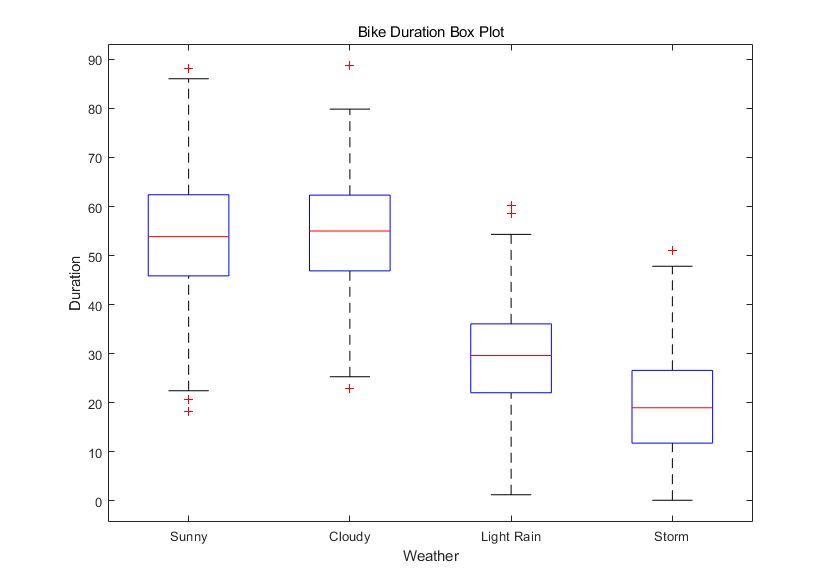
b) The summarized average bike ride duration table is listed below for reference:

|  |  |  |
| --- | --- | --- |
| mean\_duration | weather | GroupCount |
| 54.587 | 'Cloudy' | 349 |
| 395 | 'Light Rain' | 29.533 |
| 19.439 | 'Storm' | 214 |
| 54.15 | 'Sunny' | 1014 |

The boxplot is also listed below:



c) In this part, I have conducted 2-samples t-test for each pair of duration values in each weather group. The results are collected in the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sunny | Cloudy | Light Rain | Storm |
| Sunny | 0 | 0 | 1 | 1 |
| Cloudy | 0 | 0 | 1 | 1 |
| Light Rain | 1 | 1 | 0 | 1 |
| Storm | 1 | 1 | 1 | 0 |

Let me explain this table. First, I have splitted the duration data into 4 groups based on the weather type. The for each pair, e.g. Sunny and Cloudy groups, I conducted a 2-samples t-test on them, and the resulted h value is recorded in the corresponding position in the table above. Of course, this table is symmetric, and all values on the diagonal are 0's, because the same groups' mean are the same hypothesis can not be rejected. So if a value is 0, the original hypothesis that the two group means are the same can not be rejected under 95% confidence level; if the value is 1, the original hypothesis is rejected, and the group means are probably not the same under 95% confidence level.

So, from this table, we can observe that only when the weather is sunny or cloudy, the bike ride durations are not apparently different under 95% confidence level. Otherwise, the bike durations are quite different under 95% confidence level.