Shreyyas Vanarase

Victor Garcia & Tim Chu

CX 4230 Computer Simulation

Prof. Vuduc

Mini Project 1: Quantifying Segregation

Part A: Understanding and Modifying Conceptual Models

Introduction

Segregation is defined as the practice or policy of keeping people of different races or ethnic groups separate from each other (Merriam-Webster.com 2015). Segregation can occur in all communities, especially in those that maintain heavy biases towards the minor racial/ethnic groups. In our context, we define segregation as *the residential separation among individuals based on some distinctive difference*. As an example, consider two communities, A and B containing blue and red stars. A might be segregated into two completely separate groups: one section containing only blue stars and another section containing solely red stars implying complete segregation. Comparatively, B's residents might be evenly mixed throughout community implying very little segregation. Though these are extreme examples of segregation, they provide a fundamental basis for the following discussion.

In reality, as one might imagine, communities are not generally so dichotomous. In fact, there is generally a range in the amount of segregation. This project aims to provide a scheme to *measure* segregation and quantify a conceptual model for real world segregation. In the following section, we will formulate how to measure segregation and explore a widely known method for evaluating segregation called the Index of Dissimilarity.

Measurement of Segregation

The scheme we have devised to measure segregation is based on the color of the individual's neighbors. For this scheme we will assume that a neighbor is someone who is directly above, to the left of, the right of, or below an individual. The formula below describes the percentage of segregation in a community. D represents the number of individuals that have a neighbor of a *different* color and N represents the total number of individuals in the community.

$$\%\ of\ Segregation = 1 - \frac{D}{N} \qquad (1)$$

Figure 1 shows a completely segregated community according to our definition of segregation. As we can see, the red stars are completely separated from the blue stars between a row of empty slots. **Figure 2** shows an evenly mixed community in which there is no segregation. In our calculation of D, we consider each individual only once.

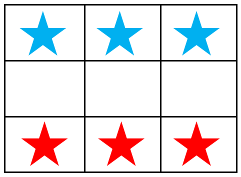


Figure 1: Complete Segregation

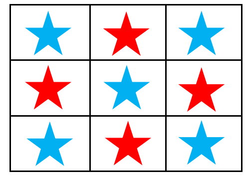


Figure 2: Evenly Mixed Community

As an example of the calculation, consider the following community of 4 stars, shown in **Figure 3**.

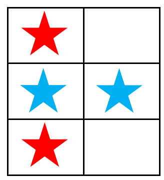


Figure 3: 4-Star Community

The top red star neighbors the middle blue star, so increment D by 1 and N by 1. The middle blue star neighbors the top red star, so increment D by 1 and N by 1. The bottom red star neighbors the middle blue star so increment D by 1 and N by 1. The rightmost blue star neighbors the middle blue star, so increment N by 1, but we do not increment D by 1. For our calculation, a neighboring star of different color will take higher precedence over a neighboring star of the same color. The percentage of segregation in this example is $1 - \frac{3}{4} = 25\%$. As we can see in the diagram, this number intuitively makes sense. The blue stars are closer together and the red stars are separated from each other, but by no means are the blue and red stars "segregated" from each other. This diagram corresponds to a low level of segregation, 25%. Adding more stars to this community will improve the accuracy of the formula we have presented since it will depend on a larger sample size.

Look at **Figure 4**. We can clearly see that this community is strongly segregated. According to our formula, we can estimate that this community is approximately 83% segregated. As a general heuristic for this formula, any community with a segregation percentage greater than 70% can be considered strongly segregated.

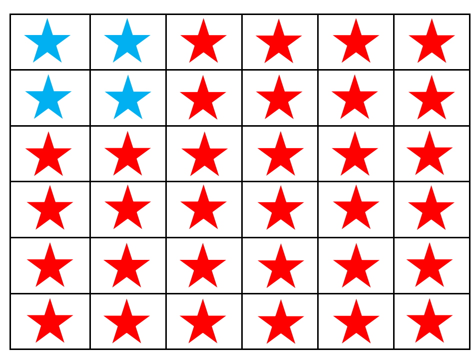


Figure 4: Segregated 36-Star Community

The second scheme we will explore is the widely known Index of Dissimilarity.

$$\frac{1}{2}\sum_{i=1}^{N}\left|\frac{b_i}{B}-\frac{r_i}{R}\right| \qquad (2)$$

This index measures how evenly two groups are distributed within various "tracts" of land across a large geographic map/community. A tract is simply a region within the entirety of the community. This index is generally interpreted as the percentage of one group that would need to move from one location to obtain the same relative number of people in the other group. Evenness is maximized (no segregation) when all units have the same relative number of minority and majority members as the community as a whole. Evenness is minimized (highly segregated) when minority and majority share no areas in common (Weinberg, Iceland, and Steinmetz 2000).

The index of similarity is a relative measure of segregation with respect to two groups. In the above formula, b_i & r_i represent the group of blue/red stars, respectively, in location i and B & R represent the total populations of blue/red stars, respectively, in the community. As an example, an index of .5 can mean that half the area is 100% blue stars and the other half 100% red stars (Miyares 2015).

In the following section, we present a list of real-life simplifications made by the conceptual model presented in class.

List of Simplifications in Conceptual Model (CM)

The current conceptual model simplifies real-world behaviors by not accounting for the following factors:

1) **The real-world difficulties of moving/swapping from one place to another**. In reality, resources such as manual labor and monetary costs of moving are taken into account while moving between locations.
    a. *Implication on CM*: Swapping would occur less frequently and, for example, the ability to swap might even be based on the income of each person.

