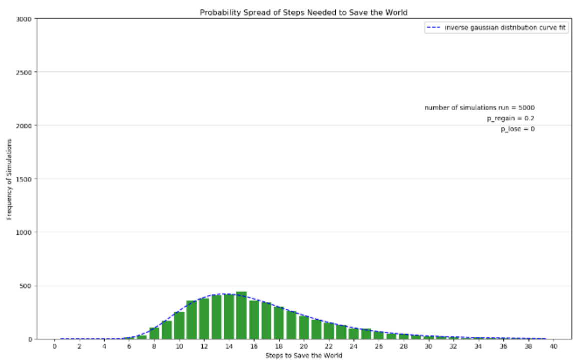
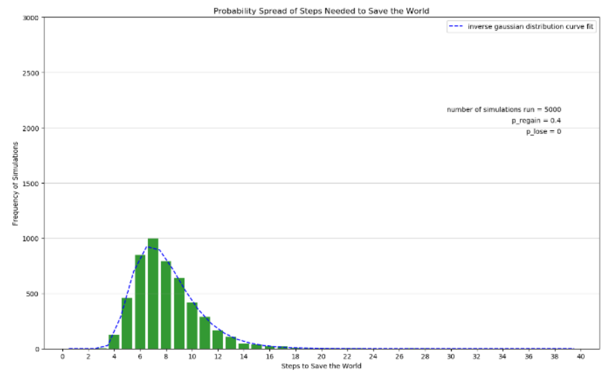
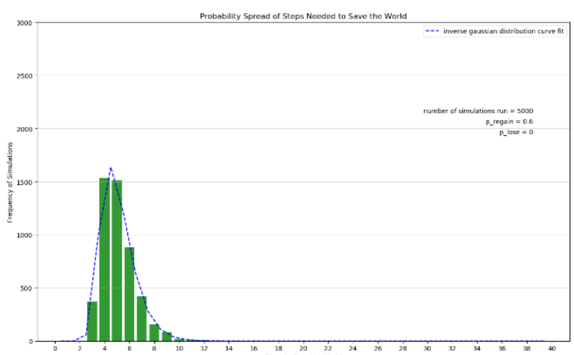
CS 2120 Assignment #2 -- Take Back Our World!
Name: Michael Tran
Student number: 251033279

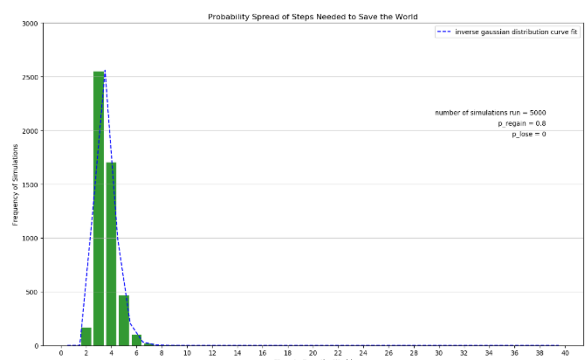
Analysis 1.

The p_lose value is set to 0, and the p_regain value is set to 0.2, 0.4, 0.6, and 0.8; and the respective histograms generated by the code are shown below:







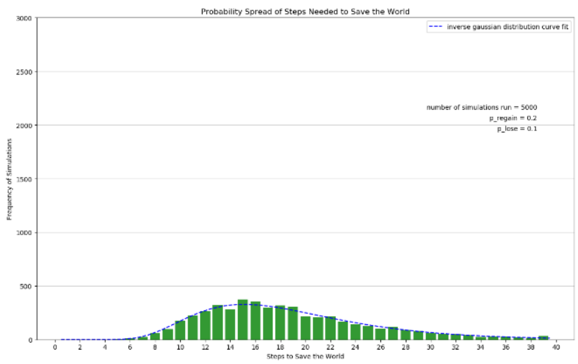


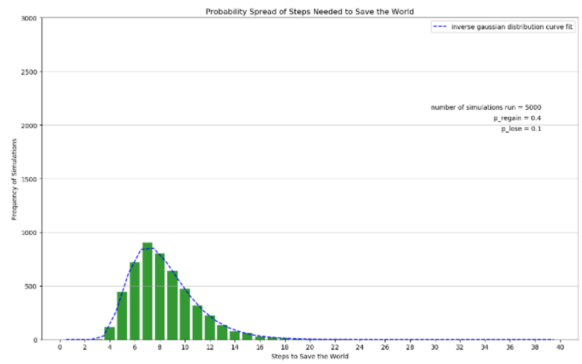
For low values of p_regain, the data is more spread out and more steps are needed to save the world. This is because there is a lower chance for a city to be saved and thus takes more steps to save all the cities.

