Purdue University Industrial Engineering

IE332 Project Report

March 10, 2020

Group 11
Michael Wang, Julia Monti, Karuna Srivastava,
Richard Amayo, Mrunmayi Dandekar

As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do.
Accountable together - we are Purdue.

Instructor
Dr. Mario Ventresca

Login Information

ITAP Account: Username: g1081391
Password: W1fihasnomeaning

MySQL: Username: g1081391
Password: Wif1hasnomeaning

Website URL: `web.ics.purdue.edu/~g1081391/Group11/index.php`

Bonus Point Considerations

1. Phase 1, Phase 2 and Final Report all finished in LaTeX
2. Workplace names randomly generated with string matching algorithm
3. Modeled in great detail disease spread simulation with 100 time steps/days
4. Animated SIR model progression of disease spread
5. 36 fully-functional results of exhaustive disease parameter combinations
6. Visual aids on population demographics displayed for user
7. Pruned 36 decision trees and recommended policies for user
8. Fully interactive decision trees that display node information, lift plots, and gain plots
9. Highlighted information regarding each of the four types of disease
10. Description of pros, cons, and economic impacts of all possible policies
11. Simulation re-run with various policies enacted
12. Contact page of website validates entries of name, email, and subject line before allowing submission.
13. SIR Diagram produced in AutoCAD
14. Social media accounts created and linked on the website
15. README.txt file outlining file structure in the deliverables folder

Contents

List of Figures

List of Tables

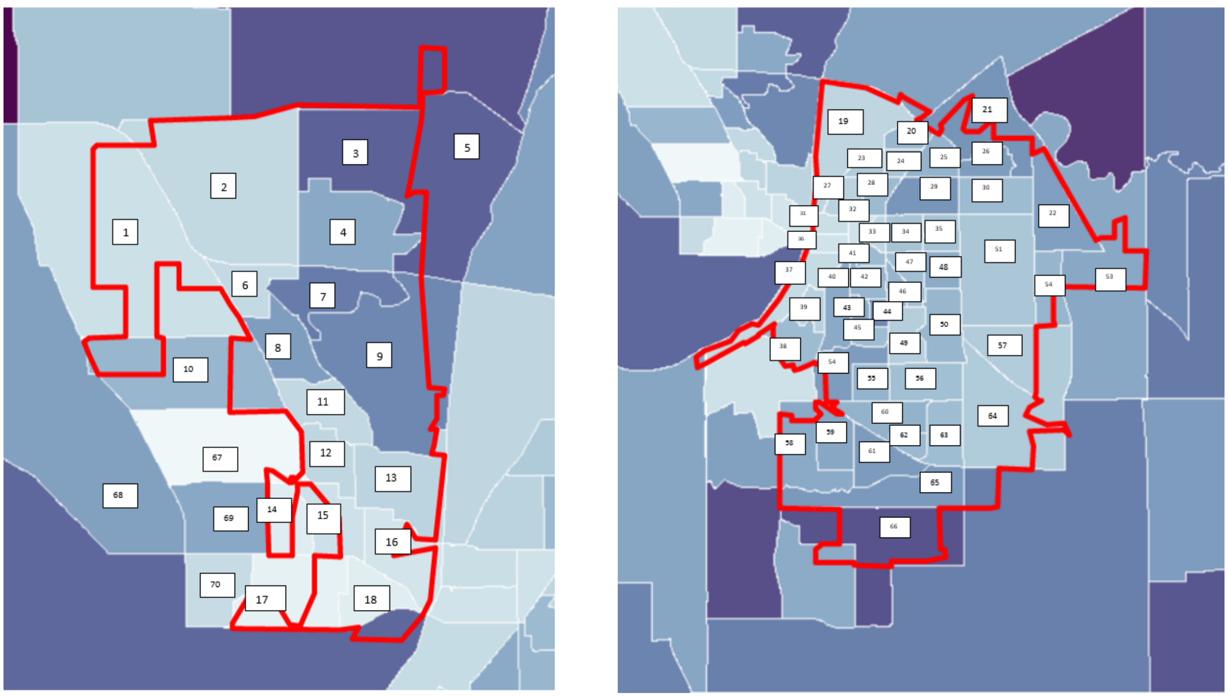
1 Executive Summary

As requested by the cities of West Lafayette and Lafayette, the Engineering Team at Disease Solutions constructed an automated process that simulates disease spread throughout the populations of both cities. The solution aids public officials in devising flexible and effective policy strategies to mitigate the spread of pandemic disease through the use of simulation and machine learning.

This project is the culmination of the following tasks: Population Simulation, Database Design, Disease Spread Simulation, and Policy Optimization. All tasks ensure a highly accurate representation of a real-world pandemic in the cities of West Lafayette and Lafayette.

2 Population Simulation

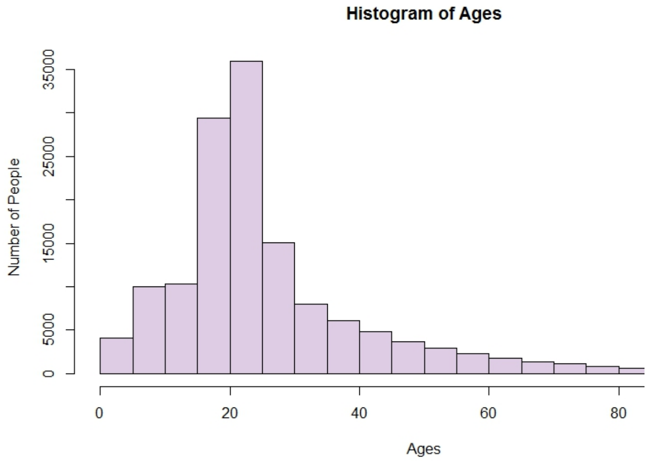
Demographic data on the population of West Lafayette and Lafayette, schools, workplaces, and locations of interest (recreational centers, stores, etc.) was used to aid in the Monte-Carlo simulation of a realistic population. Please refer to Section 8.1 for details.



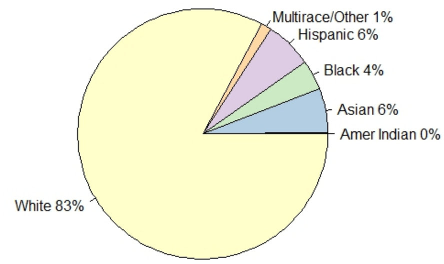
(a) West Lafayette (b) Lafayette

Figure 1: Sectioned Regions for Demographic Data Acquisition

Simulation probabilities are predominantly based on region-by-region statistics with the sectioned regions shown above in Figure 1. Age and sex probabilities are characteristics of individuals. Income, public transport use and household type are characteristics of households. Data regarding types of workplaces, bus routes and schools to model the population were collected and so was information regarding various areas of interest, such as recreational centers and shopping centers.



(a) Age Distribution



(b) Racial Distribution

Figure 2: Example of descriptive plots of West Lafayette and Lafayette simulated population

The generated population of West Lafayette and Lafayette contains 140,632 individuals who belong to 40,497 different households. Figure 2 gives two examples of descriptive statistics of the generated population. Please refer to Section 8.1 and 8.4 for more details on attributes of individuals and households modeled and the heuristics used to generate the simulated population.

3 Database Design

The database contains data procured from the cities of West Lafayette and Lafayette which includes essential information regarding the population, transportation systems, schools and businesses. Attributes of individuals and households include race, sex, age, use of public transportation, communities, and workplace. Not only are region-specific statistics available for the user (see Section 6), but the database design allows for customized statistics — some of which are utilized in Disease Solution's disease spread model. Please refer to Section 8.4 for the complete Entity Relationship Diagram (ERD) and its logical relationships.

4 Disease Spread Model

The simulation uses a SIR(D) model in which the infected (I) persons are assumed to infect susceptible (S) persons through direct interaction until the infected either recover(R) or die (D) and therefore are no longer susceptible. SIR(D) models may be used for diseases in which individuals infect each other directly, rather than through a disease vector, and when individuals who recover from illness have perfect immunity to the disease thereafter.[14] Four diseases modeled were influenza, smallpox, measles, and Ebola — each of which have different transmission rates, recovery rates, and death rates — fit the above criteria. Please refer to Section 8.2 for choice of disease and details. In Figure 3 below, this model is furthered by Individual Specific Multipliers (ISM), namely age and race. The time step used for the simulation was one day and each simulation was run for 100 days. The simulation was run with 36 unique combinations of disease parameters. More information on the simulation may also be found in Section 8.2.

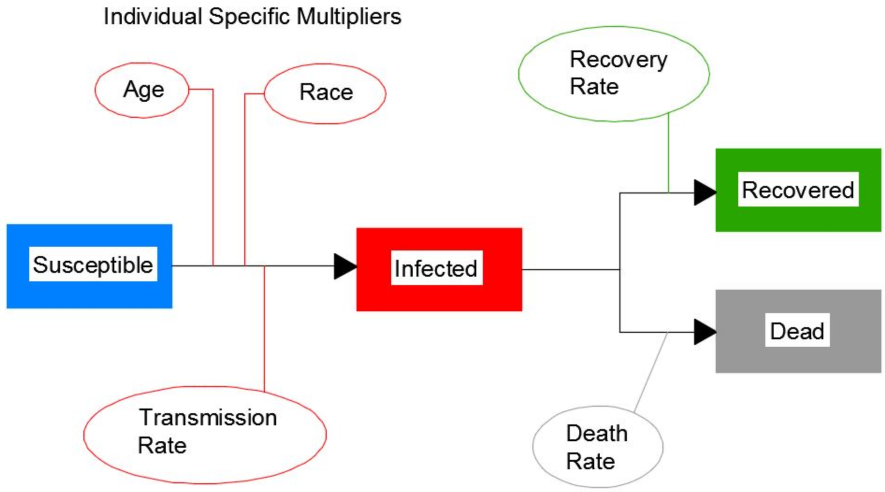


Figure 3: Diagram of SIR(D) Model Mechanics.

