GRADUATE STUDENT STAT 840 A2

Vsevolod Ladtchenko 20895137

Problem 3

a)

$$\hat{\theta} = T(\hat{F}_n) = \frac{\bar{Y}}{\bar{X}} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i}{\frac{1}{n}\sum_{i=1}^n X_i} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}$$

```
x = c(76, 138, 67, 29, 381, 23, 37, 120, 61, 387, 93, 172, 78, 66, 60, 46, 2, 507, 179, 121, 50, 44, 77

y = c(80, 143, 67, 50, 464, 48, 63, 115, 69, 459, 104, 183, 106, 86, 57, 65, 50, 634, 260, 113, 64, 58,

theta_hat = sum(y) / sum(x)
theta_hat
```

```
## [1] 1.239019
```

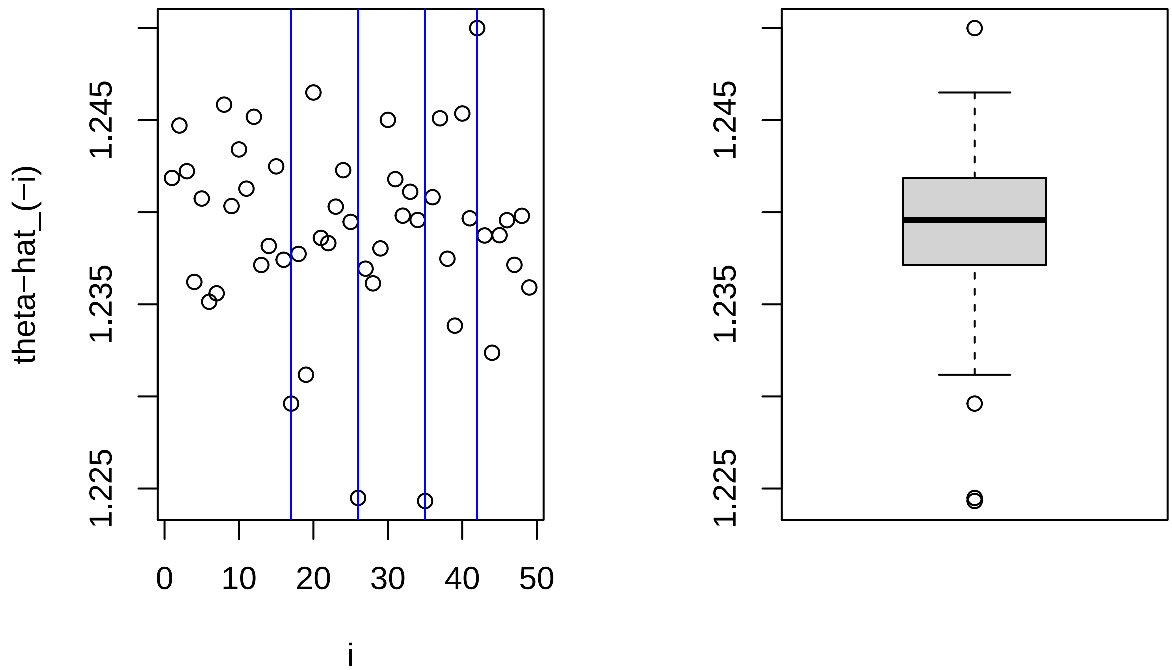
b)

```
n = length(x)

jack = rep(NA, n)
for (i in 1:n)
{
  jack[i] = sum(y[-i]) / sum(x[-i])
}
theta_bar = mean(jack)
theta_jack = n*theta_hat - (n-1)*theta_bar
theta_jack
```

```
## [1] 1.237297
```

```
par(mfrow = c(1,2))
plot(jack, xlab="i", ylab="theta-hat_(-i)")
idx = c(17, 26, 35, 42)
abline(v=idx, col="blue")
boxplot(jack)
```




```
x[idx]
```

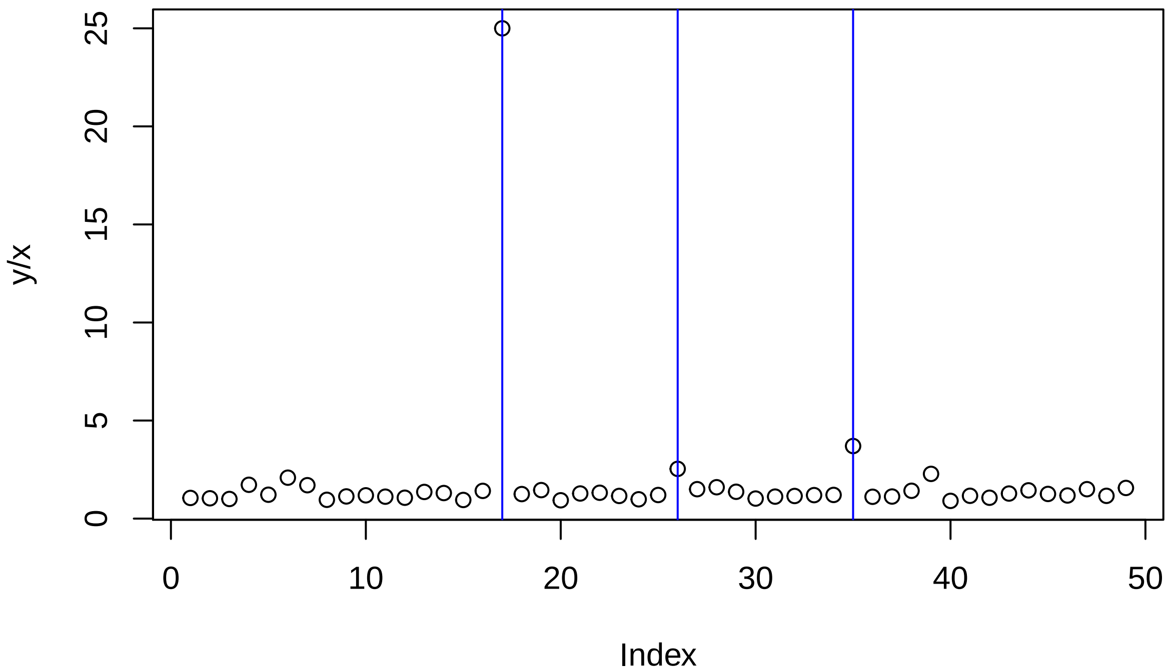
```
## [1]   2  56  30 298
```

```
y[idx]
```

```
## [1]  50 142 111 317
```

We can use a boxplot to isolate the outliers. The first 3 points are outliers from below, meaning that removing the point makes Y lower and X higher. This means the points themselves have Y high, and X low. This corresponds to cities that show large growth. Plotting Y/X confirms this. Including these points brings the average ratio much higher due to their large growth.

```
par(mfrow = c(1,1))
plot(y/x)
abline(v=c(17, 26, 35), col="blue")
```



The other outlier is from above, meaning removing it makes the estimate become higher. This city has a high population but its growth ratio is 1.06, much smaller than $\hat{\theta}$ at 1.23, so its inclusion brings the average down.

```
y[42] / x[42]
```

```
## [1] 1.063758
```

c)

```
se_jack = sqrt(((n-1) / n) * sum((jack - theta_bar)^2)) # notes 3 p 20
ci_jack = c(L=theta_hat + qnorm(0.025)*se_jack, U=theta_hat + qnorm(0.975)*se_jack)
round(ci_jack, 4)
```

```
##      L      U
## 1.1713 1.3067
```