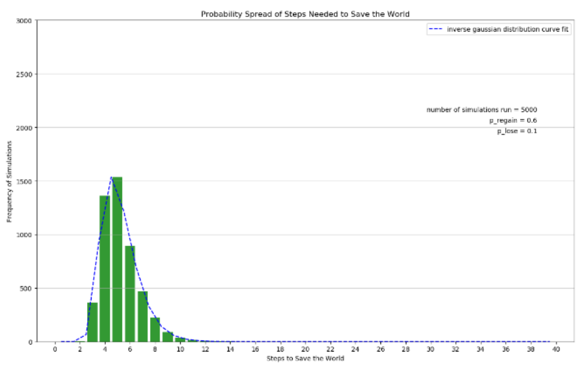
For higher values of p_regain, the data has a higher precision centred around lower numbers. This is because the chance to save a city is high and thus much less steps are needed to save all the cities.

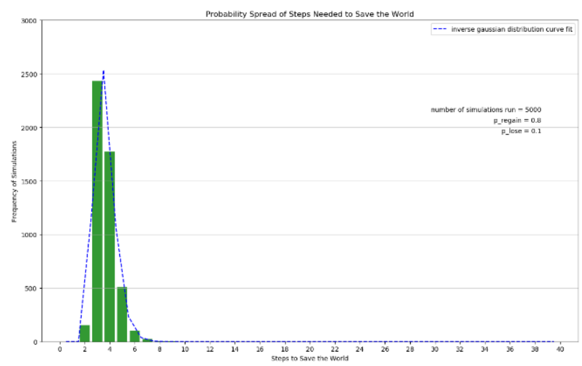
Analysis 2.

The p_lose value is now set to 0.1, and the p_regain value is set to 0.2, 0.4, 0.6, and 0.8; and the respective histograms generated by the code are shown below:







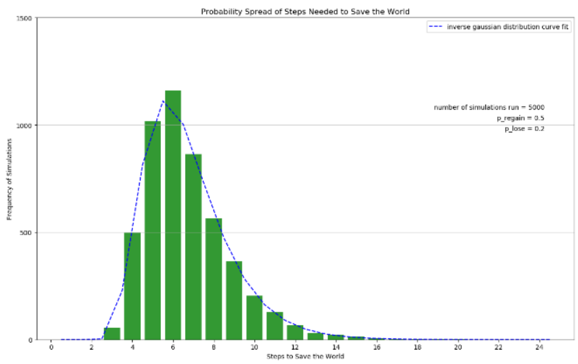


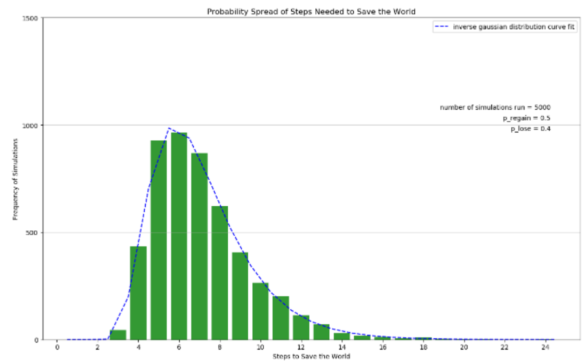
The same trends in **Analysis 1** are visible in these histograms.

