STAT1008 ASSIGNMENT 1

**1. a.**

**Rows**: Each row corresponds to the 'overall day number', with the other data in the row being the measurements of the other categories on that particular day.

**X**: This column refers to the day the data was recorded for that row, and is of the type 'count'. This is because the days are counted as a whole, increasing \*only\* by 1 each time. A half-day or some other such way of measuring it is not possible under this method of measurement, making it a 'count' type.

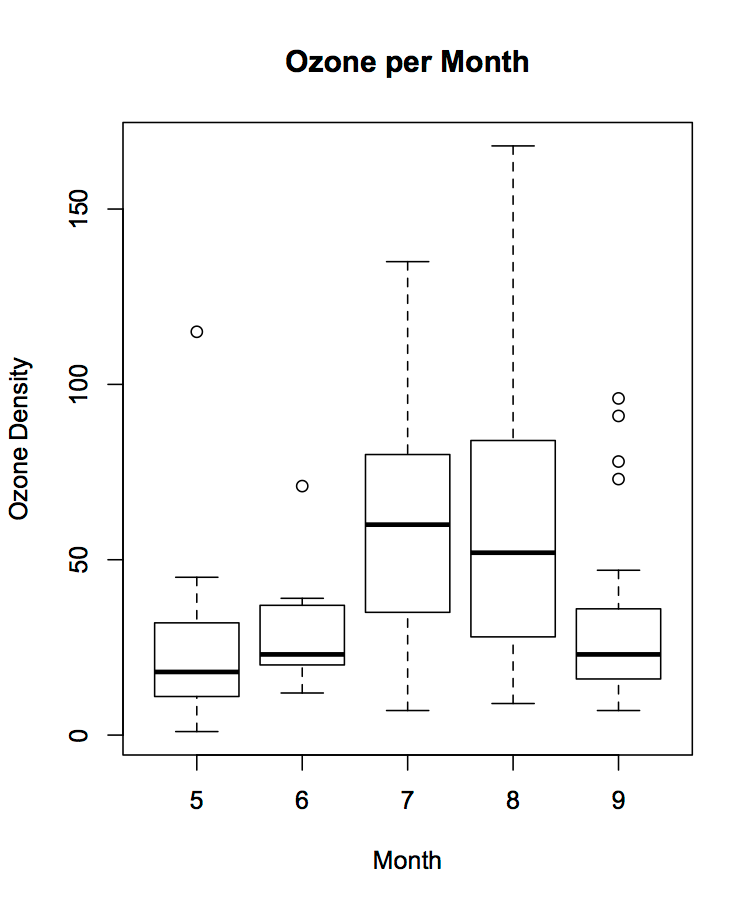
**Ozone**: This column contains the data for the Ozone density, in whatever unit this is measured in. As the data contains no decimal data, it is assumed to be 'count' data as well, as if it were not the recordings would not be whole integers. The data could be continuous and have been rounded, but this is not specified and thus it is safer to assume it as count rather than continuous.

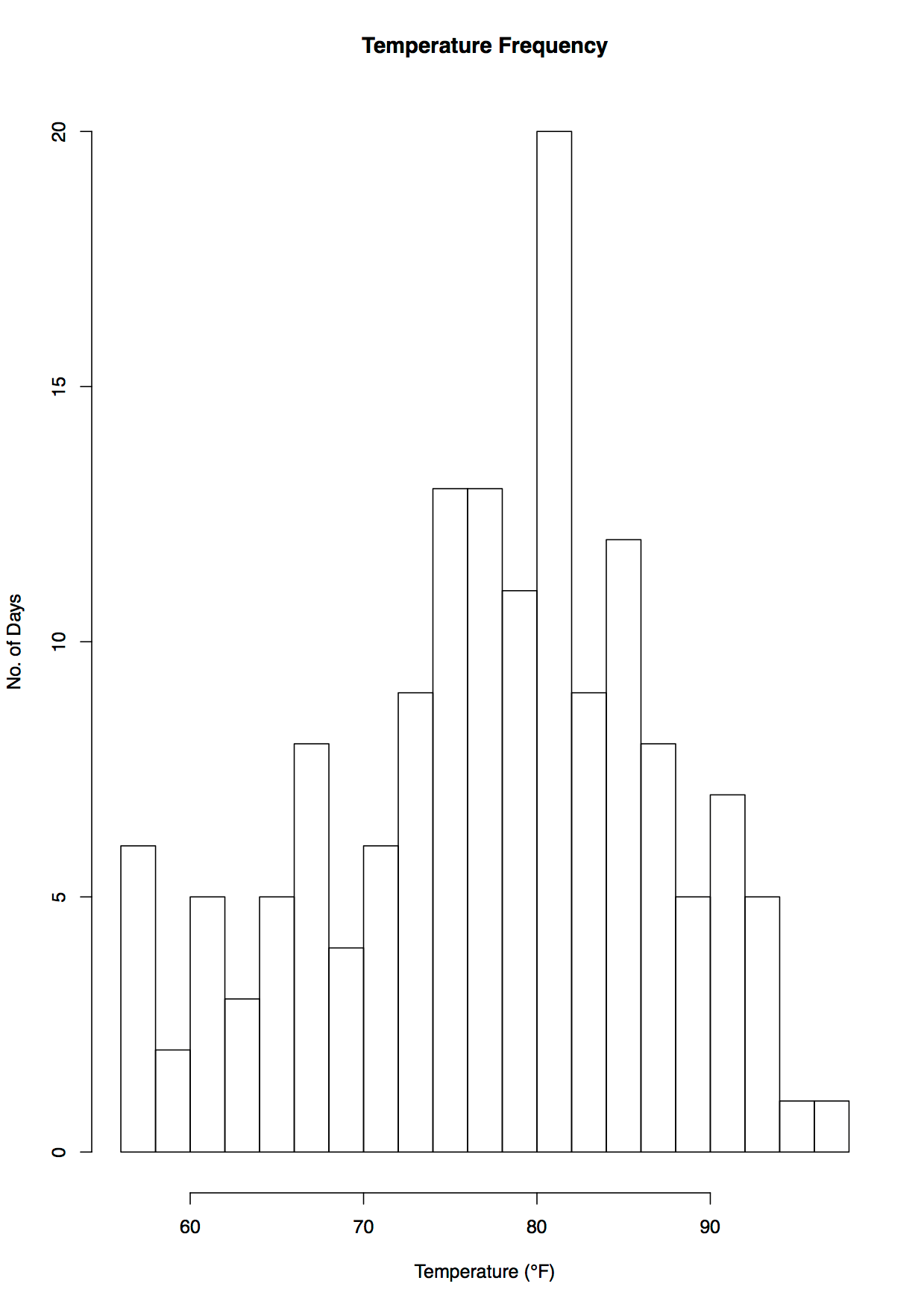
**Solar.R**: Contains data for the Solar Radiation on the particular day. Again, all of the entries are integers, so 'count' is assumed. However, due to the volatility of radiation, common sense would indicate that this data should be continuous. Nevertheless, no decimals are seen, and it is safer to assume ‘count'.

