Things that I notice that are immediately wrong

* Unknowns
  + There is a bunch of missing or ‘unknown’ values
  + Did countblanks() on each column
  + 
  + Ones that I am worried about are date and transistors
* Dates
  + Some of the dates are in general and some are in date
  + Different formats
    - 5/22/00
    - 2006-06-01
    - Empty cell
* Type
  + There’s CPU and GPU
  + Might have to break the dataset into 2