2) **Those with mixed backgrounds.**
    a. *Implication on CM:* This would imply that there is no well-defined line between blues and reds. People of mixed backgrounds would tend to be less unhappy if they had a majority blue/red neighborhood.

3) **Those who do not have preference to the color of neighbors.**
    a. *Implication on CM:* The happiness of people would not be associated with the color of the neighbors since color wouldn't be a compelling force to make people move. In this case, the CM should be derived from another factor.

4) **The variety of factors that contribute to "happiness" within a population.** Real world is dependent influences like on quality of schools, poverty rates, crime rates.
    a. *Implication on CM:* The happiness measurement would become more complex as more factors are added to determine one's happiness. For every additional factor, up to n^2 numbers of people are affected. Therefore each extra factor would contribute to $O(n^2)$ additional calculations.

5) **The tolerance levels of individuals living in a certain neighborhood.** The tolerance is the willingness of individuals to live in a neighborhood with people of same/different color. For example, in a community with blue and red stars, a blue star with medium tolerance is satisfied living with many but not all red stars.
    a. *Implication on CM:* The conceptual model would account for various tolerances of people when deciding if they would like to swap. In the current model if the sign of the sum of the values assigned to each neighbors differs from their own, they swap. In a system that considers individual tolerance, some blue stars would be highly compelled to switch even if there were 1 neighboring red star. Comparatively, highly tolerant blue stars would not switch even if there were all neighboring red stars.

6) **The irrationality of human behavior.** Humans might not make the best decision to maximize their happiness, so a small probability taken into account.
    a. *Implication on CM:* Extending from our base model, even if someone was unhappy, there would be a random probability that someone makes the incorrect choice and decides to stay at their slot instead of swapping.

The following section extends the current conceptual model through improvements to better account for real-life segregation behaviors.

Extending the Conceptual Model

The conceptual model (CM) currently contains two main simplifications of real-world segregation behaviors.

> **Simplification 1**: CM does not account for the real world difficulty of moving to a different residence.
> **Simplification 2**: CM does not consider a person's tolerance level.

We will begin by introducing two new entities into our CM: The Person entity and the Residence entity. The Person entity is a conceptual representation of the individuals in a community. The Residence entity will replace the slots we used in in the previous CM. The Person entity will have an attribute for their *Color*. As defined in the previous model, the *color* is either set to 1 or -1 and it is randomly generated (50% chance of being 1 and 50% chance of being -1).

Each person will have two new attributes called *Income* and *Tolerance*. Income is a binary attribute such that a person is 0 (Poor) or 1 (Rich). Tolerance level ranges from 0 (extremely intolerant) to 100 (extremely tolerant). In this community, 20% of the people are Rich and 80% are Poor. We decided to choose if a person is rich or poor based on **Figure 5** to the right, a right skewed distribution of incomes in the US indicating only a few people in the nation are rich. The model could be more granular by specifying the incomes each person has, but to save on resources we decided to stick with the two state system described above.

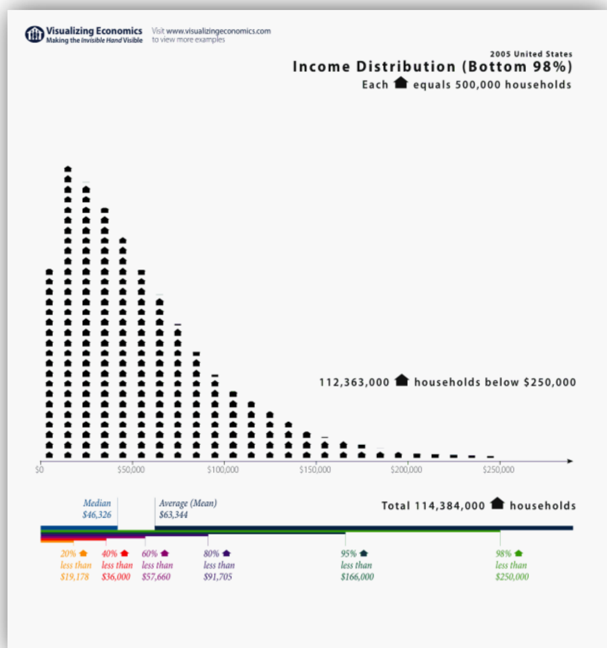


Figure 5: Income Distribution

A Residence entity will have an attribute called *Price* (P) that has a 20% chance of being Highly Priced or Rich residence (1) or 80% chance of being Lowly Priced or Poor residence (0). A rich person can reside in rich residences or poor residences. Comparatively, a poor person cannot reside in a high priced residence. There will be one residence for every previous slot on the grid, so for an $n \times n$ grid there will be up to n^2 residences. We initially place all of the rich residences into one corner of the grid, the remaining map consists of poor residences. Thus, similar to the real world, the rich and poor neighborhoods are separated from one another.

We generate *p* number of Person entities using the criteria for generating a Person as described above, such that $p < n^2$. Then, randomly add persons into residences. Assume there will be enough empty residences for future relocations. Ignore the Income and Price of each residence for the initial placing of Persons on the grid. The creation of Person entities, the initialization of each “slot” in the grid with residences, the clustering of rich residences into a corner of the grid, and the random addition of people into residences constitutes the initial state of our simulation.

