IE 332 PROJECT

FINAL REPORT

May 16, 2019

Group 3

Shaheer Ahmad
Kyle Hannon
Rieley Nelson
Michael Williams
Graden Young

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

Contents

1 Introduction **4**

2 Population Simulation **4**

3 Database **5**

4 Network Analysis and Disease Spread **5**

5 Disease Mitigation **6**

6 Economic Impact **7**

7 Website Design **7**

8 Conclusion **8**

9 Appendix **8**

9.1 Appendix A: Population . . . 8

9.2 Appendix B: *K* Set of Nodes . . . 10

9.3 Appendix C: Machine Learning . . . 10

9.4 Appendix D: Entity Relationship Diagram . . . 12

9.5 Appendix F: Website Images . . . 13

1 INTRODUCTION

Mitigating the spread of disease can have a large impact on communities around the world. However, modeling the effects can pose a very complex problem with deep networks and varied ways of spreading. One way to combat the problem of complexity is to simulate a population and analyze how disease spreads within the generated sample. In this analysis, a Monte Carlo simulation is used to create a population reflecting Tippecanoe County, Indiana. Specific factors including age, employment, and other health related statistics are simulated for each node of the population based on county and city specific data in addition to state and nationwide health data. Given the simulated population, a network of connections between individual nodes was modeled to simulate the normal interactions of life. These connections were based on three primary points: employment, education, and households. A sickness potential for each person was created using the connections that he or she has plus the susceptibility of that person. Then, strategies were identified using machine learning and assessed on their applicability and overall impact. A web application (http://web.ics.purdue.edu/~g1081383/index.php) was developed to examine and compare potential policies given possible scenarios with varying disease spread and mortality.

2 POPULATION SIMULATION

A Monte Carlo simulation was chosen to simulate a population that is representative of Tippecanoe County; data was pulled from several censuses for the comprehensive picture of the population. In some cases, data (public transit, HIV status, etc.) for Tippecanoe county was incomplete so we extrapolated ratios of a smaller set of the population (e.g. Lafayette) and applied it to the entire population. So, the assumption is that for a given factor the distribution in the subset of the population is consistent with the entire population.

A simulated population is powerful for recreating realistic conditions, but the output must be hesitantly drawn upon because of the likely variance of the real population from the simulated one. The factors for this population were chosen because they were deemed important to disease spread and/ or susceptibility of an individual [33]. The following table (Table 1) consists of a list of all the factors, how they are calculated, their source, and what other factors affect them:

3 DATABASE

In order to describe the disease spread of the population, a network of Tippecanoe County was constructed. The model consists of multiple nodes: people, houses, businesses, schools, districts, and hospitals. The vast majority of possible social interaction facilities in Tippecanoe County are left out for ease of model analysis.The entities selected for analysis were done so based on research of what is important to disease spread. The database was restricted to only pertinent and non-calculable attributes. In total, there are approximately 200,000 residents of Tippecanoe County that all have distinct attributes, features, and relationships.

Houses, businesses, and schools were all selected as important nodes due to the level of interaction present between individuals allowing the disease to spread. Households are one of the most important nodes to consider for disease spread, both in terms of predicting epidemic severity and as a target for intervention [21]. Schools inherently foster the transmission of infections from person to person because they are a group setting in which people are in close contact and share supplies and equipment [40]. Lastly, according to an article published by Chron Business, "Individuals working in close proximity to one another and utilizing the same office equipment share a likelihood of contracting communicable diseases from one another." [29]

4 NETWORK ANALYSIS AND DISEASE SPREAD

Between the options of running a stepwise disease spread simulation and a network analysis of the population, we chose the latter. While the network analysis is static, it allows connections in the population to be highlighted as main features for model identification. The model of disease spread through a given population was dependent upon four factors:

1. Connections - the number of individuals a person is in contact with
2. Intensity of Contact - the intensity of each connection
3. Indirect Contact - the number of indirect contacts a person has
4. Susceptibility - the propensity to get ill while coming in contact with a disease

We assumed each student to have ten connections to other students, each employee to have five connections to other employees, and each household to be connected with everyone else in their house [43] [18]. Potential for each individual was built using a sum of these four factors; each factor being given equal weight.

Once the potential of each individual was calculated a classification of sick, dead, or healthy was applied. Note that since this a static analysis, all of those who fall under the sick category were sick and presumed to have recovered during the outbreak. The user of the web application has the ability to select three different sickness rates and three death rates. Each of these rates are based off of three real diseases: Cholera, common Influenza, and swine Influenza (H1N1). However, to allow for in depth analysis by the application-user, the sickness rate of one disease can be paired with the death rate of another. For example, the application user can select the sickness rate of Swine Influenza (20%) and the death rate of Cholera (2.2%). Table 2 shows the sickness rate and death rate of the three diseases.

Disease	Sickness Rate	Death Rate
Cholera	1%	2.2%
Common Influenza	9%	0.1%
Swine Influenza	20%	0.45%

Table 1: Disease Spreading & Death Rates

5 DISEASE MITIGATION

Decision tree classifiers were used to identify possible population features that were highly impactful to disease spread. Features were deemed to be successful for mitigation if they selected a significant amount of sick people while disrupting as little healthy people as possible. The sick population reduction varied from about 16,000 to 46,000 people. The percentage of disrupted people varied from 21 - 66%.

