Team 10: Going Viral
ESE Senior Design Spring Demo Day

Emily Chen (SSE '16), Allison Higgins (SSE '16), Lina Huang (SSE '16), Chloé le Comte (SSE '16)
Advisor: Dr. Pappas

Project Abstract

Governments and health organizations frequently develop budgets that allocate available resources to different health-related expenditures, yet it is often challenging for officials to see how the resources would best be spent. For our Senior Design project, our team has built a website policy tool that helps users solve this resource allocation problem to optimally distribute resources over a network of cities to best contain the spread of epidemics.

We have built off of previous research in the resource allocation space to implement a conic geometric program to control a susceptible-infected-susceptible viral spreading process taking place in a directed network of major U.S. cities. The model is built on disease spread data from Google Flu Trends along with corresponding contagion parameters that have been calculated based on the data. After testing and refining the model, we have developed a dynamic policy tool that allows users to graph and compare data trends, visually measure the correlation strengths between cities' data, and ultimately determine the cost-optimal allocation of resources through a network of cities.

Explanation of the Model

For our project, we have expanded on the research of Dr. Pappas, Dr. Preciado, Dr. Nowzari, and Dr. Han of the University of Pennsylvania's School of Engineering and Applied Science. To model the spread of disease, we use a networked discrete-time SIS model as proposed by Han et al.[1] We consider a directed network of nodes in which each node represents a major U.S. city. In this model, the β_{ij} values represent the relationship weights between cities in the network corresponding to the rates at which an infection can be transmitted through an edge in the network. The δ_i values represent the recovery rates for each city. The variable $p_i(t)$ represents the fraction of city i's population infected at time t. The proportion of the population infected at time step t+1 can be modeled by the following formula:

$$p_i(t+1) = (1-p_i(t)) * \left\{1 - \prod_{j \in N_i}[1-\beta_{ij}p_j(t)]\right\} + (1-\delta_i)p_i(t)$$

Based on the data from Google Flu Trends, we were able to calculate the approximate beta values in the model according to the upper-bounded matrix representation of the above formula:

$$\begin{bmatrix} p_1(t+1) \\ p_2(t+1) \\ \cdots \\ \cdots \\ p_n(t+1) \end{bmatrix} = \begin{bmatrix} \delta_1 & b_{12} & \cdots & \cdots & b_{1n} \\ b_{21} & \delta_2 & b_{23} & \cdots & \cdots \\ b_{31} & b_{31} & \delta_3 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & b_{n-1,n} \\ b_{n1} & \cdots & \cdots & b_{n,n-1} & \delta_n \end{bmatrix} \begin{bmatrix} p_1(t) \\ p_2(t) \\ \cdots \\ \cdots \\ p_n(t) \end{bmatrix}$$

The natural recovery rates of each city, δ_i, were subsequently estimated by solving for the rates for each city that would minimize the sum of the absolute value of the difference between the predicted and actual flu data predictions in our resource allocation model under a scenario of zero dollar budget allocation.
The following geometric programming optimization function solves the resource allocation problem:

$$\min_{d^c} \quad \rho(M(B_G, d^c))$$

$$\text{subject to} \quad \sum_{i \in V_C} g_i(\delta_i^c) \le C, \qquad (1)$$

$$\underline{\delta}_i^c \le \delta_i^c \le \overline{\delta}_i^c \qquad \forall i \in V_c \qquad (2)$$

The model minimizes the spread of the flu for a given budget (constraint (1)) and a specified range of possible recovery rates, which range from the natural recovery rate to the largest feasible rate, $\overline{\delta}_i^c$ (constraint (2)). The variable g_i represents a cost function that outputs the optimal budget allocation by city.

