

Visual Analytics in Effects of Gross Domestic Product to Human Immunodeficiency Virus Using Tableau

Eunbi Kim and Ching-Yu Huang

Abstract—As data becomes more accessible, visualization methods are needed to help make it easier to understand the information. Analyzing and visualizing data makes it easier to understand a dataset without having to read through it, and elucidate connections between two or more different datasets. Tableau is one of the most popular interactive data visualization software. By using Tableau, it is easy to find correlations between datasets, reorganize datasets through pivoting or joining them, and create visualizations such as geo map charts, geo bubble charts, table charts, line charts, pie charts, and treemap charts. This project aims to show the correlation between a country's gross domestic product (GDP) and human immunodeficiency virus (HIV) through Tableau. Large data sets related to the GDP and HIV were gathered from open data sources. The data will be cleaned through Tableau and Excel, and correlations between datasets will be shown through variable charts with Tableau.

Index Terms—GDP, HIV, Tableau, dataset, visualization, correlation.

I. INTRODUCTION

A lack of socioeconomic resources can be linked to riskier health behaviors, which can lead to the contraction of human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS). These behaviors include substance use, which can reduce the likelihood of using condoms. Additionally, living in poverty can result in food insufficiency, which can contribute to HIV infection. According to research published in the national center for biotechnology information, “food insecurity, defined as persistent lack of access to adequate food in needed quantity and quality, undernutrition, and HIV/AIDS are overlapping and have additive effects. Over 800 million people worldwide are chronically undernourished and over 33 million are living with HIV infection” [1].

HIV is an epidemic in certain poor urban areas across the United States. The Joint United Nations Programme on HIV and AIDS (UNAIDS) identified general populations with more than 1 percent prevalence of HIV [2]. Within the same neighborhood in those cities, the people below the poverty line have a two times higher chance of being infected than people with incomes above the poverty line. This is not new information, since most health officials have long believed that poverty is a key driver of HIV. However, there were not many large projects to support this belief with factual data.

There may have been suspicion in the past if infection rates were correlated with racial and genetic makeup. However, this belief has been disproven as studies have shown that there were no statistically significant differences between infection rates by race or ethnicity [3]. Thus, people of color are affected by the aforementioned diseases because they are more likely to be impoverished, and not because of their race or ethnicity. According to the Black AIDS Institute, “when other racial ethnic groups face the same social determinants of health as Blacks the social and economic conditions within which they live and that impact their well-being. Their HIV rates rise to similar levels as those of Blacks, even for Whites, whose rate of infection is normally substantially lower than rates for both Blacks and Latinos.” [4] Recently published in journal Lancet, psychiatrist Wayne Fenton analyzed factors causing HIV and AIDS and concluded that reducing poverty may be the only viable long-term response to the epidemic.

This research has been augmented by recent studies based on statistical correlations of epidemiological and socioeconomic data. Studies have correlated the prevalence of the HIV infection directly with wealth. According to Dr. Chin who helped analyze data from Kenya, the “national HIV prevalence rates appeared to correlate directly with national income across sub-Saharan Africa” [5]. More recently, Mishra et al. analyzed the HIV infection prevalence in relation to a person's socio-economic status with national surveyed data for eight African countries. They concluded that there was an association between socio-economic status and the prevalence of the infection [6].

An impoverished community can face many challenges when living with HIV. Unstable housing, food deficiency, and a lack of consistent access to quality health care can make it difficult to cope with everyday life [7]. The cost of life management antiretroviral drugs can range from \$10,000 to \$15,000 per patient per year [8]. Therefore, the support that an HIV patient receives is dependent on the economy of their particular country; a poor country means poor remedy.

II. METHOD

Data was retrieved from three online open data sources: World Health Organization, The World Bank Open Data, Our World in Data, and World Population Review. The datasets for GDP were retrieved from the World Bank [9], the datasets for HIV and AIDs were retrieved from Our World in Data [10], and the datasets for the population of the world by country was retrieved from World Population Review [11]. Each dataset was either in csv or Excel format.

