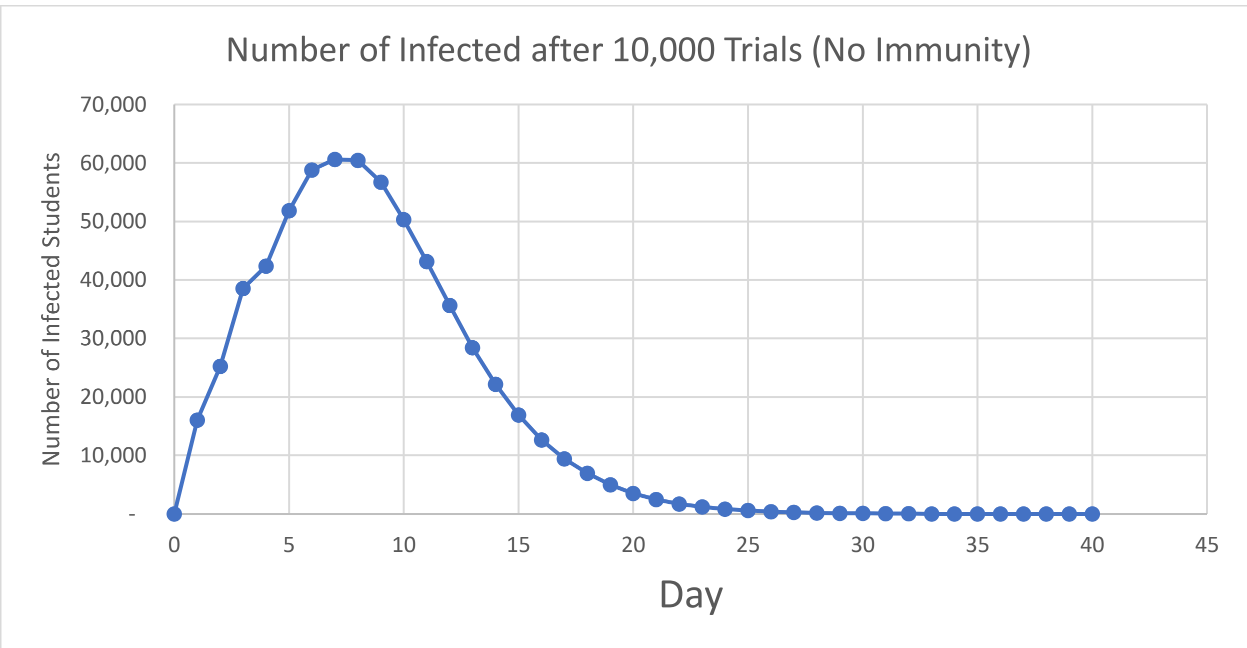
ISYE 6644: USING PYTHON TO SIMULATE A PANDEMIC FLU SPREAD WITHIN A CLASSROOM OF STUDENTS

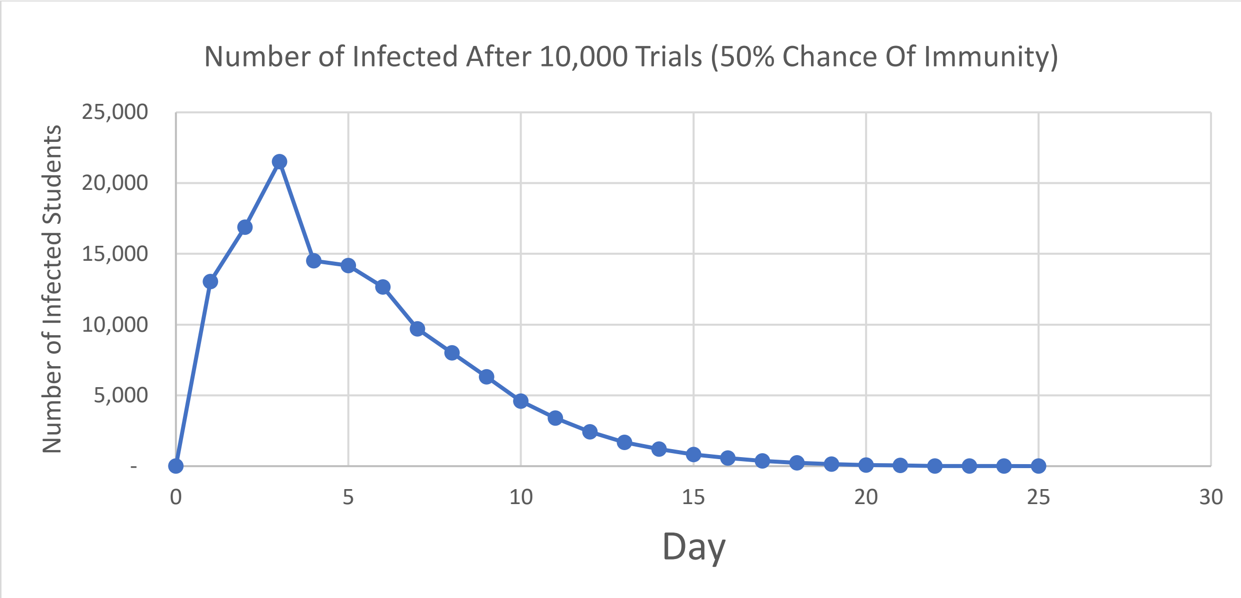
Keith Deuser (Group 193)

