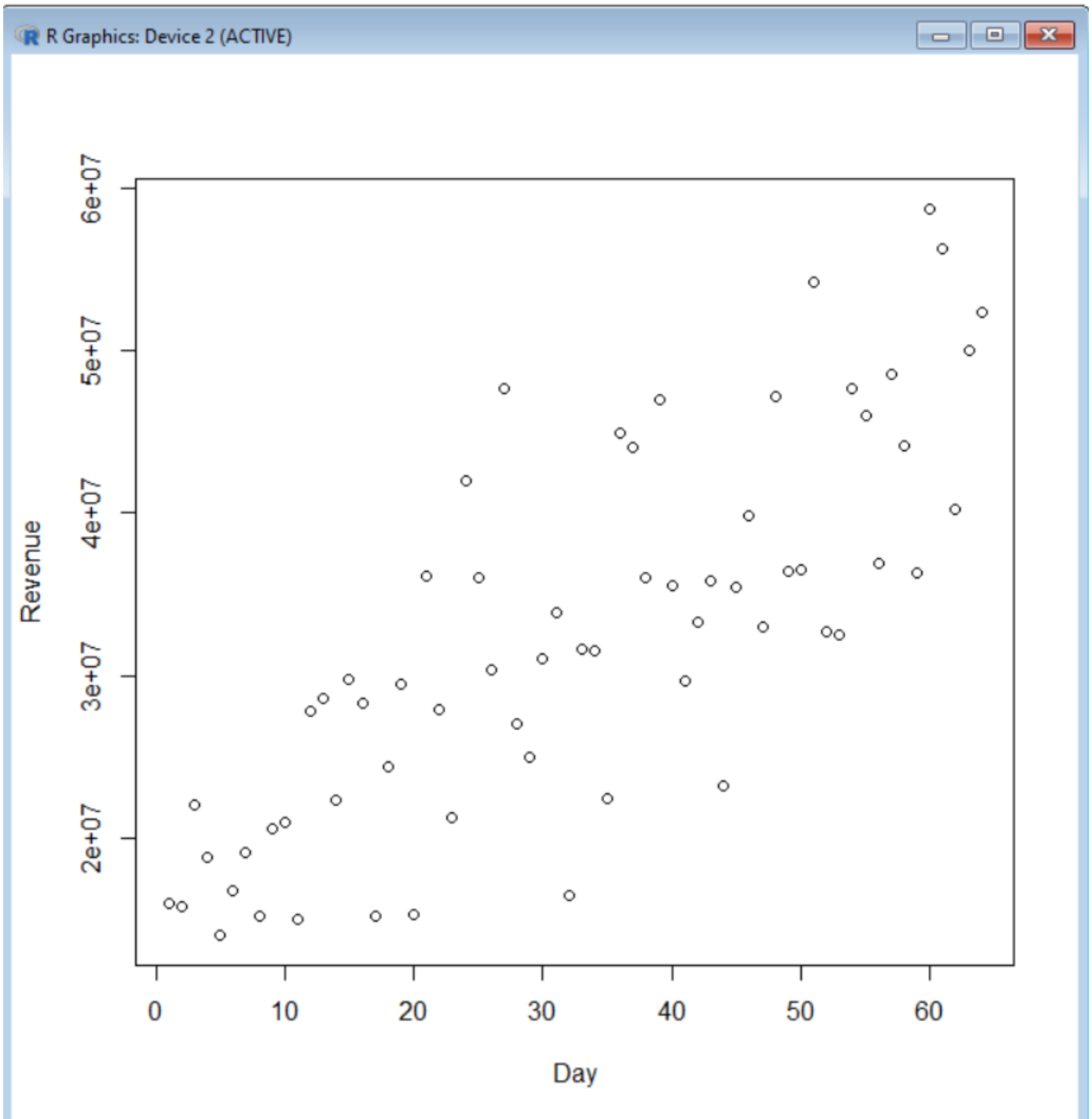
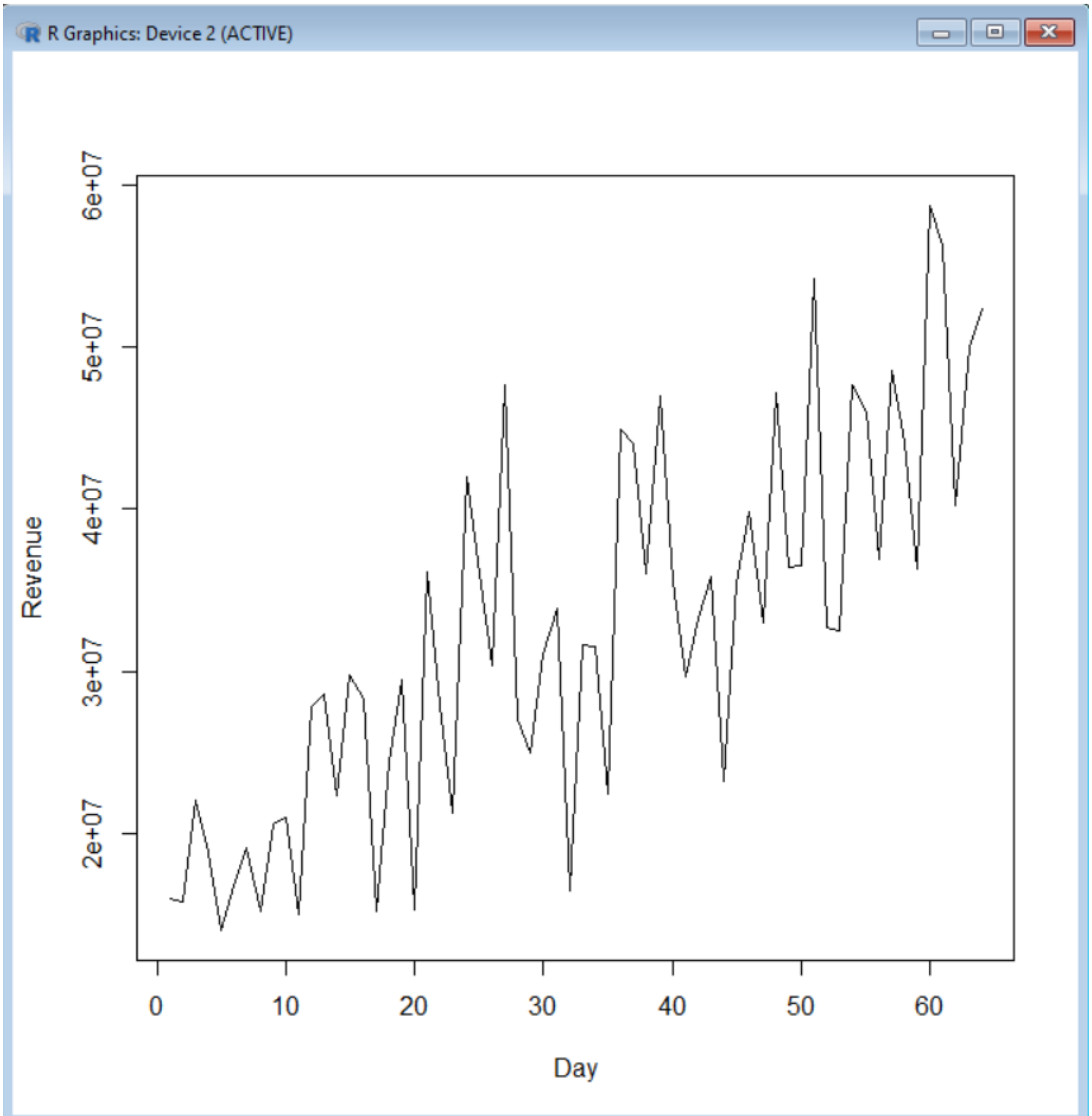
1. Download the data set “Revenue.csv” into your working directory.

