
Prasad S. Kanuparthi¹ and Asish T. Balu²

University of Pittsburgh School of Arts and Sciences, Pittsburgh, PA 15213¹
Pennsylvania State University College of Engineering, University Park, PA 16802²

Introduction

Acquired Immunodeficiency Syndrome (AIDS) exists as a disease of the immune system contracted through prior infection with human immunodeficiency virus (HIV), a retrovirus. HIV compromises the human's immune system from effectively combating opportunistic infections and tumorigenic activity. This is accomplished through the infection and destruction of leukocytes, such as helper T cells, macrophages, and dendritic cells, which normally regulate the functions of the immune system [1]. Through inhibition of this cell-mediated immunity, the body loses its capacity to combat otherwise easily preventable infections and diseases.

During the late 1970s, clinicians witnessed an outbreak in the patient populace of a variety of rare disorders normally witnessed in immunocompromised patients. During the early 1980s, through clinical research, HIV was isolated and its retroviral nature more clearly understood [2]. HIV's mode of person-to-person transmission occurs primarily via sexual contact, transfusion of infected bodily fluids, and the sharing of contaminated intravenous instruments [3]. As such, especially when considering the ease of travel and the existence of densely populated metropolitan areas, HIV presents both a global and domestic pandemic. Since the 1970s, the presence and extent of AIDS within the domestic population has grown dramatically.

Through the use of modeling software, visualization tools, and sorting algorithms, we have effectively mapped the migration and epidemiology of AIDS through the United States over the past twenty-two years.

Materials and Methods

Numerous patient datasets, containing a complete geographical and demographic record of the prevalence of Acquired Immunodeficiency Syndrome (AIDS) in the United States from the time period 1984-2005, were provided courtesy of the University of Pittsburgh Models for Infectious Disease Agent Study (MIDAS) National Center of Excellence. Census data for the years 1990 and 2000 for the United States were provided courtesy of the United States Census Bureau. Demarcation and assortment of the dataset was accomplished through sorting algorithms coded in the Java programming language through use of the integrated development environment (IDE) and Java development kit (JDK) software Eclipse™. StatPlanet Plus™, a proprietary modeling software of StatSilk, was utilized for graphical analysis and visualization of the dataset.

To ensure the proper documentation of our sorting algorithms to highlight our precise methodologies, the following will describe our various approaches. Initially the dataset was full of unsorted data arranged in a means not interpretable by the modeling software. As such, several algorithms were applied to ensure homogeneity within our dataset (Figure 1).
Figure 1 - An Algorithmic Sorting of Raw Data

(A) StatSilk provided the sample table seen here as a foundation upon which to reformat our raw datasets for compatibility with StatPlanet Plus™. The countries are listed vertically, and the identifiers (the category/dependent
variable) along with the year are listed horizontally. (B) The figure represents the raw dataset without any algorithmic modifications applied. The challenge was to rearrange this primary data into the arrangement seen in (A). The third column held the value for both year and week. Through the use of simple Java algorithms, the third column was truncated to display only the year of that particular case; *StatPlanet Plus™* does not support week-by-week analysis. In addition to cutting out the week, columns two, three, five, and six were removed because they contained superfluous information. This left the data in the format shown in (C). To go to (D) from (C) required isolation of the cases in each state each year, and a summation of the weekly cases for each year for each state. A more complicated Java algorithm handled these calculations and sorting specifications. However, for *StatPlanet Plus™* to recognize the data and perform the appropriate modeling, the data was further restructured, through an inversion of the chart format (columns transformed to rows, for example), to obtain the final dataset seen in (D). This emulates the data structure seen in (A).

**Results**

Initial modeling attempts revolved around producing a figure illustrating the raw number of AIDS cases per state per year. Our team sought to initially define and visualize simply the number of cases that occurred in each state over the course of the dataset's timeframe (1984-2005). In order to do so, our team established several guidelines and assumptions to parameterize these results. These assumptions involved the exclusion of the United States territories, due to the inability of finding a map supporting these images, in favor of the sole inclusion of the continental United States. Also, our team assumed that throughout the timeframe there were no deaths in patients who had contracted AIDS in previous years, that all cases reported each year were those of new patients, and that there was no immigration/emigration from state to state. While initially debating on the most effective display for use, the final decision rested that the images would include the sum total of cases that occurred till that point in time for that state (*Figure 2*).

*Figure 2 - A Cumulative Display of AIDS Prevalence in the Continental United States*
Figure 2 - (I) Here, the figure represents the sum total number of AIDS cases for each state in the initial year of data collection, 1984. (II) Here, the sum total numbers of AIDS cases that have occurred in each state till the year 1990 are displayed. This required the summation of all the cases in each state for those years between 1984-1990. (III) In this figure, through the same protocol as mentioned in (II), the sum total numbers of AIDS cases in each state were totaled for all the years from 1984-1995. As such, this figure shows how many AIDS cases existed in each state in the year 1995. (IV) This figure, through the same procedure as in (II) and (III), shows the number of AIDS cases present in each state in the year 2000. (V) Illustrates the same information for the year 2005, the last year of data collection.