When the simulation is running, a person will first look to his neighbors and sum up the “colors” of the people in his neighborhood. If it matches his color he is happy and nothing happens. If he is unhappy, we will use his tolerance to determine if he wants to stay at the same residence or move/swap. For example, if the tolerance of the person is 95, there is a 95% chance of him *not moving*. If a person decides to move/swap, he can become happy if he finds a residence that has a price less than or equal to his income. If he can’t find such a place, he stays at his current residence. Repeat this simulation until we develop a predictable pattern in the segregation or until the society is segregated.

With this improved model, special cases should also be taken into consideration. The following is one example of such a special case. In the case that a poor person was placed into a rich residence at the beginning of the simulation, if an intolerant rich person and the poor person are neighbors, the rich person can force the poor person to relocate to an empty poor residence regardless of the poor person’s happiness/tolerance.

Overall, as we can see, we have extended the original conceptual model by accounting for 1) someone’s income status (rich/poor) and 2) tolerance level (intolerant – tolerant).

Part B: Implementing the Conceptual Model

Baseline Model on Segregation

The conceptual model was based on three main parameters: 1) size of the world 2) ratio between populations 3) probability that a grid location is occupied. These three factors were varied among each other to analyze the effects of each on segregation. Here, we used our formula for segregation to measure the percentage of segregation at each time step. We varied each parameter 4 different times and held the remaining parameters equal to the values set in the baseline model. This allows us to correctly determine the effects of each parameter on segregation individually. For example, to measure the effect of world size on segregation, we set the size of the world to 400 (20 rows) and kept population ratio at 1 and probability of vacancy to 0.5. For the following iterations, we incremented the size of the world and kept pop ratio and probability of vacancy constant.

Before we discuss the effects on parameters, let's briefly look at the effects of the baseline model on segregation. The baseline model sets population ratio to 1.0, probability of vacancy to 0.5, and world size to 400. **Figure 6** below shows the effect of the baseline model on segregation. For each trial, we see that the initial segregation at time step 1 is between 30%–35% and the final segregation at time-step 5 levels off to around 85%–90%. The average segregation for these 5 runs = 87.8% and average time step = 5. **Figure 7** on the right displays the initial and final neighborhood states of a sample baseline simulation. With this in mind, let's delve into the analysis on the effects of segregation on the simulation parameters.

Figure 7: Increase in Segregation TS-1 to TS-4

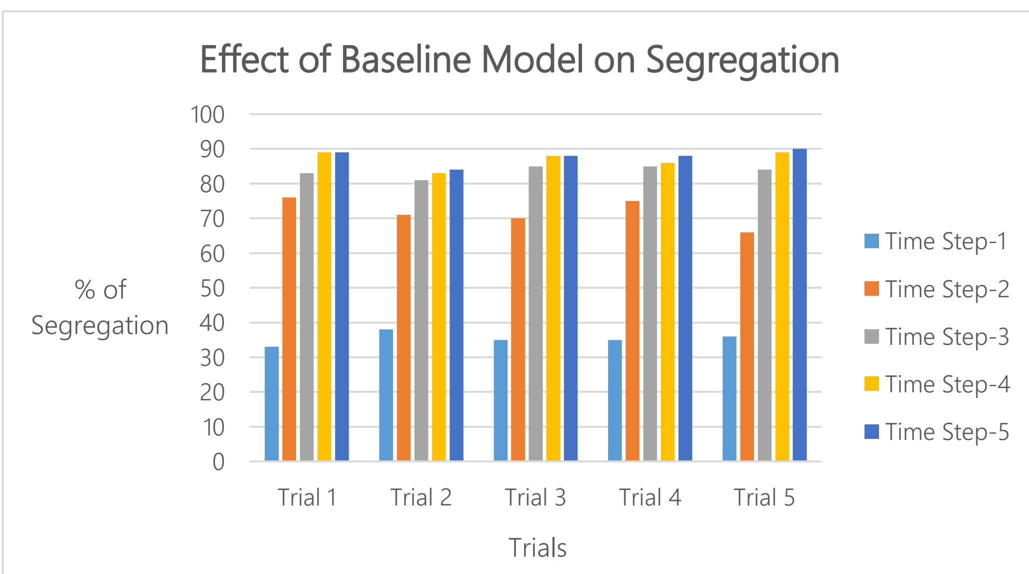


Figure 6: Effect of Baseline Model on Segregation

World Size on Segregation

Figure 8 below measures the effect of world size on segregation. Each column on the graph at a particular world size represents a different time-step. As we can see, the height of each column at a particular time step is about the same. For TS-1, time step 1, the percentage of segregation is about 32%. This means that at the start of the simulation the system was approximately 32% segregated according to our model. This makes sense if we look at **Figure 9**, which shows us the state of the simulation at Time Step 1 and Time Step 5, for a world size of 400. We can argue that this is quite evenly distributed, hence the low 32% segregation. At around TS-5 and TS-6 we can see that most of the columns begin to level off at 85%. This is because many of the simulations have reached an equilibrium or steady state, meaning they do not increase/decrease in amount of segregation after this point.