ABSTRACT

Disease and pandemics are a part of the human condition. For as long as humanity has existed, disease and sickness have had an enormous impact on all human societies. Understanding the biology of viruses and bacteria, how quickly they spread, and how long pandemics can potentially last, are key questions that need to be answered to combat the effects of pandemics. In this report, I wish to examine how quickly, and how many people, a potential flu virus can spread within a small, confined area. A classroom of 31 students could be considered a ripe potential breeding ground for a flu virus to spread.

After analyzing the results of simulating a flu virus spread for 10,000 replications, it appears that a typical flu outbreak within a classroom of 31 students could be expected to last 13 days on average, with the maximum number of infected students expected to occur on day 7. However, if each student (minus the initially infected student) has a 50% chance of being immune to the flu at the start of the simulation, the average length of an outbreak was 6 days, with the maximum number of infected students expected to occur on day 3.





BACKGROUND

How do we model a flu outbreak, and account for the random elements that contribute to the spread of viruses like the flu? If we focus our simulation on a smaller scale (For example, a classroom of 31 students), we can answer this question using discrete probability distributions and Python programming, along with basic python modules. The NumPy module can be used to simulate the flu spreading from student to student over the course of the outbreak, and the Pandas module can be used to analyze the data and draw conclusions on the results.

When building this simulation, I operated under the following assumptions. Beginning on day 0, all 31 students are healthy, and none are infected with the flu. We initiate the simulation on day 1, where a single student has caught the flu and is considered “Sick” (we’ll call this student “patient zero”), while the remaining 30 students are considered “Healthy.” Once a student is sick, they will remain sick for 3 days. Each day a student is sick with the flu, the probability that they will infect each healthy student follows a *Bernoulli* distribution. For each healthy student in the classroom, the sick student has a 2% chance of infecting the healthy student, and a 98% chance of not infecting them.

$$f(x) = \begin{cases} 0.02, & Student\ is\ Infected\ (1) \\ 0.98, & Student\ is\ not\ Infected\ (0) \end{cases}$$

Once a student recovers from the flu, they will be considered “Healthy” and “Immune,” and will not be able to catch the flu again. The simulation terminates once all students are healthy and there are no longer any infected students in the classroom.

For my implementation. I ran 10,000 trials of the simulation and collected data on the number of sick and healthy students each day of the outbreak, along with how long each trial of the simulation lasted in days.

BUILDING THE SIMULATION

Before running the script, we can decide how many trials we would like to run, and whether the students (minus patient zero) will have a 50% chance of being immune to the flu at the start of the outbreak. If the variable *Imm_50_50* is set to *False*, then none of the students are immune. If this is set to *True*, then each student (minus patient zero) has a 50% chance of being immune at the start of the outbreak.

For each trial, the simulation begins by creating our student entities. These entities are represented by a dictionary, where the keys represent each student, and the values represent a list of attributes. Student 31 always starts out as "Sick" and "Not Immune." Each day of the outbreak, the simulation detects if anyone is still sick and terminates the trial once all students are healthy and unable to spread the flu. For each day of the simulation, we will:

1. Update how many days each sick student has been infected with the flu. Students that have been sick for 3 days have their status changed to "Healthy."
2. Check if any students are still sick. Our while loop checks this as well; however, this step ensures that if all students are healthy, the simulation terminates without exposing students to the flu again (explained in step 3).
3. Expose students to the flu once per sick student.
    a. For example, if there are 2 sick students on a given day of the simulation, each healthy student will be exposed to the flu 2 times, with each of those 2 times having a 2% chance of catching the flu.
4. We update our statistical data, where we track how many students are healthy, and how many students are infected, each day of the simulation.
5. Check student entities for any unexpected attributes, or combination of attributes.
    a. For example, if a student is "Healthy," then they must have an attribute of 0 for number of days sick.
6. This process repeats until there are no longer any sick students in the simulation, at which point the trial ends.

