HW4

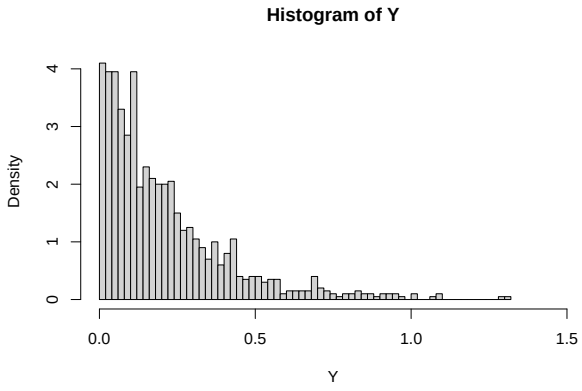
Kyu Park

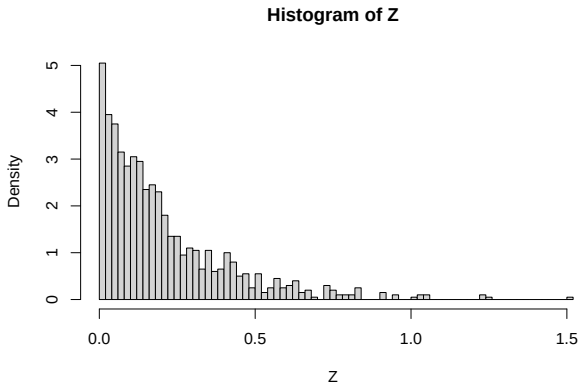
2021 2 8

2g.

```
U=runif(1000)
Z=rexp(1000,5)

hist(-log(1-U)/5, xlim = c(0, 1.5), breaks = 60,
     freq = FALSE, xlab = "Y", main = "Histogram of Y")
hist(Z, xlim = c(0, 1.5), breaks = 60, freq = FALSE)
```





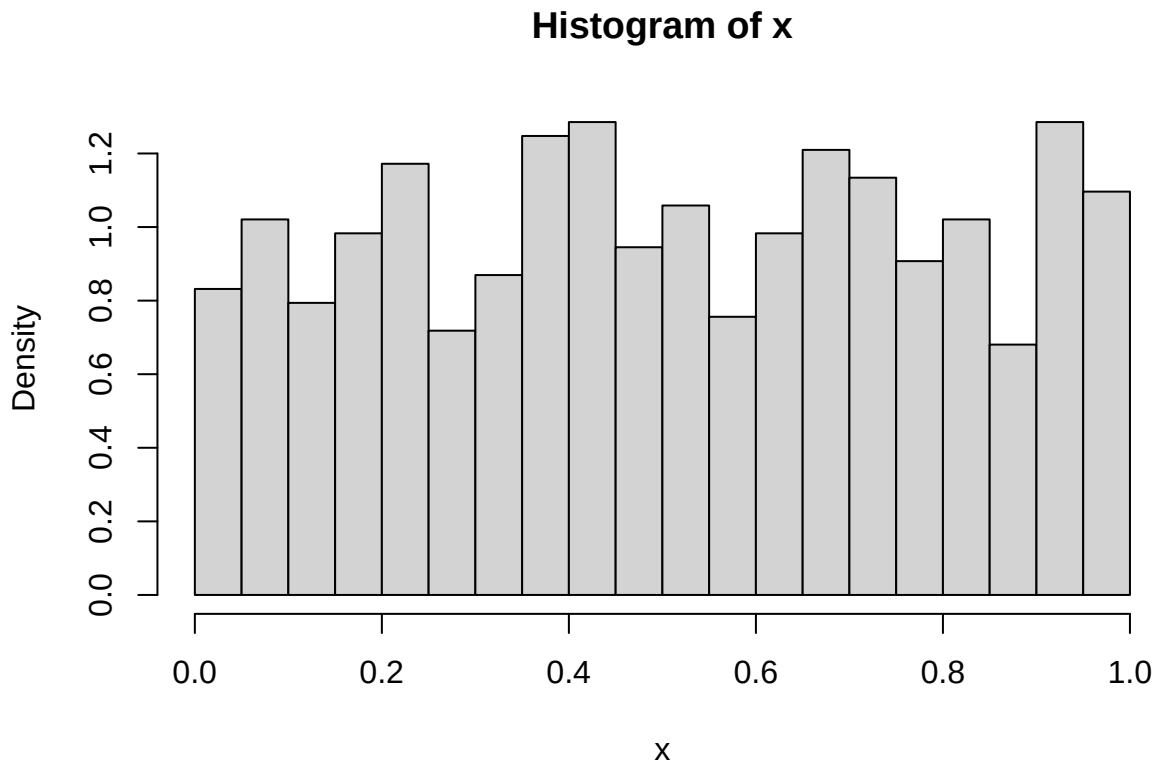
4.