Manuscript received February 7, 2020; revised October 17, 2020.

The authors are with the School of Computer Science & Technology, Kean University, Union, NJ 07083, USA (e-mail: kimeunb@kean.edu, chuang@kean.edu).

A. Cleaning the Datasets

After all the datasets were gathered, the datasets were opened with either Excel or Tableau to check if the datasets were easily understandable. For the most part, the csv files or Excel files were easily interpretable; they may have included a short explanation, title, header, or a null value for data that was not collected. It is worth noting that some of the data points required cleaning; for instance, the terms “US” and “United States” referred to the same country, or “South Korea” and “Republic of Korea”. Additionally, Tableau sometimes cannot recognize if a dataset has header or not, or if the data is related to each column.

| API_NY_GDP_MKTP_CD_DS2_en... |
|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| Data Source | World Development... | F3 | F4 | F5 | F6 | F7 |
| Country Name | Country Code | Indicator Name | Indicator Code | 1,960.00 | 1,961.00 | 1,962.00 |
| Aruba | ABW | GDP (current LCU) | NY.GDP.MKTP.CN | null | null | null |
| Afghanistan | AFG | GDP (current LCU) | NY.GDP.MKTP.CN | 24,200,001,500.00 | 24,700,000,300.00 | 24,600,000,500.00 |
| Angola | AGO | GDP (current LCU) | NY.GDP.MKTP.CN | null | null | null |
| Albania | ALB | GDP (current LCU) | NY.GDP.MKTP.CN | null | null | null |
| Andorra | AND | GDP (current LCU) | NY.GDP.MKTP.CN | null | null | null |
| Arab World | ARB | GDP (current LCU) | NY.GDP.MKTP.CN | null | null | null |
| United Arab Emirates | ARE | GDP (current LCU) | NY.GDP.MKTP.CN | null | null | null |

Fig. 1. Data before processing table.

Country	Current	2002	2003	2004	2005
Aruba	GDP (current US\$)	1,941,340,782.00	2,021,229,050.00	2,228,491,820.00	2,330,726,257.00
Afghanistan	GDP (current US\$)	4,055,176,933.00	4,515,563,414.00	5,226,775,163.00	6,209,140,044.00
Angola	GDP (current US\$)	15,285,594,828.00	17,812,705,294.00	23,552,052,408.00	36,970,918,699.00
Albania	GDP (current US\$)	4,348,070,165.00	5,611,492,283.00	7,184,681,399.00	8,052,075,642.00
Andorra	GDP (current US\$)	1,733,116,883.00	2,398,645,598.00	2,935,659,300.00	3,255,789,081.00
Arab World	GDP (current US\$)	729,000,000,000.00	823,000,000,000.00	964,000,000,000.00	1,190,000,000,000.00
United Arab Emirates	GDP (current US\$)	110,000,000,000.00	124,000,000,000.00	148,000,000,000.00	181,000,000,000.00
Argentina	GDP (current US\$)	97,724,004,252.00	128,000,000,000.00	165,000,000,000.00	189,000,000,000.00
Armenia	GDP (current US\$)	2,376,335,048.00	2,807,061,009.00	3,576,615,240.00	4,900,469,950.00
American Samoa	GDP (current US\$)	514,000,000.00	527,000,000.00	512,000,000.00	503,000,000.00
Antigua and Barbuda	GDP (current US\$)	814,615,333.30	855,643,111.10	919,577,148.10	1,022,191,296.00

Fig. 2. Final structured table.

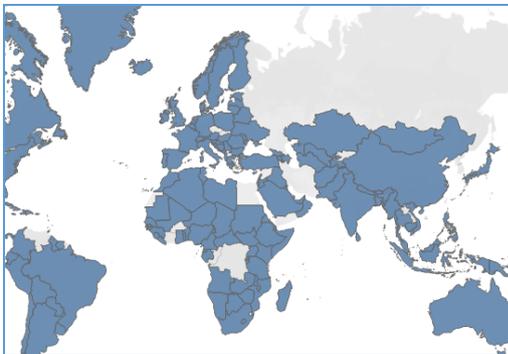


Fig. 3. Tableau country data before fixing country names.



