**Data Management and Visual Analytics**

**Group 3: Carrie Cox, Kevin Horan, Anna Thomas**

**12/4/2022**

**Country EPI Scores & Causes of Death Analysis**

**Motivations/Questions**

Our main motivation in creating this visualization analysis was to discover how deaths attributed to Acute Hepatitis, diarrheal diseases, and digestive diseases vary across the world. We picked these causes of death because they are largely preventable and linked to conditions that can arise from poor water quality and sanitation levels. We focused on poor water quality and sanitation because two billion people did not have access to safe drinking water in 2020, according to the UN. Also, approximately three billion people suffered from poor sanitation levels in 2020. Given the extent of poor water quality and sanitation across the world, we wanted to identify which countries, regions, and or continents suffer the most from these issues. From there, we wanted to identify any trends that appear between these causes of death and poor water quality and sanitation levels through visualization. To scale the total number of deaths for each country, we wanted to incorporate population to observe cause of death percentages within each country and then relate these percentages to EPI scores.

**Description of Data**

We used Kaggle.com to access a dataset on the causes of death. The dataset of the causes of death spans from 1990-2019. It contains the total number of deaths for each year, cause, and country. While the set has 32 causes of death, we focused our project on Acute Hepatitis, diarrheal diseases, and digestive diseases. The focus on these causes of death are a result of our focus on poor water quality and sanitation. According to the WHO, contaminated water and low sanitation levels leads to hepatitis, dysentery, cholera, diarrhea, and typhoid. The limiting of our causes of death to hepatitis, digestive diseases, and diarrheal diseases allowed us to correlate a country’s death rate and water/sanitation quality. We also decided to add the dataset of world population, which was retrieved from worldometer, to compare countries on a relative scale. This dataset consists of the population for 234 countries in 2019. Using the population data, we calculated the percentage of deaths per person for each of our three causes. We ranked the countries in our analysis from those with the highest percentage of total deaths per population to those with least. The last dataset we used for our project was scored from the Yale Center for Environmental Law and Policy. This dataset includes the Environmental Performance Index (EPI) score for 180 countries. The EPI scores countries by examining various measures of a country’s environmental health and sustainability. Specifically, the index uses 40 performance indicators across 11 categories to rank each country’s position on a scale of zero to 100. We examine the EPI score on unsafe sanitation and unsafe drinking water for each country to see whether a low EPI score indicates a higher death rate of hepatitis, diarrheal diseases, and digestive diseases.

While this data addressed our questions and motivations, biases or differences exist in the datasets. For instance, Yale Center for Environmental Law and Policy only assigns EPI scores to 180 countries, but we had information of causes of death and population for 234 countries. This discrepancy in our datasets suggests that some countries may have higher death rates for the three causes but could not be matched to an EPI score. Additionally, the results of the index may be skewed due to varying levels of information availability in each country. Those countries with significantly lower populations and are less industrialized may not accurately report water or sanitation quality. This lack of information availability may cause the EPI scores to be determine on incomplete data. The index also does not recognize a difference between centralized or government-controlled water and sanitation systems and decentralized or individual systems. Those countries lacking a centralized water system endure struggles that countries with government-supported systems do not to access and clean water. Overall, these datasets may contain biases as each one has estimates, which are not completely accurate.

**Rationale for Visuals**

Introduction:The first visual we created is on the introduction page of the story board. This map shows the death percentage of the country’s population from diarrheal diseases, acute hepatitis and digestive diseases. The color of red is used to show high countries with high death percentages (dark red) and countries with low death percentages (light red). The years 1990 to 2019 were used for this visual and the percentages are totaled for the given years.

Death Count: The bar chart on the right visualizes the countries with the highest numbers of deaths related to diarrheal disease, digestive diseases and Acute Hepatitis from 1990-2019. This bar chart is ranked using a created field and the function “rank unique” in order to get the countries in order by most deaths to least. There is also an animation at the bottom of the sheet that will show each year’s death count in these countries. The bar chart on the right is similar, however it is showing the top 10 countries with the highest death percentages based on their population. This percentage was found by creating a calculated field that divided the number of deaths (from the three causes) by a country's population. The color for both of these visuals is consistent with having the highest amount of death and percentages in dark red.