The pdf is valid if the area A between the two curves $y = x^2, 0 \leq x \leq 1$ and $y = x^2 + 1, 0 \leq x \leq 1$ The area A could be calculated as $\int_0^1 \left(x^2 + 1\right) dx \; - \int_0^1 x^2 \, dx$, which equals to 1. Therefore, this pdf is valid pdf.

```
##given parameters
x=runif(1000,min=0,max=1)
y=runif(1000,min=0,max=2)
dat=data.frame(x,y)
data=subset(dat,y>x*x & y<x*x+1)
```

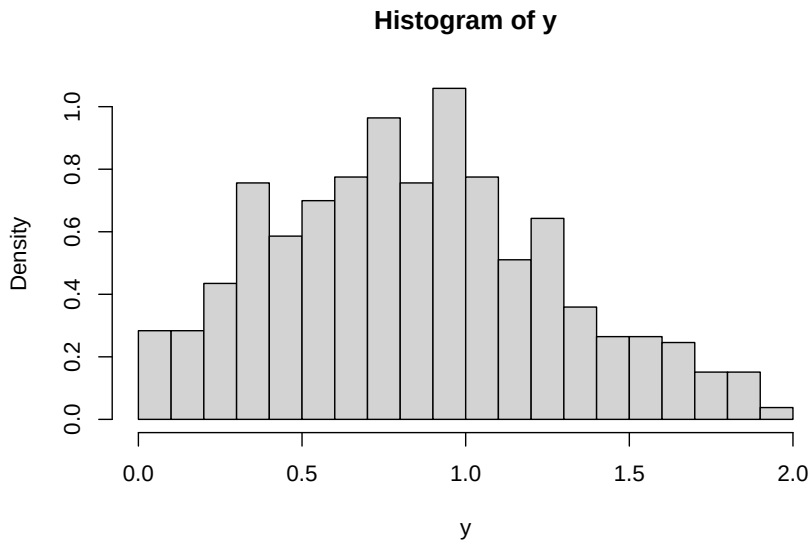
a)

```
hist(data$x, freq = FALSE, breaks = 20, main = "Histogram of x", xlab = "x")
```



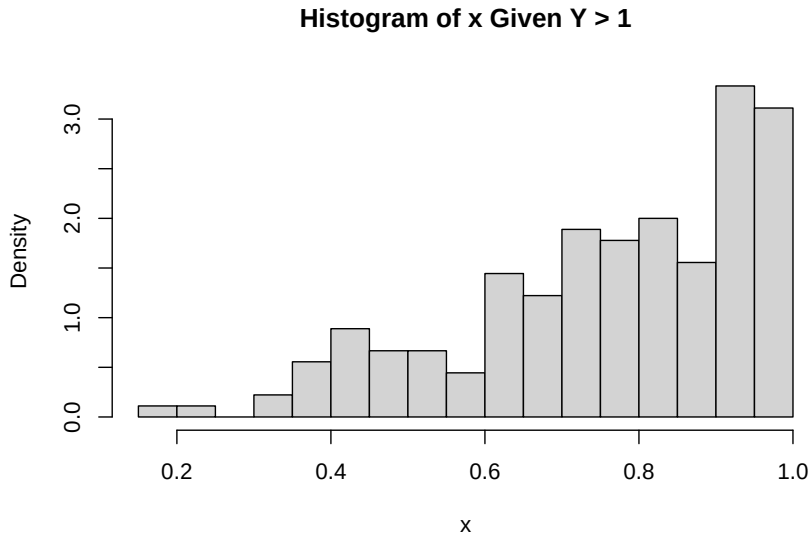
b)

```
hist(data$y, freq = FALSE, breaks = 20, main = "Histogram of y", xlab = "y")
```