Fig. 4. Tableau country data after fixing country names.

Therefore, the datasets must be cleaned, and the country’s names must be standardized before one can insert it into Tableau. Additionally, if a country’s name has been changed before, Tableau will skip the country when it tries join the

data with other datasets. Fig. 1 shows the data before it was cleaned. Since Tableau cannot recognize if it is a header or not, there are many gaps and formatting errors. Fig. 2 shows the result after cleaning up the data. Fig. 3 shows a map of the data before it was processed, including missing countries for those that have undergone name changes. Fig. 4 is a world map after standardizing all country names.

B. Joining Datasets

After cleaning up all the datasets, the next step was to join the tables to find a correlation between the datasets. An inner join is one such type of join that returns all rows from both participating tables where the key record of one table is equal to the key record of another. This join requires a comparison operator to match rows from the participating tables based on a common field or column in both tables. Fig. 5 shows inner joining the three tables by country and year.

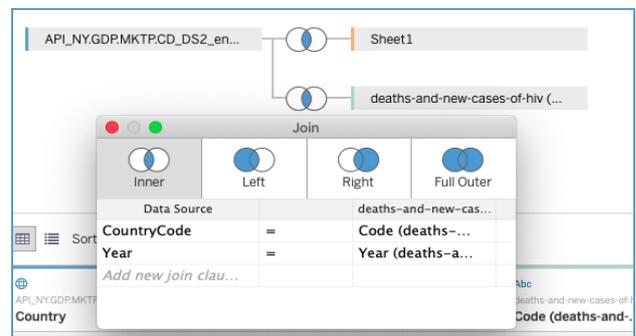


Fig. 5. Two Tableau data inner joining for country code and years.

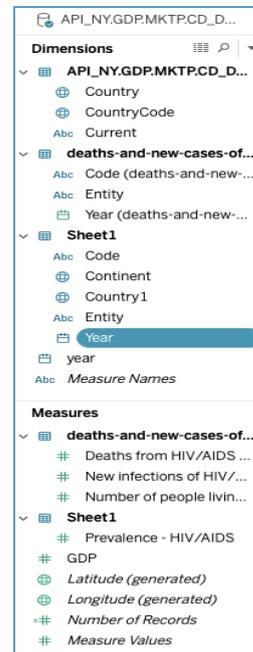


Fig. 6. Selectable data from the list.



Fig. 7. Kinds of data visualization.

C. Choosing Data for Visualization

The Tableau software provides data visualization for multiple datasets. After joining the tables, one can use the software to view the data. On the left side of Tableau, there are dimensions and measures to choose which data the user wants to see, as shown in Fig. 6. If the user clicks on some data, Tableau automatically shows which charts are available for the selected section on the right hand side,

shown in Fig. 7. To use geographical data, it must to include country or specific latitude or longitude.

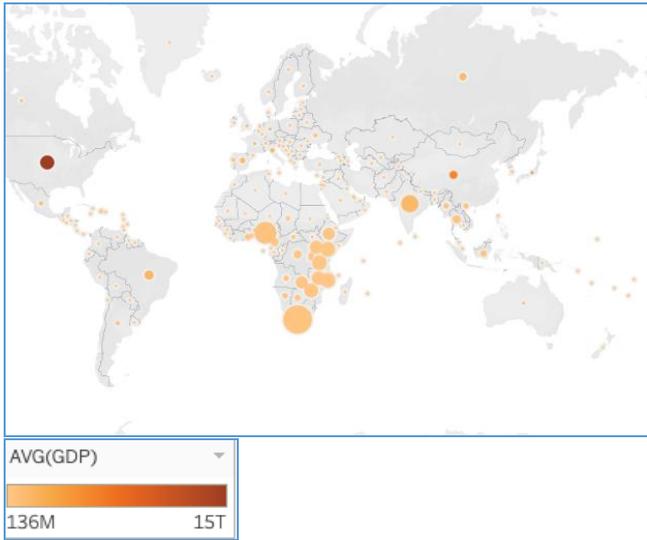


Fig. 8. Map visualization of GDP and HIV/AIDS; color is GDP and circle size is HIV/A.