Susceptible individuals may be infected by direct contact with infected individuals based on a weighted probability. Factors of this probability include age, race, duration of contact, and disease type. The infected subset of the population may either recover or die, each of whose rates depend on the disease and immune system of the individual.

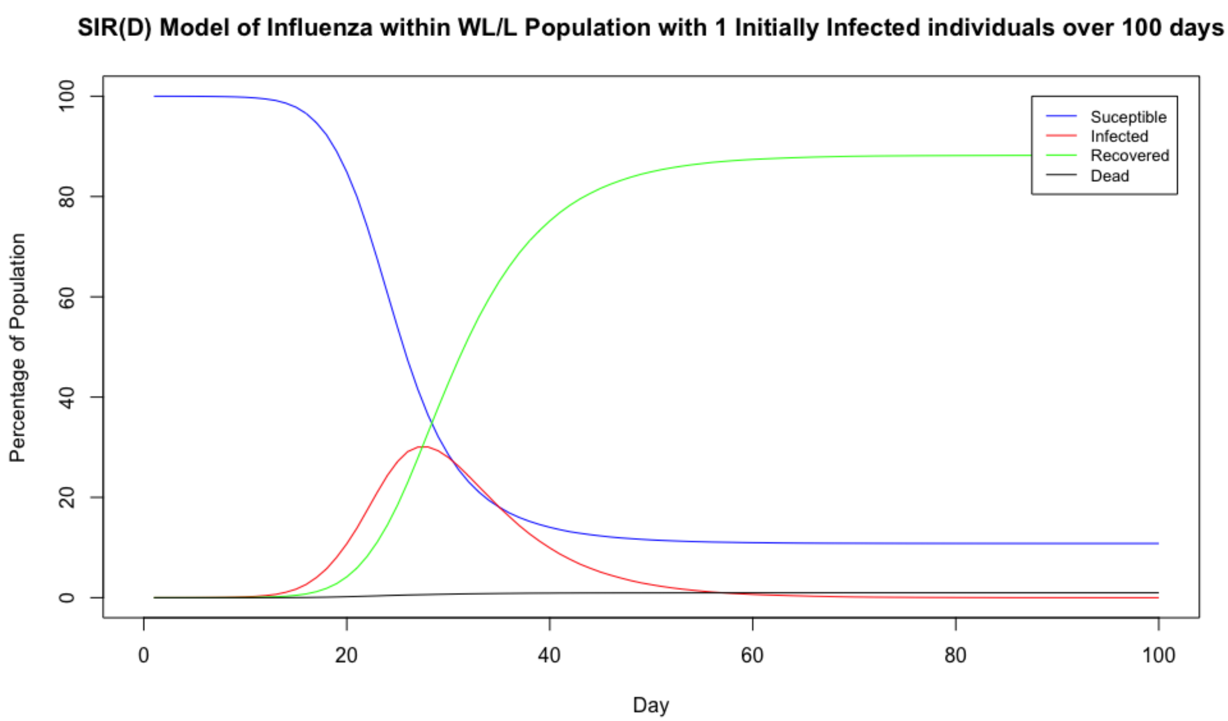


Figure 4: Example SIR(D) Model: Influenza, 1 Initially Infected, 0.1 K node Percentage

5 Policy Optimization

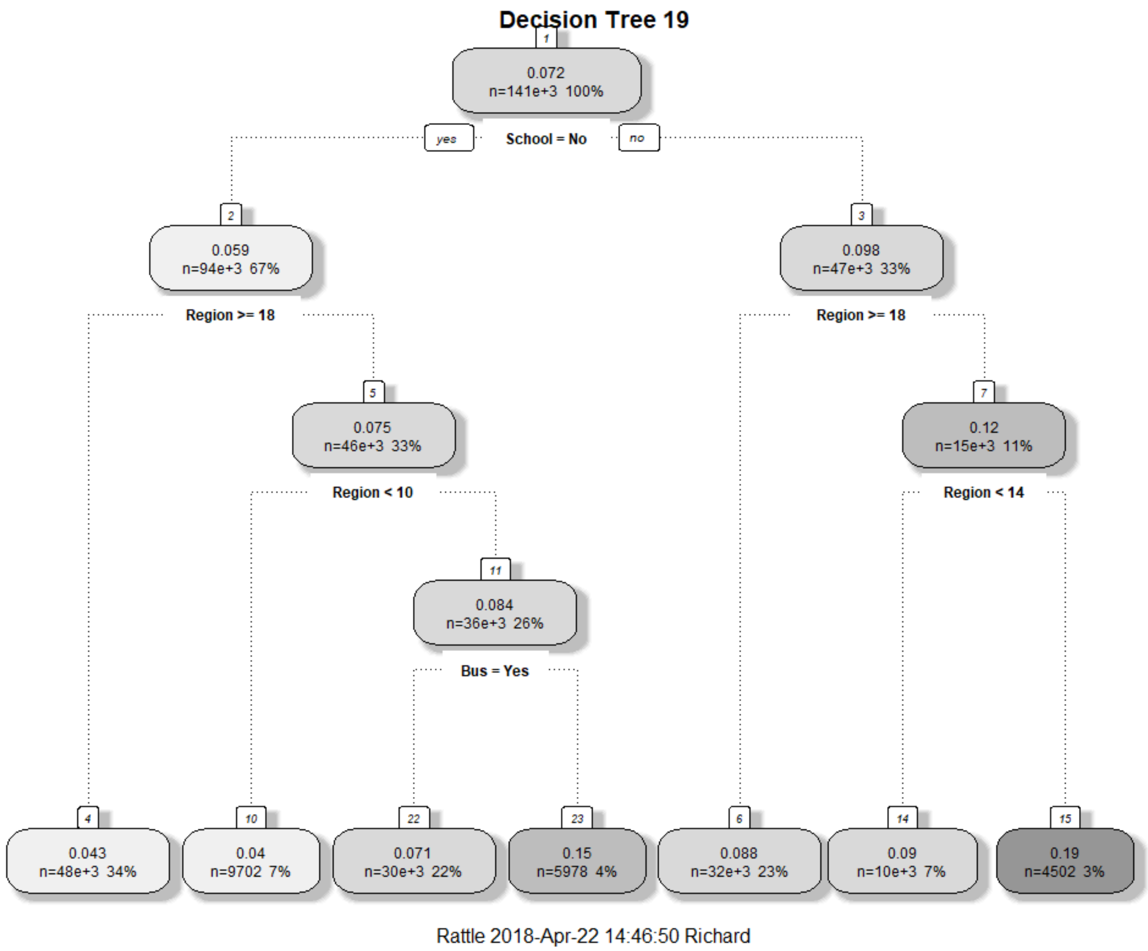


Figure 5: Example Decision Tree for Measles with 1 Initially Infected and 0.1 K Node Percentage

To ensure the most effective policy for each combination of disease parameters, a policy optimization algorithm was created based on machine learning. Along with disease type and the number of people initially infected, public officials may input a percentage (10%, 30%, 60%) of infected individuals to target (K node percentage). The "targeted individuals" were classified to be the subset of individuals who had infected the most people during the simulation. These individuals are labeled and input into the decision tree algorithm which is furthered in Section 8.3. Please refer to Section 8.2 for the technical details of the disease spread model.

The decision tree returns attributes of the these target individuals that most importantly classify them as a K node. Complexity of the trees can be customized for the user. These attributes such as school attendance and public transportation usage assist public officials in their policy decision. Please refer to Section 8.3 for details on the implementation. To aid decisions, Table 1 from the website summarizes the costs associated with each choice.

Once recommended a policy, the user may view the results of re-running the simulation with the policy enacted. The SIR model is updated with the removal of target individuals and mixing groups dependent on the policy. The percentage of K nodes corresponds to the scope of the policy. Please refer to Section 8.5 for more details.

Table 1: Average Cost of Implementation of Possible Policies to Consider

Policy Name	**Average Implementation Costs**
Flu Vaccination	$12.30 per person receiving treatment
Smallpox Vaccination	$3.00 per person receiving treatment
Measles Vaccination	$42.12 per person receiving treatment
Ebola Vaccination	$135.90 per person receiving treatment
Closing Public Schools	$222 per student per week closed
Closing Workplaces	$192,250 per business per week of closure
Closing Bus Routes	$1164 per route per week
Quarantining Individuals	$160 per person per day
Quarantining Households	$555 per household per day
Quarantining Regions	$321,445 per region per day
Isolating Infected People	$1,254 per person per day

6 Website Design

To provide users with background on the population simulated in the disease spread model, the demographics page includes the table shown in Figure 6 to show regional data about population parameters.