c)

```
data_subset = subset(data, y > 1)
hist(data_subset$x, freq = FALSE, breaks = 20,
     main = "Histogram of x Given Y > 1", xlab = "x")
```



5.

```
load("~/STAT23400/Wealth.Rdata")
```

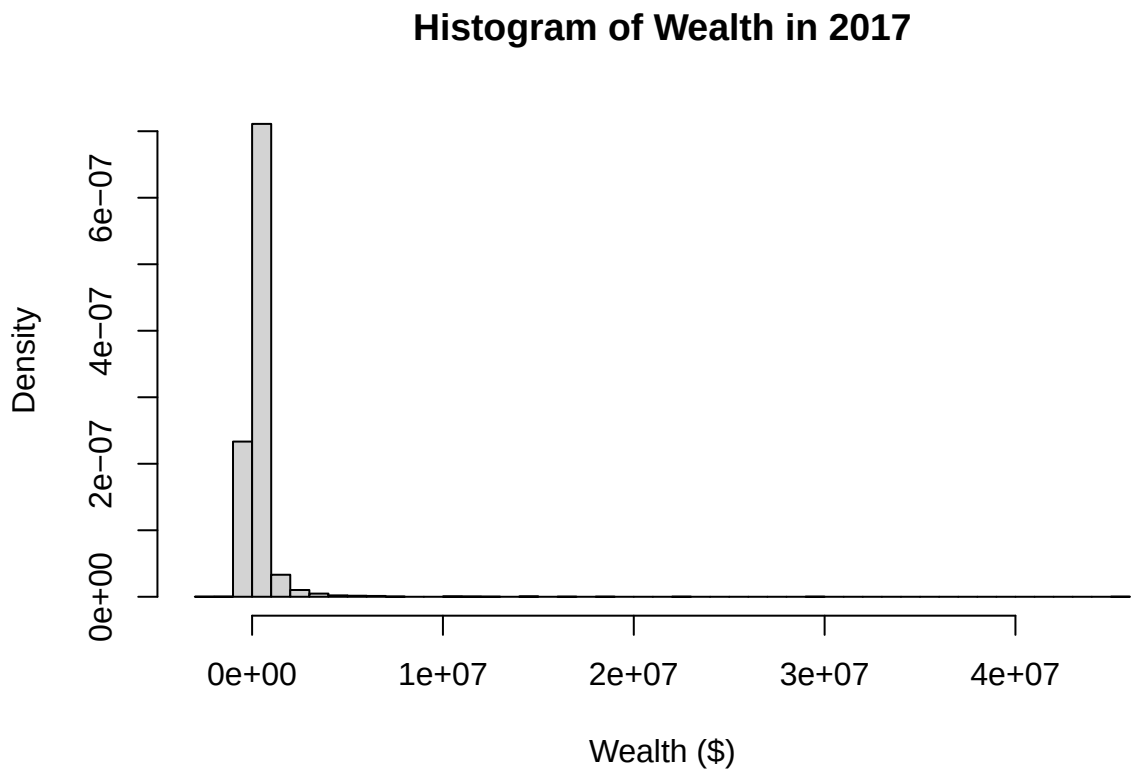
a)

```
average <- mean(Wealth$Y2017, na.rm = TRUE)
average
```

```
## [1] 240327.6
```