EPI Scores: We created this set of visuals to focus on a country’s EPI and to see what areas in the world have the lowest scores. By utilizing the EPI scores for water quality and sanitation levels, we were able to geographically visualize countries, regions, and continents with low and high EPI scores. Countries that have an EPI score of 50 or above are shown in green, while countries with a score of 49 or below are shown in red. To the right of the world map, is a bar chart showing countries with the lowest 10 EPI scores. This visual shows countries that we would like to analyze further, to see if they have high death percentages in the three causes we are looking at. At the bottom of this sheet, there is a cluster analysis of countries, EPI scores, and the 10-year growth in EPI scores. There are four clusters that group countries into different categories. The cluster categories are: high EPI & low change, medium EPI & high change, low EPI & medium change, and medium EPI & low change. We included this cluster analysis in order to identify groups of countries that have the worst water quality conditions.

Digestive Diseases, Diarrheal Diseases, Acute Hepatitis: Moreover, we used tree maps and world maps to illustrate those countries with highest number of deaths in 2019 for each cause by population size. For our three causes, the tree maps identifies the top five countries by size who had the highest percentages in 2019 for Acute Hepatitis, diarrheal diseases, and digestive diseases. The tree map is also colored to show the ranking of EPI scores among the five countries from dark to light red (i.e., low to high scores). We paired these tree maps for each cause with a world map to identify whether there was a geographical pattern between high death rates and location. The maps follows the same coloring as the tree map and also display the country’s EPI score. This map helps visualize the concentration of countries with high death rates for these diseases and whether the levels of water/sanitation quality are similar in these areas. The last type of graph we used in our project was a line graph for each disease. These line graphs present the rate of growth or decline in the total number of deaths caused by digestive diseases, diarrheal diseases, or hepatitis from 1990-2019. The line graphs can also be used to show an individual country’s pattern of total deaths by one of the causes of time. The line graphs are colored by whether the number of deaths have increased (light to dark red) or decreased (dark to light red). We included these line graphs for each cause to identify trends in the death rates over time in total and by the top five countries.

**Results**

Death Count: The two bar charts on this page show the prevalence of death from diarrheal disease, digestive disease, and Acute Hepatitis deaths in different ways. The visual on the left shows the total amount of deaths in each country, ranked in order, from 1990 to 2019. While this is interesting to look at it, it does not consider each country’s population. The chart on the right shows countries death rate (total deaths divided by country population). From this visual we can more accurately see how problematic these causes of death are for each country. India has the highest overall number of deaths but is third when compared based on the death rate. Nigeria has the highest percentage based on its population, and therefore is at the worst state based on these causes of death.

EPI Scores: After creating a map of EPI scores across the world, we discovered that regions of South America, Africa, Middle East, and Southern Asia suffered from the worst water quality and sanitation levels compared to the rest of the globe. The bar chart on the top right shows the countries with the lowest EPI scores. The bottom cluster analysis shows that countries can be divided into four separate groups. The first group of countries have a high EPI score with relatively low change. These countries are not concerning as they have EPI scores close to 100 and have not changed much in the past 10 years because of already having high scores (marked in green to show good quality). The second group is countries that have a medium to low EPI score, but high 10-year change. While these countries are still of concern, they are not viewed as poorly because of the water quality and sanitation being improved. The third cluster group shows countries that have a low EPI score and relatively medium change in the past 10 years (marked in dark red to show worst condition). Marshall Island has the highest EPI score of this group at 32.3 while Niger has the lowest at 1.5. This group is the cluster we are focusing on the most because of the extremely low EPI scores without a lot of improvement in the past 10 years. Lastly, the fourth cluster group of countries have a relatively average EPI score and low amount of change. This group is still concerning, but not at the same level as cluster group 3.