III. EXPERIMENT RESULTS

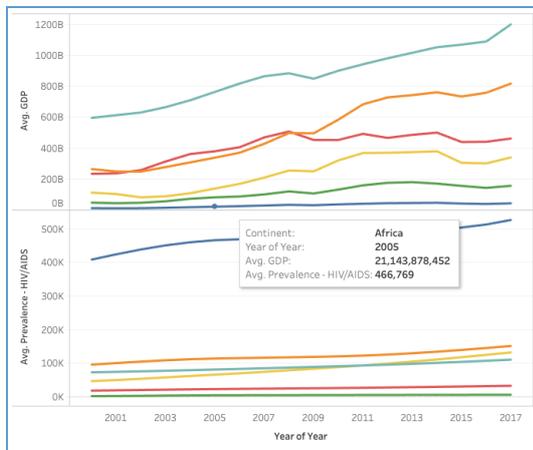
The final data set consists of 2987 rows. Each country contains data from 2000 to 2017, and the countries span 6 continents. The HIV/AIDS datasets include deaths from HIV/AIDS, new infection of HIV/AIDS, and number of

people living with HIV. To see the results for HIV/AIDS and GDP in the map chart, one must select the country from the dimensions and select the number of people living with HIV/AIDS and GDP. The circle size represents the prevalence of HIV/AIDS and the color represents the GDP of the country. A bigger circle size represents a higher number of HIV/AIDS, and a darker color represents a higher GDP (Fig. 8).

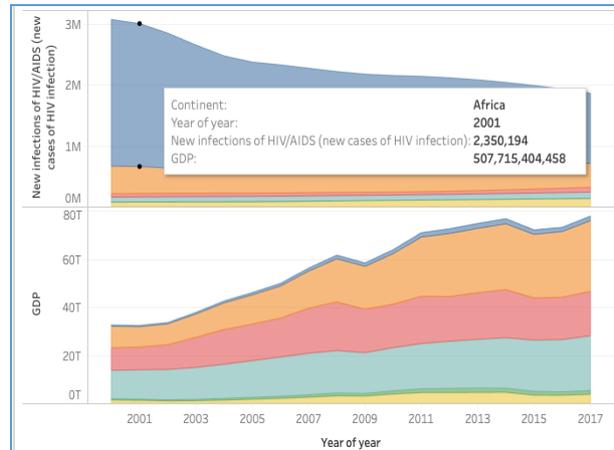
A. Visualization in GDP and HIV/AIDS

Tableau also supports line charts. A line chart is commonly used to display change over time as a series of data points connected by a straight line over two axes. The line chart therefore helps to determine the relationship between two sets of values, with one data set being dependent on the other set. For this dataset, the independent variable is the range of years from 2000 to 2017, while the dependent variable is the GDP of a continent. Fig. 9 compares GDP with HIV and AIDS and as the graph represents, Africa has the lowest GDP but the highest number of cases of HIV and AIDS.

Tableau allows a user to filter for a specific continent or country. Fig. 10 is filtered for Africa. As GDP increases, the rate of new infections drops. However, some other countries do not exhibit a similar effect on infection rate from a similar increase in their GDP. For example, Asia's GDP increases, but the infectious rate does not drop as much as the increase in GDP. Also, the correlation is not perfectly negative.



(a)



(b)

Fig. 9. GDP and HIV/AIDS year of year line chart.