The features for classification were selected from data theoretically available to the Tippecanoe County government. This is the same set of parameters on the database. 6 policies are introduced as a part of disease mitigation. These policies are:

- Policy 1 - Randomly vaccinate 27,804 students in Wabash schools
- Policy 2 - Close all businesses that have 150 or more employees
- Policy 3 - Close all schools
- Policy 4 - Randomly assign 27,804 vaccines to individuals of Tippecanoe County
- Policy 5 - Quarantine everyone within their households
- Policy 6 - Close all food service, retail, arts, entertainment, and recreation establishments

The limit of vaccinations– 27,804 –was found from the fraction of the total federal budget put toward immunization in 2016-17, which was 0.02%. Given the Tippecanoe county's budget is $171 million, the estimated amount they will be willing to spend on immunization is $34,200. This divided by an average cost of $1.23 per vaccine gives the vaccination amount [**?**] [46].

6 Economic Impact

The economic impact of the disease spread was estimated based upon people being removed from work and otherwise displaced from normal economic activity. In addition, in cases where vaccines were dispersed, the overall cost of vaccination was added to lost man hours. The average daily wage in Tippecanoe County was $182.50; this was used as a proxy for total economic loss per person per day of an effective disease spread scenario[44]. While estimating devastation from industry, there is a base loss for each sick worker in the county and an additional loss for any workers that were displaced by a potential policy. For example, if there is a policy to shutdown 10 businesses with 1500 total workers then there is a $182.5*n base loss (where n is the number of workers sick) plus $182.5*1500 from the policy. In addition there is a cost to shutting down a school when the children affected are under the age of 12; this represents a parent taking off work (this an assumption that children under 12 must have adult supervision). We did not factor in an economic multiplier effect for Tippecanoe County because it behaves less like an isolated economic community than normal candidates for multipliers, like countries (i.e. many persons may bank, work, or shop outside of the county).

7 Website Design

The Carbon Solutions team designed the website (http://web.ics.purdue.edu/~g1081383/index.php) with the user in mind. It consists of a total of 5 main pages, varying results pages and common navigation bars and footers. Their names and functionalities are listed below:

- Navigation Bar - The navigation bar consists of buttons leading to the five main pages and a link to our twitter. Having this at the top of every page allows the user to go back and forth between pages. (Refer to appendix F)
- Footer - The footer contains the company name along with a contact email.
- Home - The home page includes a description of the problem at hand, allowing a user to understand how the website is to be used.
- Population - The population page has an about section which explains how the population was simulated along with a link to a further detailed report. The page includes graphs showing the population by various distributions.
- Simulation - The simulation page allows a user to select the disease spread rate, death rate, availability of vaccination, mitigation policy to be used, and also explains the 6 different policies at hand.
- Results - Once all items have been selected in the Simulation page, the user is redirected to a results page. This outputs the amount of deaths, sick people, healthy people, and the estimated

economic cost if the policy was implemented. This allows a user to compare / contrast different policies rather quickly.

- Contact - The contact page includes a form that allows a user to input their name, email address, subject of the message, and the message. Once submitted, an email is sent to Carbon Solution's gmail account for further usage.
- About - The about page gives a brief description of Carbon Solution's motto and background information. This allows users to get to know the company.

8 CONCLUSION

Carbon Solutions has worked hard to provide the public policy officials of the West Lafayette / Lafayette areas with an interactive tool to examine different pandemic scenarios. We hope that with information about the effects of different policies on both a societal and economic level, Tippecanoe County officials will be able to make better decisions about disease spread preparation and mitigation. In addition, our aim is that this report and website can serve as a stepping stone for others investigating disease spread and possible mitigation strategies, so communities everywhere can benefit from a growing understanding.

9 APPENDIX

9.1 Appendix A: Population

ID	Factor	Rate	Source	Influential Factors
1	Gender	51.8% - Male	Suburb Stats & Purdue	None
2	Age	varies	City Data / DataUSA	Gender
3	Ethnicity	varies	DataUSA	Gender
4	Pregnancy	12.09%	CDC	Age, Gender
5	Smoking	18.6%	DataUSA	Age
6	Drug Use	3%	Samhsa	Age
7	Alcoholic	17%	DataUSA	Age
8	Student	varies	DataUSA	Age
9	HIV	0.0848%	City Data / Purdue Data	None
10	Obesity	26.4%	DataUSA	None
11	Diabetes	6.9%	DataUSA	None
12	Public Transportation	4.33%	DataUSA	None
13	Employment	varies	DataUSA	Student, Age
14	Industry	varies	DataUSA	Student, Age, Employment
15	Poverty	varies ≤ 17	DataUSA	Employment, Industry, Age
16	Group Living	varies	City-Data	Employment, Student, Age
17	Housing	varies	City-Data	Age, Group Living

Table 2: Population Factors & Rates

Some of the factors have multiple rates and are described as "varies" in Table 2. The following are descriptions of the variability within those factors.

- Age - We assigned ages 1 through 85 to the population according to the intervals given from City-Data and DataUSA and assumed uniformity within given intervals (for example, we were given n amount of 19-22 year olds and assumed that these ages were distributed uniformly).