Digestive Diseases, Diarrheal Diseases, Acute Hepatitis: As shown by the tree and geographical maps for deaths by digestive diseases, countries in Eastern Europe suffer from the highest total deaths per person. Specifically, Moldova had the highest rate in 2019 at about 1% of deaths caused by digestive diseases. With an EPI score of 50 for water and sanitation quality, this result indicates that poor water/sanitation levels do not have a significant effect on the digestive tract. This result is further supported by Croatia’s EPI score of 70.3 as the country ranked fifth with a rate of .05%. The line graph also shows that the number of deaths for digestive diseases is increasing, but it has slightly decreased in recent years in Moldova. For diarrheal diseases, the results align with the WHO’s research. The countries with the highest percentages of death from diarrheal diseases in 2019 were concentrated mainly in Western Africa with one of the top five countries in Eastern Africa. The EPI scores for each of these countries (Eritrea, Niger, Nigeria, Mali, and Guinea-Bissau) range from approximately one to eight. Niger, the country ranked second with a rate of about 1%, has an EPI score of only one-and-a-half. The correlation between the rate of deaths caused by diarrheal diseases and the country’s EPI score indicates poor water quality and sanitation may be the source of these deaths. While the number of deaths caused by diarrheal diseases has consistently declined year-over-year, Niger has experienced periods of increasing deaths with declines only in the most recent years. Similarly, the countries with high percentages of Acute Hepatitis deaths by population size also have low EPI scores. Hepatitis deaths were concentrated in South Asia with one of the top five countries in Eastern Africa. The scores for the top five countries (Afghanistan, Pakistan, Bangladesh, Sudan, and India) range from 17 to 28. Overall, the number of deaths caused by hepatitis has decreased every year, but Afghanistan suffered from steep increases in these deaths until 2003. Pakistan’s number of deaths from hepatitis has been increasing slightly year-over-year since 2007.

**What We Learned/Takeaways**:

Overall, this project identified the countries and regions that suffer from the highest number of deaths caused by Acute Hepatitis, digestive diseases, and diarrheal diseases. To compare countries of different sizes, we calculated the number of deaths for each cause per person. We used these percentages to analyze the top five countries with the highest rates and their EPI scores for water and sanitation quality. We learned that deaths from diarrheal diseases and Acute Hepatitis are correlated with low EPI scores, but deaths from digestive diseases are not. While the Eastern European countries with the highest death rates for digestive diseases have high EPI scores, this cause of death is the only one increasing of the one examined. This increasing death rate may indicate new factors are causing people to contract deadly digestive diseases. Additionally, we learned that countries in Eastern and Western Africa, and South Asia endured the most deaths from Acute Hepatitis and diarrheal diseases in 2019. This cluster of countries also have the worst EPI scores as shown on the map. This finding can help direct funding from the IMF and World Bank to establish access to safe drinking water and proper sanitation in these regions. Given the low EPI scores and high death rates, these countries require additional support from these organization to fund centralization water and sanitation systems. With this funding, countries, like Niger and India, can reduce their death rates caused by these preventable diseases.

**Contributions**

|  |  |
| --- | --- |
| Group Members: | Contributions: |
| Carrie | I attended every group meeting and participated in conceptualizing our project/visualization by finding the main data source of death causes. I created the death count and percentage bar charts, the EPI cluster analysis, and the tree maps for each death cause. I put the story board and everyone’s visuals together and helped write the report. |
| Kevin | I attended every group meeting and participated in conceptualizing our project/visualization. I created the world map for EPI scores. I identified the problem of not incorporating population into the analysis which led to us using cause of death percentage of the population in some parts of our analysis. I also helped with outlining and writing the report. |
| Anna | I attended every group meeting and participated in conceptualizing our project/visualization. I found the EPI dataset and suggested we correlate EPI scores and the causes of death. I created the world maps for our three causes of death with their EPI scores. I took notes to structure our report and helped with writing the report. |

**References**

Datasets:

<https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world>

<https://epi.yale.edu/epi-results/2022/component/h2o>

<https://www.worldometers.info/world-population/population-by-country/>

Report Sources:

<https://www.who.int/news-room/fact-sheets/detail/drinking-water>

<https://epi.yale.edu/epi-results/2022/component/epi>

<https://www.un.org/sustainabledevelopment/water-and-sanitation/>