Once all trials have completed, we are able to organize the data we've collected and begin our analysis.

MAIN FINDINGS IF STUDENTS START WITH NO IMMUNITY

On the first day of the simulation, patient zero has a 2% chance of infecting each student in the classroom with the flu. If we consider each of these 30 interactions to be independent and identically distributed (i.i.d.), we can model the expected number of students patient zero infects using a Binomial distribution, which is a discrete probability distribution representing the number of successes in a fixed number of Bernoulli trials, where *n* represents the number of trials, and *p* represents the probability of success. If we round our expected value to the nearest integer, we expect patient zero to infect one student on the first day.

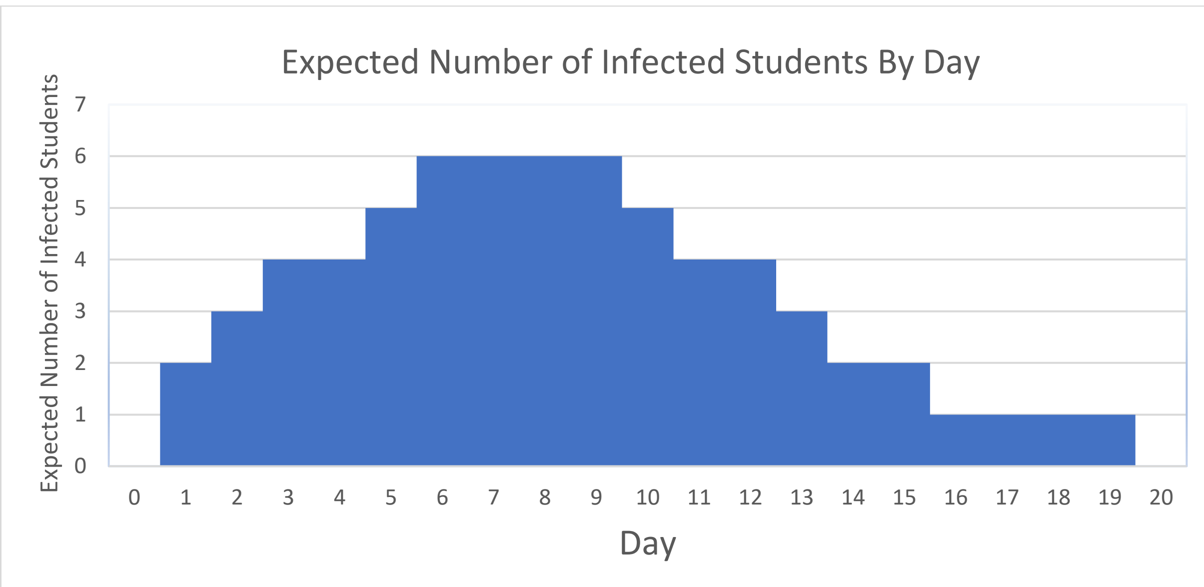
$$E[X] = np = 30(0.02) = 0.6 \qquad Var(X) = npq = 30(0.02)0.98 = 0.588$$

Our *simulated* expected value for the number of students infected by the end of Day 1 was calculated to be two students, so our simulation appears to line up with what we might expect the theoretical value to be.

If we assume that patient zero has infected one other student by the second day of the simulation. We can calculate the expected number of students infected by the end of day 2. On day 2, we have 29 healthy students in our classroom, and 2 infected students. This means that each healthy student will be exposed to the flu twice on Day 2. We can model this with a Binomial distribution, and our expected value ends up being 1 when rounded to the nearest integer. This means we would expect 3 students to be sick by the end of day 2.