Region-Specific Demographics

Please select a region ▾

Select

You have selected: Region 11

Parameter	Data
Region	11
Population Density (people/sq mile)	7118
Number of Households	318
Total Number of Residents	799
Number of Infants	30
Number of Children	261
Number of Adults	449
Number of Senior Citizens	59
Average Age	29.36426179
Most Common Race	White
Number of Caucasian Individuals	505
Number of Asian Individuals	163
Number of African American Individuals	99
Number of Hispanic Individuals	9
Number of Multirace Individuals	0
Number of American Indian Individuals	505
Average Income per Household ($)	44162.0818
Homes Below Poverty Level	51
Individuals Living Below Poverty Level	128
Number of People Regularly Using Public Transportation	181

Copyright © 2018 **Disease Solutions**

Figure 6: Region-Specific Population Statistics

To provide users with the most flexibility possible, the website's simulation page includes three separate drop down menus as shown in Figure 7 to customize the simulation results displayed.

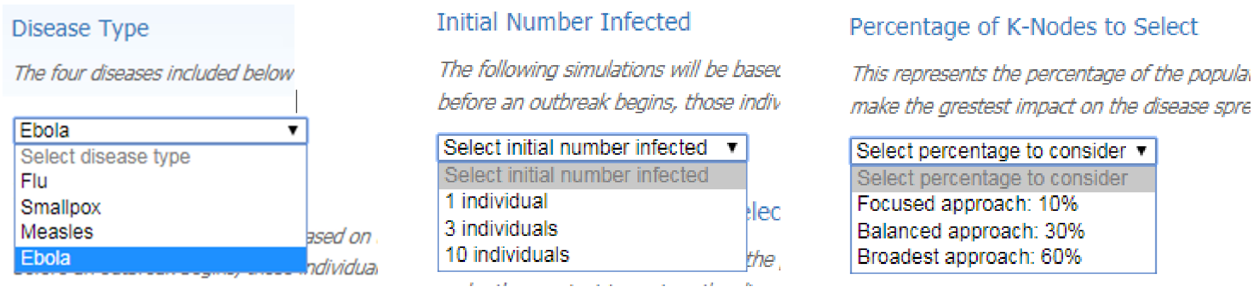


Figure 7: Flexible Disease Spread and Policy Optimization Parameters

After the user customizes their desired output, in addition to the graphics shown in Figures 4 and 5, they will receive additional information about the disease spread and costs of vaccination based on their choices as in Figure 8. Additional information is available via the "click here" link.

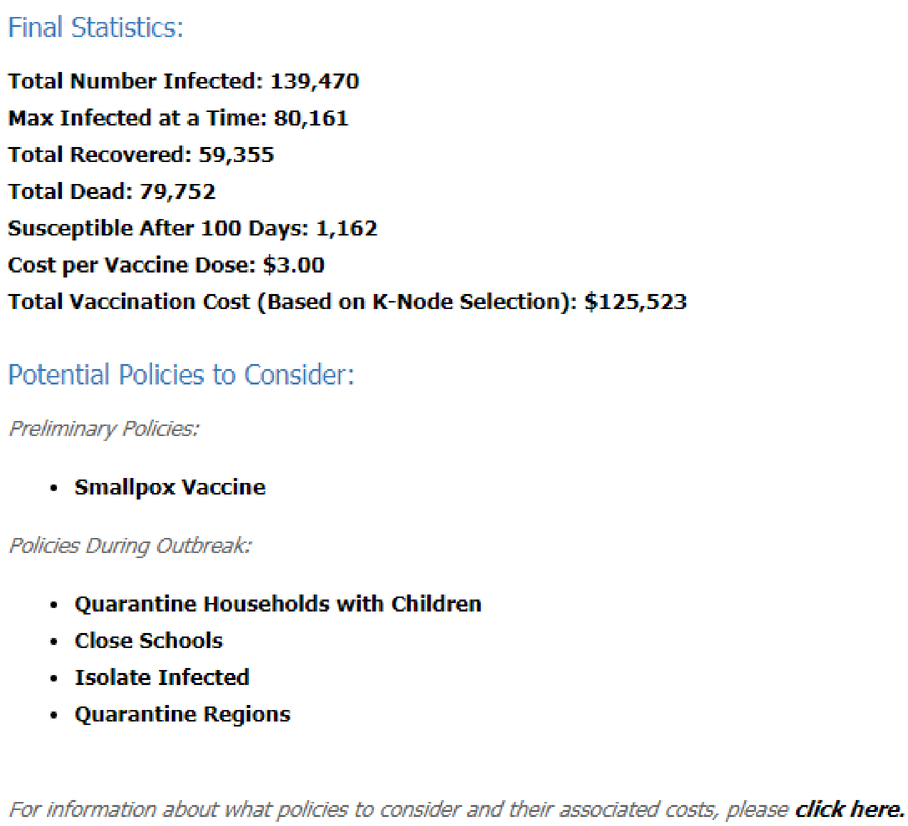


Figure 8: Results from Disease Spread Simulation

The link takes the user to the policy web page, shown in Figure 9, which extrapolates on the costs summarized in Table 1. For public officials, information such as cost of implementation per infected person per day, specific to the cities of West Lafayette and Lafayette, are displayed on this page. Disease-specific information on vaccination costs are also available.

Figure 9: Economic Costs of Possible Policies for Implementation

Please refer to Section 8.7 for more features of the Web User Interface.

7 Future Development

This project can be expanded to suit the needs of the users. For example, it may be useful for the government officials to customize the point in the disease spread simulation where the simulated population has conditions similar to real-time numbers in a pandemic scenario. Another feature that can be added is showing the economic savings of re-running the simulation after choosing a policy, or running all the different policies to determine which is the most effective. Animations showing the disease spread visually would be another way to improve the user interface. This, in particular, would improve the utility of this product as a model-based decision support system. Given more time and knowledge these are developments that can realistically be implemented.

8 Appendix

8.1 Population Generation Heuristics

The simulated population by Disease Solutions is represented both by Households and by Individuals. Region-dependent demographic data of West Lafayette and Lafayette were gleaned from `http://www.city-data.com` during the data acquisition phase.[15 16] Please see Figure 1 for a representation of sectioned regions for both cities. From this source, Disease Solutions procured region-by-region probabilities of attributes of the generated population such as Race, Poverty Status, and Public Transportation Usage. These probabilities are assumed to be dependent on the regional distributions.[15 16]

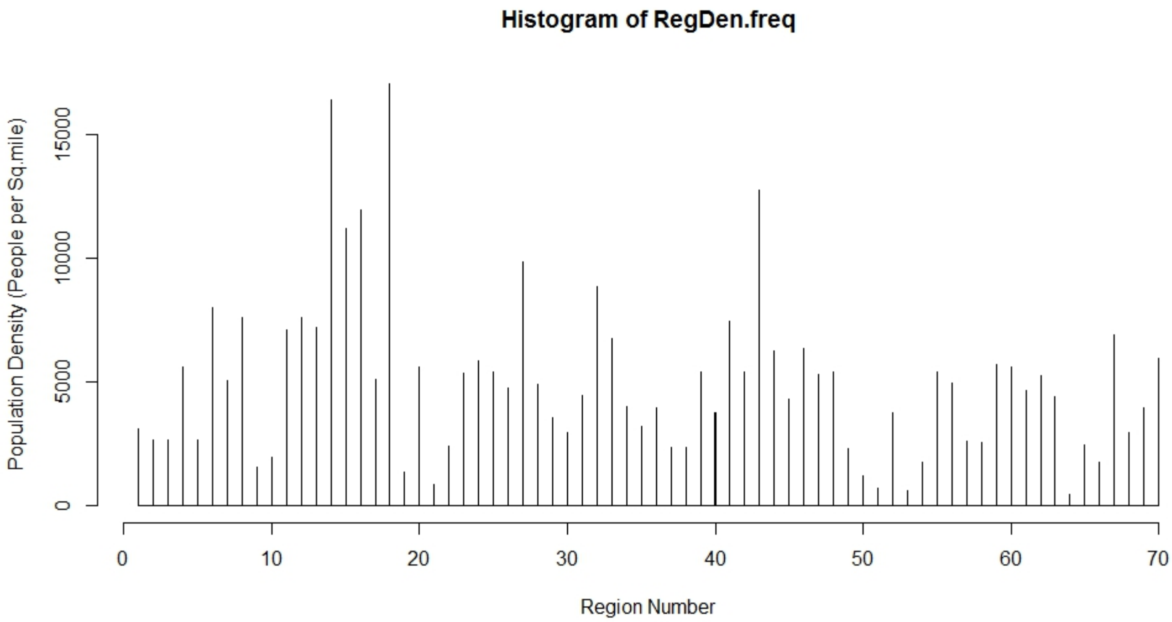


Figure 10: Population Density per Region

One such region-dependent attributes is exemplified above in Figure 10, the population density for all 70 regions of the city. Households that are in Poverty Status are all assumed to take public transportation.[53] Further dependencies between attributes such as Age and School/Workplace will be addressed in the population of individuals. Other utilized statistics include each region's total population and population density. The normalization of the total populations of each region allows for the probabilistic sampling of the region in which each generated household resides. The 40,932 households within West Lafayette and Lafayette thus have region-dependent probabilities for the mentioned attributes and are assigned accordingly. Household Type was determined for West Lafayette and Lafayette separately, which then number of people within each household is sampled accordingly.[56]

The population of individuals is derived from the population of households by expanding along the number of persons in each household. Disease Solutions assigned individual attributes such as Age and Sex based on the population pyramid of both West Lafayette and Lafayette.[55] From the population pyramid it is assumed that up to age 18, the distribution of age is uniform. After the age of 18, the population is modeled with an exponential distribution with $\lambda = 0.05$ after applying non-linear regression to the age distribution of West Lafayette and Lafayette.[57] Then, each individual is assigned a school (if applicable) or a workplace based on occupational statistics gleaned from West Lafayette and Lafayette city data.[40]

