Due: 22nd January 2024,10pm

2. Shamsheen has a log book of time spent by them on Instagram. In the log book they keep track of the 24-hour reading before each time they log in. The last 10 readings are:

$$8, \quad 8, \quad 3, \quad 12, \quad 13, \quad 10, \quad 1, \quad 8, \quad 7$$

(a) Enter these numbers into R as a variable `instasee`. Use the function `diff` on the data. What does it give?

```
> instasee = c(8, 8, 3, 12, 13, 10, 1, 8, 7)
> differences = diff(instasee)
```

Write down, x, the number of hours between each time Shamsheen logs into Instagram.

Solution: Recomputing the above commands:

```
> instasee = c(8, 8, 3, 12, 13, 10, 1, 8, 7)
> differences = diff(instasee)
> differences
[1] 0 -5 9 1 -3 -9 7 -1
```

The command `diff` computes the iterated differences with a default `lag=1`. For example if

```
> differences2 = diff(instasee, lag =2)
> differences2
[1] -5 4 10 -2 -12 -2 6
```

x is the vector `differences`. So Shamsheen logs into Instagram successively $0, -5, 9, 1, -3, -9, 7, -1$ hours. ■

(b) Use the `max` to find the maximum number of hours, the `mean` function to find the average number of hours and the `min` to get the minimum number of hours for Shamasheen between two logins.

Solution: As indicated in the question the execution of the respective commands on the vector differences will yield the answer.

```
> MaximumHours=max(differences)
 > MaximumNumberHours
 [1] 9
```

Maximum number of hours for Shamasheen between two logins is 9

```
> AverageNumberHours=mean(differences)
 > AverageNumberHours
 [1] -0.125
```

Average number of hours for Shamasheen between two logins is -0.125

```
> MinimumNumberHours=min(differences)
 > MinimumNumberHours
 [1] -9
```

Minimum number of hours for Shamasheen between two logins is -9

3. Super-Suppandi scores in various quizzes in test of a lifetime are given below

   7, 6, 10, 8, 7, 9, 9, 6, 4, 10, 8, 6, 9, 10

   (a) Enter this into R as a variable `scoreSS`. Use the function `max` to find the highest score, the function `mean` to find the average and the function `min` to find the minimum.

   Solution:

```
> scoreSS = c( 7, 6, 10, 8, 7, 9, 9, 6, 4, 10, 8, 6, 9, 10)
 > HighestScore = max(scoreSS)
 > HighestScore
 [1] 10
```

Super-Suppandi highest scores is 10

```
 > AverageHighestScore = mean(scoreSS)
 > AverageHighestScore
 [1] 7.785714
```

Super-Suppandi average scores is 7.785714

```
 > MinimumScore = min(scoreSS)
 > MinimumScore
```

```
[1] 4
```

Super-Suppandi minimum scores is 4

(b) When confronted by instructor, SS realises that entry 4 was a mistake. It should have been 5. How can you fix this? Do so, and then find the new average.

Solution: We can fix this by changing the entry 4 which appears at position 9 to 5 and then recomputing the average time

```
> Newscore = scoreSS
> Newscore
 [1] 7 6 10 8 7 9 9 6 4 10 8 6 9 10
> Newscore[9]=5
 > Newscore
 [1] 7 6 10 8 7 9 9 6 5 10 8 6 9 10
```

(c) What does the below command provide in R ?

```
> sum( scoreSS >= 9)
```

Solution : The command counts the number of entries in the vector scoreSS that are greater than or equal to 9

```
> sum( scoreSS >= 9)
 [1] 6
```

(d) What do you get? What percent of your scores are less than 17 ? How can you answer this with R?
Solution: We see that all times are less than 17. If we were to do it in R then to calculate the percent of scores that are less than 17 minuteswe can first count the number that are below 17 and divide by the length, times a 100.

```
> count = sum( scoreSS < 17)
> Length = length(scoreSS)
> Percent = count/Length*100
```

The percent of your scores are less than 17 is given below

```
> Percent
 [1] 100
```

4. Shreelakshmi's cell phone bill varies from month to month. Suppose in their first year of online Data Science program, under the Drop-atmost 10-calls monthly plan, the following monthly amounts were incurred:

   460, 330, 390, 370, 460, 300, 480, 320, 490, 350, 300, 480

   (a) Enter this data into a variable called `Shreelakshmibill`. Use the `sum` command to find the amount spent by Shreelakshmi that year on the cell phone.

```
> Shreelakshmibill = c(460, 330, 390, 370, 460, 300, 480, 320, 490, 350, 300, 480)
> SumShreelakshmibill = sum(Shreelakshmibill)
> SumShreelakshmibill
[1] 4730
```

   Solution: The amount spent by Shreelakshmi that year on the cell phone is 4730

   (b) Using R find out what is the smallest amount they spent in a month and the largest amount they spent in a month ?

```
> SmallestBillAmount = min(Shreelakshmibill)
> SmallestBillAmount
[1] 300

> LargestBillAmount = max(Shreelakshmibill)
> LargestBillAmount
[1] 490
```

   Solution: The smallest amount they spent in a month

```
[1] 300
```

   The largest amount they spent in a month

```
[1] 490
```

   (c) How many months was the amount greater than Rs. 400? What percentage was this?

