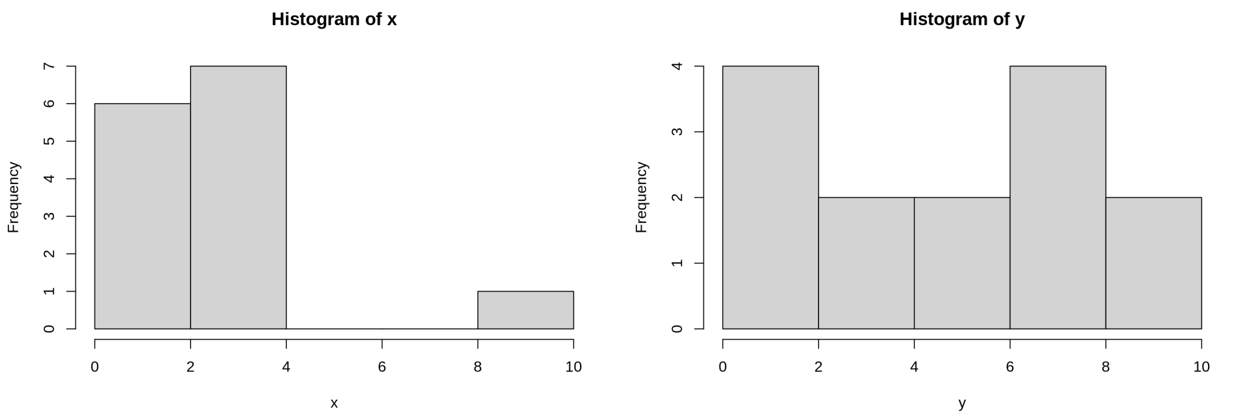
Homework 1

Kevin Ha

September 2nd, 2021

Problem 1

Problem 1.1



Problem 1.2

```
summary(x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.200   1.275   2.250   2.407   2.900   9.000
```

```
var(x)
```

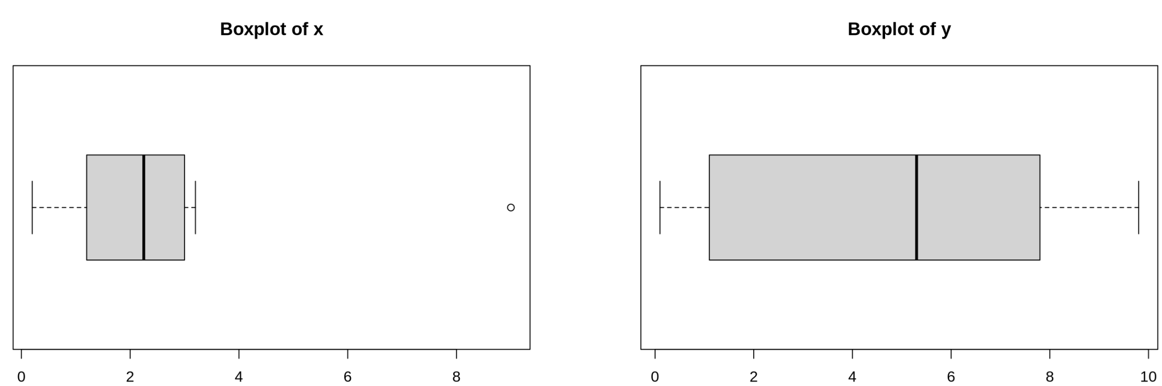
```
## [1] 4.568407
```

```
summary(y)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.100   1.400   5.300   4.871   7.575   9.800
```

```
var(y)
```

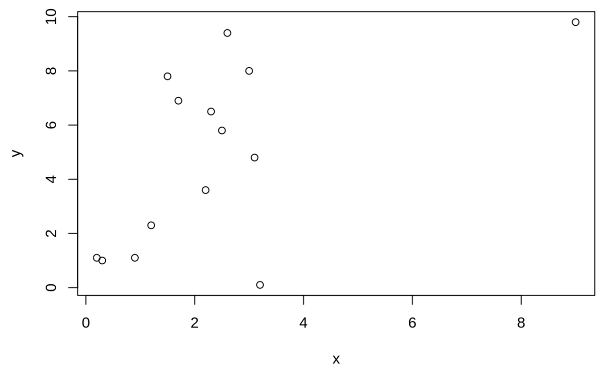
```
## [1] 11.17143
```



The only outlier is 9

There are no outliers

Problem 1.3




```
cor(x,y)
```

```
## [1] 0.5679153
```

The correlation coefficient presents that there is neither a weak or strong linear association between x_i and y_i

Problem 1.4

Yes, the outlier is the pair **(9.0, 9.8)** since the first coordinator (x) was previously identified as an outlier for all x_i's.