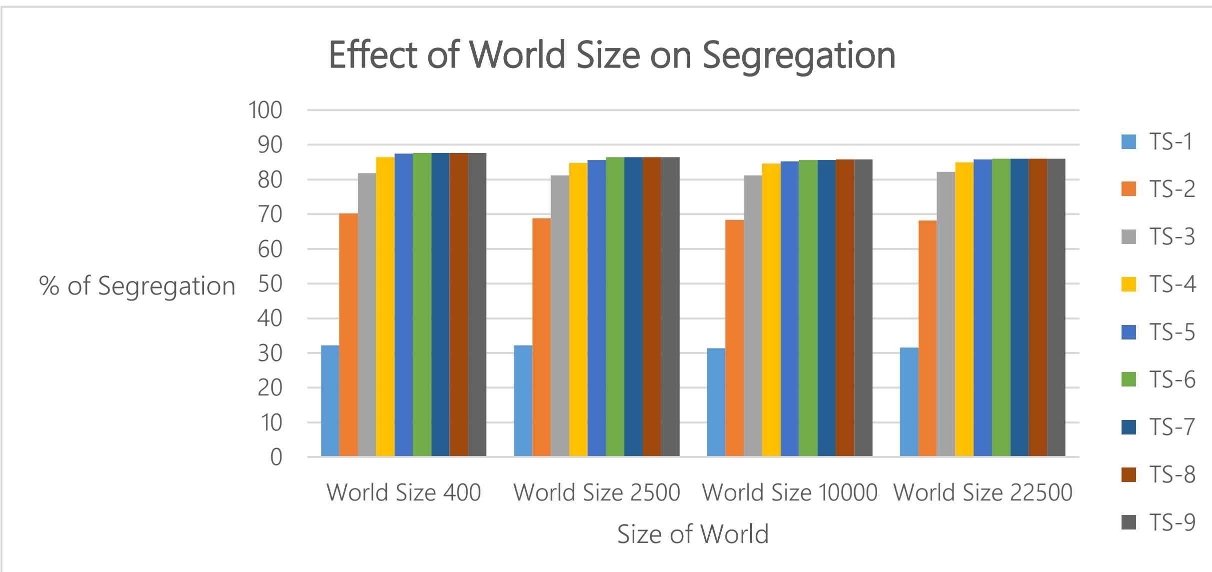


Figure 8: Effect of World Size on Segregation

Note that the additional time steps were added to maintain consistency with the graphs of simulations of the other parameters, which did take up to 9 time steps. Overall, we can say that given the consistency of the percentage of segregation across world sizes, *the size of the world does not make much difference on segregation*. This might be since the population ratio and the relative number of people occupying a world are approximately the same across world sizes, so segregation would act in the same way.

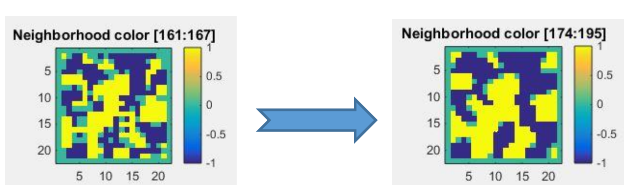


Figure 9: Increase in Segregation TS-1 to TS-5

Population Ratio on Segregation

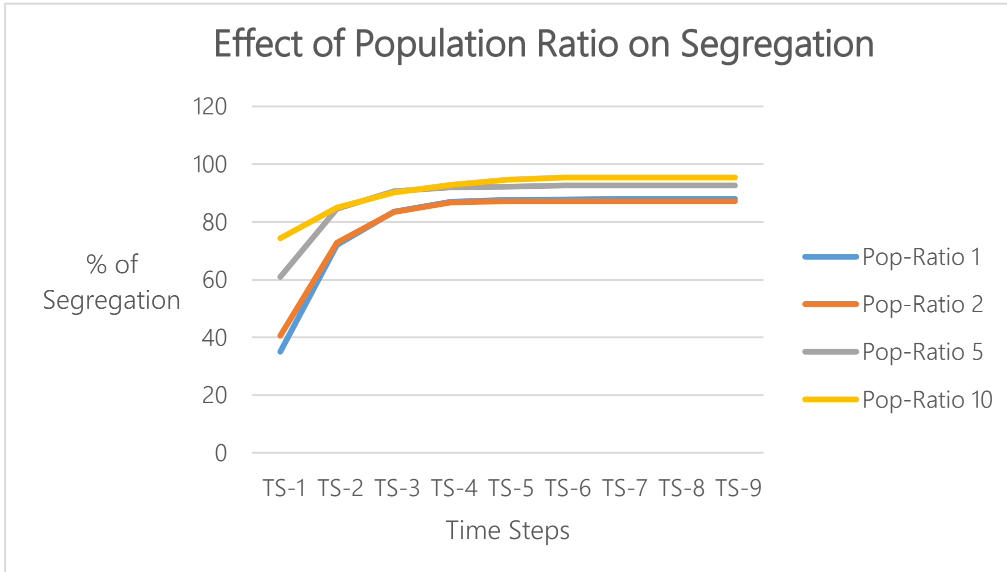
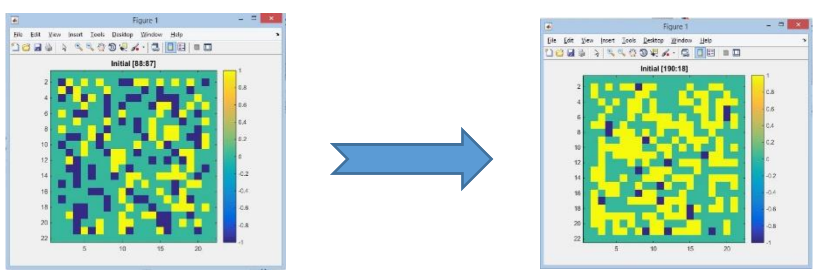


Figure 10: Effect of Population Ratio on Segregation

Next, let's look at the effect of population ratio on segregation. Here we measured the population ratio as being 1:1, 2:1, 5:1, or 10:1. The yellow population dominates in ratios greater than 1:1, otherwise yellow and blue populations are equal. We can see below in Figure 10 that with increasing population ratio the amount of segregation at the initial time step increases as well. This is expected since there will be more yellow individuals than blue individuals. This graph is a slightly different from the previous two since it plots percentage of segregation with respect to time steps. This is to emphasize that the simulation reaches steady state over time at approximately the same level of segregation, i.e. around time step 6/7. Figure 9 displays a comparison of the initial time step 1 at *population ratio 1* to the initial time step at *population ratio 10* as we would see in the simulation.