- Ethnicity - Ethnicities are distributed according to data from DataUSA.

- Student - All individuals from ages 5-17 are considered students, and 61.7% of individuals aged 18-24 are considered students. This rate was calculated by adding up the number of college students aged 18-24 within Tippecanoe County (Purdue, Harrison College-Lafayette, and Ivy Tech Community College-Lafayette), and dividing it by the total population of people aged 18-24 within Tippecanoe County.

- Employment - For employment, we determined whether a person was employed or not using West Lafayette and Lafayette's combined unemployment rate of 4.3% . Ages of the individuals were also factored into employment to create a more realistic data set. We then placed each employed individual into an industry using employment by industry data.

- Industry - The rate of employment by industry is distributed according to City-Data.

- Poverty - The rate of poverty is taken from City-Data which was broken up by age groups.

- Group Living - This is broken up by different types including group homes, jails, and correctional facilities. These are distributed according to City-Data.

- Housing - This is broken up by the amount of people living in a household and is distributed according to City-Data.

- Pregnancy - The rate was calculated using the CDC standard for estimating the number of women pregnant at any given time The equation used was pulled from the CDC and is as follows: WRA/1000 * ((B*Pb) + (A*Pa) + (D*Pd)) where WRA is the women of reproductive age (15-44); B, A, and D are fertility rate, abortion rate, and fetal loss rate respectively; and Pb, Pa, Pd are the ratio of the year that the respective pregnancy outcome happens. This gave the 12.09%.

- An assumption was made that the poverty status of an individual will affect the individual's chances of catching the disease rather than the individual's level of income.

- Another assumption made is that anyone in the population who smokes must be above the age of 18 and to be considered an alcoholic you must be at least 21 years old.

- A final assumption made for public transit was that the only way to get around was by bus.

9.2 Appendix B: *K* Set of Nodes

In order to mitigate disease spread we used a couple optimization heuristics to rate machine learning output. The first heuristic was maximization of removal of nodes labeled sick. However, we wanted to optimize this maximization with the constraint of minimizing the number of healthy nodes removed from the population. Healthy nodes taken out of the population were considered unnecessary economic damage. We followed these heuristics to rate decision trees predicting sickness given features about the population. This way a branch of the decision tree represents a potential policy that can be evaluated using the above heuristics.

9.3 Appendix C: Machine Learning

Decision tree classification was used to find which publicly available parameters can be used to muster or close off certain parts of the population for disease mitigation. The highest impact classifier of sick people was students (this is presumably because of the high number of connections). Following from this the highest impact publicly available parameter are the following business types: food services, retail, arts, entertainment, and recreation. This is presumably because of the high proportion of students that work there.

