Assignment 1 Question 1

Sheen Thusoo

Part c)

```
data <- read.csv("EconomicMobility.csv")
```

```
powerfun <- function(x, alpha) {
  if(sum(x <= 0) > 1) stop("x must be positive")
  if (alpha == 0)
    log(x)
  else if (alpha > 0) {
    x^alpha
  } else -x^alpha
}
```

```
L <- function(theta) {
  x <- data$Population
  y <- data$Commute
  power_x <- powerfun(x + 1, theta[1])
  power_y <- powerfun(y + 1, theta[2])
  attribute <- 1 - cor(power_x, power_y)^2
}
```

```
optimal_theta <- nlminb(start=c(1,1), objective=L)
optimal_theta$par
```

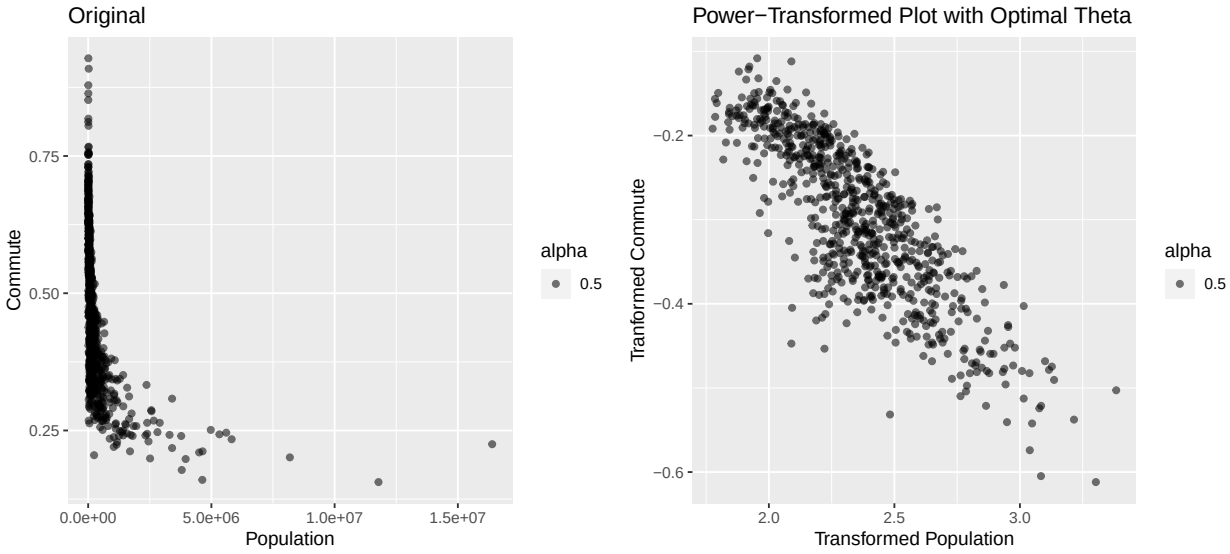
```
## [1]  0.07337527 -3.38815908
```

The optimum value of α_x is 0.07337527 and α_y is -3.38815908.

```
library("ggplot2")
library("gridExtra")
plot1 <- ggplot(data) +
  geom_point(
    aes(x = Population,
        y = Commute,
        alpha = 0.5)
  ) +
  labs(
    title = "Original",
    x = "Population",
    y = "Commute"
```

```
  )

transformed_pop <- powerfun(data$Population +1, optimal_theta$par[1])
transformed_commute <- powerfun(data$Commute +1, optimal_theta$par[2])
plot2 <- ggplot(data) +
  geom_point(
    aes(x = transformed_pop,
        y = transformed_commute,
        alpha = 0.5)
  ) +
  labs(
    title = "Power-Transformed Plot with Optimal Theta",
    x = "Transformed Population",
    y = "Tranformed Commute"
  )
grid.arrange(plot1, plot2, nrow = 1)
```




```
original_correlation <- cor(data$Population, data$Commute, method = "pearson")
message("The correlation coefficient for the original data is ", original_correlation)
```

```
## The correlation coefficient for the original data is -0.414150334705171
```

```
transformed_correlation <- cor(transformed_pop, transformed_commute, method = "pearson")
message("The correlation coefficient for the transformed data is ", transformed_correlation)
```

```
## The correlation coefficient for the transformed data is -0.825299449710719
```