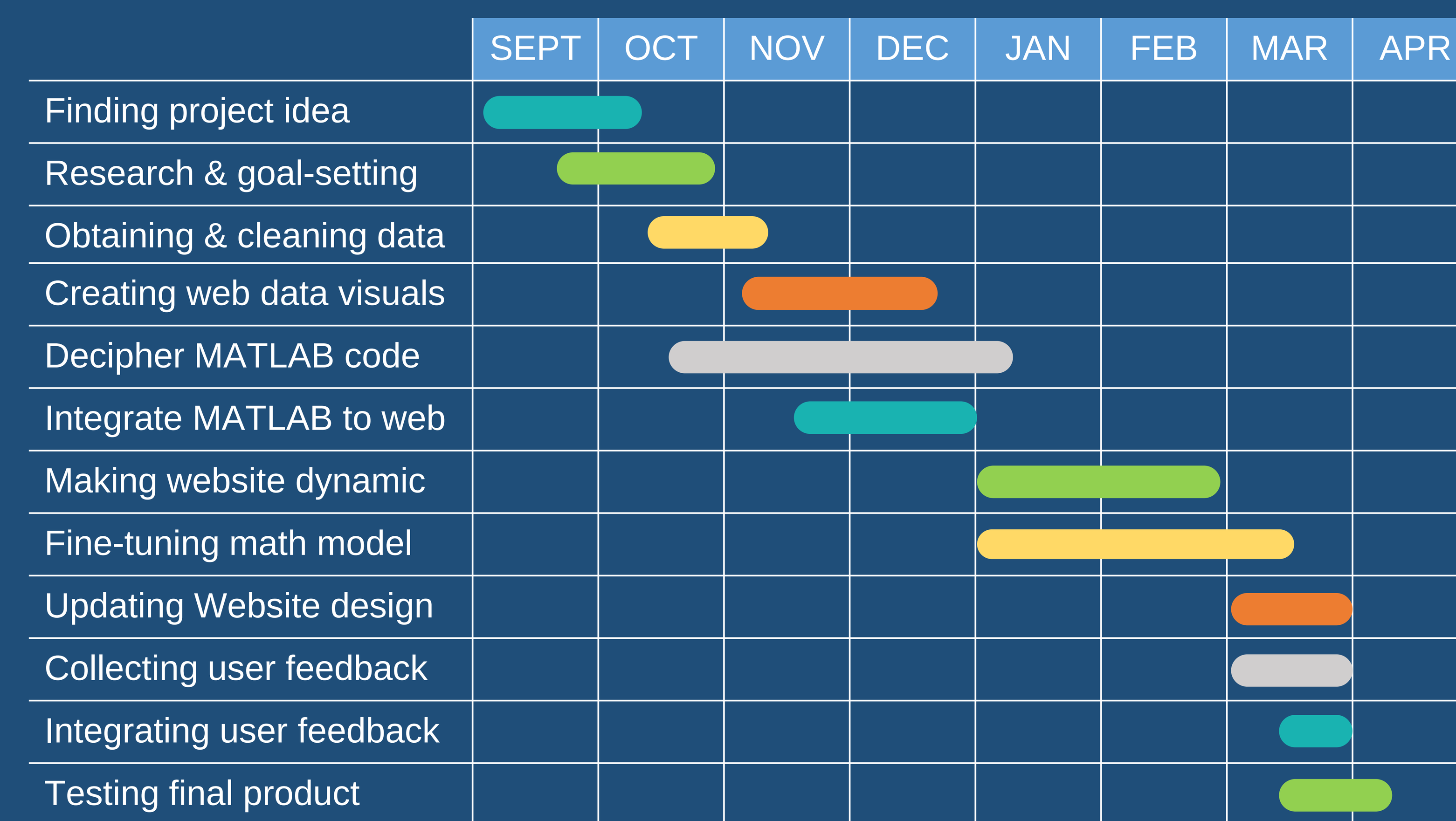

Figure 1: Project Timeline from September until April

Results & User Testing

Since the SIS model is an approximation of reality, validating results involves the prediction capability of the formulas used. With a budget of $0, the graphs of people infected with and without allocation should be about the same. This can be seen in Figure 2.
Second, we validate the results of the resource allocations given generalized statistics about the flu taken from the CDC website. First, 5-20% of the US population gets the flu in a given flu season, giving a range of about 16 to 64 million infected individuals each year.[2] These 16 to 64 million people lead to direct costs from hospitalization of $10.4 billion,[3] meaning each person infected costs anywhere from $162 per person to $650 per person. A reasonable cost per person saved should therefore not exceed $162 to provide net economic benefit. The cost of a vaccine dosage for the CDC is $16,[4] but since preventing one infection can save multiple people from ultimately being infected, there is no distinct lower bound to the cost per person saved. The suggested budgets resulting on the website reflect these realistic limits, even showing the phenomena of diminishing marginal benefit of each dollar invested, as seen in Figure 3.
Given that the goal of our project is to be a public policy planning tool, we sought out public health professionals at Penn to get feedback on our website. Because of these meetings, we added helpful features such as showing data about age distributions of cities, providing the user with suggested networks in the Heat Map tab, and presenting resource allocation results in a tabular form with additional statistics. They also identified non-intuitive features of our website so that the current version is much more visually appealing and user-friendly.

