Topic Modeling  
Farmer’s Market Game

**Beginning assumptions:**

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

**“Authors” / “Shoppers”** –

* 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 farmer’s market
  + About how much of each kind of produce was there at the farmer’s 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?