Although the above data provides an insight into the magnitude of AIDS epidemiology, the figure in its totality only represents growth at various arbitrary time points evenly spaced over the dataset's timeframe. In order to truly display the growth in the number of AIDS cases amongst the patient population, the next logical step involved the construction of various line graphs over the whole timeframe interval. In addition to simply seeing the trends involved in the number of AIDS cases in the United States, our team wanted to provide a greater level of detail and sought to isolate the various different regions of the United States to view how AIDS growth and development varies in each sector of the continental United States (Figure 3).

Figure 3 - A Depiction of AIDS Growth in Four Defined Regions of the Continental United States

Figure 3 - (I) Here, the Northeast Region of the continental United States is displayed. The states included in this region include New York, New Jersey, Pennsylvania, Massachusetts, Connecticut, Rhode Island, New Hampshire, Maine, and Vermont. The state with the most rapid growth and largest number of cases was New York (the data for New York City and Upstate New York were combined). (II) In this figure, the Midwest Region of the continental United States is displayed. The states included in this region include Illinois, Michigan, Ohio, Missouri, Indiana, Kansas, Wisconsin, Minnesota, Iowa, Nebraska, and both North and South Dakota. The state with the most rapid growth and largest number of cases was Illinois. This process was repeated for the states in the Southern United States (in III) and in the Western United States (in IV). The states with the largest growth and numbers were Florida and California, respectively.
Figure 4 - Identifying States Possessing the Greatest Incidence of AIDS Cases with Respect to Population
Upon inspection of the line graphs, our team initially saw that in each of the four regions (Northeast, South, Midwest, and West), the states with the largest growth and greatest number of AIDS cases within its population were New York, Florida, Illinois, and California, respectively. However, the question arose from this new information: do these states, which possess large populations and several urban centers within their boundaries inherently lead to a greater prevalence of AIDS? In short, does the data seen in Figure 3 depict the states with the largest proportion of individuals with AIDS with respect to the population of the whole state, or does the figure simply show those states with the largest populations (which may inherently lead to a greater incidence of AIDS cases)? To further investigate this idea, our team isolated the number of cases for all the states in the years 1990 and 2000 - the two years within the timeframe of the dataset when the national census was taken. Afterwards, the number of AIDS cases in each state was divided by the population of that state in those two years, illustrating the proportion of individuals with AIDS with respect to the whole population (Figure 4).

Upon inspection of the figure, some of the results of Figure 3 were supported, while others were refuted. For example, in the Northeast Region, Figure 3 displayed New York as possessing the largest number of cases and the most rapid growth. However, as seen in Figure 4, New Jersey possessed the greatest number of AIDS cases with respect to population size. In the other regions (South, Midwest, and West), the states with the greatest incidence of AIDS cases with respect to population were Florida, Illinois, and California, as also shown and supported by Figure 3. However, novel information revealed in Figure 4 shows that states such as Maryland (for the South region), and Missouri (in the Midwest region) also have a high incidence of AIDS with respect to their populations. On a more generalized scale, the state-by-state proportion of individuals possessing AIDS all dramatically increased (for every state in every region) when comparing the 2000 levels with those of the year 1990.

Discussion

On a global level, AIDS exists as a threat to modern society. Over the past twenty or so years, as indicated in the figures above, the prevalence and existence of AIDS within the continental United States has markedly increased. Through modeling software and algorithmic decomposition, our team has modeled the epidemiological expansion of the disease within the nation. States with barely any cases at the beginning of the data collection possessed an immensely larger incidence of patient cases by the end of the collection timeframe.

As seen from the results above, the states with the largest number of AIDS patients and the fastest rise of new AIDS patients were New York, Illinois, Florida, and California. These states, as of the 2010 Census provided by the United States Census Bureau, constitute four of the top five most populous states. As such these states are potential grounds for a rapid increase of AIDS prevalence should no further action be undertaken. Also, as seen from Figure 4, other states, such as New Jersey, Maryland, and Missouri, possess a large proportion of patients with AIDS with respect to their population. This suggests that although large urban centers found in those states with the most cases of AIDS might be the primary target for public health education, these other states should also be considered as potential targets.

For future work, our team has considered that should the death rates also be included of those patients who possessed AIDS, a more accurate model could be constructed
representing how the current AIDS patient populace exists and its dispersion throughout the nation.

Bibliography


MIDAS Data Palooza Competition 2013