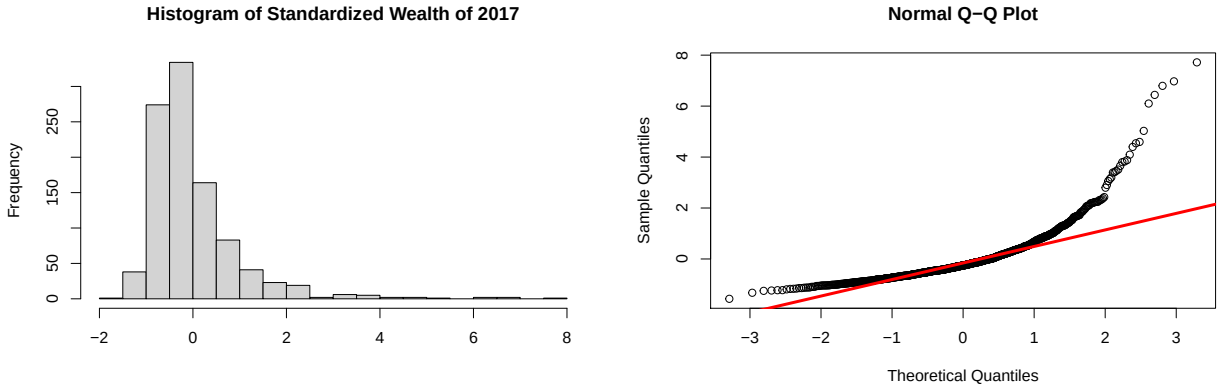
The mean wealth of the population is 240327.6. The histogram could be plotted by

```
hist(Wealth$Y2017, breaks = 60, freq = FALSE,
     main = "Histogram of Wealth in 2017", xlab = "Wealth ($)")
```



b)

```
samp = replicate(1000, sample(Wealth$Y2017, 100, replace = TRUE))
samp_average = colMeans(samp, na.rm = TRUE)
samp_standardized = (samp_average - mean(samp_average))/sd(samp_average)
hist(samp_standardized, xlim = c(-2, 8), breaks = 20,
     main = "Histogram of Standardized Wealth of 2017", xlab = NULL)
qqnorm(samp_standardized)
qqline(samp_standardized, col = "red", lwd = 3)
```

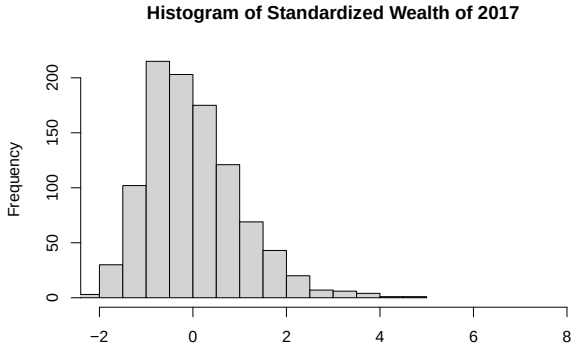


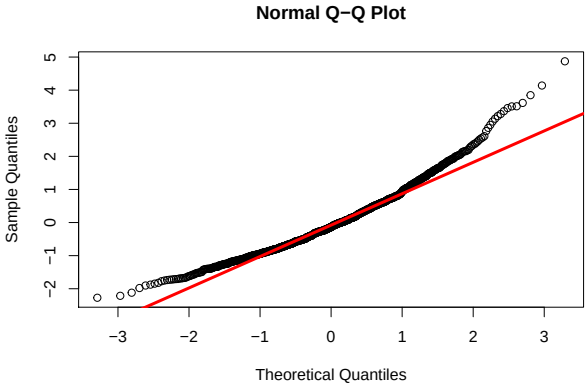
From the histogram, it appears that the sample is not normally distributed, rather skewed to the right.

c)