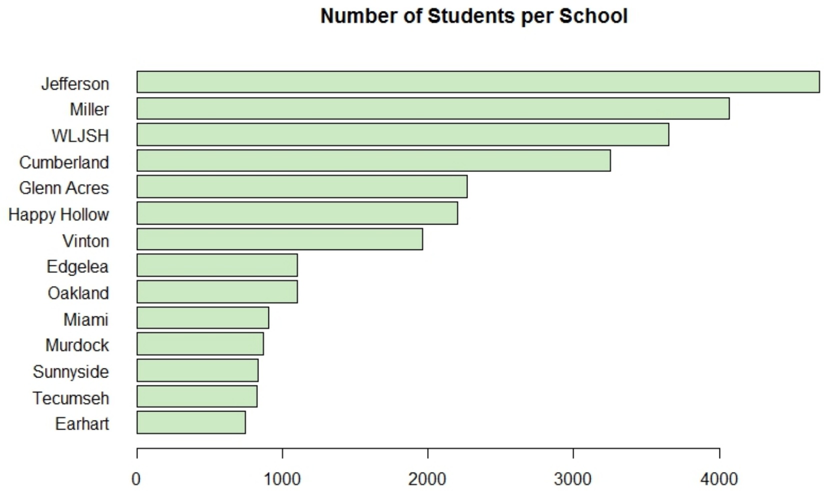


Figure 11: Number of students per school

If the individual is under the age of 18 years old, they are assigned one of the fifteen workplaces shown in Figure 11. Workplace IDs 1-15 correspond to these schools within the West Lafayette and Lafayette areas. For the purpose of the simulation, it is unimportant in which industry the individuals work; only the fact that they share the same business within an occupation matters for disease spread simulation. Therefore, each individual is assigned a Workplace ID number corresponding to a business. Due to the absence of Purdue University students census data on various demographic websites, a population of Purdue students is generated based on enrollment data.[42] These 41,573 students are placed in one of the four regions (67, 68, 69, 70) that encompass the Purdue University campus and surrounding residential areas and are assigned a Workplace ID of 1000. Each region is considered a household, but the number of contacts is limited by the average contacts per day Poisson distribution described in Section 8.2.[40] If the individual is above the age of 64, they are considered ”retired” and are assigned a Workplace ID of 1001.[12]

8.2 Disease Spread Model and K Node Heuristic

The disease spread model uses the data stored in the city database to perform a simulation. The approach is, for each individual, to compile all possible contacts per day for each person. This is done by determining other individuals who share the same mixing groups. The mixing groups utilized in the simulation are those who share bus routes, workplaces, households, and local communities (shared region). The base weight of disease transmission is based on the time typically spent in each of these mixing groups.[6] Bus commute times average 45 minutes per day.[1] Time spent at home, consisting of helping household members, leisure, and household activities, average 303.9 minutes per day.[3] community time (defined as sports, organizational, civic, and religious activities, and helping non-household members) average 255.9 minutes[3] per day. Time spent at work per day is modeled as the standard 8 hour workday, or 480 minutes per day.[2] These weights are then multiplied by a race factor determined by the racebased infection rates for Influenza.[5]

The probability P that transmission occurred during a contact of d minutes was modeled as

$$P(d) = 1 - e^{-\lambda d}$$

where λ refers to the transmission rate.[34] λ was calculated from the transition probability matrix Λ which was summed up columnwise and normalized with the sum of the columns for each age group.[34] This columnwise summation and normalization considers the differing strengths of the immune systems between age groups.

Table 2: SIR(D) Model Transmission, Recovery, and Death Rates of Ages 19-64 for Various Diseases

Disease Name	**Transmission Rate** (Probability per Minute)	**Recovery Rate** (Probability per Day)	**Death Rate** (Probability per Day)
Influenza	0.0003175	0.167	0.001
Smallpox	0.0003175	0.05	0.15
Measles	0.000762	0.125	0.03
Ebola	0.001905	0.2	0.5

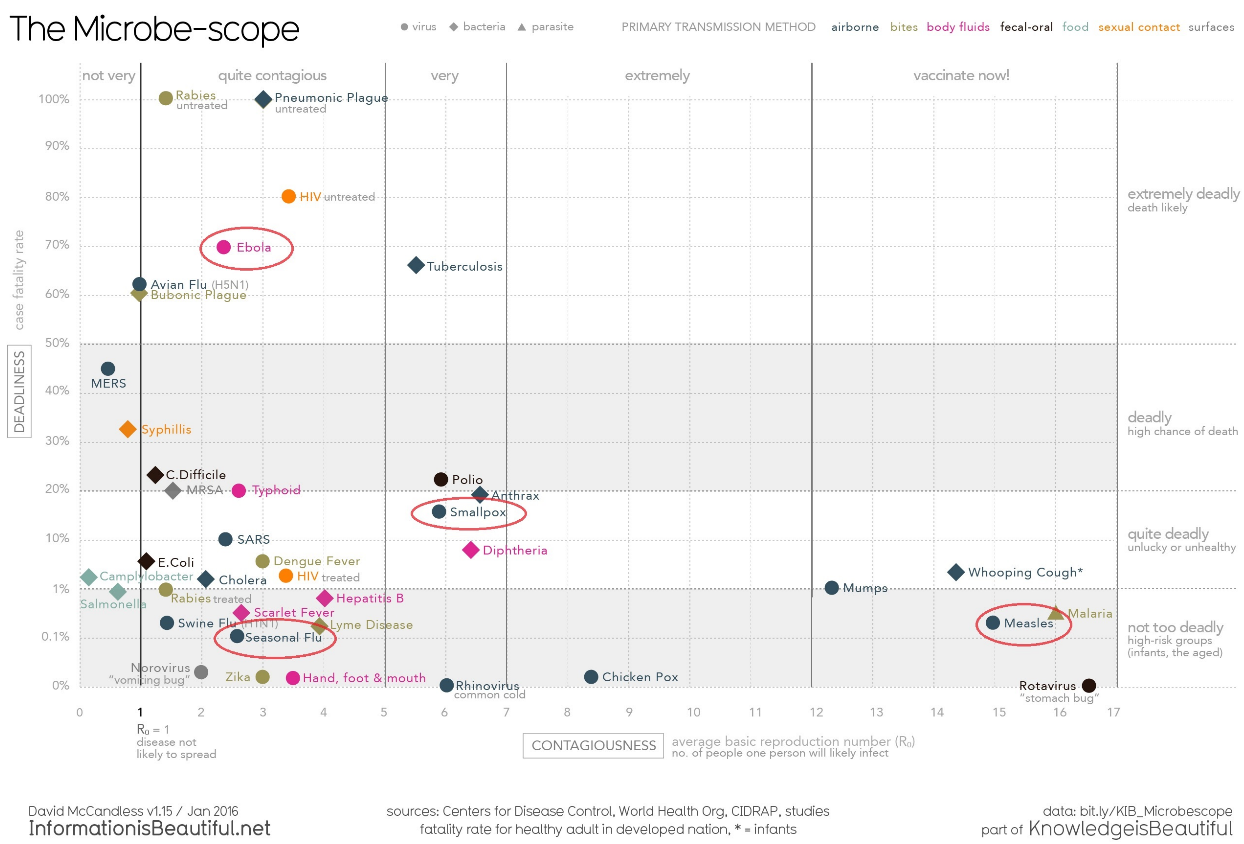


Figure 12: Infectiousness and Deadliness of Disease

Influenza, smallpox, measles, and Ebola were chosen because they have varying infection and death rates as shown in Figure 12 and therefore provide the user with a wide range of options. Users may input their choice of disease, number of initially infected individuals (1, 3, 10) and K node percentage (0.1, 0.3, 0.6). Data on the four diseases, specifically the contagiousness, recovery rate and death rate were integral to the disease spread model to ensure flexibility and accuracy of the SIR(D) simulation, per contact.[60] Recovery rate R per day is calculated by

$$R = \frac{1}{T_s}$$

where T_s is the time spent in infectious class in minutes, per day.[7] These were calculated for all four diseases as shown in the table below.[8,9,10,11] The individuals initially infected are randomly selected from all of the individuals within the population. K-node percentage has no bearing on the results of the simulation itself, but is integral in the decision tree algorithm in Section 8.3.

