DATA PROCESSING

Instructions:

Given the following semicolon-separated dataset, clean it according to the following rules:

   If A or B is mis-represented as a string, use the numeric component (that is, 3.14a should be treated as 3.14)
   If either A or B is missing, drop the row entirely
   A + B is >= 9.5

After it is cleaned:

   Report the number of rows that satisfy the above conditions
   Create a linear regression plot of X vs. Y for those rows
   Report the pearson R value for the regression.

Time Allocation: 30 to 60 minutes (we won't be timing you, just for estimation purposes)

Submitting Instructions:
Please send your answer and a picture of your graph to camille@soullabs.com