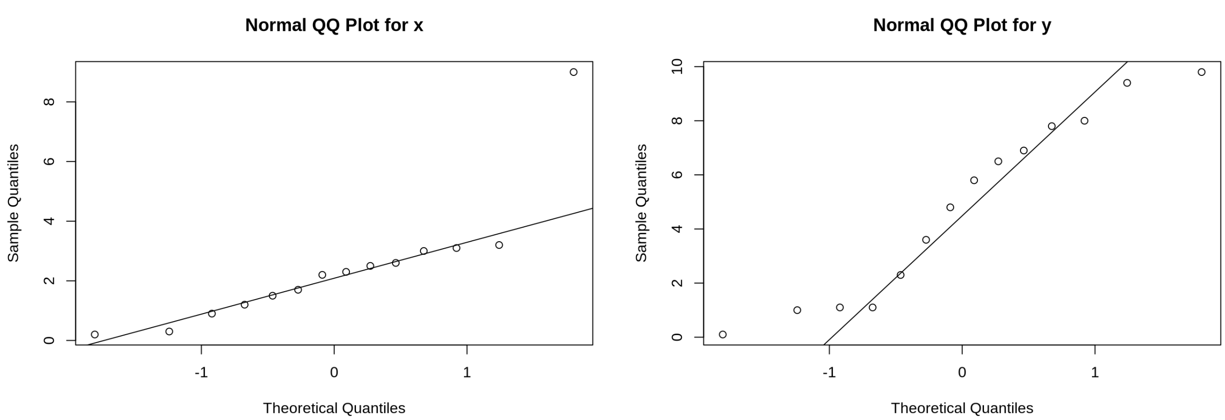
Removing the pair produces the following correlation coefficient:

```
## [1] 0.4586256
```

Problem 1.5

Upon removing the outlier in Problem 1.4, the correlation coefficient decreased and **weakened** the linear association between x_i and y_i.

Problem 1.6



The QQ plot of x_i's is most likely of normal distribution as there is minimal deviation from the straight diagonal line in contrast to the plot of y_i's

Problem 2

Given that F is the cdf for the random variable X, $F^{-1}(\alpha)$ is the inverse cdf or quantile function which determines the x value that is needed to achieve the probability α. With this in mind, the following probabilities can be determined:

$P(X \leq F^{-1}(\alpha/2)) = \alpha/2$
$P(X \leq F^{-1}(1-\alpha/2)) = 1-\alpha/2$
$P(F^{-1}(\alpha/2) \leq X \leq F^{-1}(1-\alpha/2)) = (1-\alpha/2-\alpha/2) = 1-\frac{2\alpha}{2} = 1-\alpha$

Given that $X \sim N(0,1)$ this establishes that X is normally distributed with a $\mu = 0$ and $\sigma = 1$, which is a standard normal distribution.

To calculate the following `qnorm()` and `pnorm()` will be used with $\mu = 0$ and $\sigma = 1$

$P(X \leq z_{0.05})$

```
pnorm(qnorm(0.05, 0, 1))
```

```
## [1] 0.05
```

$P(X \leq z_{0.05}) = 0.05$

$P(X \leq z_{1-0.05/2})$

```
pnorm(qnorm(1-0.025, 0, 1))
```

```
## [1] 0.975
```

$P(X \leq z_{1-0.05/2}) = 0.975$

$P(z_{0.05/2} \leq X \leq z_{1-0.05/2})$

$P(X \leq z_{0.05/2})$

```
pnorm(qnorm(0.05/2, 0, 1))
```

```
## [1] 0.025
```

$P(z_{0.05/2} \leq X \leq z_{1-0.05/2}) = 0.975 - 0.025 = 0.95$

Problem 3

Verify using algebraic manipulation:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

1. Expand out $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ to $\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x})$
2. $\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\mu + \overline{x})$ is equivalent to $\frac{\sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + \overline{x}^2\sum_{i=1}^{n} 1}{n}$
3. $\sum_{i=1}^{n} 1$ is the same as adding 1, n times and can be reduced to just n.
4. Simplified more: $\frac{\sum_{i=1}^{n} x_i^2}{n} - \frac{2\overline{x}\sum_{i=1}^{n} x_i}{n} + \frac{\overline{x}^2 n}{n}$
5. $\frac{\sum_{i=1}^{n} x_i}{n}$ is the sample mean and can be rewritten $\overline{x}$
6. Lastly, the simplifications result in: $\frac{\sum_{i=1}^{n} x_i^2}{n} - 2\overline{x}^2 + \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$, verifying equivalency.