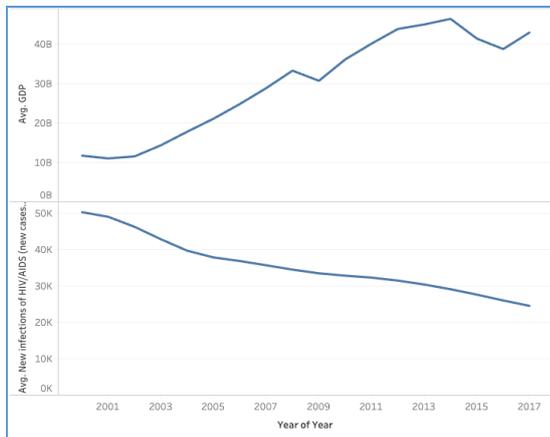


Fig. 10. Africa GDP and HIV/AIDS line chart.

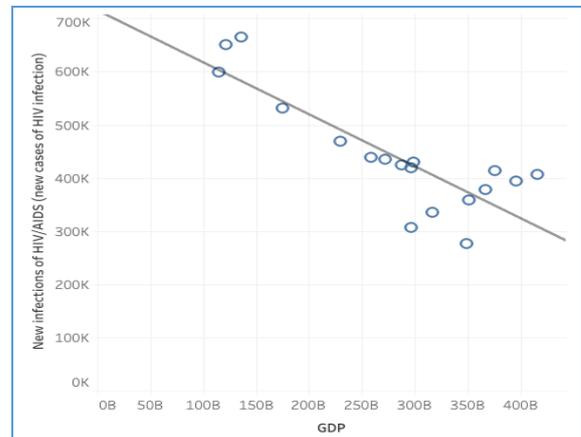


Fig. 11. Correlation trend line between GDP and new infections of HIV/AIDS in South Africa.

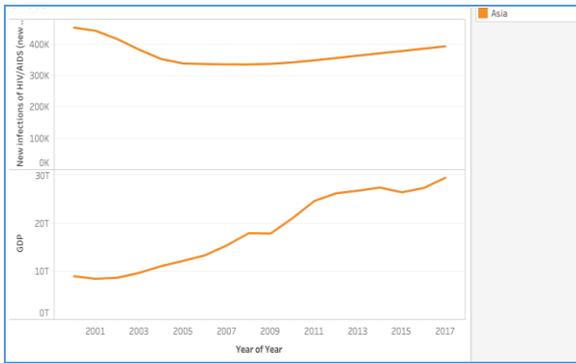


Fig. 12. Africa GDP and HIV/AIDS year of year line chart.

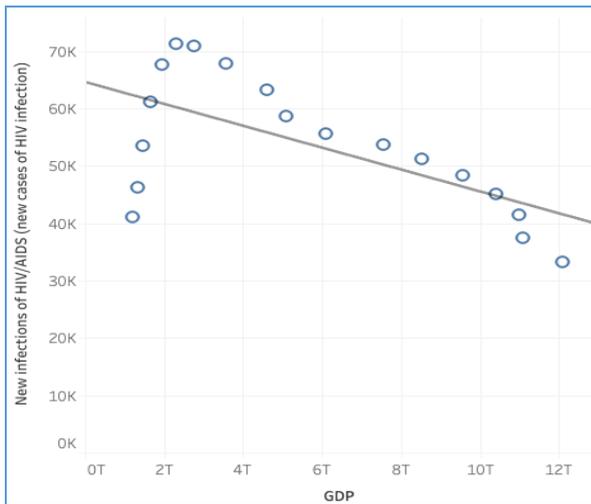


Fig. 13. Correlation trend line for China.

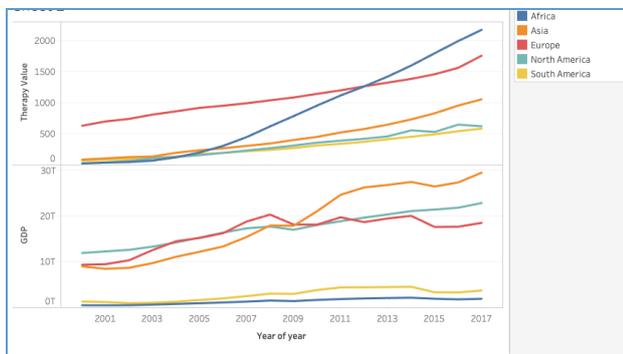


Fig. 14. GDP and HIV/AIDS Therapy value year of year line chart.

