Raul Martinez - Ivan Ulloa

Final Project Report: Environmental Radiation Monitoring in the US

Abstract

Gamma radiation monitoring is of national importance due to background fluctuations induced by the environment and the potential damaging effects to the human body. Many air radiation monitors located around the US have been collecting data since 2006 in near-real-time with 1 hour resolution, and data is publicly available via the RadNet program. To facilitate the understanding of spatio-temporal radiation patterns and correlations within the data, a Dash-Plotly Python based dashboard application for visualizing environmental radiation data is presented in this report. The solution includes a series of plots to make comparisons across different states and cities over time for aggregated and raw data. Every element is interactive and data can be transformed with user inputs. To illustrate the usability of this visualization, some sample findings are discussed such as identifying areas with the highest gamma count rates like Raleigh, Colorado Springs, and Little Rock; in addition to areas with lowest levels like Puerto Rico, Anchorage, and Seattle. Other observations from Knoxville, TN show a large spike in gamma counts per minute for November 2016 likely due to a large scale wildfire in the Smoky Mountains. The correlation between elevation and radiation metrics is also discussed. All in all, better analysis and visualization tools are needed for radiation monitoring across the country in order to understand fluctuations in both background and abnormal levels, and how these changes relate to the environment, human health, and national security.

Introduction

Radiation detection and measurement has been an important field of research for medical professionals as well as scientists and engineers in many fields such as oil extraction, space exploration, and food safety among others. Radiation or more specifically ionizing radiation is a form of high energy photons that can be harmful to the human body. The term "ionizing" signifies the capacity of a high energy photon to interact with electrons in an atom and transfer energy such that they may be ejected and therefore leave the atom in an ionized state. Events like these may create charge instability in the near field of the atom which will initiate a series of changes to ultimately regain stability. Health effects may arise when sufficient ionizing events occur in sensitive locations such as DNA which may alter its sequence beyond repair. Gamma radiation is a form of ionizing radiation emitted specifically from the nucleus of an atom such as when an isotope goes through the process of radioactive decay. Radioactive decay is a naturally occurring process and is a great contributor to background radiation. The type of soil, building materials, and elevation among other features contribute differently to the amount of background radiation in the environment.

The US Environmental Protection Agency (EPA) provides environmental radiation data accessible to the public via the RadNet program. The RadNet program monitors environmental radiation in air using 140 air radiation monitors scattered across the US. Currently there isn't a streamlined method to view the information provided by the RadNet program and data for every sensor must be downloaded individually by year. This project aims at delivering an accessible and integrated view of the RadNet environmental radiation data for exploration and comparison.

Dataset

Information used for the visualization was obtained from the EPA's RadNet program website. Data consisted of 1742 files in CSV format totalling over 1.08 GB of data. Each of the data files represents individual sensor readings for one year and contains the following information: Location Name, Time in UTC, Dose Equivalent Rate, and Gamma Count Rates spread over 8 channels. The location name contains the name of the city and state where the sensor is located, the collection time is presented in UTC time with a resolution of 1 hour, for the purposes of this visualization all time's were adjusted to local time by adding the UTC adjustment depending on the geographical location of the sensor. The Dose Equivalent Rate is presented in units of nSv/h (nanosieverts per hour), this dose rate represents the rate of radiation damage specifically to the human body. The Gamma Count Rates represent individual radiation particle interactions per minute, these are binned by their energy detected over 8 bins or channels, here particle energy was not incorporated and therefore all channel counts were aggregated and presented as a total gamma count rate for each sensor.

Data was consolidated into one dataframe using python's pandas library and all timestamps aligned in local time. Latitude, longitude, and elevation (meters) information of the city were merged with the dataset for mapping and analysis purposes. Specific coordinates for each sensor weren't available and therefore general city coordinates were used to generate the US radiation map. The augmented dataset was ultimately stored as a pickle file for improved performance.