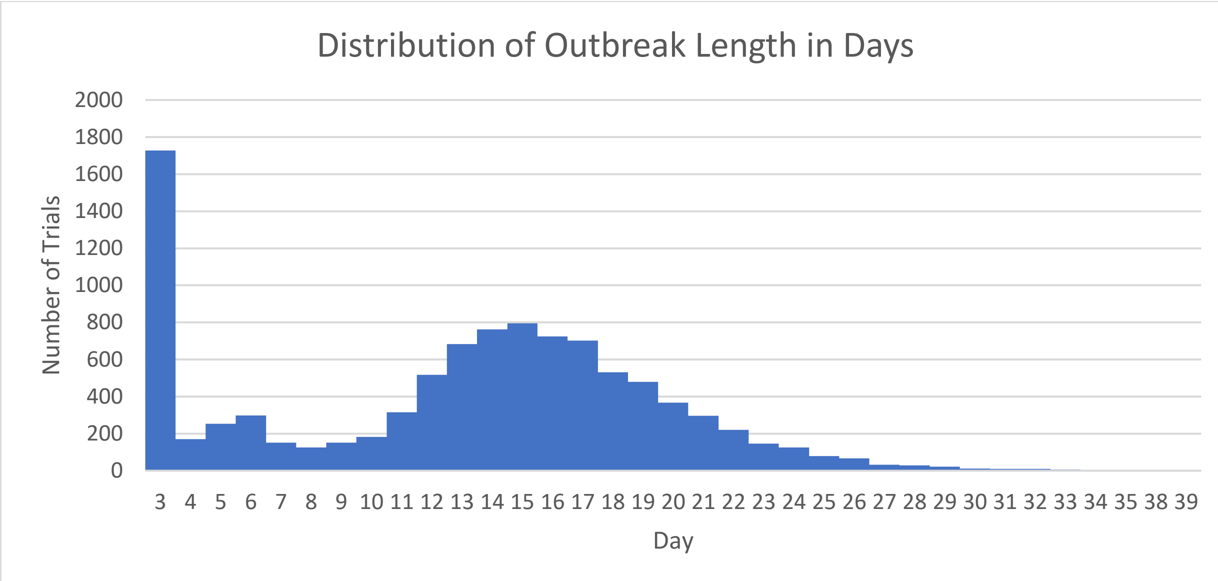
$$E[X] = np = 58(0.02) = 1.16 \qquad Var(X) = npq = 58(0.02)0.98 = 1.14$$

Our *simulated* expected value for the number of students infected by the end of Day 2 was calculated to be three students, so our simulation appears to show what we might expect the theoretical value to be.



In our classroom of 31 students, how many students should we expect to be infected each day of the outbreak? The values in the graph above were calculated by applying the ratio of students infected for all 10,000 trial runs of the simulation to the total number of students in the classroom and rounding this number to the nearest integer. Based on this visualization, we'd expect to have the maximum number of infected students on days 6 through 9 (6 students), with sick students likely recovering by day 20.

How long can we expect the flu outbreak to last? We can visualize this by viewing a histogram of the number of days each outbreak lasted in our 10,000 trials.



When viewing the distribution of the number of days the outbreak lasted, the first thing you will probably notice is that there appear to be many trials where the simulation terminated after 3 days (exactly 1,727 trials). This occurs if patient zero does not infect any student in the first 3 days of the simulation. 17.27% of our trials ended without any other student catching the flu.

Another way to phrase this is that there were 1,727 simulations where the first 90 Bernoulli(0.02) trials all ended in failure. We can see if this is close to what we'd expect in theory using the following calculation below. The expected probability of the first 90 Bernoulli trials ending in failure is 16.23%, while our simulated probability is 17.27%.