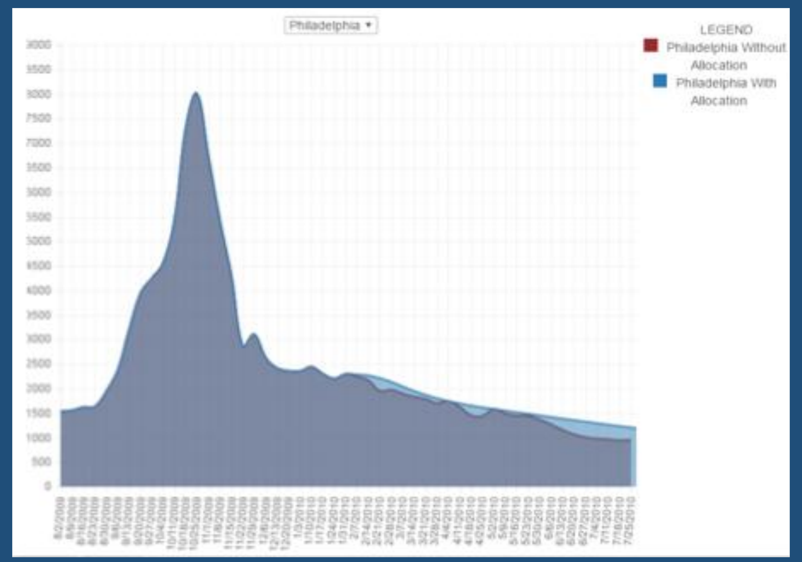

Figure 2: Comparison of Actual vs. Predicted

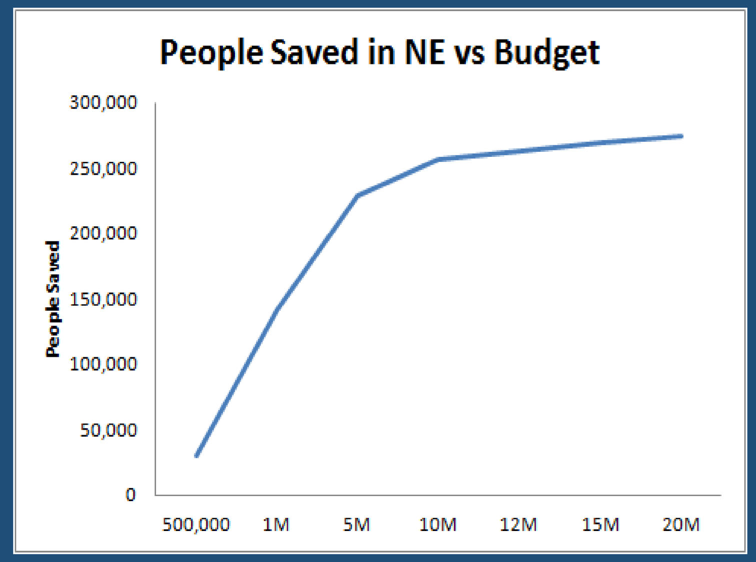

Figure 3: People Saved vs. Budget

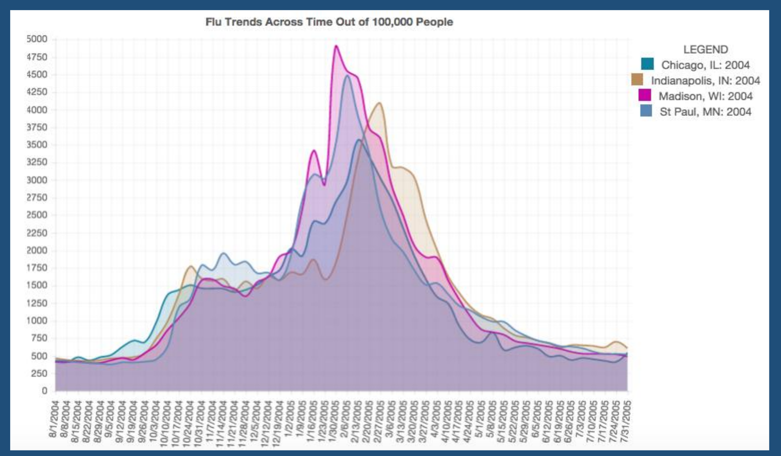

Figure 4: Infected Populations Across Cities

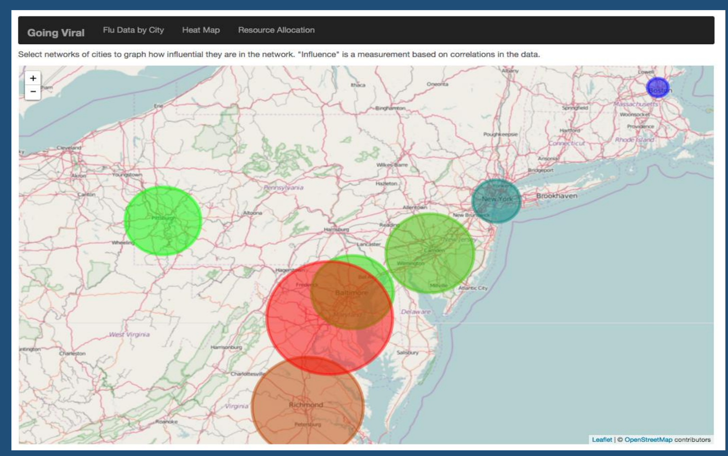

Figure 5: Heat Map of Northeast Cities

Website

Our website visualizes the data from Google Flu trends in three different ways. The first tab of our website is Flu Data by City, which graphs the proportion of a population infected by the flu out of 100,000 in selected cities during the selected flu season. Users can graph multiple cities or multiple flu seasons on the same graph to compare the trends in outbreaks. Demographic information on age and gender breakdowns for the cities is also available at the top of the page for the user to better interpret the data. This can be seen in Figure 4.
The next tab is Heat Map, which allows users to visualize the correlations between the flu data of cities in a selected network over a given period of time. The relative size of the circles over each city and their corresponding influence value increase as the city's data becomes more strongly correlated with that of other cities in the selected network during the specified time period. Correlation is affected by each city's infection rate, recovery rate, and proportion of population infected in a given time step, and the correlations are calculated according to the following formula:

$$\frac{1}{n}\sum_{j=1}^{n} \beta_{ij} p_j(t) \le 1 - \left(1 - \frac{p_i(t+1) - p_i(t)(1-\delta_i^0)}{1-p_i(t)}\right)^{1/n}$$

Finally, the main portion of our website is the resource allocation tool. Users can input their budget into the site, select a region and a flu season, and see the results of the resource allocation, as seen in Figure 6. The table displays the optimal budget allocations to each city in the network, the natural and improved recovery rates for each city, the number of people saved, and the average cost per life saved. Additionally, users can graph a particular city to compare the infected populations if the resource allocation were applied compared to allocating no additional resources, which is seen in Figure 7. The red graph depicts the number of people infected without the resource allocation, and the blue graph shows the results of the allocation.