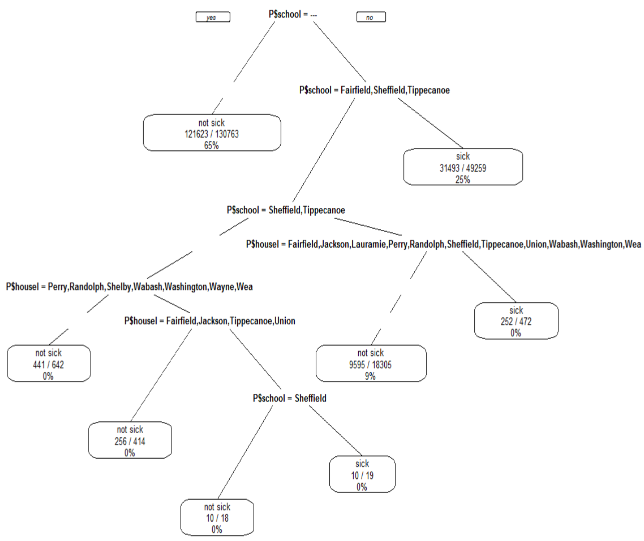

Figure 1: Decision Tree for Policy 1

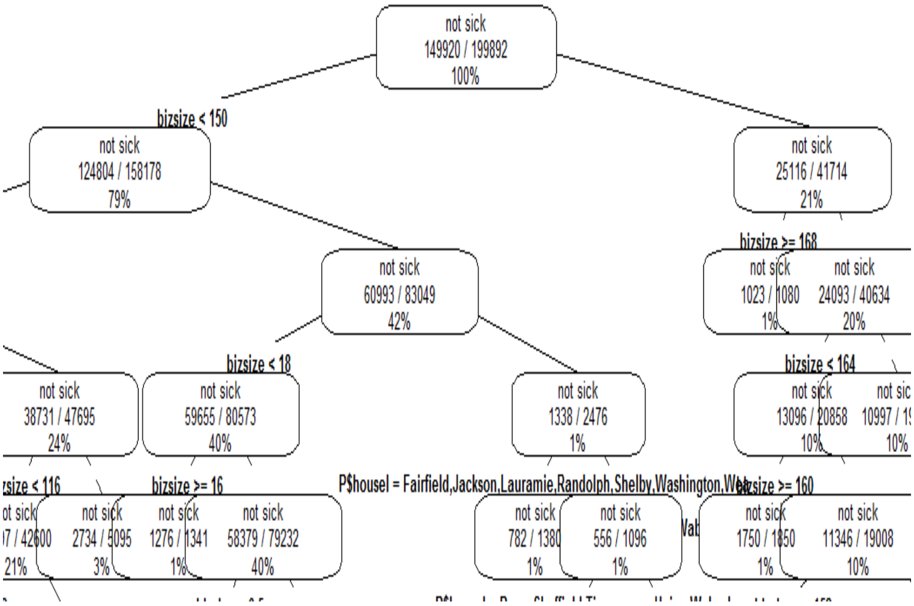

Figure 2: Decision Tree for Policy 2

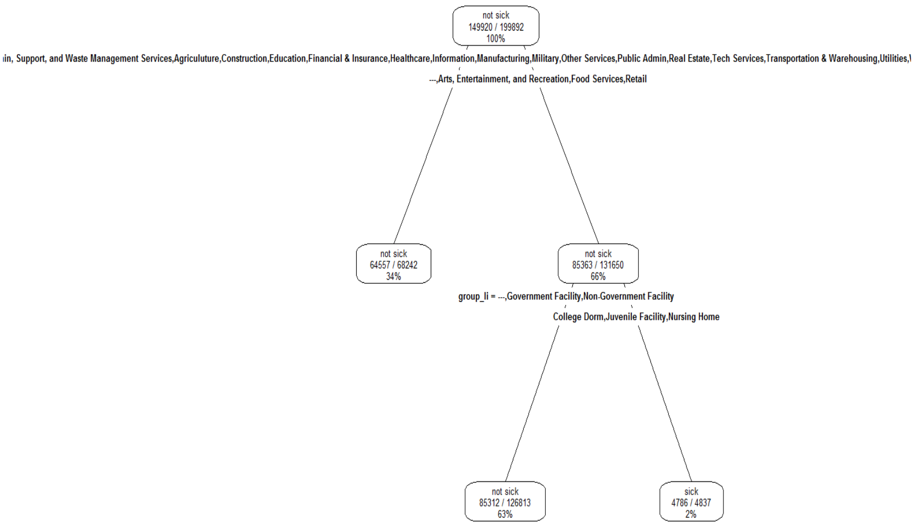

Figure 3: Decision Tree for Policy 6

9.4 Appendix D: Entity Relationship Diagram

The entity relationship diagram was created in a way that centered around people. The relationship between people and homes is one or many to one or many. However, the relationship between people to businesses, schools, and hospitals is one or many to one or none. Going off of this, every home, business, school, and hospital is located in one district so a person must go to their respective hospital / school / etc.. in their district. The final object in the ERD is called SickData, which is what stores data about each individual's sickness. The relationship between people to SickData is one to one. This can be either healthy, sick, or dead.

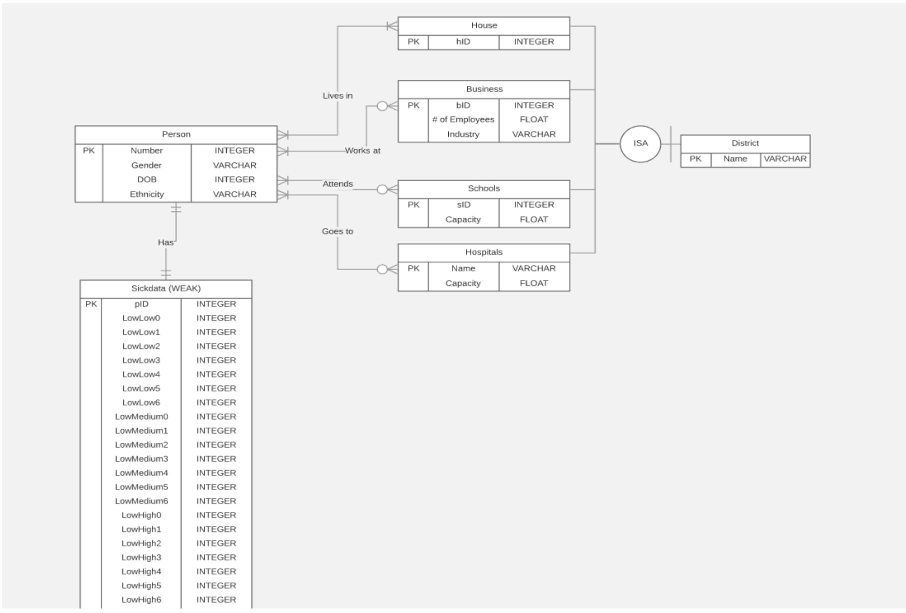

Figure 4: Entity Relationship Diagram

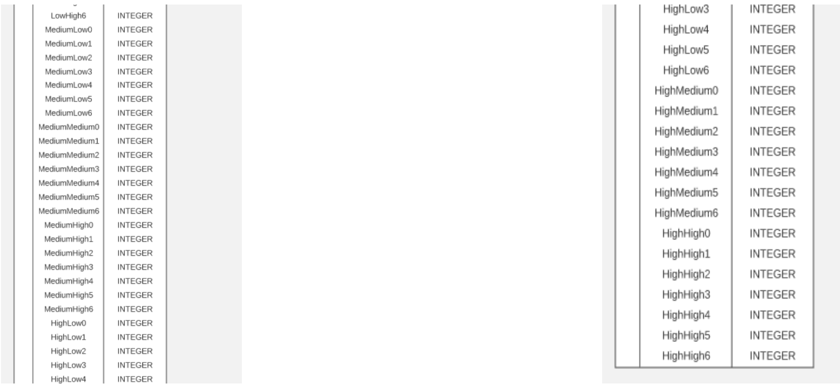

Figure 5: Entity Relationship Diagram Cont.