During the simulation, every infected person on any day is modeled to directly contact a subset of all the possible contacts he/she possesses. The number of contacts per day, per individual, X, is given by a Poisson distribution

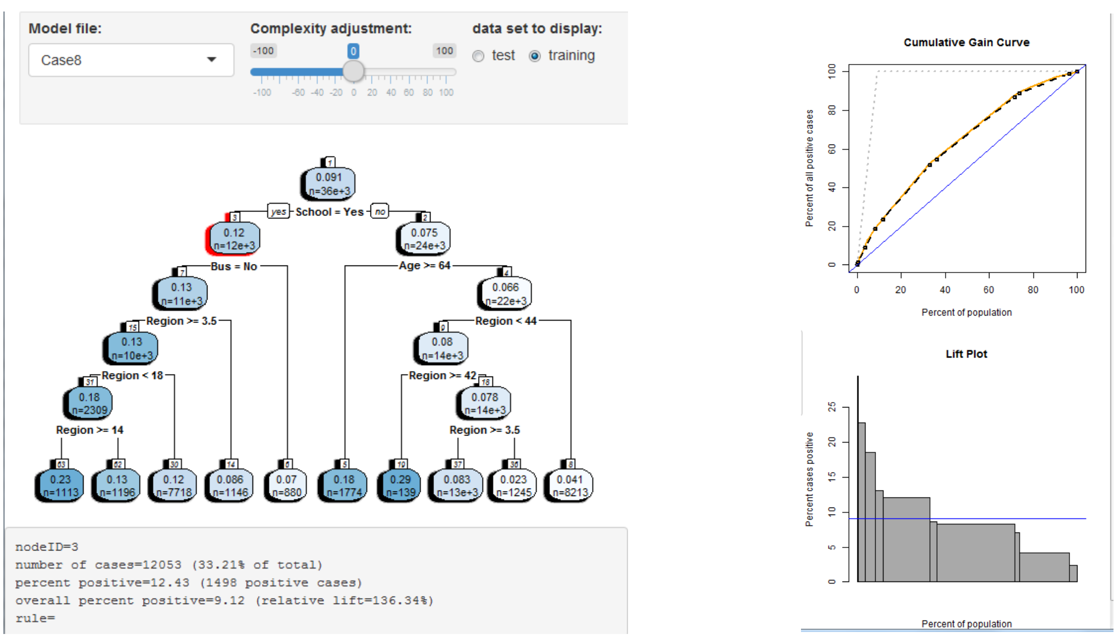
$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where $\lambda = 13.4$.[4] These contacts may be by the means of any of the mixing groups mentioned. Then, the simulation decides whether or not each recipient of the contact initiated by the infected individual contracts the disease based on the transmission probability. On each day, all infected individuals have a chance to infect those who they have direct interaction with. Furthermore, once the day has finished, the infected individuals, excluding those who only got sick during the current day, have a chance at recovery or death based on the disease parameters.[8] The mechanism for deciding recovery and death are similar to the decision of disease transmission. After the simulation duration of 100 days, an SIR(D) model is produced that visualizes the change in the number of people susceptible, infected, recovered, or dead. This is an example of an SIR(D) model, animated: `https://gfycat.com/DimFlawedAnt`

Those who are labeled as a K node are individuals who, in network terms, are central to the population network. They are most likely to infect others in shared mixing groups and are targets of the pandemic mitigation policies. The K nodes heuristic constructed by Disease Solutions is based on a simple scoring of the number of individuals each person infected during the simulation. The logic is that those who infected the most people should be targeted and removed through the enactment of policy. The user-defined K node percentage decides the size of the subset of individuals to be labeled as a K node. All necessary attributes such as Age, Workplace, School, etc. are passed on to the Decision Tree algorithm in Section 8.3.

8.3 Machine Learning Approach: Decision Tree Algorithm

The machine-learning algorithm uses the statistical technique Analysis of Variance (ANOVA) to "learn" about the K nodes to generate a predictive model.[54] The attributes of the K nodes (Age, Sex, Workplace, School) are used as predictors for the model. The entire set of K nodes are inserted into the first node and a binary decision (yes/no) is made about the current predictor. This process repeats until each branch reaches its terminal node.



(a) Dynamic decision tree for each of 36 cases

(b) Lift and gain plots

Figure 13: Customizable complexity of machine learning outputs in user interface

The K nodes that provide the most information relevant for policy action are those with the most correct K nodes. In Figure 13 above, the most prevalent combination of predictors are individuals that go to school, take the bus, and are located in key regions.

Complex models require pruning to allow for prediction to be applied to several situations. Pruning refers to the method used to simplify the trees. The statistical significance is checked for any significance between the attributes and class of the current node. If no significance is found, the tree growth halts and that node becomes a Leaf/Terminal node. After the tree is completed, the complexity parameter is evaluated. The ideal complexity parameter corresponds to the leaf node with the least amount of error. For this example, the user is able to adjust the complexity parameter, which eliminates any nodes that do not meet the minimum complexity parameter specified. The Lift and Gain plots in Figure 13 measures the performance of the model at classifying the choices correctly. A higher value of lift corresponds to a node more impacted by policy.

8.4 Entity Relationship Diagram

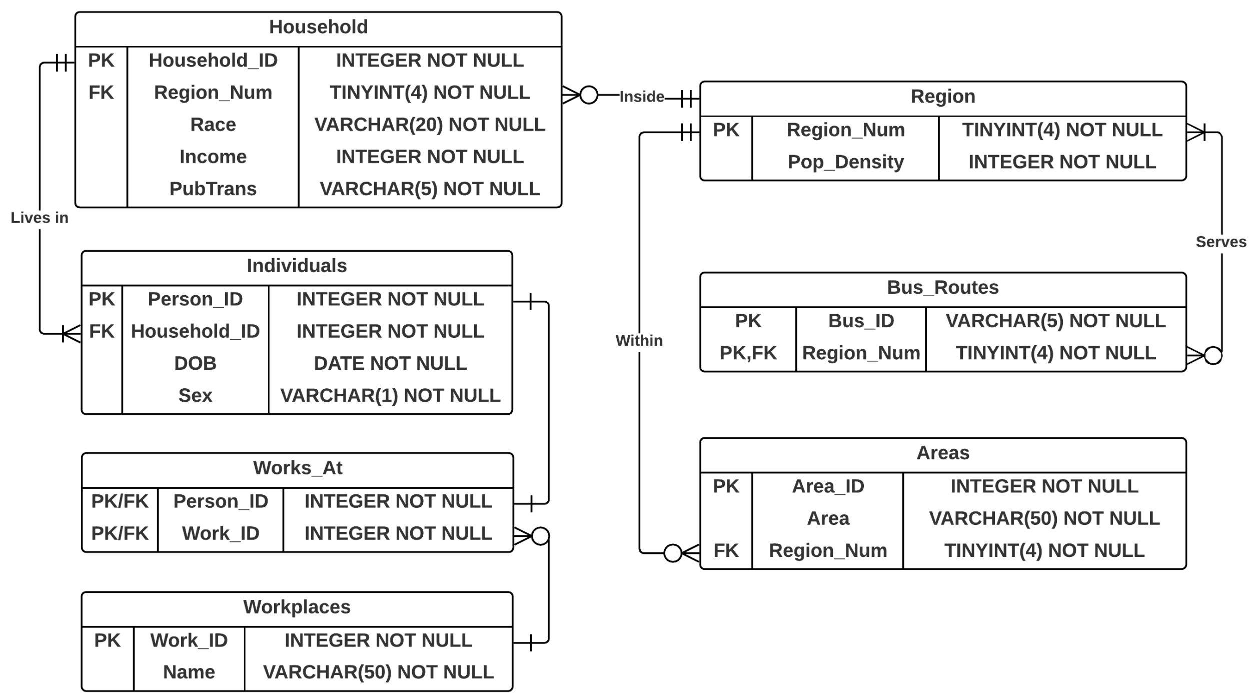


Figure 14: Entity Relationship Diagram

This Entity Relationship Diagram (ERD) ensures fast query performance and good database design. Table *Individuals* represents the population along with immediate characteristics, while the other tables including *Workplaces*, *Regions*, *Household*, and *Bus_Routes* act as mixing groups. Further locations of interest including recreational centers, shopping malls, and restaurants are contained in the *Areas* table but are unused in the disease simulation due to the established mixing group of "Community". Data normalization is achieved through the *Works_At* table and other relationships are labeled along the arrows accordingly.

To ensure relationship logical integrity, the following steps were taken. Each individual may be in one and only one household, whereas a household can contain one or many individuals. An individual can work at one and only one workplace or have no job. However, a workplace can have zero or many employees. Each region contains zero to many households but a particular household can be located inside one and only one region. There are several bus routes that connect different areas within the regions. A particular region can have many buses serving it or have zero connectivity. In conjunction with data integrity constraints, the database structure contains the tables that are shown in the ER Diagram below:

8.5 Calculation of Policy Costs

This is a description of the calculations made to derive the costs summarized in Table 1. The averages for vaccination costs[25,33,47] were taken directly from the Center of Disease Control (CDC) and World Health Organization's (WHO) estimates of the cost for vaccinating adults for those diseases.

To estimate the cost of closing a school in West Lafayette or Lafayette, including Purdue, the National Center for Biotechnology Information's (NCBI) estimate that closing all schools nationwide would cause an economic loss of \$45 billion over 4 weeks[36] and the National Center for Education Statistics' estimate that 50.7 million students attended public school in the United States[37] were used to compute the average cost of \$222 per student per week of closure cost cited in Table 1.

The cost of quarantining and isolating individuals[43] were derived from a study on different economic costs of quarantining, with the definitions defined as stated by the US Department of Health and Human Services (HHS).[44]

The average cost of closing a workplace was derived by averaging the yearly national GDP contribution of the Lafayette area[29] amongst the 1001 companies included in the simulation as replicated from the Greater Lafayette Commerce Business Directory[21] to calculate that the average business would contribute a \$192,250 economic loss per week of closure.