Sample size of 1000

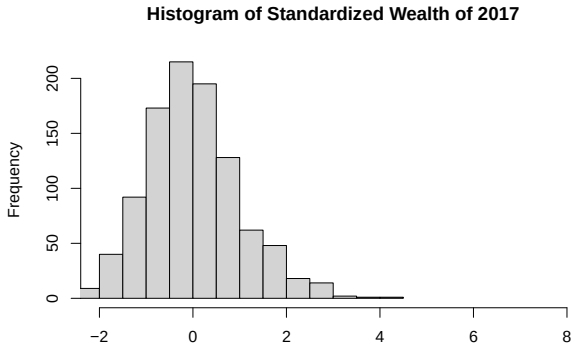
```
samp = replicate(1000, sample(Wealth$Y2017, 1000, replace = TRUE))
samp_average = colMeans(samp, na.rm = TRUE)
samp_standardized = (samp_average - mean(samp_average))/sd(samp_average)
hist(samp_standardized, xlim = c(-2, 8), breaks = 20, main = "Histogram of Standardized Wealth of 2017"
qqnorm(samp_standardized)
qqline(samp_standardized, col = "red", lwd = 3)
```

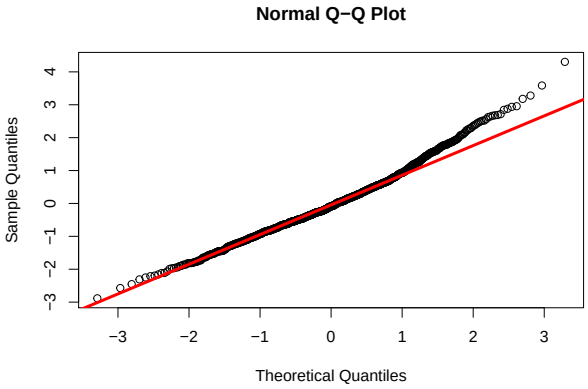




Sample size of 2000

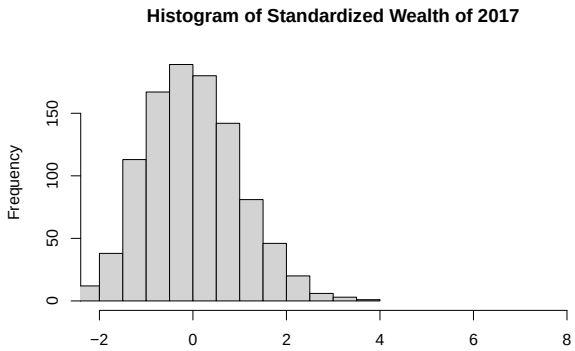
```
samp = replicate(1000, sample(Wealth$Y2017, 2000, replace = TRUE))
samp_average = colMeans(samp, na.rm = TRUE)
samp_standardized = (samp_average - mean(samp_average))/sd(samp_average)
hist(samp_standardized, xlim = c(-2, 8), breaks = 20, main = "Histogram of Standardized Wealth of 2017"
qqnorm(samp_standardized)
qqline(samp_standardized, col = "red", lwd = 3)
```

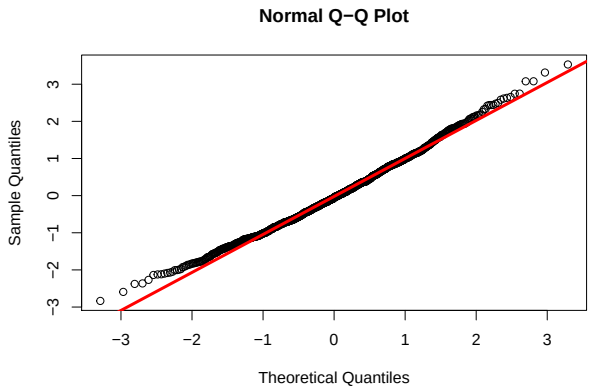




Sample size of 4000

```
samp = replicate(1000, sample(Wealth$Y2017, 4000, replace = TRUE))
samp_average = colMeans(samp, na.rm = TRUE)
samp_standardized = (samp_average - mean(samp_average))/sd(samp_average)
hist(samp_standardized, xlim = c(-2, 8), breaks = 20, main = "Histogram of Standardized Wealth of 2017"
qqnorm(samp_standardized)
qqline(samp_standardized, col = "red", lwd = 3)
```





At around 4000 samples, the sample averages begin to look normal.

It is clear that the bigger the sample is, the more the sample averages look normal. As the sample size increases, the histogram displays normal distribution and the points allign on the straight Q-Q plot line.