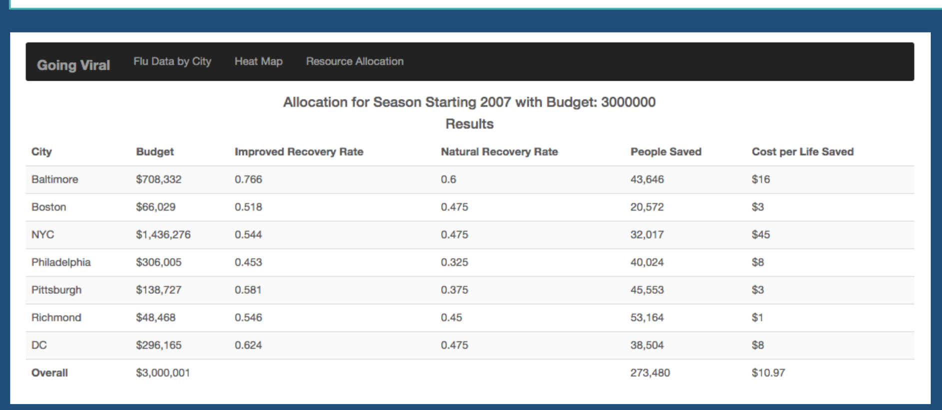

Figure 6: Website Results Table

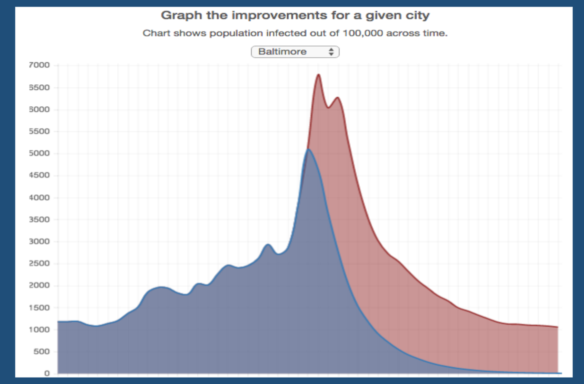

Figure 7: Graph of Allocation Results

Future Steps

While we made a lot of progress on our project this year, there are still other features that we have not included due to data and time constraints. Some additional steps that could be taken in the future if the project were to continue include:

- Working with the CDC to obtain more accurate flu infection data (as opposed to using Google Flu Trends)
- Providing more historical information such as vaccination rates to provide context for observed data
- Obtaining more accurate estimates of historical flu budgets for each city
- Investigating how transit routes affect the epidemic spreading
- Expanding the model to analyze resource distribution for other diseases or regions

Contact

Emily Chen: chenem@seas.upenn.edu
Chloé le Comte: chloeml@seas.upenn.edu
Allison Higgins: ahigg@seas.upenn.edu
Lina Huang: hlina@seas.upenn.edu
George Pappas: pappasg@seas.upenn.edu

Methods and Materials

- Bootstrap CSS templates
- JavaScript libraries: Chart.js, Leaflet maps
- Python Flask framework
- Python MATLAB Engine
- MATLAB CVX library
- Ubuntu server through Digital Ocean
- Domain bought from Namecheap.com

References

1. "Data Driven Allocation of Vaccines for Controlling Epidemic Outbreaks" by Shuo Han, Victor M. Preciado, Cameron Nowzari, and George J. Pappas
2. Bland, William, MD. "Flu Statistics: What Are Your Odds of Getting the Flu?"*WebMD*. WebMD, n.d. Web. 04 Apr. 2016.
3. "Make Your Trading a Business." *Myths, Realities, and Strategies That Really Work No-Hype Options Trading* (2015): 157-67. *CDC*. Web.
4. "CDC Vaccine Price List." *CDC*. N.p., 1 Apr. 2016. Web.

Website

http://seniordesign-goingviral.com/