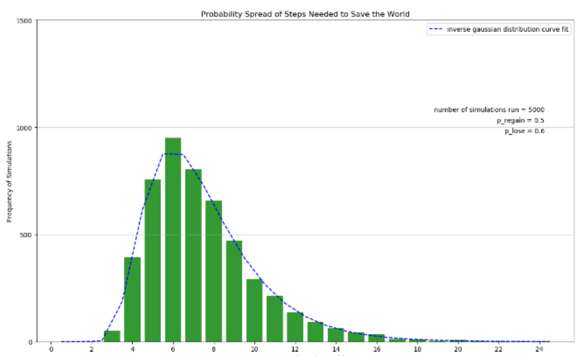
The difference between this simulation and the first one where p_lose = 0 is minimal. Although hard to tell, the histogram peaks are slightly lower, and the curves extend further when p_lose = 0.1. This is because there is a slight chance for cities to be lost now, and this causes the steps needed to save the world to increase, and the distribution to be more spread out.

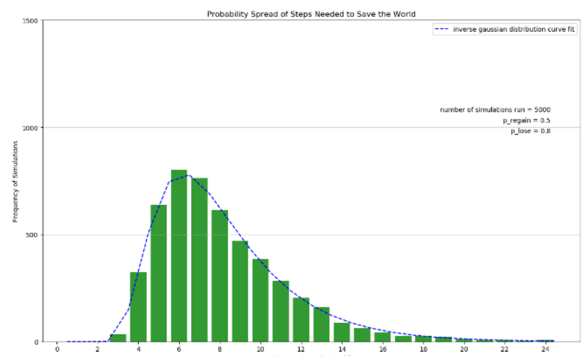
Analysis 3.

The p_regain value is now set to 0.1, and the p_lose value is set to 0.2, 0.4, 0.6, and 0.8; and the respective histograms generated by the code are shown below:







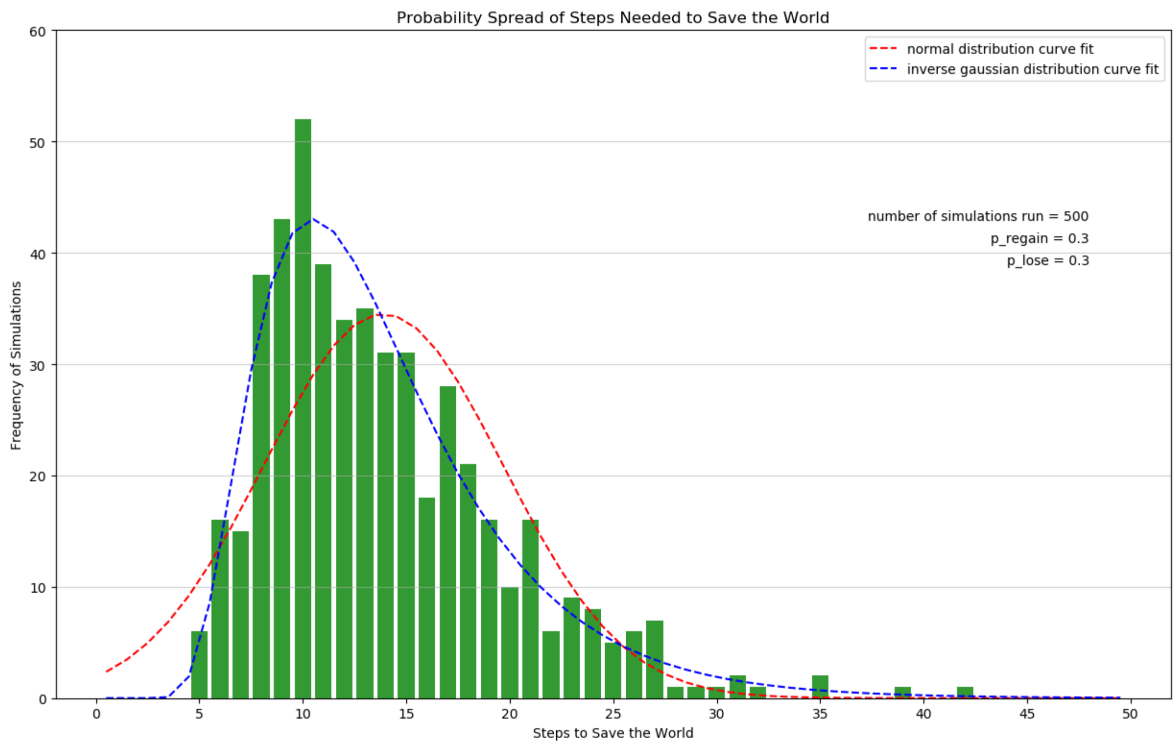


Note the x-axis limit is now 25 and the y-axis limit is now 1500 when comparing the graphs from previous analyses. The same trends visible in **Analysis 1** and **Analysis 2** are visible in these histograms.