9.5 Appendix F: Website Images

Figure 6: Home Page

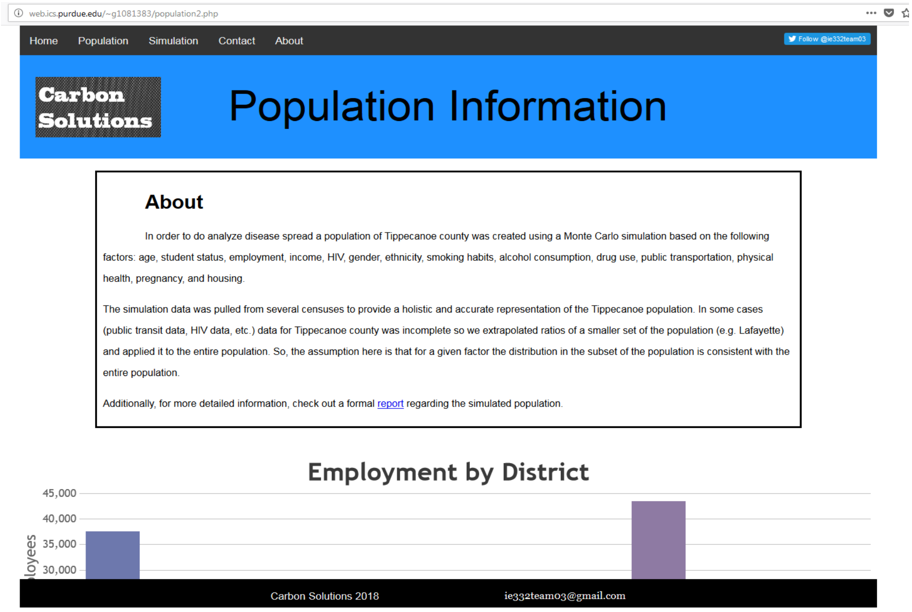

Figure 7: Population Page

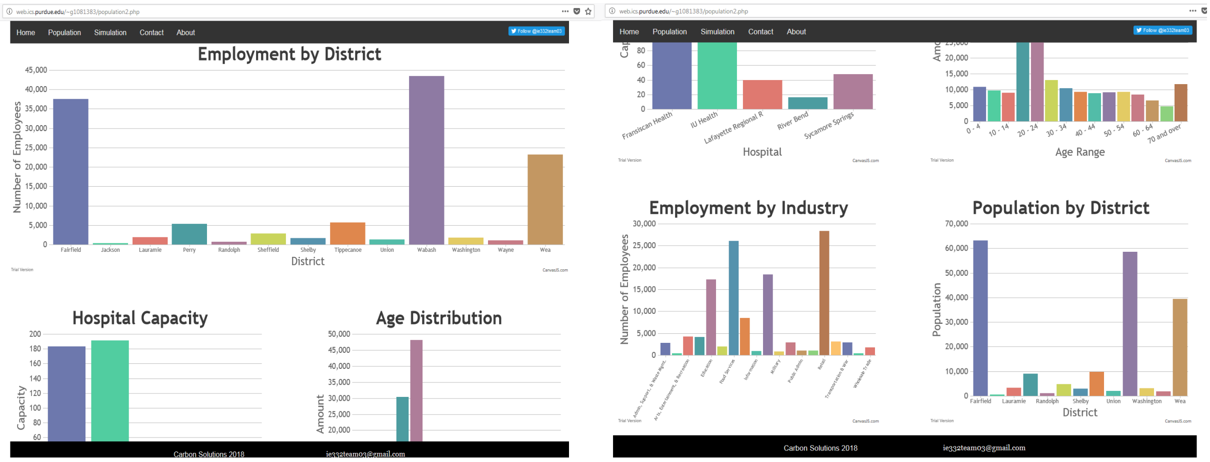

Figure 8: Population Cont.

Figure 9: Simulation Page

Figure 10: Simulation Output Cont.