Population Ratio 1:1 **Population Ratio 10:1**

Figure 11: Comparison of Initial Time Steps for Pop Ratio 1:1 and 10:1

Probability of Vacancy on Segregation

The probability of vacancy is the probability that a location (*i*, *j*) on the grid is free. Therefore, as the value of P(Vacancy) increases, the number of empty spots on the grid increases. Figure 12 shows the effect of the probability of vacancy on segregation. On the graph we see clearly see that the increase in empty spots causes a stark increase in the percentage of segregation.

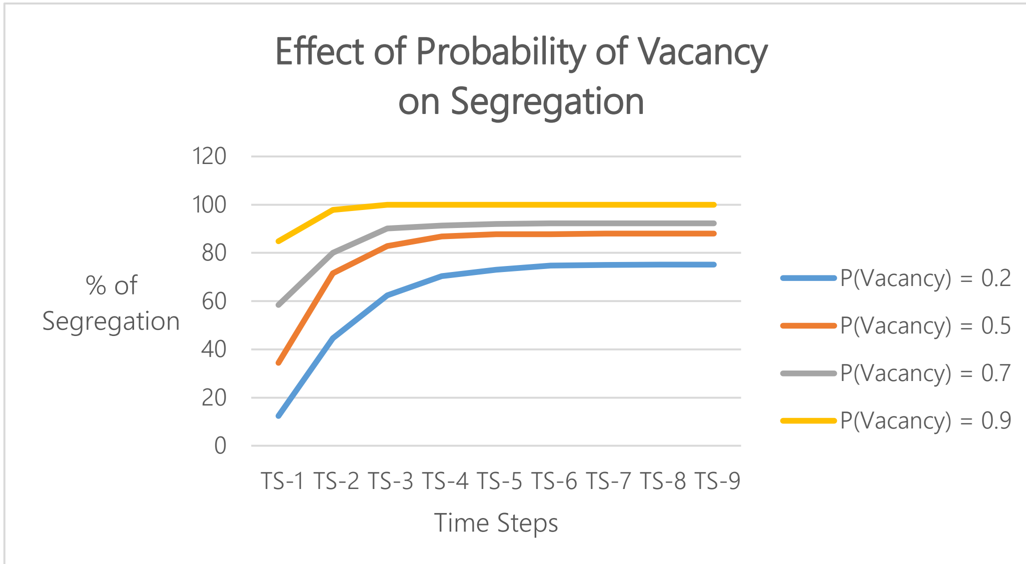


Figure 12: Effect of Probability Vacancy on Segregation

This is because there are fewer neighbors for any individual so there are generally more gaps between individuals of varying color. *Let's note the correlation that segregation decreases as the closer people get, especially for people of different color, and as the number of free spots decreases.* This will be important in making our segregation reduction policy in the next few sections.

Figure 13 provides a comparison of the simulation of the initial time step for P(Vacancy) at 0.2 and 0.9, and the final time step at 0.2 and 0.9. The segregation for P(Vacancy) = 0.2 was 83 and for P(Vacancy) = 0.9 it was 100. This makes sense since there is a fine line of empty spaces in between the two populations in P(Vacancy) = 0.2 and in P(Vacancy) = 0.9 there are no blues that neighbor yellows.

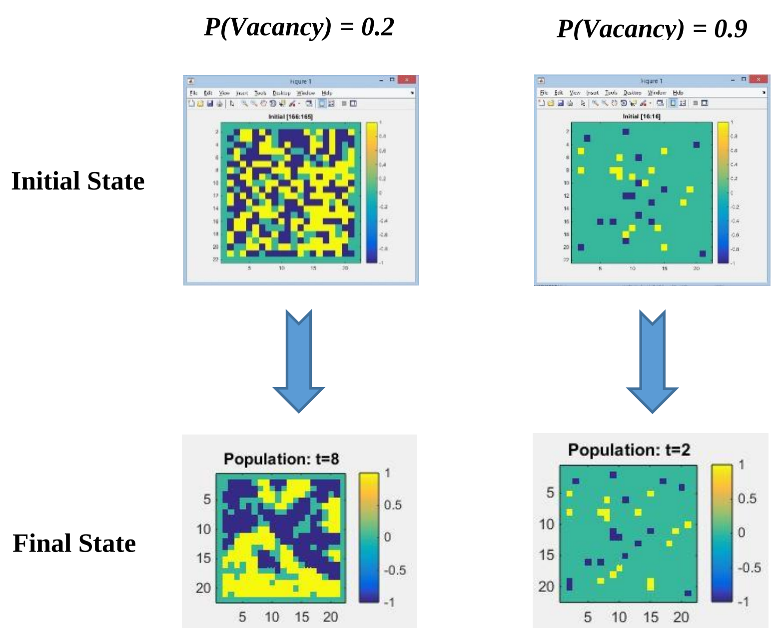


Figure 13: Comparison of Initial and Final Time Steps for P(Vacancy) at 0.2 and 0.9

Tolerance on Segregation

Finally, let's view the effect of adding tolerance to segregation of individuals. As mentioned earlier in the report, tolerance is a 0-100 value associated with someone to represent the amount of 'lenience' he has with respect to the color of his neighborhood. So a person with 95 tolerance is regarded to be highly tolerant and has a 95% chance that he will stay at his current location and has 5% probability he will move to another location. People are given tolerance based on a distribution of tolerance values with mean of approximately 50 and max at most 100.