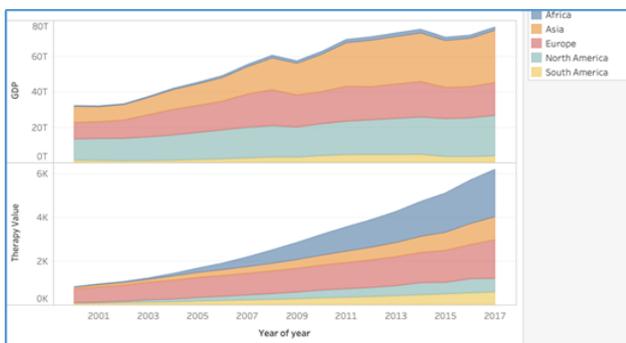


Fig. 15. GDP and HIV/AIDS Therapy value yearly area chart.

According to this chart, GDP has an affect on HIV and AIDS; however, GDP is not the only reason for the reduction of the new infections. There are other issues have can an effect on which the rate of infections increases or decreases, such as a country’s therapeutic support in the prevention of the disease. GDP affects a country’s economy,

which in turn has an effect on their ability to support patients and mitigate the spreading of the disease. From the data of each country’s therapy rates, Africa increases their therapy support at the same rate that their GDP increases.

IV. CONCLUSION

The charts and maps that were created using Tableau show that not only does GDP have an effect on the rate of HIV and AIDS cases, but it also impacts a country’s ability to care about advancement and treatment for the disease. Fig. 9a and 9b show that all continents are affected by their GDP, especially in the case of Africa where the number of HIV and AIDS cases were visibly reduced. The line chart, in Fig. 10, shows there was a positive correlation between a decrease in the number of infections and an increase in GDP. Fig. 11 is an example of one of the countries in Africa where one can see the correlation, and it shows that there is a decrease in the number of cases when the GDP begins to rise. However, there are also some exceptions to the rule. Fig. 12 is a line chart for Asia. Asia shows big increase in GDP, but the number of HIV and AIDS cases didn’t drop as much in relation to their GDP growth. Also, Fig. 13 is a correlation trend line between the GDP of China and the number of infections. China, the biggest country in Asia, shows a downward trend, but not as much when compared to Africa. Many similar countries with a large population show just a slight downward trend when comparing their GDP to their cases of HIV and AIDS.

In June 1981, the United States received their first report of AIDS. Soon thereafter, the number of cases and deaths among people with AIDS increased rapidly during the 1980s, followed by a substantial decline in new cases and deaths in the late 1990s. The CDC analyzed the reported cases from 1981 through 2000 and they found that the number of cases rapidly declined starting in 1990. They also noted no relation between infection rates and race, age, and gender. Also, many of the infections were caused from misuse of needles and cross contamination of blood. At that time, it was not known that infected blood could transmit HIV, which caused a rapid increase in the infection rates. According to the CDC, “in 1985, the first federal resources dedicated to HIV prevention were made available to all state and local health departments nationwide. In 1987, a national effort to educate the public about HIV and AIDS was launched and the CDC created a comprehensive AIDS information resource, the CDC’s national AIDS hotline and National AIDS Information Clearinghouse. Comprehensive school-based HIV education to inform and educated young people began in 1987, and funding for national, regional, and community-based organizations began in 1988” [12].

From this experience in the US, it shows that a government’s efforts to prevent spreading and provide therapy is effective in reducing HIV and AIDS. Fig. 14 and Fig. 15 show that when Africa’s GDP increased, their ability to provide therapy increased as well. However, Asia didn’t increase their therapy at the same rate as Africa; thus, their infection rate did not see as much of a reduction.