Finally, the cost of closing bus routes was derived from the annual revenue CityBus earns from passenger transport[22] of \$908,109 in 2016. From there, using population use of public transportation[40] and routes in each region,[20] the following table was constructed with the cost of closures:

Table 3: Traffic and Cost of Closing per CityBus Route

Route	Regular Traffic per Year	% of Total Traffic	Yearly Revenue (\$)	Weekly Revenue (\$)
10	13,804	9.89	89,819	1,727
1A	808	0.58	5,257	101
1B	14,454	10.36	94,048	1,809
23	26,101	18.70	169,832	3,266
2A	704	0.50	4,581	88
2B	672	0.48	4,373	84
3	1,440	1.03	9,370	180
4A	849	0.61	5,524	106
4B	38,459	27.56	250,242	4,812
5	14,081	10.09	91,621	1,762
6A	327	0.23	2,128	41
6B	1,057	0.76	6,878	132
7	739	0.53	4,808	92
8	25,130	18.01	163,514	3,145
9	940	0.67	6,116	118
Total	**139,565**	**1**	**908,109**	**17,464**

8.6 Simulation with Policy Enacted

The user is presented several possible policies for disease spread mitigation at the conclusion of the simulation results. Such policies remove a specific subset of the contacts during the disease spread model. For the closures of schools and bus routes, the contacts of each infected individual is truncated to these respective mixing groups. The assumption of isolation and vaccination is that any individual may be targeted at any given point in time.[47] Thus, those who are vaccinated or isolated correspond to the K nodes population subset produced the disease spread model outlined in Section 8.2.

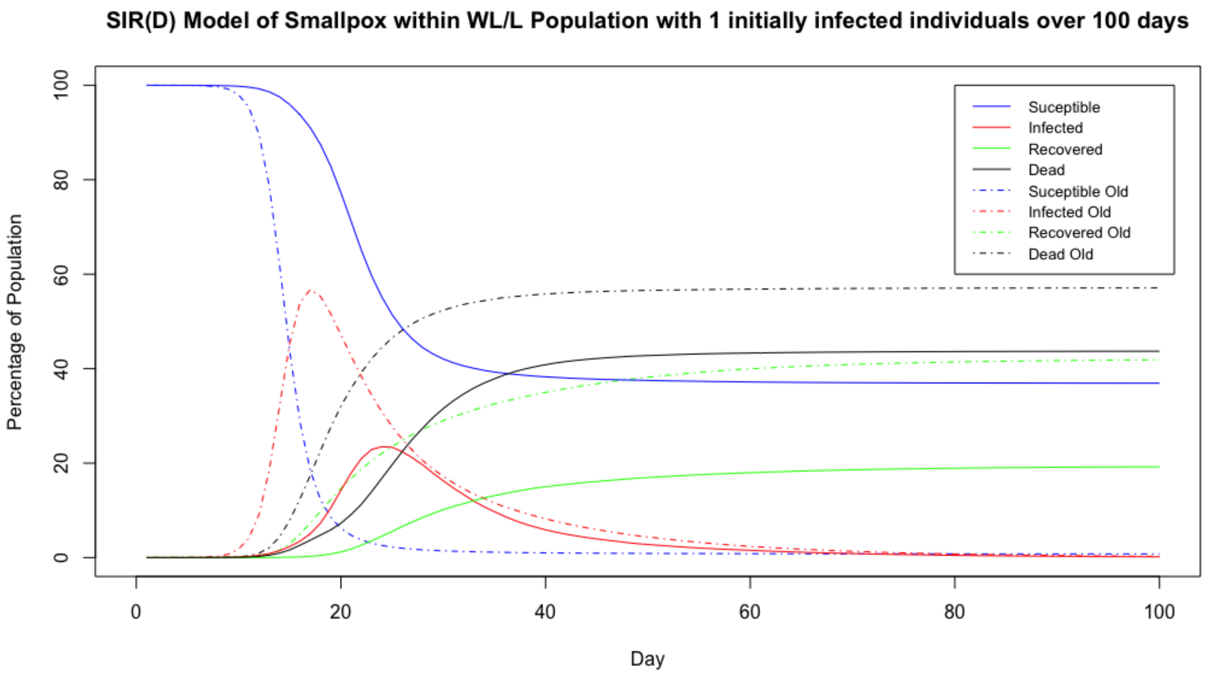


Figure 15: SIR Model of Smallpox with and without School Closure and 0.6 K node Vaccination

Most combinations of disease parameters suggest that vaccination and school closure are most effective in pandemic disease mitigation. Furthermore, for deadly diseases such as ebola, isolation is suggested. The earlier public officials detect the disease (low infection number), the more effective isolation is relative to vaccination due to its high cost.

8.7 Additional Website Features

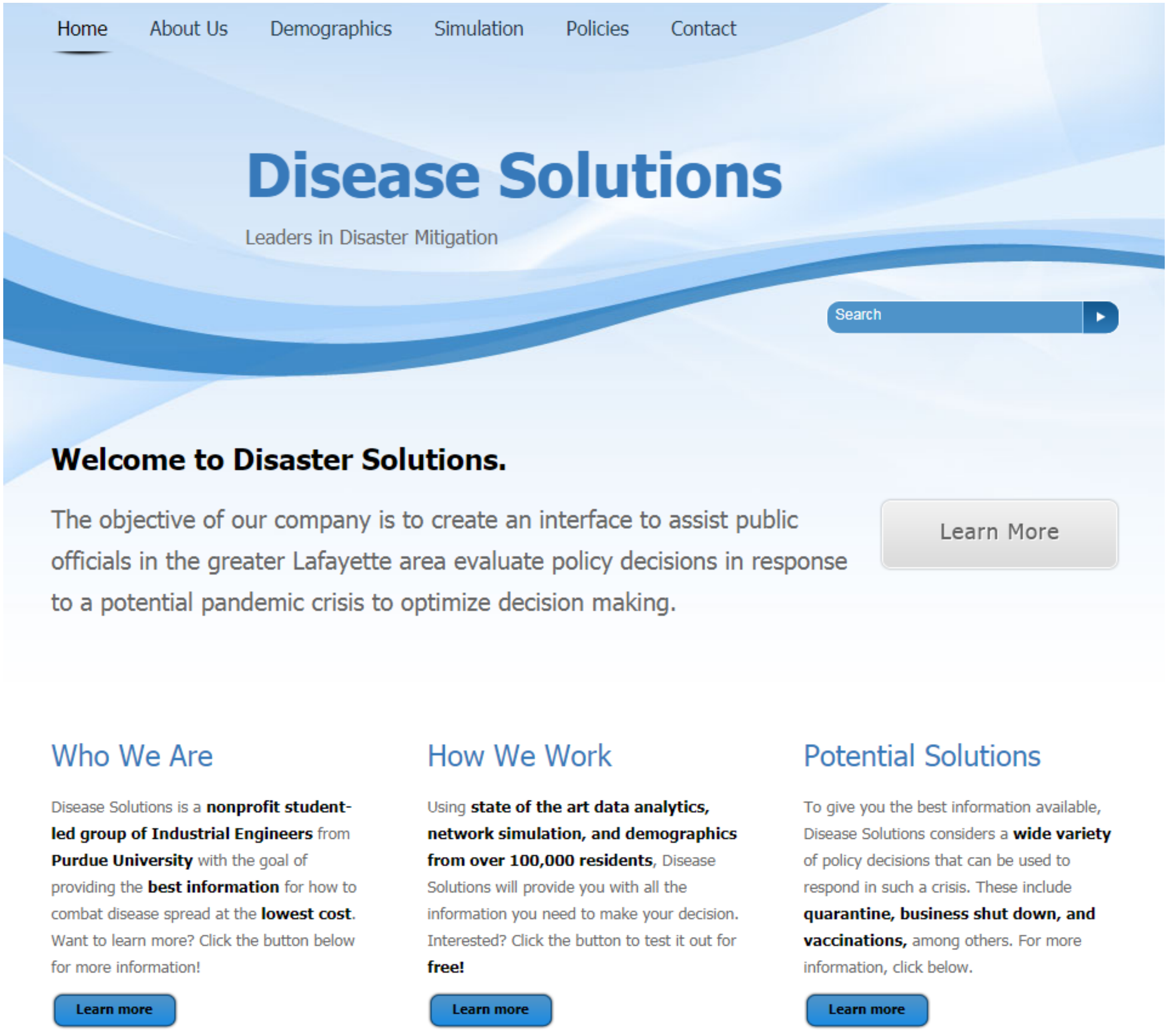


Figure 16: Homepage of Disease Solutions

Above is the homepage of Disease Solutions where the user may redirect to any of the features of the website by navigating the tool-bar and icons.