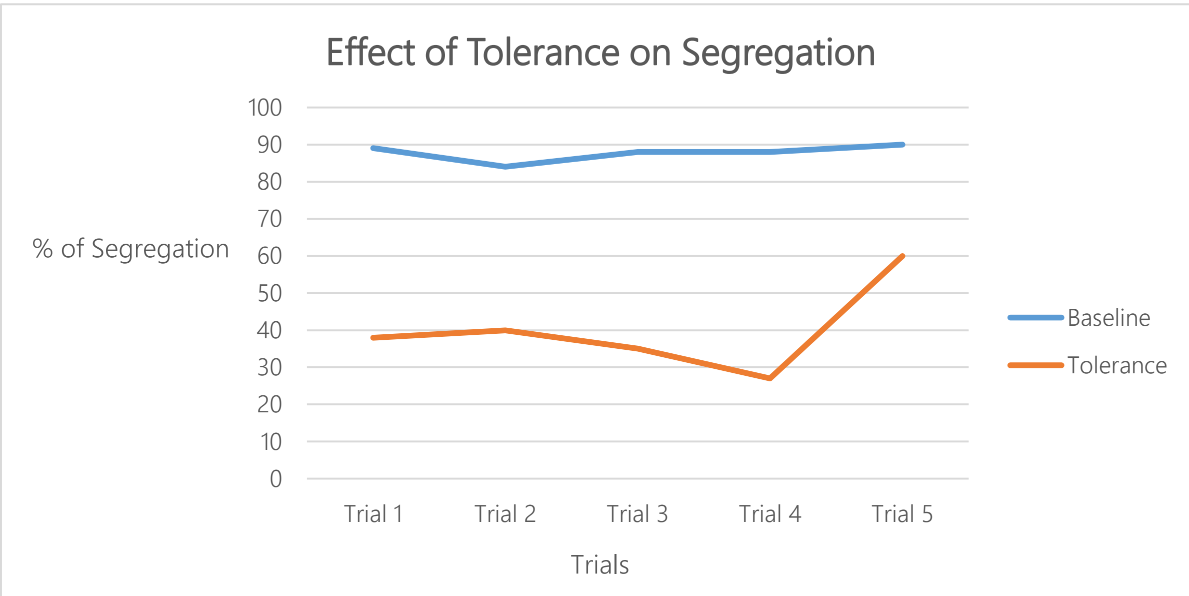


Figure 14: Effect of Tolerance on Segregation

In the five trials we see above in **Figure 14**, we can see that implementing tolerance has lowered segregation levels considerably. With tolerance, segregation has been reduced to an average of 40%. From successive trials, we have concluded that most people do not move from their locations and reach steady state after 2-3 trials. Comparatively, the baseline model provided an 87.8% level of segregation. In **Figure 15** to the right, we see an example of tolerance, in which the system reaches steady state after one iteration with 31% segregation. This low level of segregation makes sense with the relatively mixed population. We see 87 movable cells in the bottom left hand corner, none of which desire to move (likely because they have a high tolerance). Thus we can see that implementing tolerance significantly decreases segregation.

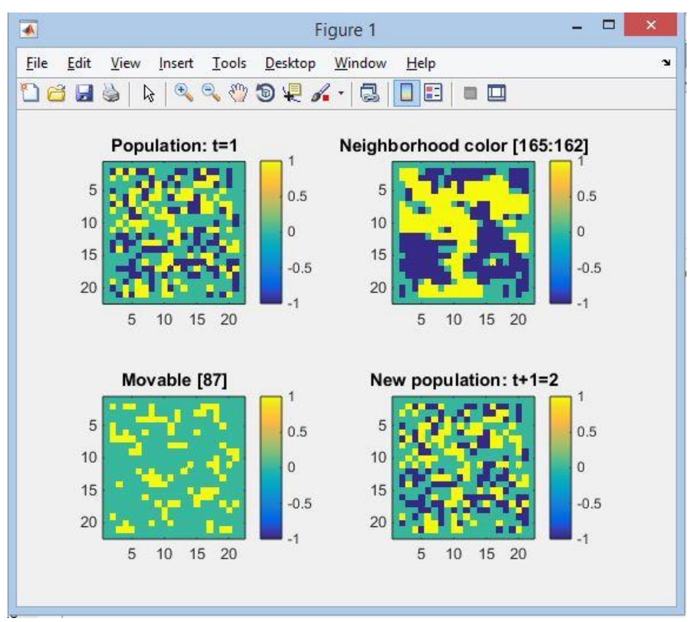


Figure 15: Decrease in Segregation

Segregation Reduction Technique – Unbiased Realtors

There are two main segregation reduction techniques we will explore. As mentioned previously, we are currently interested in dealing with residential segregation based on color. Keeping this in mind, we will first describe the technique in the context of a solution to a real world problem. Then using computer simulations we will show that this technique does in fact reduce segregation assuming the constraints placed within our current conceptual model. The following will provide historical context to the first technique.

The United States has a very deep history of discrimination against individuals who matched certain racial profiles. Discrimination, up until the early 1980s, used to be very outwardly towards these individuals. Signs distinguishing between white and colored public places were common. Housing was no exception; white communities wanted only whites in their neighborhoods. Today, such outward discrimination has significantly decreased. Physical signs displaying preference to one's racial profile is rare. However, that is not to say discrimination has become eradicated. Discrimination still persists. In our context, it persists in a hidden form, especially through the housing market.