   Solution: First we count the number of months where the bill amount was greater than Rs. 400.

```
> BillCount = sum( Shreelakshmibill > 400)
> > BillCount
> [1] 5
```

To calculate the percent of bill amount that are greater than Rs. 400, we have to first count the number that are above Rs. 400 and divide by the length, times a 100.

```
> BillCount = sum( Shreelakshmibill > 400)
> Length = length(Shreelakshmibill)
> Percent = BillCount/Length*100
> Percent
[1] 41.66667
```

The number of months, Shreelakshmi's bill amount was greater than Rs. 400 is

```
[1] 5
```

The percent is

```
[1] 41.66667
```

(d) If her monthly loan from NOmoney Bank was Rs 3000. Using R store their balance(after paying their phone bill) in a variable called `freemoney`. Find the average amount available each month for other expenses.

Solution: We can use the vector subtraction effect in R to calculate freemoney

```
> MonthlyLoan = 3000
> freemoney = 3000-Shreelakshmibill
> AverageFreemoneyPerMonth= mean(freemoney)
```

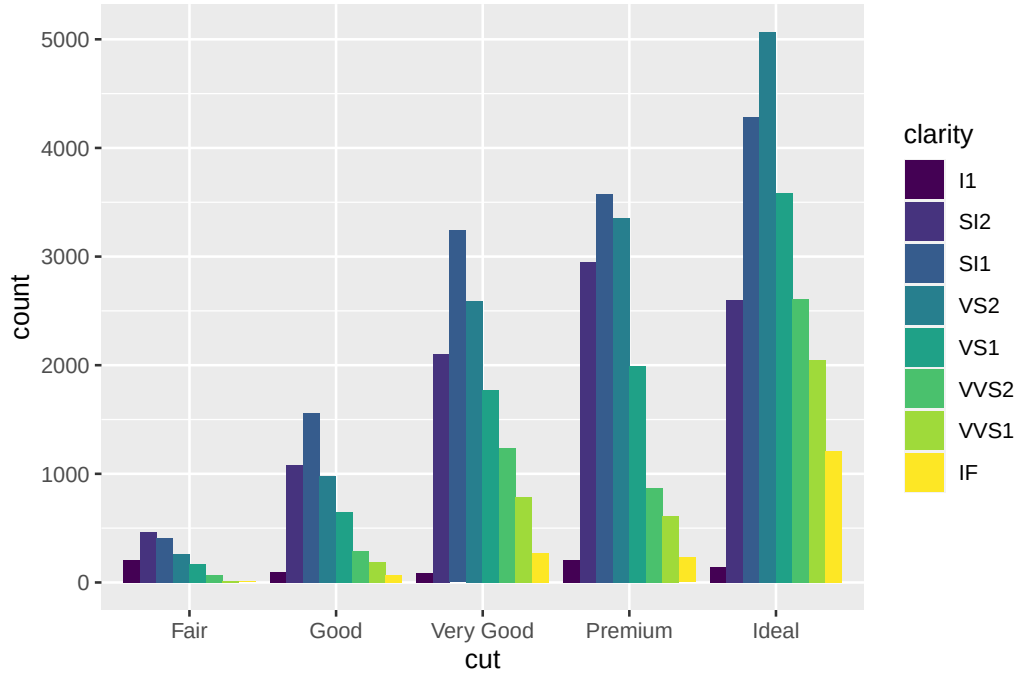
Then the Freemoney is

```
> freemoney
[1] 2540 2670 2610 2630 2540 2700 2520 2680 2510 2650 2700 2520
> AveragefreemoneyPerMonth
[1] 2605.833
```

5. The below figure is from the dataset `diamonds`.

(a) Write a couple of sentences describing the data set.
Solution: A dataset containing the Prices of over 50,000 round cut diamonds. Here the variables are diamonds, price, carat, color, clarity, length, width, depth.

(b) Write an R-command that uses `ggplot` and `geom_bar` to produce the following figure.



Solution: The R-code is given below

```
> ggplot(data = diamonds) +
+ geom_bar(mapping = aes(x = cut, fill = clarity), position="dodge")
+ scale_fill_viridis_d()
```

(c) Write (one sentence) an observation you can draw from the bar graph.

Solution: Good clarity diamonds will be increased for Fair cut.