$$Expected\ Probability = 0.98^{30(3)} = 0.162$$

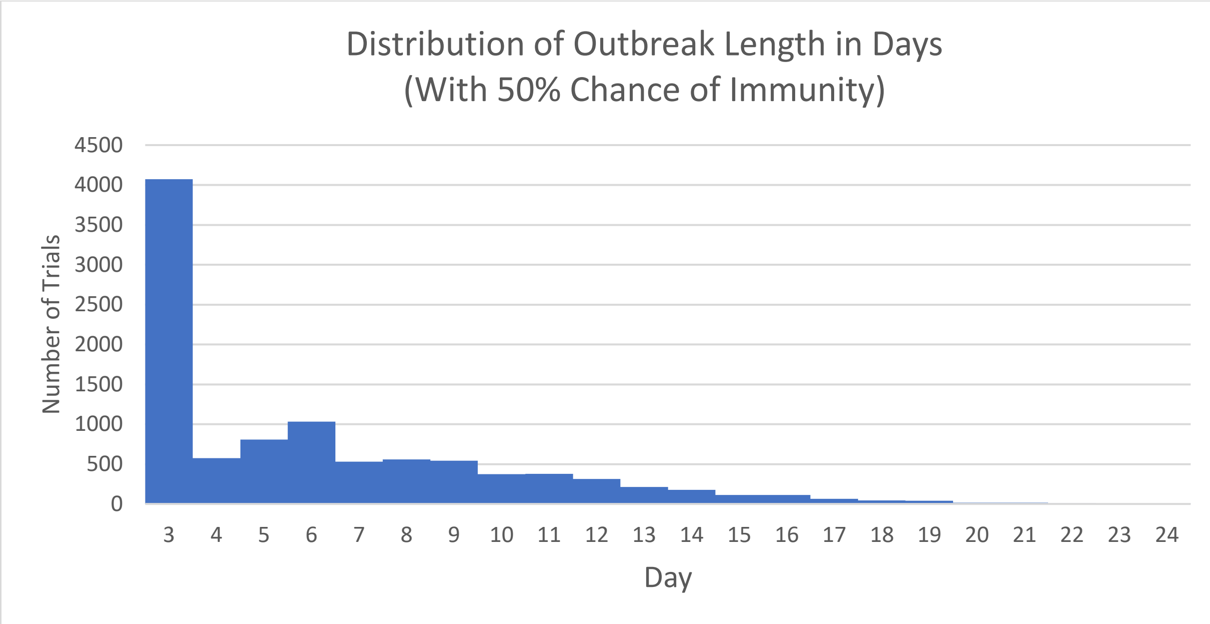
$$Simulated\ Probability = \frac{1727}{10000} = 0.172$$

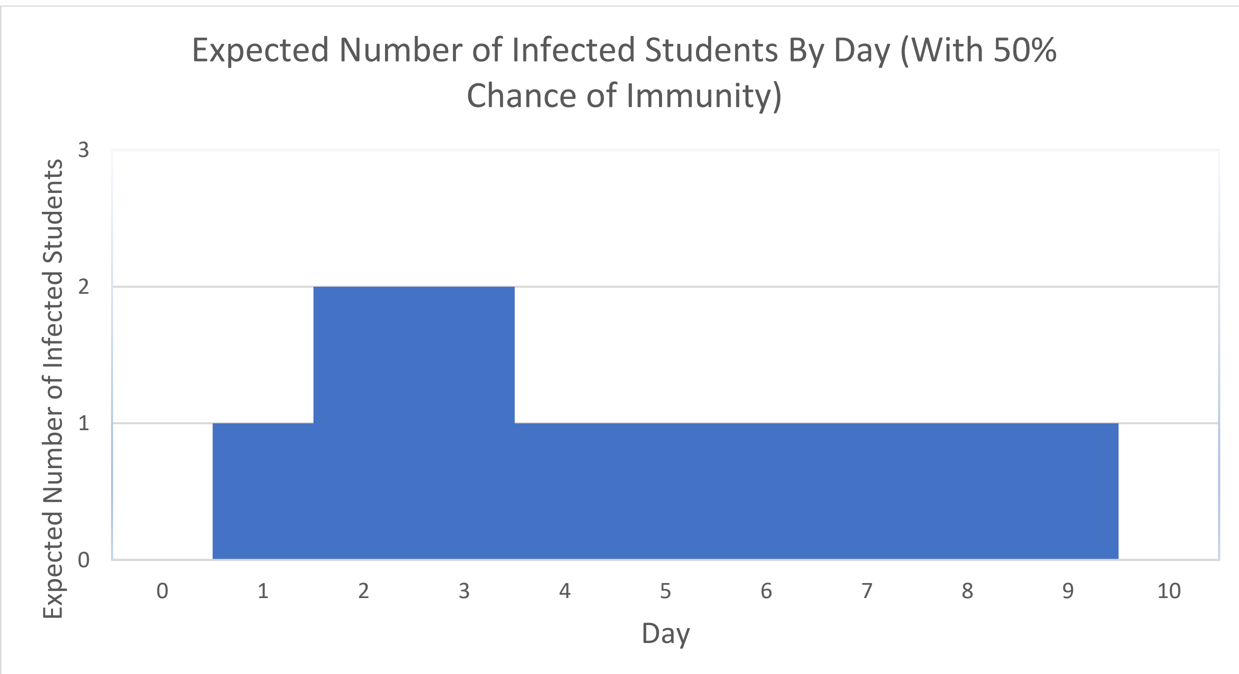
On average, the flu outbreak lasts for 13 days, with 94.55% of simulations terminating before day 23. We can see that most flu outbreaks will not last more than 22 days, except in rare cases. Only 5.45% of trials lasted longer than 22 days. While the variance for the 10,000 replications is high at 42.608 days, we can still say with 95% confidence that a typical flu outbreak will last 13 days if we round to the nearest integer