There are less dramatic changes between the p_lose = 0.2 - 0.8 than between p_regain = 0.2 – 0.8 shown in the previous two analyses. This is because altering the p_lose value has a lesser effect on the shape of the histogram than altering the p_regain value. Nevertheless, increasing the p_lose value spreads values out and shifts all values down to the right of the histogram as there is a higher chance of losing a city which extends the simulations.

Analysis 4.

Simulation A: p_regain = 0.3, p_lose = 0.3

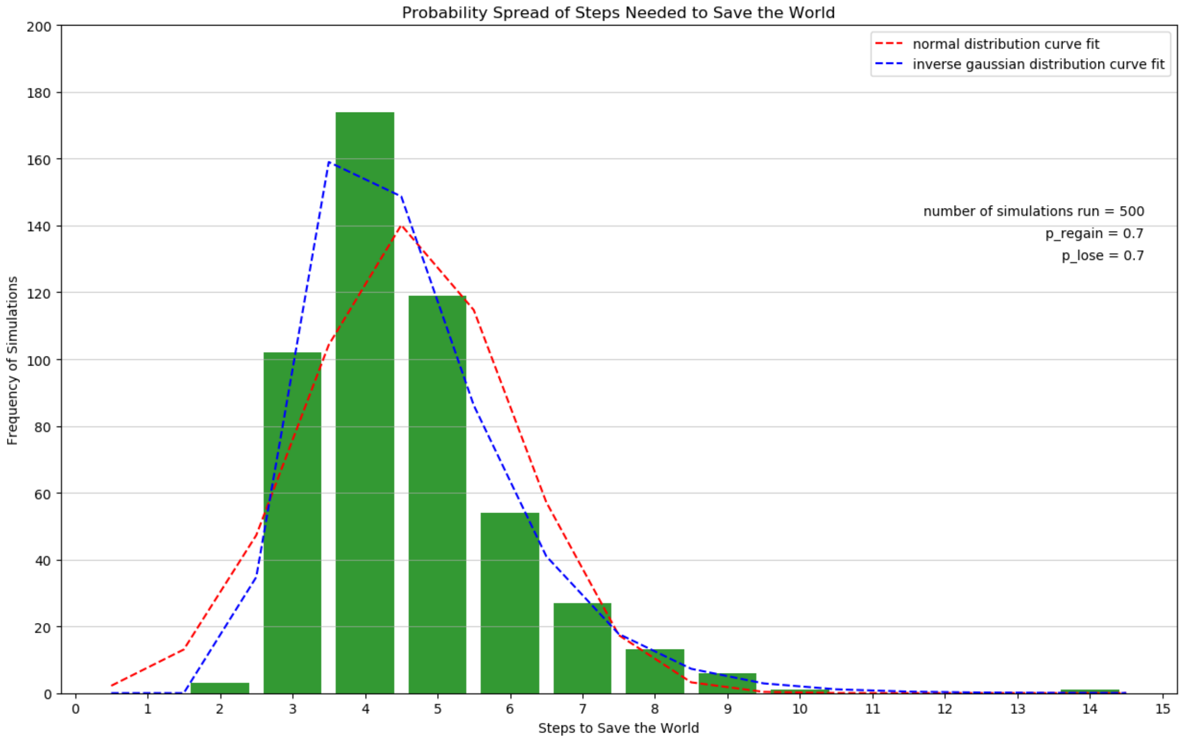


This histogram shows a nice distribution of data over a large range of steps needed to save the world. This is because both the p_regain and p_lose values are relatively low and extend the simulations for longer.

Another interesting point is that the data fits the inverse gaussian curve much less than previously. This is because 500 simulations are run instead of 5000 previously. As more simulations are run, the histogram bars converge upon the trend curve.