Tasks

Multiple visualizations and data transformations presented in this report are used to explore the different levels and distribution of gamma radiation collected from RadNet air sensor data across the US. This effort aims to enhance geospatial-temporal analysis of the data for radiation professionals. First to understand what parts of the country, at a state and city level, have the highest or lowest amounts of background radiation and at which rate it expresses given that these effects have the potential to damage the human body to a certain extent; and can also induce long lasting effects to future generations of family members. In addition to looking at single time snapshots geographically it is also important to raise questions about how radiation patterns move overtime and how fast they move in a relative time scale of either days, months, or years. It is also of interest to analyze trends over time for multiple cities in conjunction by aligning their respective time scales with time zone correction, as well as looking at multiple rolling window sizes for smoothing fluctuations and other unwanted effects in the near-real-time data. While it is important to look at the data available, another goal is to compare this data with known events where radiation deviated from normal levels, with the intent of providing relative comparisons as radiation measurements might be domain specific, some examples include natural phenomena or other human induced disasters that eventually resulted in radiological incidents in history. And lastly, these visual elements also aim to evaluate if there is a noticeable correlation between elevation and radiation levels and how this correlation might change temporally.

Solution

The visualization makes use of a dark mode interface with a combination of yellow and some shades of blue to display the different main elements, the colors are chosen to provide the user a link with the commonly observed radioactive sign, which is also included on the top-right. The layout is well-structured with multiple cards to hold the different elements within each such as plots, text, and controls; they are sized and separated accordingly. The components added for interactivity include a date range for filtering time, multiple dropdowns with multi and single value options to select the gamma radiation metrics as well as grouping and filtering the different cities or states, an input field for holding the numerical value for the window size, and a button for triggering data load. All these components contain a dark blue background with lighter blue edges to separate the object visual encoding from the background, and lastly, when hovering along the dropdown options there is a change in font color to white with a light blue highlight for highlighting the selection process.

