

$\rho = 0$

$A = 1$
$B = 1$

$A = -2, 2$
$B = 4$

$A = -4, 4$
$B = 16$

value of A determines

↳ clearly B dependent on A

As A increase in magnitude (−/+), so does B.

Empiracle Proof

$u = 0 \Rightarrow \rho(A, A^2) = 0$ then A and A^2 uncorrelated

$\rho > 0 \quad \rho > 0$

$(A > 1)(A^2 < 1) = 0$

$P(A>1, B<1) = 0$
$P(A>1) > 0$
$P(B<1) > 0$

Independence mean

$P(A,B) = P(A) - P(B)$
$0 \neq >0 \quad >0$

$P(A>1, B<1)$
$= P(A>1) \times P(B<1)$
$0 = (>0)(>0)$
$\hookrightarrow$ **not** independent ; therefore, dependent