In terms of the normal curve (shown in red), the distribution does not seem to fit as effectively as the other curve. This shows that the data is not normal nor symmetrical around the peak, but rather skews to the right.

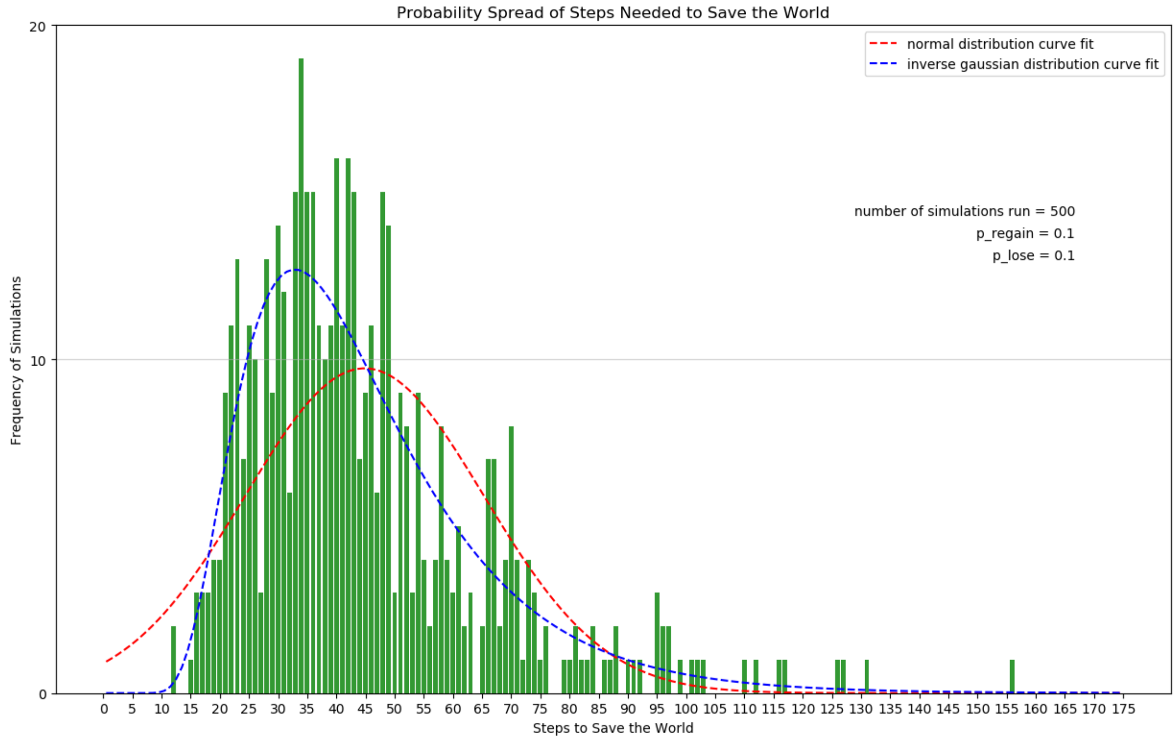
Simulation B: p_regain = 0.7, p_lose = 0.7



This histogram contrasts Simulation A in that the p_regain and p_lose values are still equal but they are much greater now. The resulting distribution is much narrower in range as well as has much higher peaks. Note the change in scale between the two histograms when comparing.

The data still does not fit the normal curve well, and again skews to the right.