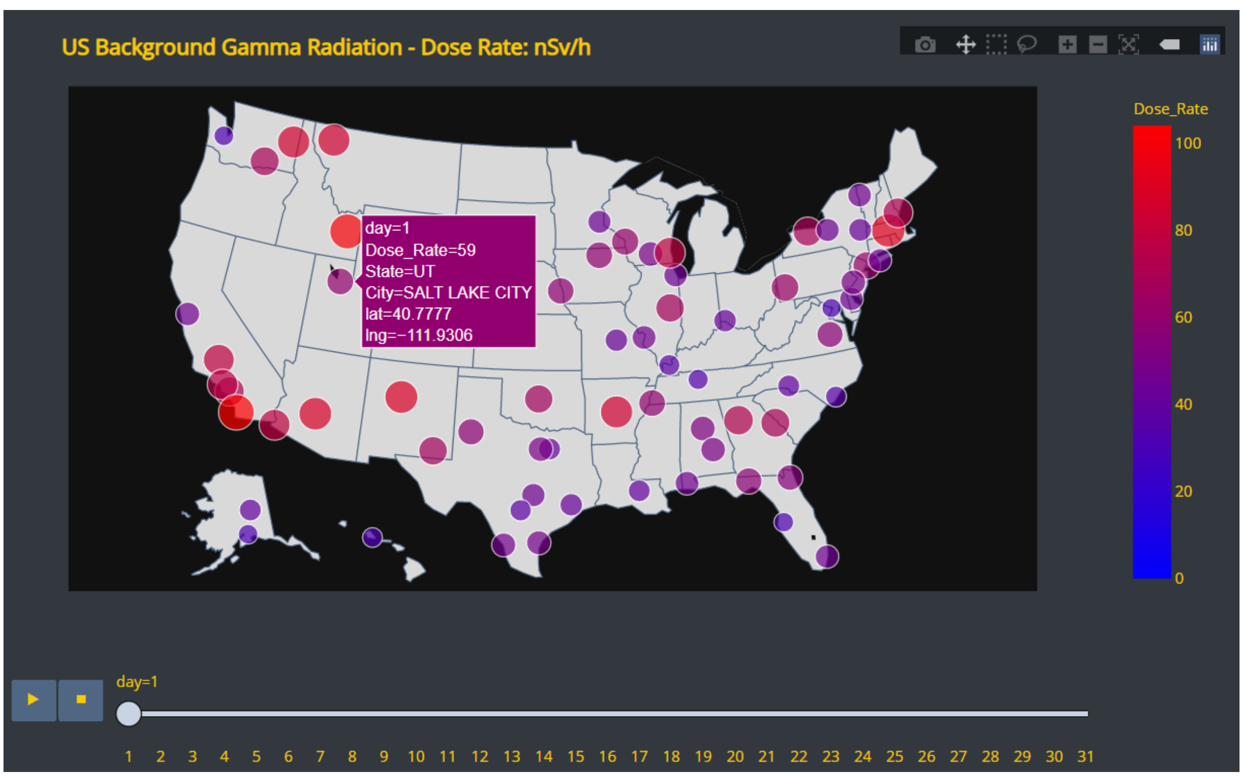

Figure 1. Geo Scatter plot, shows sample data for Dose Rate metric by day for Salt Lake City.

The first chart for the visualization is a geo-spatial scatter plot for the US map (*see Figure 1*), including Alaska and Hawaii, where a red-blue color map is shown to display high and low concentration levels of radiation respectively. The plot allows the user to change the marker colors by playing and stopping temporal data iterations when scanning across multiple animation frames as defined by the group by option where the lowest resolution for the data is an hour, for example, when aggregating by day or year a total of 24 or 12 values are shown accordingly; month values might differ from 28 to 31 but only relevant data is included. In other words, aggregations are done positionally as defined by the time scale. Min and max values are calculated for the entire dataset to avoid re-defining the color scale values on each animation frame and allow fair comparisons. This design gives the user the ability to explore spatio-temporal trends across the country or by expanding in regions of interest within the map.

Next to the map, radiation metrics are displayed with horizontal bar plots sorted in descending order by the selected metric in order to easily find the areas with highest and lowest values. Bars color is defined to match the yellow selected in the other elements and light gray vertical lines for the x-axis are used to simplify relative comparisons between bars. Then, a treemap is used to show the relative proportions of radiation within the country, it displays the state and city hierarchy with values in a well compacted visual channel (*see Figure 2*). It not only shows the cumulative contribution for each state but also the number of components. The chart's interactivity allows the user to expand on every single object while the color scale is kept the same to compare across. Additionally, a scatter plot is included to display the correlation between elevation and radiation metrics. Horizontal and vertical distribution box plots are included along both x-axis and y-axis as sub-plots to show distribution characteristics like min, first quantile, median, third quantile, and max; in addition to providing some insight about possible outliers. All elements within the plot are colored yellow to provide a single aspect and maintain focus when finding correlation patterns, and a light grey border is used on the scatter plot markers to allow differentiation when superimposed.

Figure 2. State and city hierarchy illustration with Treemap chart. All US states (left) and California selection (right).

At the end of the visualization there is a large time series plot to display line charts for either radiation metric per city. The design includes dropdowns for state and city selections, where cities options are auto-populated after selecting or removing every state. An input cell is included to allow the user to find the moving average of the time-series for any integer value lower than the data size. The date range used at the beginning of the visualization is also connected with this plot for filtering. Due to the high number of values for each and multiple time series lines the data load step is controlled, meaning that data will only be updated when the button "PLOT" is clicked.

The implementation provides great diversity of visual encodings and data structure to explore data patterns and perform analysis. Every object within the interface is interactive for selecting, re-scaling, highlighting, and hovering in order to display relevant details. It also consolidates large amounts of information within the visualization channels. Some of the key features for handling a lot of data are not only to filter and select but also to give the user the ability to transform raw data on the fly, as opposed to reporting only raw data, by utilizing the different components available such as temporal aggregations and computing the moving average for either radiation metric.

Lastly, the visualization solution is built with a dash plotly web application software which is basically an open-source python library powered with flask and react frameworks. It makes use of callback functional decorators with inputs, states, and outputs to update and modify the interactive components. The usability of this application is asynchronous and scalable, especially when doing back-end data processing because the front-end and back-end are decoupled, therefore, users are not expected to wait for processes to complete and they are free to explore both the user interface and already loaded data in the front-end.