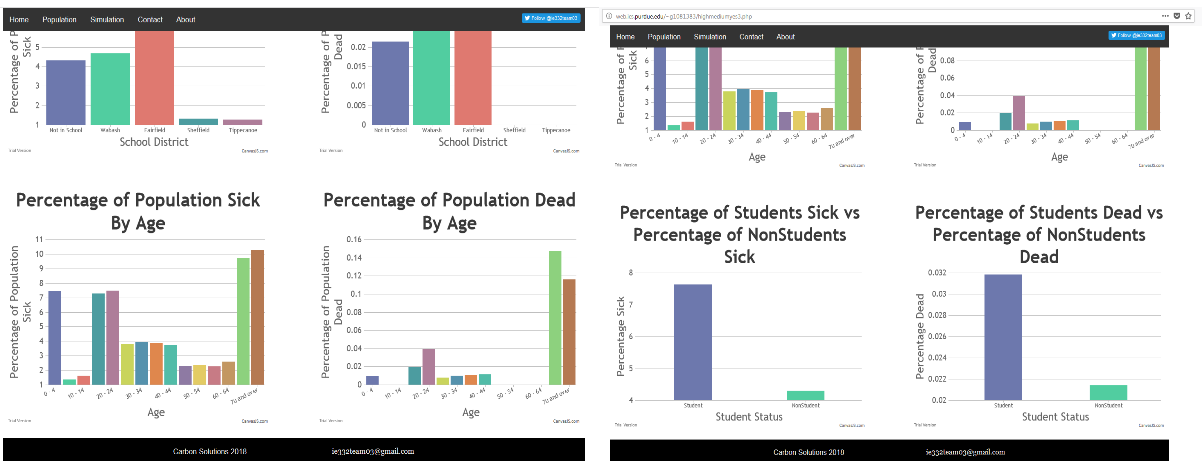

Figure 11: Simulation Output Cont.