The Urban Institute, a D.C. based think tank that carries out social policy research, carried out a study which showed that in 17% of the cases, whites were offered housing units while blacks were told none were available. Though everyone received kind treatment, regardless of the race, the realtors presented fewer options to the minority groups. According to the NYTimes, prospective black home buyers and Asian home buyers were presented 17% and 15%, respectively, fewer homes than whites. Prospective black renters were presented 11% fewer rentals, Hispanics 12% fewer rentals and Asians 10% fewer rentals than whites. In the study, white testers were offered lower rents compared to the minority testers (Dewan 2013). These examples show the main effect biased realtors can have: biased realtors can limit the number of locations a person can move to. In fact, it's often the case that a colored person may be offered a home near other colored people and a white person offered a home near other whites.

Considering this, we deem that our baseline conceptual model is currently biased. Currently, the model gets free locations and presents desirable locations to the moveGroup() method. The moveGroup() method then moves to a location from a list of *desirable* locations. The desirable locations vary between the different individuals based on if they are a 1 or -1. Those who are 1s are assumed to want to move to locations with a majority of 1s in the neighborhood. Similarly, those who are -1s are assumed to want to move to locations with a majority of -1s in the neighborhood. Such biased locations are called desirable locations. The individuals themselves do not have the free choice in deciding which community they like best. Similar to a biased realtor, the current model limits the location to which people move into by assuming desirable locations for them. We can create an unbiased realtor that provides all free spots and allows users to move into them regardless of the color of the neighborhood. This is what we call the unbiased realtor technique.

The unbiased realtor technique implements the following ideas:

1) Every person is offered the full listing of available locations to move into. No bias is presented about desirable locations. This means that a -1 can decide to move beside a 1 and vice versa. After moving, the individual can make a decision about whether if moving to their current spot was a good decision or bad decision.
2) At the start of the simulation, each person is granted a 0-100 value representing whether he likes his decision or not. 0 means a person really dislikes his decision and 100 means he loves his decision. These decisions are based on a normal random probability distribution. Every time a person moves to a different location, he makes a new decision about that place and therefore obtains a new decision value.
3) Happiness is no longer judged by color of neighborhood. The color of the neighborhood is simply used to determine which people might want to move. For example, line 47 in segDemo `M_t = (X_t ~= 0) & ((X_t.* C_t) <= 0);` originally meant to create a mask of people who were unhappy. This is now interpreted as the mask of people who are in a neighborhood of majority different color. These are the people who make a decision about whether they want to move.
4) Note one major difference between this mechanism and our previously proposed version of tolerance. With decision based movement, the person makes a new decision every time step about their current location. With tolerance, we have a set tolerance level per person, which does not change over the course of the simulation.

Having understood this, we can see through **Figure 16** that this proposed method of segregation reduction actually does decrease segregation over time. We have presented 5 trials and with this data we can conclude that the trials have an average time step = 1.5 and average segregation level is 37.2%.

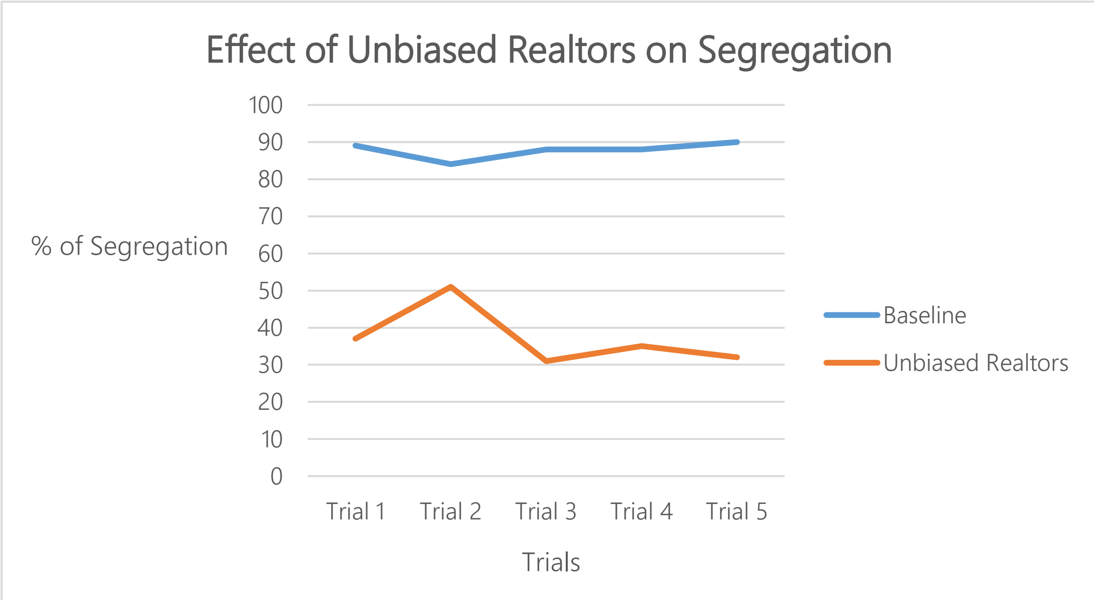


Figure 16: Effect of Unbiased Realtors on Segregation

As mentioned earlier, this technique is not only implementable in simulations but also in reality. In reality, The Urban Institute, or some similar institution, can conduct a long term study in which it tests the effect of unbiased realtors in segregation in a community over a long period of time. The testers they use can be the unbiased realtors in their study to determine how segregation decreases with lack of discrimination to minority groups. Because this unbiased realtor provides people all the options and doesn't filter based on discrimination, we expect that, similar to the simulation, we will find a decrease in discrimination to minority groups who aim to buy/rent households. The incentive for this group to provide unbiased realtors is valuable evidence on the effect of decreasing biased realtors. The government could also create jobs for people who agree to provide unbiased/holistic information to individuals of all racial profiles.