Simulation C: p_regain = 0.1, p_lose = 0.1



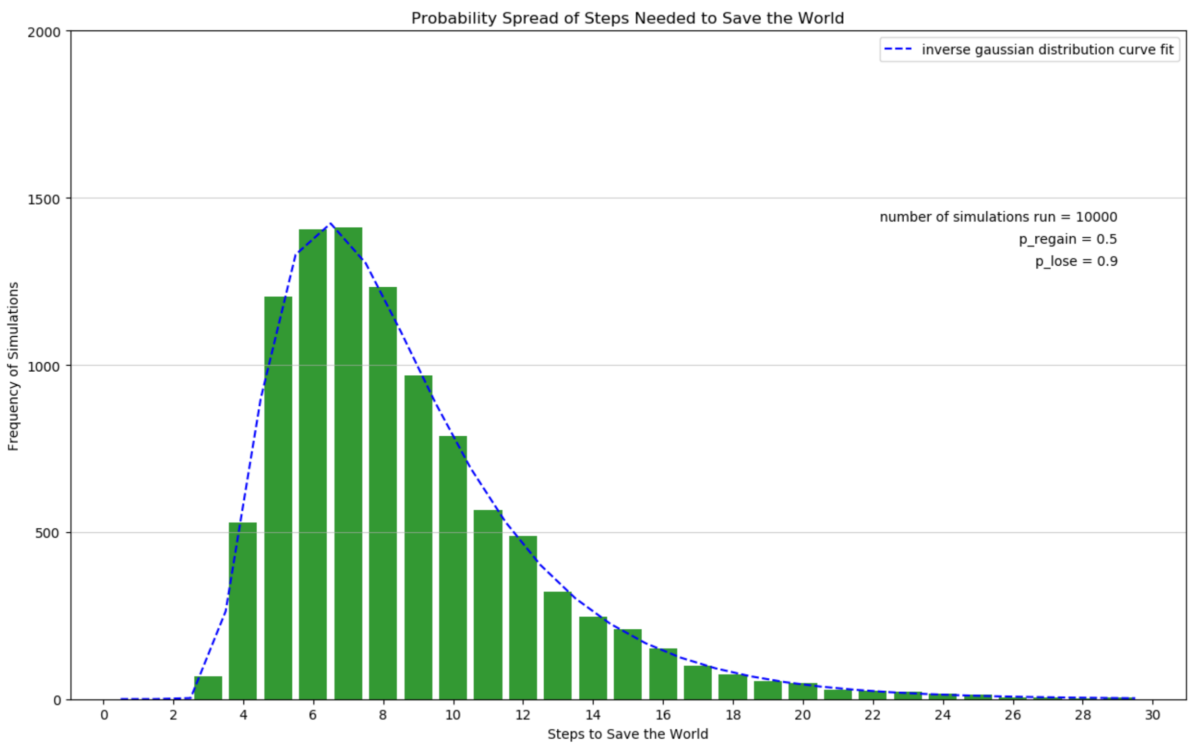
This simulation is interesting in that the p_regain and p_lose values are minimized to 0.1. This results in the histogram with the lowest height/peak and the broadest range of steps needed to save the world. Again, there are visible outliers from the inverse gaussian distribution curve fit because only 500 simulations were run. The data is unsurprisingly skewed to the right again and does not match the normal curve.

From this data, we can see that when the simulation is pushed to one of its lowest probability values, it can require up to around 156 steps for the world to be saved in this data set. From the curve fit, the distribution minimized and converges on zero at around 135-140 steps.

Analysis 5.