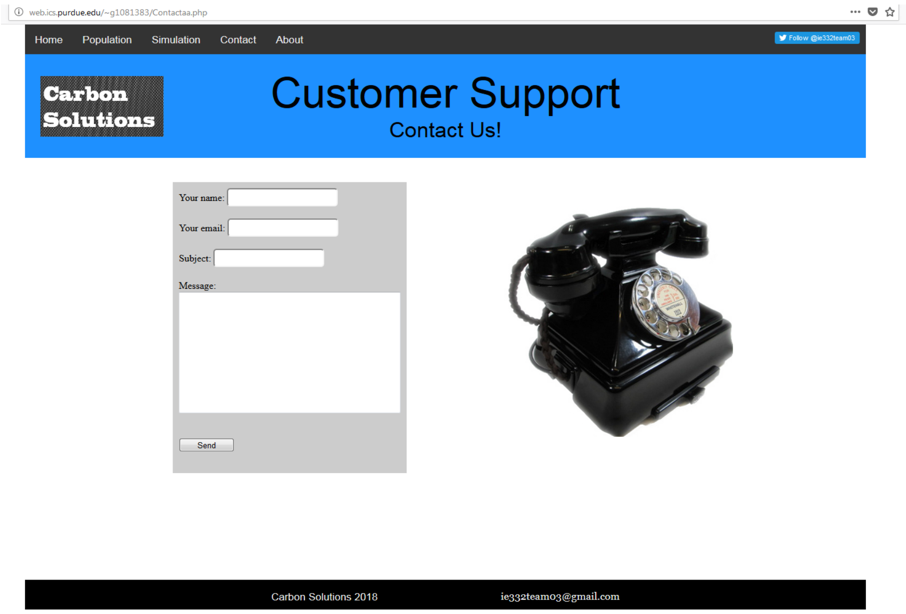

Figure 12: Contact Page

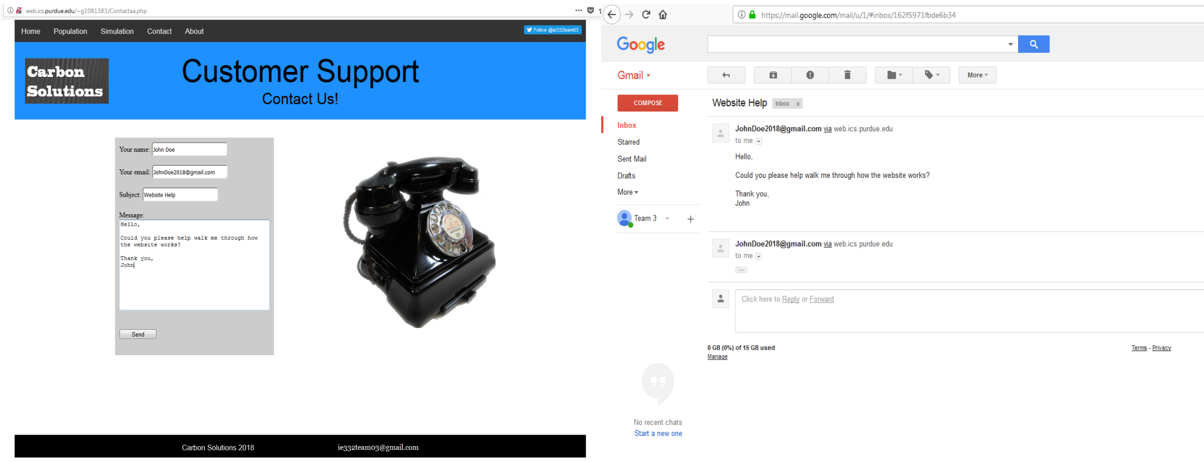

Figure 13: Contact Form Input / Output

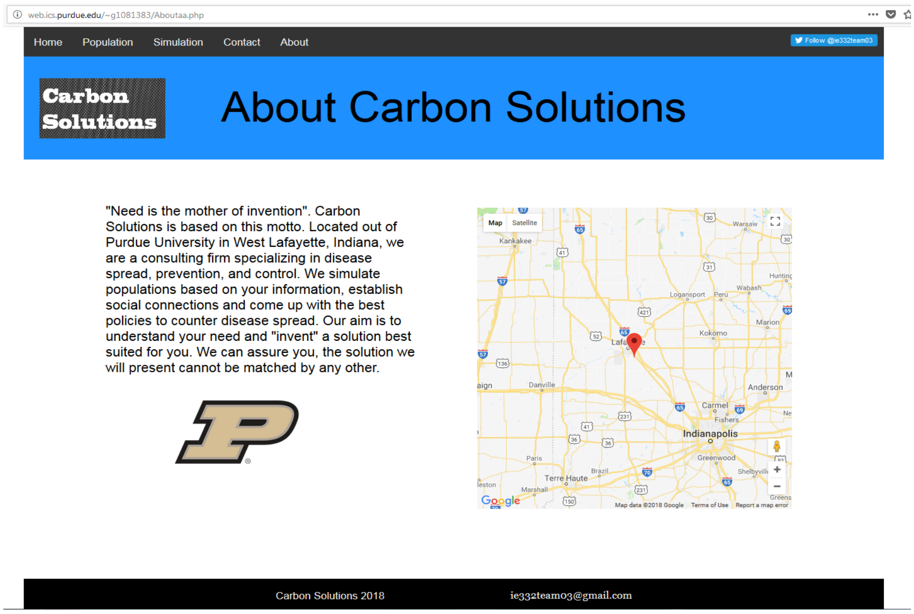

Figure 14: About Page

References

[1] Abdy Antique Telephones. ”Old Vintage Bakelite GPO Telephone Sale.” Abdy Antique Telephones, abdyantiquetelephones.co.uk/.

[2] ”Academic Logo Guidelines.” Academic Logo Guidelines - Brand Toolkit - Purdue University, www.purdue.edu/brand/elements/logos/academic.php.

[3] ”Beautiful PHP Charts & Graphs.” CanvasJS, Fenopix, canvasjs.com/php-charts/.

[4] “Behavioral Health Barometer Indiana, 2014.” SAMHSA, SAMHSA, www.samhsa.gov/data/sites/default/files/State_BHBarometers_2014_1/BHBarometer-IN.pdf.

[5] Bernard, Helen et al. "Nurses' Contacts and Potential for Infectious Disease Transmission." Emerging Infectious Diseases 15.9 (2009): 1438—1444. PMC. Web. 24 Apr. 2018.

[6] “Beyond Hangovers:” National Institute on Alcohol Abuse and Alcoholism, U.S. Department of Health and Human Services, pubs.niaaa.nih.gov/publications/hangovers/beyondhangovers.htm.

[7] Black, Jessica. “The State of Tobacco Control.” Tippecanoe County, www.in.gov/isdh/tpc/files/local_policy_brief_tippecanoe.pdf.

[8] ”Calculate Age in MySQL (InnoDb).” Sql - Calculate Age in MySQL (InnoDb) - Stack Overflow, stackoverflow.com/questions/5773405/calculate-age-in-mysql-innodb.

[9] “Chapter 10: Controlling the Spread of Infectious Diseases.” Advancing the Right to Health: the Vital Role of Law , 2017, pp. 151-164., www.who.int/healthsystems/topics/health-law/chapter10.pdf.

[10] “College Student Employment.” College Student Employment, nces.ed.gov/programs/coe/pdf/coe_ssa.pdf.

[11] “Current Tippecanoe County, Indiana Population, Demographics and Stats in 2017, 2018.” Suburban Stats, suburbanstats.org/population/indiana/how-many-people-live-in-tippecanoe-county.

[12] Davis, Jessica. “School Enrollment and Work Status: 2011.” U.S. Census Bureau, www.census.gov/prod/2013pubs/acsbr11-14.pdf.

[13] “Digest of Education Statistics, 2014.” National Center for Education Statistics (NCES), U.S. Department of Education, nces.ed.gov/programs/digest/d14/tables/dt14_503.10.asp.

[14] “Estimating the Number of Pregnant Women in a Geographic Area.” CDC Division of Reproductive Health, www.cdc.gov/reproductivehealth/emergency/pdfs/pregnacyestimatobrochure508.pdf.

[15] Fenopix, canvasjs.com/assets/script/canvasjs.min.js.

[16] HHS Office of the Secretary,Office of Budget, and OB. "FY2016 Budget in Brief - CDC." HHS.gov, US Department of Health and Human Services, 2 Feb. 2015, www.hhs.gov/about/budget/budget-in-brief/cdc/index.html.

[17] Holloway, Rachel, et al. “Updated Preparedness and Response Framework for Influenza Pandemics.” Centers for Disease Control and Prevention, CDC, 26 Sept. 2014, www.cdc.gov/flu/pandemic-resources/pdf/mmwr-rr6306.pdf.

[18] Holmes, Lindsay. "Science Says You Can Really Only Have 5 Close Friends At A Time." The Huffington Post, TheHuffingtonPost.com, 4 May 2016, www.huffingtonpost.com/entry/dunbar-layers-friendship-study_us_5728d4c5e4b016f37893ac14#.

[19] “Home.” Tippecanoe County, IN Community Colleges | CommunityCollegeReview.com, www.communitycollegereview.com/indiana/tippecanoe-county.