Segregation Reduction Technique – Custom Eminent Domain

Having thoroughly discussed the unbiased realtors technique let us now consider the second technique, eminent domain. At a high level overview, eminent domain refers to the ability of the government to appropriate private property for public use such as constructing parks, buildings, landmarks, etc. In our custom implementation of eminent domain, we will allow the government or any company to appropriate *free* spaces in the community. This implies that fewer free spots will be available at any given time since those free spaces will be held by some company/government. Intuitively, individuals cannot move into a spot that has been taken control of by the government/company. Implementing functionality to appropriate free locations is simple. There is a 0-100 value associated to free values, with 0 being very low chance of being appropriated and 100 being very high chance of being appropriated. At every time step, a certain number of free spots are appropriated. Additionally at the end of teach time step, free spots are updated with randomly generated probabilities to signify a certain probability of being appropriated. So, for example, if a free spot has a value of 95 that means it is 95% chance of being appropriated and 5% chance of not being appropriated. That same free spot on the next iteration may have a different value (say 23%) meaning 23% of the time that spot will be taken over by the company or government and 77% of the time that spot will not be taken by gov/company.

Looking at **Figure 17**, after performing 5 trials, we can see that there is a strong decrease in the segregation levels. The average segregation is 32% and the average time step is 3. The reasoning for this decrease is similar to reasoning for the decrease in segregation when *P(Vacancy)* decreased. That is, segregation decreases with fewer free locations since people are generally closer together. This is especially true for people of different color. Thus, custom eminent domain decreases the level of segregation.

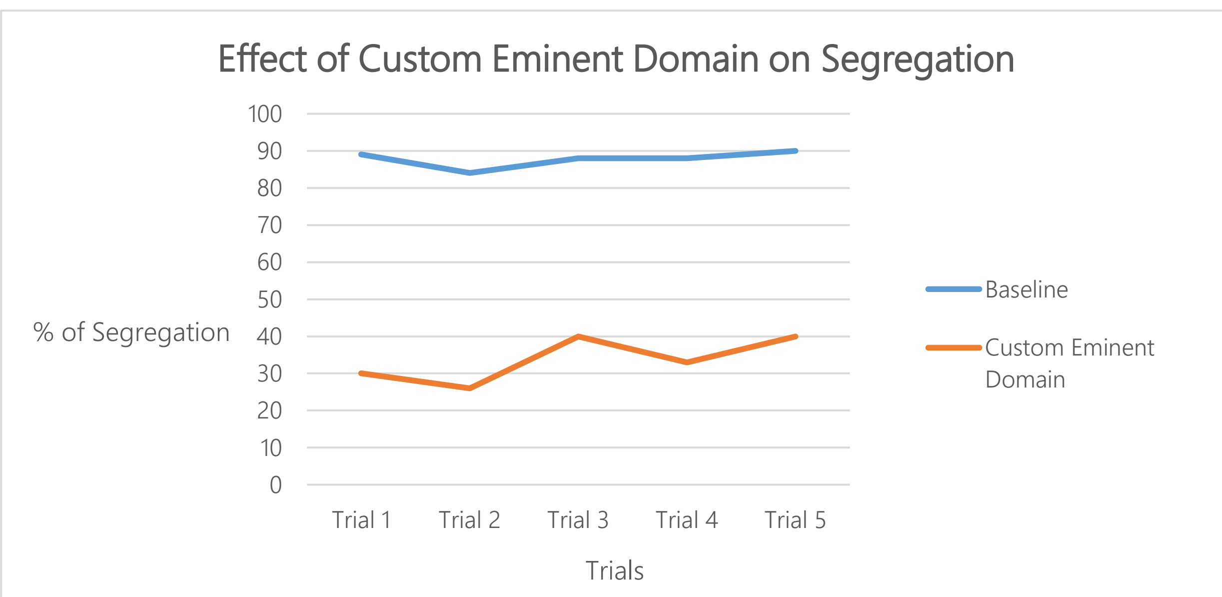


Figure 17: Effect of Custom Eminent Domain on Segregation

References

Dewan, Shaila. "Discrimination in Housing Against Nonwhites Persists Quietly, U.S. Study Finds." *The New York Times*. The New York Times, 11 June 2013. Web. 25 Jan. 2015. <http://www.nytimes.com/2013/06/12/business/economy/discrimination-in-housing-against-nonwhites-persists-quietly-us-study-finds.html?_r=0>.

Merriam-Webster.com. "Segregation." Merriam-Webster. Web. 16 Jan. 2015. <http://www.merriam-webster.com/dictionary/segregation>.

Miyares, Ines. "Segregation: Index of Dissimilarity." *Segregation: Index of Dissimilarity*. Web. 16 Jan. 2015. <http://www.geo.hunter.cuny.edu/~imiyares/Segregation.htm>.

Weinberg, Iceland, and Steinmetz. "Measurement of Segregation by the US Census Bureau." Census.gov. 1 Jan. 2000. Web. 16 Jan. 2015. <https://www.census.gov/hhes/www/housing/resseg/pdf/massey.pdf>.