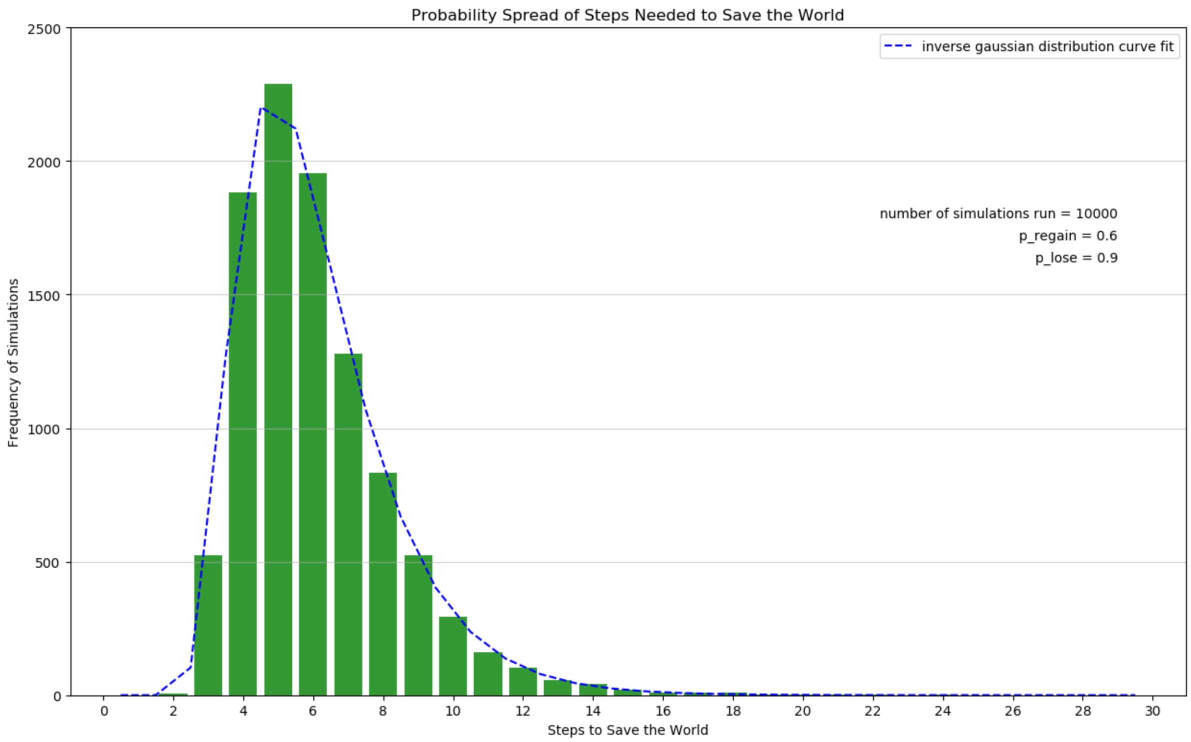
The world must be saved within 2 years, which is equivalent to 24 steps. The p_lose value is set to 0.9 and the p_regain value must be altered so that the right edge of the histogram does not pass 24. In order to obtain a more accurate probability spread, 10000 simulations will be run for data collection.

As a benchmark, we set p_regain = 0.5:



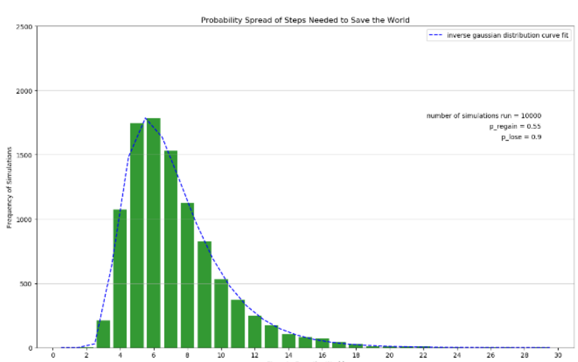
From this histogram we can see that the data is already very close to the desired spread, however there is a still a very low chance that it will take more than 24 steps to save the world. In order to shift the data points to the left, the p_regain value should be increased (This trend was discussed in **Analysis 1.**

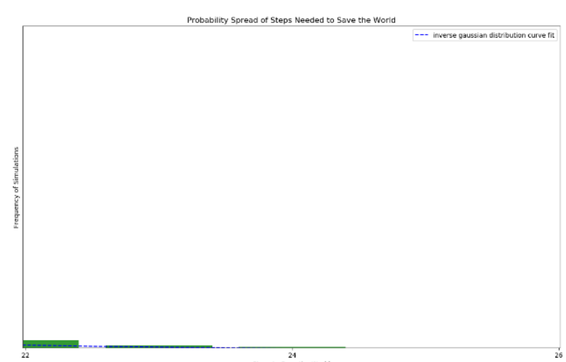
Now we set p_regain = 0.6:



Luckily, this data already fits the desired specifications. All data points fall under 24 steps to save the world; in fact, they seem to all fall below 20. This means that p_regain ≥ 0.6 will ensure that all cities are saved within 24 steps, or 2 years.

EXTRA:
Expanding this further out of curiosity, the p_regain value was set to 0.55 as a midpoint between the two previously examined simulations:





Pictured on the left is the complete histogram generated by these values, and on the right is a zoomed in view of the 22 – 26 range of steps needed to save the world. We seem to be lucky again, as there seem to be no simulations that required past 24 steps to save the world. From this data we can conclude that a minimum of p_regain = 0.55 ensures the world will be saved within 2 years with a high degree of certainty, and as it is increased to p_regain = 0.6 or higher, the certainty will be further reinforced and virtually be 100%.