Assignment 1: Heavenly Chocolate Website Transactions

1.

| Summary measure	Time (min)	Pages Viewed	Amount Spent ($)
Mean	12.81	4.82	68.13
Median	11.4	4.5	62.15
Standard deviation	6.06	2.04	32.34
Variance	36.76	4.15	1046.12
Range	28.6	8	140.67
Minimum	4.3	2	17.84
Maximum	32.9	10	158.51

Range = Maximum - Minimum

Range (Time(min)) = [4.3,32.9] = 32.9 - 4.3 = 28.6

Range (Pages Viewed) = [2,10] = 10 - 2 = 8

Range (Amount Spent ($)) = [17.84,158.51] = 158.51 − 17.84 = 140.67

These descriptive statistics provide insights such as:

1. Customers spending more time on the website are likely to spend more.
2. Customers spend an average of 12.81 min on the website before making a purchase.
3. Minimum time spent by the customer on the website is 4.3 minutes, whereas the maximum time spent by the customer on the website is 32.9 minutes.
4. The below histogram for time spent shows that it is skewed to the right. Hence the mean is more than the median.

Histogram of Time Spent

Time in Minutes
Column Time(min)
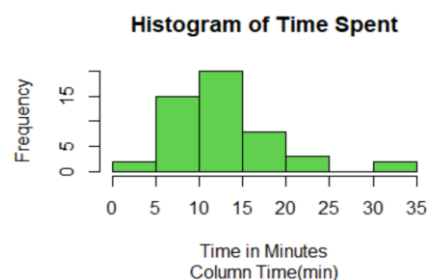
5. Each customer on the website spends a minimum of $17.84.
6. The median amount spent by the customer on the website is $62.15, whereas the average amount is $ 68.13.
7. The range of the amount spent on the website is $17.84 to $158.51.
8. The below histogram for the amount spent shows that it is skewed to the right. Hence mean is more than the median.

Histogram of Amount spent

Frequency vs Amount in $ / Amount spent ($)
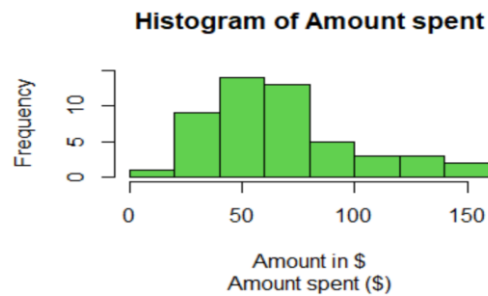
9. Each customer on the website visits at least 2 pages before making a purchase.
10. Customers on the website view 4.82 pages on average, whereas the maximum number of pages viewed by the customers is 10.
11. The below histogram for the amount pages shows that it is skewed to the right. Hence mean is more than the median.

Histogram of Pages Viewed

Frequency vs Pages / Pages Viewed
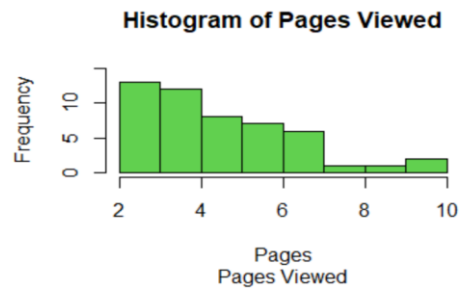
2. Justification

1. It can be clearly seen from the below scatter plot that there is a positive linear relationship between pages viewed by the customers on the website and the amount spent ($) by the customers on the website.

Scatter Plot

Amount Spent vs Pages Viewed
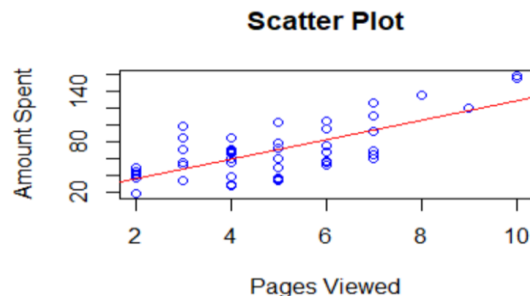
2. The covariance is 47.685, this indicates that there is a positive relationship between pages viewed and the amount spent as the value is positive.

3. The correlation coefficient is 0.724, this indicates that there is a positive relationship between pages viewed and the amount spent as the value is positive.

Hence, we can conclude that the number of pages viewed by the customer on the website is related to the amount spent on the website. This infers that the customers who view more pages on the website also spend more money on the website.

3.

