The Topic Modeling Game  
LDA @ the Farmers’ Market

**Materials needed:**

* Something to topic model. Ideally, plastic fruit with a few colorful blocks (think children’s toys) might work, since sorting them depends less on verbal sense-making. The first time this activity was attempted, we used small pieces of paper with types of produce: (20 slips with a type of red apple, 20 slips with a type of green apple, 10 pear of assorted variety, 10 types of squash of assorted variety, 10 melons of assorted types, 5 beans of assorted types, 5 types of leafy lettuce, 5 broccoli, 5 tomatoes, and 5 miscellaneous other types of produce)
* Something to collect the produce in. If you are playing with sheets of paper with produce word types, they can be taped to paper as if the paper were the “basket/document.” If it’s plastic fruit and/or blocks, you may want to use an actual basket for collecting purposes and then an additional sheet of paper to keep tallies.
* You’ll need a chalk or white board for whomever decides to be the record keeper for the “MALLETers”

**Keep in mind that this will be an uncomfortable process for students. That’s ok. It takes time and some frustration. As students become “stuck” it is a good idea to offer some encouragement and prompts, such as “You’re on the right track” or “Remember that you may have to try this process several times before you can “name” the topics. The point is for students to make mistakes, to revise their predictions.**

**\*\* You will also have to explain that this is a simplified version of LDA… \*\***

**Beginning assumptions:**

* There are 10 possible varieties of produce at the farmers’ market
* Everyone at the farmers’ market buys approximately the same amount of produce
* There are 5 shoppers at the farmers’ market.

**“Shoppers”/ “Authors”** –

* Choose from at least 5 produce types (more if possible) to “add to your basket” (aka. Create your document).
* When you have selected 20 items from at least 5 groups, tape them to your colored piece of paper (Basket/document). Order is not important.

**MALLETers-**

1. Figure out what the 10 types of produce should be. Use the board to write down your most likely “produce types” and figure out how to winnow it down to 10.
2. Figure out what the probability of finding each topic is in each document (Remember, in MALLET there must be some percentage of every topic type… For this exercise, you can have 0 be a proportion).
3. Figure out what proportion of each produce type there was at the market.

**Together, we will try to answer the following questions:**

* + What are the 10 “types” of produce at the farmers’ market
  + About how much of each kind of produce was there at the farmers’ market (in percentages)?
  + What are the proportions of types of produce in each basket?
    - Written as basket color: proportion of type of produce in descending order
  + What is the list of words most likely to be found in each produce type?
  + How did the group handle the “outlier” produce?
  + What skills as a human do you have to help you answer these questions?
  + What are the differences between the way you answered each question and the way MALLET does it with actual texts?
  + What other kinds of questions could we ask with this data?