A conclusion can be made from this study, that a country’s GDP has some effect on the rate of HIV and AIDS transmission. This is because when their GDP increases, the country can prevent, or at least minimalize, the rate at which

the disease spreads through therapy and education about the virus. The initial expectation of the study was that only GDP has an effect on HIV and AIDS; however, the final outcome shows that what a country does with their GDP is also a determining factor in the trend of the infectious disease.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

E.K. collected and analyzed the raw data, transform the raw data into processed data, applied data mining and analytical techniques to create different graphs using Tableau software and wrote the paper. C.H. guided the overall research and revised the paper. Both authors approved the final version.

REFERENCES

- [1] C. I. Louise *et al.*, "HIV/AIDS, undernutrition, and food insecurity," *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, U.S. National Library of Medicine, vol. 1, Oct. 2009.
- [2] The Global HIV/AIDS Epidemic, *The Global HIV/AIDS Epidemic*, KFF, Sept. 2019.
- [3] J. H. Geiger, "Racial and ethnic disparities in diagnosis and treatment: A review of the evidence and a consideration of causes," *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.*, U.S. National Library of Medicine, 1 Jan. 1970.
- [4] C. Alexandra, "Poverty is driving an HIV epidemic," *Center for American Progress*, July 21, 2010.
- [5] J. O. Parkhurst, "Understanding the correlations between wealth, poverty and human immunodeficiency virus infection in african countries," World Health Organization, Mar. 4, 2011.
- [6] N. Rachel *et al.*, "The cost of HIV treatment," *Healthline*, Mar. 2019.
- [7] P. Piot, R. Greener, and S. Russell, "Squaring the circle: AIDS, poverty, and human development," *PLoS Med.*, vol. 4, no. 10, e314, Oct. 2007.
- [8] A. A. Aidala *et al.*, "Housing status, medical care, and health outcomes among people living with HIV/AIDS: A systematic review," *American Journal of Public Health*, American Public Health Association, Jan. 2016.
- [9] APL_NY.GDP.MKTP.CD_DS2_en_csv_v2_247793. [Online]. Available: <https://Data.worldbank.org/Indicator/NY.GDP.MKTP.CD>
- [10] R. Max and H. Ritchie, "HIV / AIDS," *Our World in Data*, Apr. 2018.
- [11] Total population by country, *World Population Review*.
- [12] R. S. Janssen *et al.*, "HIV and AIDS --- United States, 1981—2000," *Centers for Disease Control and Prevention*, Centers for Disease.

Copyright © 2021 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Eunbi Kim is a former student of Kean University where she graduated Magna Cum Laude with a degree in computer science in May of 2020. Before studying at Kean University, she was living in her home country of South Korea.

During her time at Kean University, she was the treasurer of the Association for Computing Machinery Women. She was accepted into and participated in the IEEE MIT undergraduate research technology conference and she was accepted into the Council on Undergraduate Research conference in order to present her sentiment analysis.

Based on her academic excellence, she was accepted into the mentorship program as a tutor and advisor for the computer science department. She also became the recipient of the international student scholarship for the 2019-2020 academic school year.



Ching-Yu Huang is an assistant professor of the School of Computer Science & Technology at Kean University since September 2014. He received a Ph.D. in computer & information science from New Jersey Institute of Technology, Newark, New Jersey, USA.

Prior to joining Kean University, he had more than 16 years of experience in the industry and academics in software development and R&D in bioinformatics. His research focuses SNP genotype calling and cluster detection; image processing and pattern recognition, especially in microarray and fingerprint; geotagged images and location information reconstruction; database application development; data processing automation; E-learning, educational multimedia, methodology, and online tools for secondary schools and colleges. Dr. Huang has more than 50 publications in journals and conferences and more than 40 presentations in workshops and invited lectures.

Dr. Huang received the Faculty Research Mentor of the Year, Kean University on April 24, 2019. He also received the Teaching Award, Excellent Performance in Computer Science Education, 3rd Annual NJ/NY CS Teaching Awards, Newark, NJ on June 14, 2018.