Histogram of Time Spent

Time in Minutes
Column Time(min)
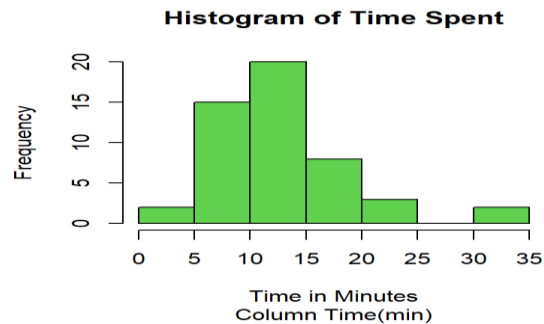
1. The time spent on the website is distributed in the range of 4.3 min to 32.9 min.
2. The below histogram for time spent shows that it is skewed to the right. Hence the skewness is positive also this indicates that the mean is more than the median.

4.

1. The below table shows the joint frequencies of the two categorical variables in the dataset using cross tabulation.

	Chrome	Firefox	Other	Sum
Fri	4	6	1	11
Mon	5	2	2	9
Sat	3	3	1	7
Sun	3	1	1	5
Thu	4	1	0	5
Tue	4	2	1	7
Wed	4	1	1	6
Sum	27	16	7	50

1. Out of 7 days of the week Fridays have the highest number of customers and they use the Firefox browser the most.
2. Usage of other browsers is 0 on Thursdays.
3. Number of customers visiting the website is less on Sundays and Thursdays.

2. The below table shows frequencies as percentages of row total.

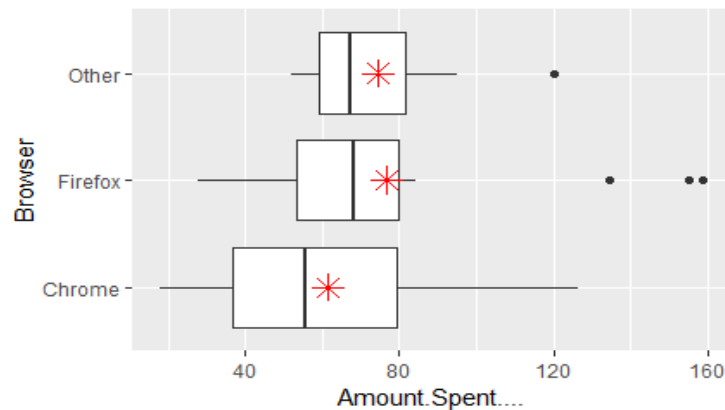|	Chrome	Firefox	Other
Fri	36.364	54.545	9.091
Mon	55.556	22.222	22.222
Sat	42.857	42.857	14.286
Sun	60.000	20.000	20.000
Thu	80.000	20.000	0.000
Tue	57.143	28.571	14.286
Wed	66.667	16.667	16.667

1. On Friday, the Firefox browser is used more by the customers followed by chrome and then other browsers.
2. On Monday, Tuesday, Wednesday, Thursday, and Sunday the Chrome browser is used more by the customers followed by Firefox and then other browsers.
3. On Saturday, chrome, and Firefox are used equally by the Customers, and then another browser.
4. Out of 7 days of the week, Chrome browser is used the maximum by the customers, i.e., for 5 days of the week followed by the Firefox browser and then other browsers.

3. The below table shows frequencies as percentages of column total

```
        Chrome Firefox   Other
Fri 14.815  37.500 14.286
Mon 18.519  12.500 28.571
Sat 11.111  18.750 14.286
Sun 11.111   6.250 14.286
Thu 14.815   6.250  0.000
Tue 14.815  12.500 14.286
Wed 14.815   6.250 14.286
```

1. Chrome is used maximum on Mondays and minimum on Saturdays and Sundays
2. Firefox is used maximum on Fridays and minimum on Sundays, Wednesdays, Thursdays.
3. Other browsers are used maximum on Mondays and minimum on Thursdays.

5.

	Chrome	Firefox	Other
Minimum	17.840	27.910	52.150
First Quartile	36.825	51.660	59.100
Median	55.580	67.985	67.440

Third Quartile	79.275	81.375	81.675
Maximum	126.400	84.170	94.900

1. **Location**: The above Box Plot shows that the median amount spent by chrome users is less than firefox and other browser users.
2. **Dispersion:** The Interquartile range of chrome users is evidently greater (as shown by the length of the boxes). The overall range of the data set shows the similar trend indicating that there is a greater variability in the amount spent for chrome users as compared to firefox and other browser users.
3. **Skewness:** It can be seen from the box plot shown above that the data on amount spent for chrome and other browser users is slightly positive skewed (mean > median).
4. **Potential Outliers:** It can be seen in the box plot shown above that there are no far out values in the amount spent by chrome users that require a closer look. On the other hand, there are visible outliers for Firefox and Other browser users. These values are considered outliers as it is higher than (Q3 + 1.5IQR).