$$12.978 \pm 1.96\sqrt{42.608/10000} = [12.850, 13.106]$$

MAIN FINDINGS IF STUDENTS HAVE 50% CHANCE OF BEING IMMUNE

To simulate this scenario, we can use the exact same python program with one key difference. We set our variable *Imm_50_50* to *True*, and run the simulation again. We get the following results on page 6.





We can see that almost half of the trials (4,072 exactly) did not make it past the third day, and that by day ten of the outbreak, we can expect the number of infected students in a classroom of 31 students to be zero.

Using the same methods established in our original simulation, we can calculate the expected number of students patient zero infects on Day 1. With 30 students having a 50% chance of being immune to the flu at the start of the simulation, we can expect 15 students to start out with immunity to the flu. We can see that based on our expected values, we're less likely to infect a student on day 1. This is reflected in our simulation results, where on day 1 the expected number of infected students is 1, which means patient zero has not infected any other students (if we round our expected values to the nearest integer).

$$E[X] = np = 15(0.02) = 0.3 \qquad Var(X) = npq = 15(0.02)0.98 = 0.294$$

Using the same theory, we can expect patient zero to infect one other student by day 2 (if we round to the nearest integer), which is what we observe in our simulation results.

$$E[X] = np = 30(0.02) = 0.6 \qquad Var(X) = npq = 30(0.02)0.98 = 0.588$$

We can calculate the probability that patient zero will infect no students on the first 3 days of the simulation using the following calculation and compare that to the simulated probability that patient zero infects no students.

$$Expected\ Probability = 0.98^{15(3)} = 0.402$$

$$Simulated\ Probability = \frac{4072}{10000} = 0.407$$

When taking the average length of all trials where students start out with immunity, it appears we can expect an outbreak to last 6 days on average, with 94.01% of outbreaks lasting 13 days or less. Only 5.99% of trials made it past day 13. While the variance for the 10,000 replications is high at 14.860, we can still say with 95% confidence that a typical flu outbreak will last 6 days if we round to the nearest integer

$$6.194 \pm 1.96\sqrt{14.860/10000} = [6.118, 6.270]$$

CONCLUSION

When no students have immunity to the flu, there is a 94.55% chance that the outbreak will last a maximum of 22 days, with the percentage of infected students most likely reaching its maximum on day 7 with 19.56% of students infected. We can predict with 95% confidence that a typical flu outbreak will last 13 days, with 100% of simulations terminated by day 40.

When students start with a 50% chance of being immune to the flu, there is a 94.01% chance the outbreak will last a maximum of 13 days, with the percentage of infected students most likely reaching its maximum on day 3, with 6.94% of students infected. We can predict with 95% confidence that a typical flu outbreak will last 6 days, with 100% of simulations terminated by day 25.

When students had a 50% chance of being immune to the flu at the start of the simulation, there was a 41% chance that the flu never spread beyond patient zero, while the trials where students had no immunity only had a 17% chance that the flu never spread beyond patient zero. The average length of the outbreak was reduced by 47.73% when compared to the trials where no students started out with immunity.

While we have received valuable insights from these simulations, there are areas that could be improved for future projects. One possible way the simulation could be modified to better reflect reality, would be having the probability of being infected be a random variable instead of a constant. For example, when the student entities are created, the probability they will be infected with the flu when exposed could be a uniform distribution between 0.01 and 0.05, since some people are more resistant to viruses and bacteria than others. This is just an example, however, and you would need to consult a medical professional to determine what the distribution for the probability of being infected should be.

In conclusion, this simulation demonstrates the variability in how much a virus can spread, and how long an outbreak can last, even when modeling the scenario using simple Bernoulli probability distributions. These results also demonstrate the value of immunity and/or vaccinations in helping prevent the spread and severity of a flu outbreak.