1. To improve or quantify segregation, you need a way to measure it. Design a scheme to measure segregation in the baseline model. Formalize, explain, and justify your scheme.
2. Make a list of (at least) six ways in which the conceptual model presented in class might be simplifying real-world segregation phenomena. For each simplification, explain how you think the results of baseline model would change if you were to account for it.
3. Choose two of the simplifications you listed above. Explain how you might extend the baseline model to remove or reduce each simplification. (Try to be as precise as you can, but feel free to use a mix of both "plain English" and formal notation as needed. Let the IPython notebook's conceptual model description be thy guide.)

1. In the baseline model to define individuals prefer to be in neighborhoods that have more members of the same tribe, the baseline model defines cij(t) of its neighborhood at time t. Low values of color correspond to dissimilarity, and high values similarity. Define the neighborhood Nij of cell (i,j) to be the collection of cells right next to it, including (i,j). The raw color cij(t) of cell(i,j) at time t is the sum of the population values in its neighborhood. If the neighborhood has the same number of +1 and -1 values, then its color is zero. Otherwise, the sign of cij(t) indicates whether the neighborhood has more members of tribe A (sum is positive) or B (sum is negative).

   In the class example, we define the happiness to be the hij(t) = gij(t)*cij(t) . The small h value means the occupant is unhappy. To measure segregation, we can simply sum up the happiness value of all occupants in the graph (S(t)=sum_ij(hij(t))). Happiness of one cell(i,j) is actually the measurement of the number of identical tribe neighbors around it. Higher happiness value means more identical tribe neighbors are around it. With higher overall happiness, cells cluster more closely together and in other words occupants of the same tribe live more closely together.

2.

   a) Cellular automaton

      Cellular automaton in nature is a simplification of real world. Each real world location is simplified as cells. In the real world, locations are not discrete cells. If I would account for the real world, the results of grid will not be represented as cells therefore the baseline model will not exist.

   b) Dimensions of the "physical" world

      In the baseline model, the dimension of the "physical" world is 10x10 cells. However, the real world dimension as we know is much more complicate and enormous. If I would account for the real world dimension, the results of baseline model would have a large scale because the population grid will become enormous and complicate.

   c) The number of Tribes

      In the baseline model, people are separated into two groups Tribe A and Tribe B. However, in the real world people would not only be separated into two groups. If I would account for it, the result grid will have more than two tribes because we will have more population cell states than before.

d) Occupation probability

To decide whether one cell is occupied and occupied by which tribe, the baseline model use probability. In the real world, whether people occupy one location is hard to determine and it is mainly dependent on individual cases. If I would account for it, the result would not change much since the probability is sufficient to describe the behavior of occupants in a general scale.

e) Neighborhood

In the baseline model, the neighborhood is defined as the cells right next to the center cell. However, the radius of neighborhood is defined to be one. In the real world, the neighborhood cells may not only be the cells right next to the center and not all neighbor cells will have an impact on the center cell. If I would account for it, the result values will be different because the algorithm is changed due to the definition of neighborhood, although the main structure remains the same.

f) Unhappy threshold

The algorithm in the baseline model defines that the threshold to be zero. This constant is purely defined by human for simplicity. In the real world, the threshold is complicated because different people may have different threshold and it can also evolves through time. If I would account for it, the result values will be more close to the real world since a constant determined by human is always a lazy choice for simplicity.

3.

a) Dimension of the "physical" world

The dimension of the baseline model is two-dimensional and it has very limited scale. I will extend the dimension scale from 10x10 to a large scale so that it is large enough to simulate real world city as the graph shown in the notebook. Moreover, the real-world is three-dimensional. Therefore, I would add one more dimension to the grid. People are not only lived on a plane of ground but they also live in high buildings.

b) Neighborhood

The neighborhood in the baseline model is defined as the cells right next to the center cell. Therefore, the happiness only relates to the cells that have radius equal to one around the center cell. In the real world, this is not always the case. On the one hand, the radius might not be only one, and on the other hand, not all the cells have an impact on the center cell. For example, you might not know your neighbor if he or she is very introvert and do not want to talk to people. Moreover, you might have a greater probability to know the people that live more close to you. Therefore I would extend the baseline model by increasing the radius of neighborhood and adding probability scheme for choosing the cells that have impact on the center cells.