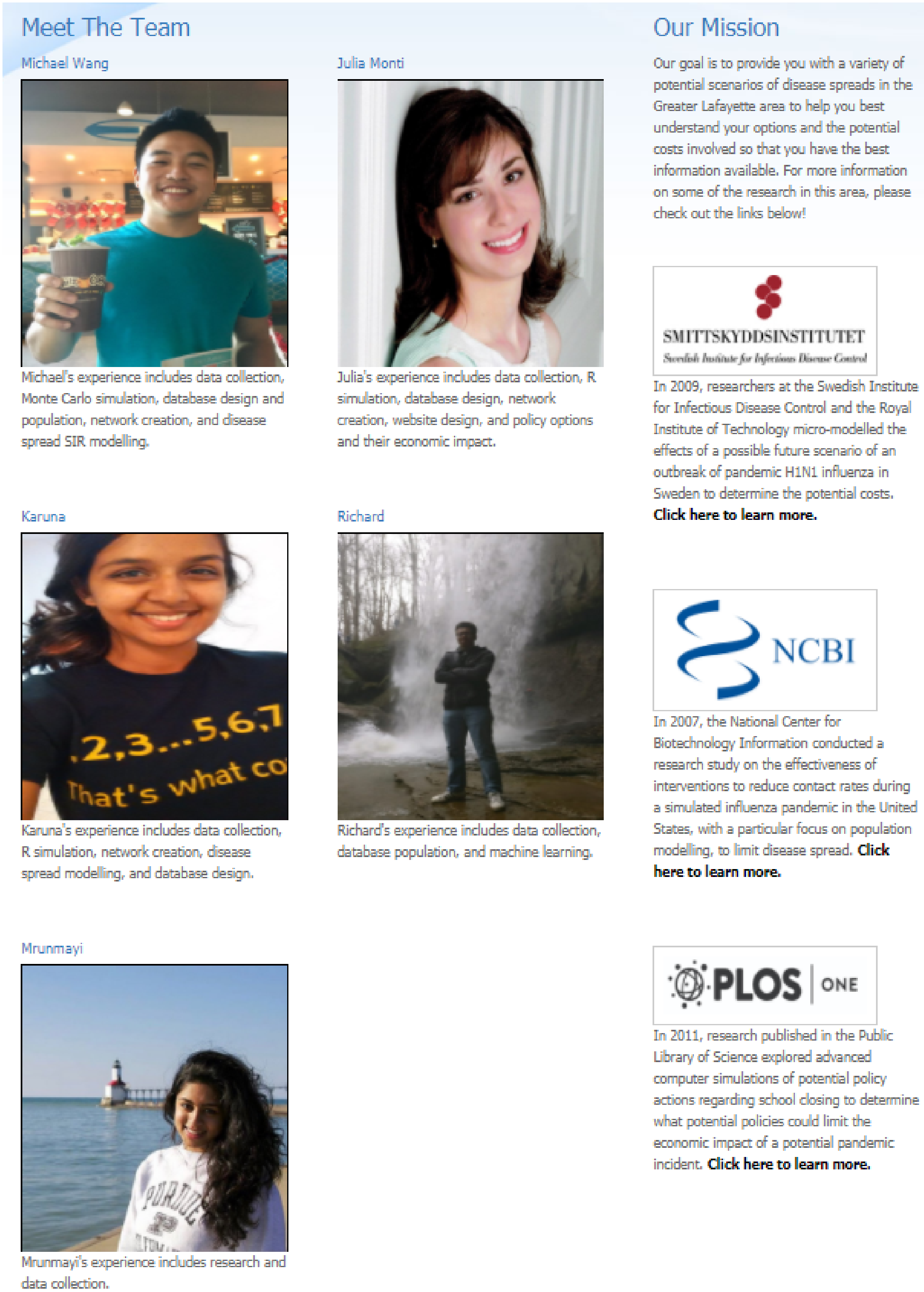


Figure 17: The Disease Solutions Team

Meet the Disease Solutions team! Here, information regarding each team member as well relevant scientific studies can be found.

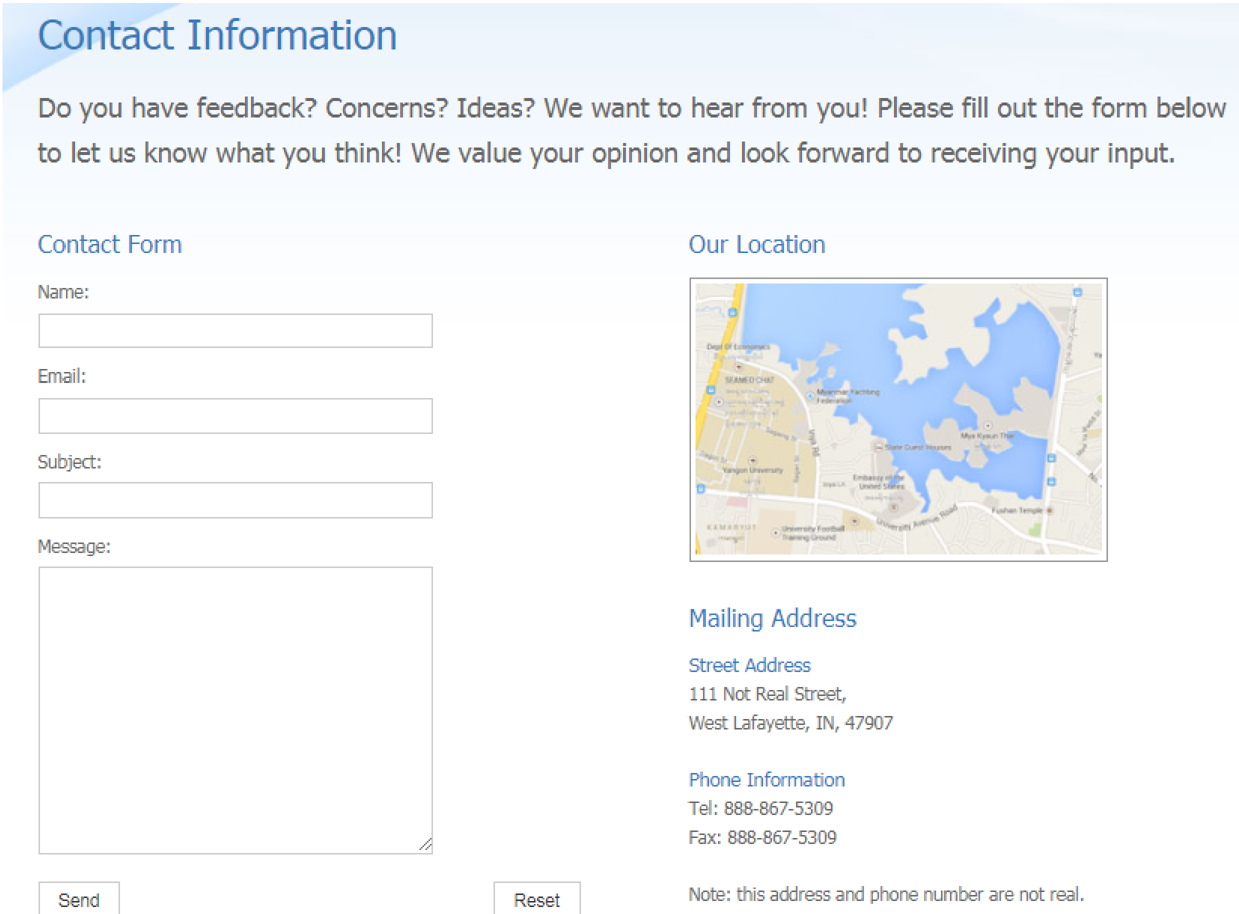


Figure 18: Contact Page

Contact Disease Solutions! The contact form verifies the integrity of user inputs including name, email and message and sends the respective message to the Disease Solutions team. We have contact information available on the right hand side.

References

[1] *Bus Commute Time*: Average Commute Times for Metro Areas. (2017, January 27). Retrieved April 15, 2018, from http://www.governing.com/gov-data/transportation-infrastructure/commute-time-averages-drive-public-transportation-bus-rail-by-metro-area.html

[2] *Work Time*: Average hours employed people spent working on days worked by day of week. (n.d.). Retrieved April 15, 2018, from https://www.bls.gov/charts/american-time-use/emp-by-ftpt-job-edu-h.htm

[3] *Household Time*: Average hours per day spent in selected activities by sex and day. (n.d.). Retrieved April 15, 2018, from https://www.bls.gov/charts/american-time-use/activity-by-sex.htm

[4] *13.14 Contacts per day*: Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. (n.d.). Retrieved April 15, 2018, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2270306/

[5] *Race Multiplier*: CDC Novel H1N1 Flu — 2009 H1N1: Overview of a Pandemic - Impact of 2009 H1N1. (2010, December 9). Retrieved April 15, 2018, from https://www.cdc.gov/h1n1flu/yearinreview/yir5.htm

[6] *Transmission Rate of Different Diseases*: McCandless, D. (2018, February 14). Visualised: how Ebola compares to other infectious diseases. Retrieved April 15, 2018, from https://www.theguardian.com/news/datablog/ng-interactive/2014/oct/15/visualised-how-ebola-compares-to-other-infectious-diseases

[7] *Recovery Rate*: Retrieved from http://www.springer.com/cda/content/document/cda_downloaddocument/9781489976116-c1.pdf?SGWID=0-0-45-1544270-p177336748

[8] *Measles Recovery Period*: Measles - Symptoms and causes. (2018, March 9). Retrieved April 15, 2018, from https://www.mayoclinic.org/diseases-conditions/measles/symptoms-causes/syc-20374857

[9] *Influenza Infection Period*: Clinical Signs and Symptoms of Influenza — Seasonal Influenza (Flu) — CDC. (2016, May 26). Retrieved April 15, 2018, from https://www.cdc.gov/flu/professionals/acip/clinical.htm

[10] *Smallpox*: Smallpox - how long, body, last, viral, contagious, causes. (n.d.). Retrieved April 15, 2018, from http://www.humanillnesses.com/Infectious-Diseases-Sk-Z/Smallpox.html

[11] *Ebola*: EBOLA VIRUSE DISEASE. (n.d.). Retrieved April 15, 2018, from http://wethechemist.blogspot.com/2015/01/ebola-viruse-disease.html

[12] *Oldage Homes*: Search Results Near 47906 — SeniorLiving.org. (n.d.). Retrieved from https://www.seniorliving.org/?geodir_search=1&stype=gd_place&s=+&snear=47906&sgeo_lat=40.4542360022181&sgeo_lon=-86.9429416999434

[13] *Shutdown of Schools*: Economic Cost and Health Care Workforce Effects of School Closures in the U.S. (n.d.). Retrieved April 15, 2018, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2762813/

[14] *SIR model*: (n.d.). Retrieved April 21, 2018, from http://www.public.asu.edu/ hnesse/classes/sir.html

[15] *West Lafayette Demographic Data*: (n.d.). Retrieved March 12, 2018 from http://www.city-data.com/city/West-Lafayette-Indiana.html

[16] *Lafayette Demographic Data*: (n.d.). Retrieved March 12, 2018 from http://www.city-data.com/city/Lafayette-Indiana.html

[17] *Azure Decision tree*: (n.d.). Retrieved April 25, 2018 from https://github.com/Azure/DecisionTreeExplorer

[18] *Age and Sex Composition*:Age and Sex Composition: 2010. (n.d.). Retrieved from https://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf

[19] *Age distribution past 18 is exponential*:Lafayette-West Lafayette, IN Metro Area. (n.d.). Retrieved from https://datausa.io/profile/geo/lafayette-west-lafayette-in-metro-area/

[20] *Bus rides per year*:(n.d.). CityBus of Lafayette Indiana Advertise. (n.d.). Retrieved March 29, 2018, from https://www.gocitybus.com/advertise

[21] *Business Directory*: Business Directory Search. (n.d.). Retrieved March 8, 2018, from http://www.greaterlafayettecommerce.com/list

[22] *Closing Bus Routes*: Crowe Horwath LLP. (2016, December 31). Greater Lafayette Public Transportation Corporation Financial Statements [PDF].