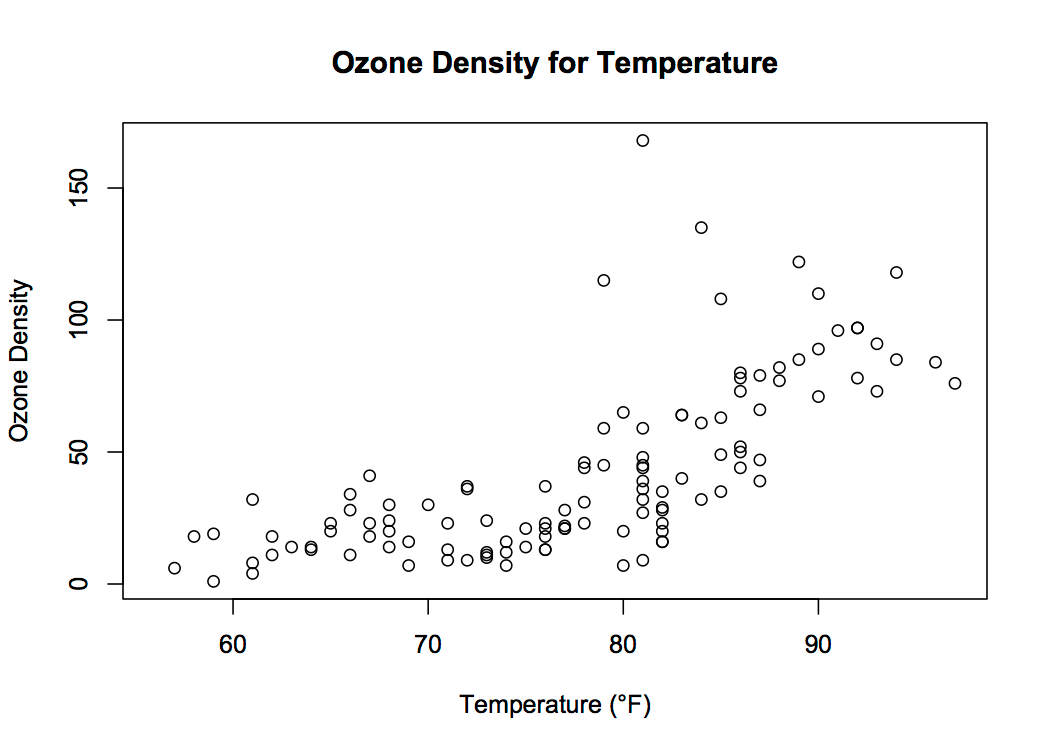
**Wind**: The data for the wind speed in miles per hour. There are an infinite number of speeds wind can travel at, and speed must be measured rather than counted, so the data is of type 'continuous'.

**Temp**: The temperature measured in Fahrenheit, again continuous.

**b.** The shape of this histogram centres around the 80°F area, and infers that the mean temperature is between 80°-85°F. Temperatures below 80°F are also appear to be more frequent than temperatures above 80°F.



**c.** The box plots show that the ozone density was much higher in the months of July and August than in other months, both in mean and variance. The ozone densities in May, June and September are all very similar, with those in June covering a much smaller range than the other months. However, the lower quartiles of all months are much more similar than the upper quartiles. July had the smallest variance, and August the largest.

****

**d.** This scatterplot shows the density of ozone increasing along with the increasing temperature. From this it can be concluded that ozone density and temperature are positively related, as the magnitude is positive for both variables.

**e.** Covariance is -70.9385.

**f.** See attached working.

**g.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Month | May | June | July | August | September |
| Correlation | 0.5541 | 0.6683 | 0.7227 | 0.5979 | 0.8281 |

The person is correct that the correlation between these variables is different for July and August, but the correlation for ALL months is different, so it is of little worth to note especially these two months.