a) Draw a scatter plot. X-axis is the “Day” and Y-axis is the “Revenue”.

```
> df <- read.table("C:\\Users\\Udaya Vijay Anand\\Downloads\\Revenue.csv", header=T)
> head(DF)
      Period Day  Revenue Sales_quantity Average_cost The_average_annual_payroll_of_the_region
1 01.01.2015   1 16010072          12729     1257.764                                 30024676
2 01.02.2015   2 15807587          11636     1358.507                                 30024676
3 01.03.2015   3 22047146          15922     1384.697                                 30024676
4 01.04.2015   4 18814583          15227     1235.607                                 30024676
5 01.05.2015   5 14021480           8620     1626.622                                 30024676
6 01.06.2015   6 16783929          13160     1275.375                                 30024676
> |
```



b) Draw a line plot.



c) Looking at the plots in a) and b) do you observe any trends or patterns?

Observing the scatter plot and line plot in a) and b), it can be observed that the revenue exhibits a positive trend as time passes, although there are noticeable fluctuations in the data.

Question 2

The following data set shows the heights and the arm spans of 10 persons.

Height (cm) 178.5 | 176 | 185.6 | 179 | 172 | 167.7 | 174.3 | 136.5 | 164.5 | 165.5

Arm Span (cm) 183.5 | 170 | 194.1 | 181 | 181 | 171.7 | 171 | 135.5 | 164 | 160.5

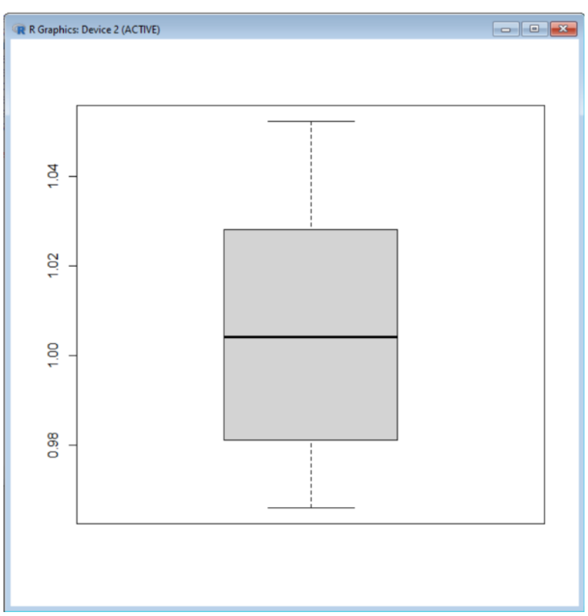
Create a data frame named mydata using “mydata <-data.frame(height,armspan)”. Here note that you have to create two vectors height, armspan manually.

a) Add a new variable, “apeindex”, as the ratio between armspan and height, to your data. Show your new data set.

```
> height <- c(178.5, 176, 185.6, 179, 172, 167.7, 174.3, 136.5, 164.5, 165.5)
> armspan <- c(183.5, 170, 194.1, 181, 181, 171.7, 171, 135.5, 164, 160.5)
> mydata <- data.frame(height, armspan)
> mydata$apeindex <- mydata$armspan / mydata$height
> mydata$apeindex <- mydata$armspan / mydata$height
> mydata
   height armspan  apeindex
1   178.5   183.5 1.0280112
2   176.0   170.0 0.9659091
3   185.6   194.1 1.0457974
4   179.0   181.0 1.0111732
5   172.0   181.0 1.0523256
6   167.7   171.7 1.0238521
7   174.3   171.0 0.9810671
8   136.5   135.5 0.9926740
9   164.5   164.0 0.9969605
10  165.5   160.5 0.9697885
> 
```

b) Attach a boxplot of apeindex. Are there any outliers?

```
> boxplot(mydata$apeindex)
> 
```



c) Find the min, Q1, Q2, Q3, and max of apeindex.

```
> summary(mydata$apeindex)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.9659  0.9840  1.0041  1.0068  1.0270  1.0523 
> 
```

d) Find IQR, Q1-1.5*IQR, and Q3+1.5*IQR.

```
> IQR(mydata$apeindex)
[1] 0.04300259
> Q1 <- quantile(mydata$apeindex, 0.25)
> Q3 <- quantile(mydata$apeindex, 0.75)
> lower <- Q1 - 1.5*IQR(mydata$apeindex)
> upper <- Q3 + 1.5*IQR(mydata$apeindex)
> 
```

```
> lower
     25% 
0.919465 
> upper
     75% 
1.091475 
> 
```