[20] “Hoosiers by the Numbers.” Unemployment Rates in Indiana MSAs: Infographics, 11 Oct. 2016, www.hoosierdata.in.gov/infographics/msa-unemployment.asp.

[21] HOUSE, T., and M. J. KEELING. "Household Structure and Infectious Disease Transmission." Epidemiology and Infection 137.5 (2009): 654—661. PMC. Web. 24 Apr. 2018.

[22] Hynes, Lori. “Fall Term Enrollment Trends.” Ivy Tech, www.ivytech.edu/files/Fall_2015_Enrollment_Trends.pdf.

[23] ”In SQL, How Can You Group by in Ranges?” Sql Server - In SQL, How Can You "Group by" in Ranges? - Stack Overflow, stackoverflow.com/questions/232387/in-sql-how-can-you-group-by-in-ranges.

[24] Jatlaoui, Tara C., et al. “Abortion Surveillance - United States, 2014.” Centers for Disease Control and Prevention, CDC, 24 Nov. 2017, www.cdc.gov/mmwr/volumes/66/ss/ss6624a1.htm?s_cid=ss6624a1_w.

[25] “Lafayette, IN.” Data USA, DataUSA, datausa.io/profile/geo/lafayette-in/.

[26] “Lafayette, IN Demographic Data.” Neighborhood Scout, Location Inc, www.neighborhoodscout.com/in/lafayette/demographics.

[27] MacDorman, Marian F., and Elizabeth C.W. Gregory. “Fetal and Perinatal Mortality: United States, 2013 .” National Vital Statistics Reports , CDC, 23 July 2015, www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_08.pdf.

[28] Martin, Joyce A., et al. “Births: Final Data for 2016.” National Vital Statistics Reports, CDC, 31 Jan. 2018, www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_01.pdf.

[29] McQuerrey, Lisa. "Communicable Diseases in the Workplace." Small Business - Chron.com, http://smallbusiness.chron.com/communicable-diseases-workplace-13091.html. Accessed 23 April 2018.

[30] ”MySQL Query to Group Data into Different Ranges.” Stack Overflow, stackoverflow.com/questions/30335203/mysql-query-to-group-data-into-different-ranges.

[31] “Pandemic Influenza.” Centers for Disease Control and Prevention, CDC, www.cdc.gov/flu/pandemic-resources/index.htm.

[32] Parker, Kim, et al. “6 Facts about the U.S. Military and Its Changing Demographics.” Pew Research Center, Pew Research Center, 13 Apr. 2017, www.pewresearch.org/fact-tank/2017/04/13/6-facts-about-the-u-s-military-and-its-changing-demographics/.

[33] “People at High Risk of Developing Flu-Related Complications.” Centers for Disease Control and Prevention, CDC, www.cdc.gov/flu/about/disease/high_risk.htm.

[34] Purdue Marketing and Media, and Purdue University. ”Purdue University.” About Purdue’s History, Traditions, Achievements & More - Purdue University, www.purdue.edu/purdue/about/.

[35] “QuickFacts.” U.S. Census Bureau, www.census.gov/quickfacts/fact/table/US/PST045216.

[36] “Real-Time Collaborative Writing and Publishing Tools with Integrated PDF Preview.” Overleaf, www.overleaf.com/.

[37] ”Redirect in PHP without Use of Header Method.” Stack Overflow, stackoverflow.com/questions/27123470/redirect-in-php-without-use-of-header-method.

[38] ”Run a SELECT Statement Based on Values in a Checkbox Array.” Php - Run a SELECT Statement Based on Values in a Checkbox Array - Stack Overflow, stackoverflow.com/questions/39394779/run-a-select-statement-based-on-values-in-a-checkbox-array?rq=1.

[39] “Schools and Child Care Centers.” Centers for Disease Control and Prevention, CDC, www.cdc.gov/flu/pandemic-resources/archived/schools-child-care-planning.html.

[40] "School Health." DPHHS Home, dphhs.mt.gov/schoolhealth/communicabledisease.

[41] ”Simple Question: How to Split Date (08-17-2011) into Month, Day, Year? PHP.” Variables - Simple Question: How to Split Date (08-17-2011) into Month, Day, Year? PHP - Stack Overflow, stackoverflow.com/questions/7103617/simple-question-how-to-split-date-08-17-2011-into-month-day-year-php.

[42] “Student Enrollment, Fall 2017.” Undergraduate Admissions - Purdue University, Purdue University, www.admissions.purdue.edu/academics/enrollment.php.

[43] "The Laws of Friendship." PBS, Public Broadcasting Service, 16 Aug. 2012, www.pbs.org/parents/education/going-to-school/social/laws-of-friendship/.

[44] “Tippecanoe County, IN.” Data USA, DataUSA, datausa.io/profile/geo/tippecanoe-county-in/.

[45] “Tippecanoe County, Indiana (IN).” Tippecanoe County, Indiana, Onboard Informatics, www.city-data.com/county/Tippecanoe_County-IN.html.

[46] Tippecanoe County, IN, www.tippecanoe.in.gov/Archive.aspx?AMID=40.

[47] “West Lafayette, IN.” Data USA, datausa.io/profile/geo/west-lafayette-in/.

[48] Wix, Wix.com, Inc, www.wix.com/.