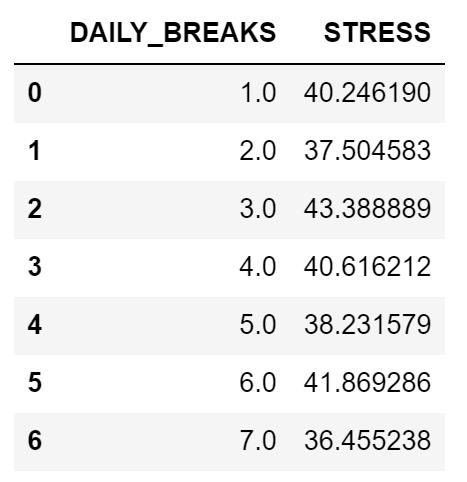
**Team 1**

Deliverable 2

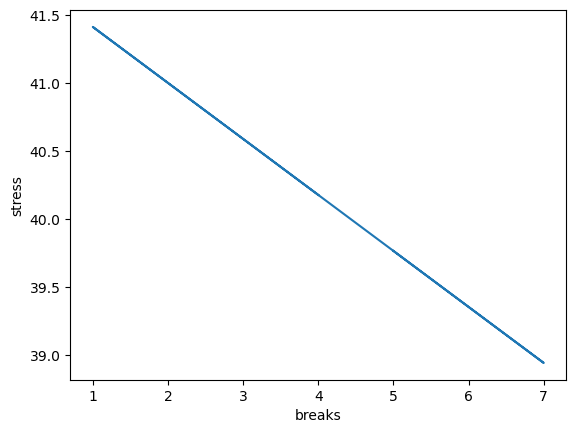
Name: Ming-Han Hsieh, Wei-TSE KAO

**Hypothesis d:** Participants' stress algorithm will be inversely correlated to their number of breaks.

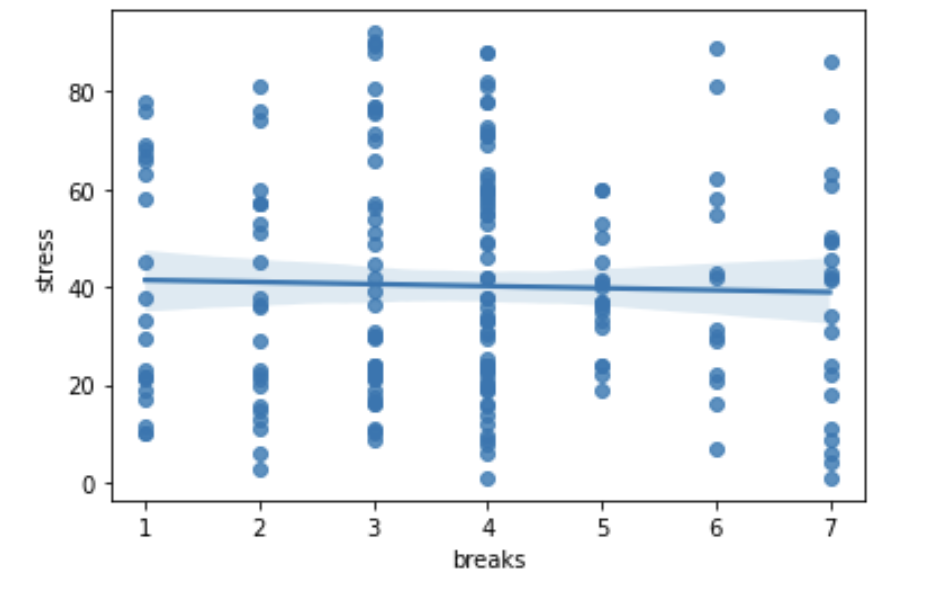
**Preprocessing:**

We dropped all the 0 stress value since we can’t confirm that the value is actually 0 or is missing. 

Then we plot the linear regression line when daily breaks in the x axis and stress in y axis.



We can clearly see the inverse relationship between daily breaks and stress value. However, if we take a deeper look at the data, we would find out that the data is pretty distributed. That is, the correlation is not as strong as we thought.



**The conclusion:**

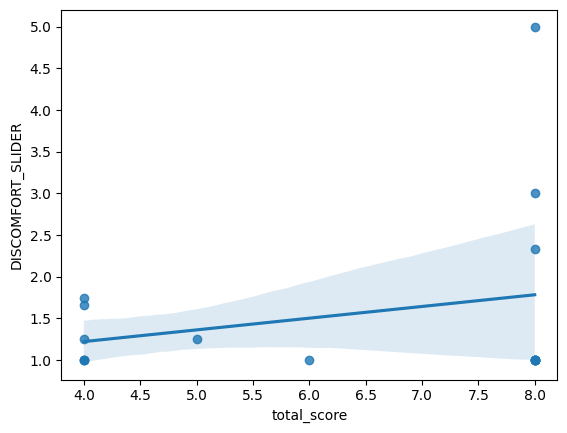
The correlation between participants’ stress levels and the number of breaks is slightly inverse.

Name: Jessica Woo and Vani

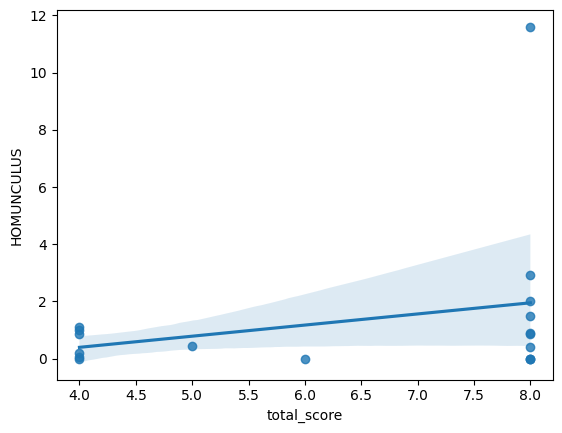
**Hypothesis e**: Based on question #15 in the Computer Workstation Checklist (with 4 responses regarding ergonomics training), participants with lower scores will report less pain at 6-months.

Pre-processing steps:

1. Extracted the six month data from *Daily PM, Friday AM,* and *Computer Workstation Checklist*
2. Extracted the columns that represent question 15 in *Computer Workstation Checklist*
3. For *Daily PM*:
   1. Categorized the entries in “OTHER\_DISCOMFORT” to locations the participants experienced pain. For example, if the entry was “Stomach and chest” it would have the value True under the columns “stomach” and “chest”.
   2. In the “HOMUNCULUS” column, I extracted all of the integers into a list and got the sum of that list.
   3. Dropped rows if all of the custom pain location categories had NaN values or if the “HOMUNCULUS” column had a NaN value.
   4. Merged with the question-15 data
4. For *Friday AM*:
   1. Grouped the data by “mbl\_cod” and calculated the mean discomfort level for each participant (since some participants had more than one entry for “DISCOMFORT”)
   2. Dropped NaN values
   3. Merged with the question-15 data



The regression line shows that there is a direct relationship between scores on question 15 and average discomfort level. Therefore, participants with lower scores will report a lower discomfort level.



The regression line shows that there is a direct relationship between question-15 scores and average homunculus scores. Therefore, participants with lower scores will report a lower homunculus score (i.e. less pain)