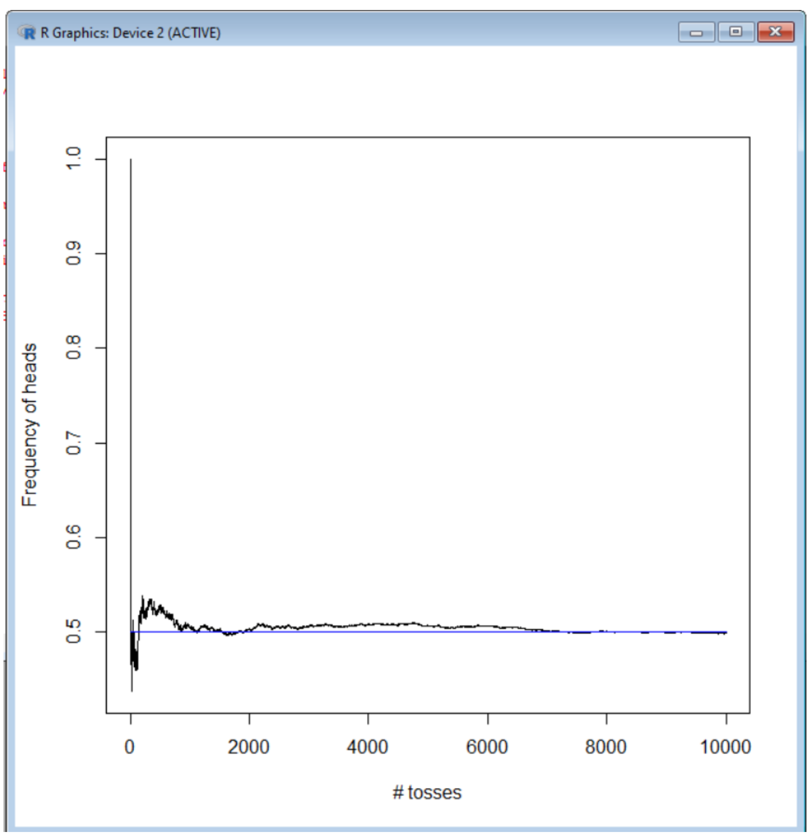
e) What is the lower adjacent value (LAV) and upper adjacent value (UAV)? (Note that they're not necessarily equal to Q1-1.5*IQR or Q3+1.5*IQR!!!)

```
> LAV <- min(mydata$apeindex[mydata$apeindex > lower])
> UAV <- max(mydata$apeindex[mydata$apeindex < upper])
> 
> LAV
[1] 0.9659091
> UAV
[1] 1.052326
> 
```

3.Consider an unfair coin with a chance of 1/5 to get head and 4/5 to get tail in each toss

a). For 3 independent tosses, what are the outcomes in terms of heads (H) and tails (T) and their relative frequencies? Draw a line plot.

```
> # Simulate 3 coin tosses
> A <- sample(c(0, 1), 3, replace = TRUE)
> A
[1] 0 0 0
>
> # Calculate relative frequency of heads in A
> Rpre <- sum(A == 1) / 3
> Rpre
[1] 0
>
> # Simulate 10000 coin tosses
> X <- sample(c(0, 1), 10000, replace = TRUE)
> Rpreq <- sum(X == 1) / 10000
> Rpreq
[1] 0.4989
>
> # Calculate relative frequency of heads as tosses increase
> Headnumbs <- cumsum(X == 1)
> Proportions <- Headnumbs / (1:10000)
>
> # Plot the relative frequencies
> plot(1:10000, Proportions, type = "l", xlab = "# tosses", ylab = "Frequency of heads")
>
> # Add a horizontal line at y=0.5
> lines(1:10000, rep(0.5, 10000), col = "blue")
>
```



b). For 1000 independent tosses, draw a plot to see how the relative frequency converges

```
> # Define the probabilities of getting heads and tails
> p_H <- 1/5
> p_T <- 4/5
>
> # Simulate 1000 coin tosses with the given probabilities
> outcomes <- sample(c("H", "T"), size = 1000, replace = TRUE, prob = c(p_H, p_T))
>
> # Calculate the relative frequency of heads as the number of tosses increases
> r_freq <- cumsum(outcomes == "H") / (1:1000)
>
> # Plot the relative frequency as a function of the number of tosses
> plot(r_freq, type = "l", xlab = "Toss Number", ylab = "Relative Frequency of H", main = "Relative Frequency Convergenc$
> |
```