[23] *Cumberland Enrollment*: Elementary School in West Lafayette, Indiana. (n.d.). Retrieved March 29, 2018, from http://public-schools.startclass.com

[24] *Deaths of Old people during epidemics above 65 yrs*:Pneumonia and Influenza Deaths During Epidemics. (1982, January 1). Retrieved from https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/601671

[25] *Ebola Vaccine Cost*: World Health Organization. (2016, May). GEVIT Guidance Appendix [PDF].

[26] *Emergency Planning*: Strategies for Identifying and Engaging At-Risk Groups. (n.d.). Retrieved from https://www.cdc.gov/nceh/hsb/disaster/atriskguidance.pdf

[27] *Factors that affect outbreak controllability*:Factors that make an infectious disease outbreak controllable. (n.d.). Retrieved from http://www.pnas.org/content/pnas/101/16/6146.full.pdf

[28] *Footer before table*: PHP/MySQL: Footer displayed before table (powered by a database). (n.d.). Retrieved March 25, 2018, from https://stackoverflow.com/questions/9694508/php-mysql-footer-displayed-before-table-powered-by-a-database

[29] *GDP Contribution*: US Department of Commerce Bureau of Economic Analysis. (n.d.). Bureau of Economic Analysis. Retrieved April 16, 2018, from https://www.bea.gov/itable/index.cfm

[30] *HTML Statements from MySQL*: Walsh, D., & Coyier, C. (2010, July 22). Build HTML Tables From MySQL Tables with PHP. Retrieved March 21, 2018, from https://davidwalsh.name/html-mysql-php

[31] *HTML Tables*: HTML Tables. (n.d.). Retrieved March 22, 2018, from https://www.w3schools.com/html/html_tables.asp

[32] *iGraph*: N.a (31 Dec. 2017.). igraph R package. Igraph.org. Retrieved from http://igraph.org/r/

[33] *Influenza and Measles Vaccine Costs*: Vaccines for Children Program (VFC). (2018, April 01). Retrieved April 15, 2018, from https://www.cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html

[34] *Likelihood of people getting affected by contagious persons*: Haber, M. J., Shay, D. K., Davis, X. M., Patel, R., Jin, X., Weintraub, E., . . . Thompson, W. W. (2007, April). Retrieved March 29, 2018, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2725959/

[35] *Modeling Disease Spread*: Do, T.& Gatica-Perez, D. (n.d.). The Places of Our Lives: Visiting Patterns and Automatic Labeling from Longitudinal Smartphone Data. Retrieved March 29, 2018, from https://infoscience.epfl.ch/record/192391/files/Do_TMC_2013.pdf

[36] *National School Closing Cost*: Lempel, H., Epstein, J. M., & Hammond, R. A. (2009, October 05). Retrieved April 16, 2018, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2762813/

[37] *Number of School Children in US*: The NCES Fast Facts Tool provides quick answers to many education questions (National Center for Education Statistics). (n.d.). Retrieved April 16, 2018, from https://nces.ed.gov/fastfacts/display.asp?id=372

[38] *Number of Times Americans Visit Their Local Parks Annually*: Retrieved from https://www.nrpa.org/blog/29-number-of-times-americans-visit-their-local-parks-annually/

[39] *Online website template*: Blue Wave. (n.d.). Retrieved March 20, 2018, from http://www.tooplate.com/view/2015-blue-wave

[40] *Population Data*:West Lafayette, Indiana (IN 47906) profile: population, maps, real estate, averages, homes, statistics, relocation, travel, jobs, hospitals, schools, crime, moving, houses, news, sex offenders. (n.d.). Retrieved from http://www.city-data.com/city/West-Lafayette-Indiana.html

[41] *Probability distribution of Income is log-normal*: RFabio Clementi & Mauro Gallegati. (2005, February 2). Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States. Retrieved from http://ideas.repec.org/p/wpa/wuwpmi/0505006.html

[42] *Purdue Enrollment*: Purdue News Service. (n.d.). Purdue University sets records for largest enrollment and highest graduation rates ever - Purdue University. Retrieved March 29, 2018, from https://www.purdue.edu/newsroom/releases/2016/Q3/purdue-university-sets-records-for-largest-enrollment-and-highest-graduation-rates-ever.html

[43] *Quarantine and Isolation Cost*: Mubayi, A., Kribs Zaleta, C., Martcheva, M., & Castillo-Chavez, C. (2009). A Cost-Based Comparison of Quarantine Strategies for New Emerging Diseases [PDF]. Mathematical Biosciences and Engineering.

[44] *Quarantine and Isolation Definition*: Digital Communications Division. (2015, August 21). What is the difference between isolation and quarantine? Retrieved April 15, 2018, from https://www.hhs.gov/answers/public-health-and-safety/what-is-the-difference-between-isolation-and-quarantine/index.html

[45] *Select and Radio Option*: PHP: Get Value of Select Option and Radio Button. (2018, March 22). Retrieved March 29, 2018, from https://www.formget.com/php-select-option-and-php-radio-button/

[46] *Select Count*: Select count(*) from table of mysql in php. (n.d.). Retrieved March 26, 2018, from https://stackoverflow.com/questions/6907751/select-count-from-table-of-mysql-in-php

[47] *Smallpox Vaccine Cost*: Willman, D. (2011, November 13). Cost, need questioned in $433million smallpox drug deal. Retrieved April 17, 2018, from http://articles.latimes.com/2011/nov/13/nation/la-na-smallpox-20111113

[48] *Summary of some attributes*:Populations and Vulnerabilities - CDC Tracking Network. (2018, February 13). Retrieved from https://ephtracking.cdc.gov/showPcMain

[49] *Swedish Influenza Simulation*: Brouwers, L. (2009, September 17). Economic consequences to society of pandemic H1N1 influenza 2009 – preliminary results for Sweden. Retrieved April 15, 2018, from https://www.eurosurveillance.org/content/10.2807/ese.14.37.19333-en

[50] *Volume of People per Store*: Retrieved from http://distributech.net/Demographics.aspx

[51] *Vulnerable Groups*: Vulnerable groups. (n.d.). Retrieved March 29, 2018, from http://www.who.int/environmental_health_emergencies/vulnerable_groups/en

[52] *Websites Resources*: Add "border" to table within echo statement. (n.d.). Retrieved March 25, 2018, from https://stackoverflow.com/questions/23768051/add-border-to-table-within-echo-statement

[53] *Poverty, Policy and Public Transportation*: Thomas W.sanchez (20 Feb. 2008.). Poverty, policy, and public transportation. Sciencedirect.com. Retrieved from https://www.sciencedirect.com/science/article/pii/S0965856408000359

[54] *ANOVA*: N/A. Retrieved from http://trevorstephens.com/kaggle-titanic-tutorial/getting-started-with-r/

[55] *Population Pyramid of Lafayette*: N.a (11 Mar. 2018.). Lafayette Indiana Population 2018 (Demographics, Maps, Graphs). Worldpopulationreview.com. Retrieved from http://worldpopulationreview.com/us-cities/lafayette-in-population/

[56] *Household Type of West Lafayette*: N.a (18 Apr. 2015.). Household Types in West Lafayette, Indiana (City) - Statistical Atlas. Statisticalatlas.com. Retrieved from https://statisticalatlas.com/place/Indiana/West-Lafayette/Household-Types

[57] *West Lafayette Demographics*: Data USA (n.d.). West Lafayette, IN. Data USA. Retrieved from https://datausa.io/profile/geo/west-lafayette-in/

[58] *Decision Tree* : A. (n.d.). Azure/DecisionTreeExplorer. Retrieved from https://github.com/Azure/DecisionTreeExplorer/blob/master/decision_tree_explorer.Rmd

[59] *Stack Overflow*: Stack Overflow - Where Developers Learn, Share, & Build Careers. (n.d.). Retrieved from https://stackoverflow.com/

[60] *Infectiousness and Deadliness of Disease*: Google Docs (n.d.). Microbe-scope. Google Docs. Retrieved from https://docs.google.com/spreadsheets/d/1kHCEWY-d9HXlWrft9jjRQ2xf6WHQlmwyrXel6wjxkW8/edit?usp=embed_facebook