Lab 8 submission

================

1. 5377 of the values are blank. Total states are 61 (+blank) which is wrong. To solve this I uploaded a list of states to a new tab (named "states") and I added a calculated column with the following GREL grel:cell.cross("states", "Abbrev").cells["Abbrev"].value[0]. In the new column I can see the total number of values that failed to lookup: 6648.

2. 4362 rows are missing zip code.

3. The zipcodes were already loaded as Text. To answer the question, I used:

   grel:if(value.length() > 4, value, '99999').

   After applying the transformation there are:

     4,362: Blanks

     34,961: 99999 (invalid, of which 17 were in the original data the rest had less than 5 digits)

     345,192: Good values (5 digits less than 99999)

4. It is offering to merge Tajikistan with Pakistan and Indonesia with Micronesia, in addition to a few good matches for California and Alaska.

5. California <-> Calfiornia <-> Cailfornia

   Alaska <-> alaska

   Canada <-> Candaa

6. When I try to cluster by "place" because of the field length the algorithm needs to compare a lot more combinations. The complexity of the comparison is roughly O(len(place)^2). The average length is roughly triple that of "location"

   so the computation time will be about nine times longer.

7. The Levenshtein distance between "gumbarrel" and "gunbarell" in 3.

		G	U	M	B	A	R	R	E	L
	0	1	2	3	4	5	6	7	8	9
G	1	0	1	2	3	4	5	6	7	8
U	2	1	0	1	2	3	4	5	6	7
N	3	2	1	1	2	3	4	5	6	7
B	4	3	2	2	1	2	3	4	5	6
A	5	4	3	3	2	1	2	3	4	5
R	6	5	4	4	3	2	1	2	3	4
E	7	6	5	5	4	3	2	2	2	3
L	8	7	6	6	5	4	3	3	3	2
L	9	8	7	7	6	5	4	4	4	3