Results

There are three objectives that have been set to be answered using the visualization, determining the highest and lowest levels of background radiation in the US, searching for correlations or changes in the amount of radiation after large events have occurred such as natural disasters and human made accidents, and lastly, the correlation between altitude and radiation levels.

To answer the first question the visualization makes use of a horizontal bar chart which contains the list of sensors, or in this case the names of the cities sorted from high to low either by radiation dose rate or gamma counts per minute as shown in *Figure 3*. The expressiveness of this plot is successfully met in the sense that the order of the list from top to bottom matches the magnitude of the information being shown which for this purpose is the radiation metric. This allows for the user to quickly find and compare metrics between cities. The effectiveness of the plot is met by using spatial alignment and length of the bars to easily compare between them. To answer the objective question, it was determined that the most background radiation was seen in the cities of Raleigh, Colorado Springs, and Little Rock and the lowest in Puerto Rico, Anchorage, and Seattle in terms of gamma count rates.

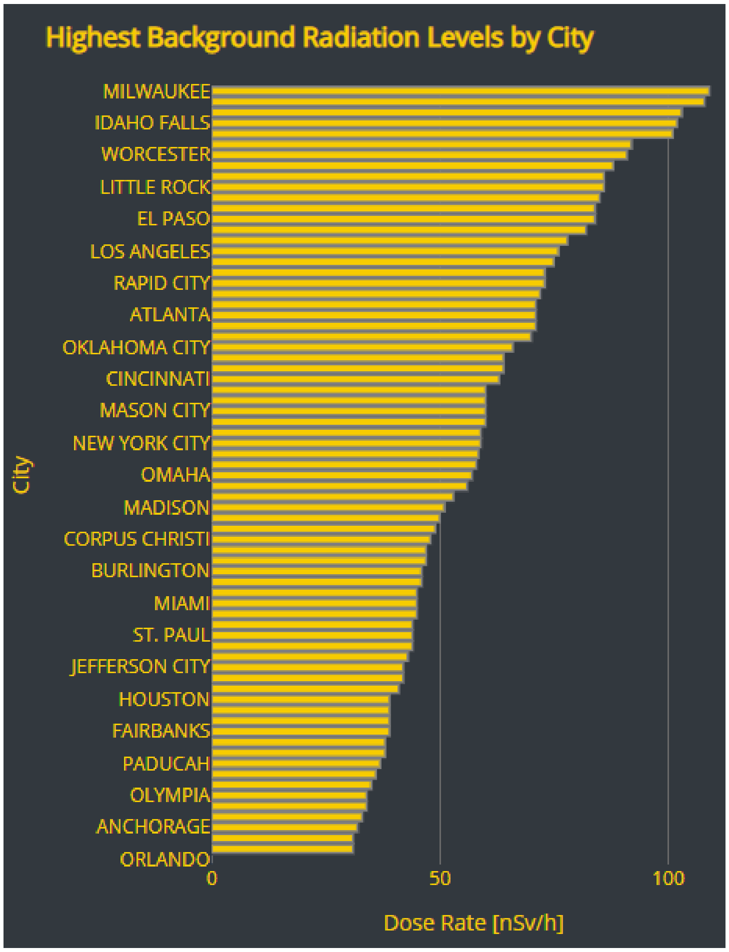

Figure 3. City Background Radiation Bar Plot. Shows average radiation levels by city sorted from highest to lowest.

The second objective was to determine if changes in background radiation can be seen throughout time, after natural disasters, or after nuclear events occur. For the US, no recent nuclear event has occurred and therefore only natural disasters and time related patterns have been investigated. For this, a time series graph which maps a radiation metric against time can be used to detect changes in the amount of radiation over time. Here the expressiveness of the visualization is accomplished by only displaying the information needed with the additional hover overlay of additional information. The information presented is categorical and therefore data has been separated using color to identify the different sensors or cities. Further, the moving average feature allows the user to smooth the data and identify trends more easily.

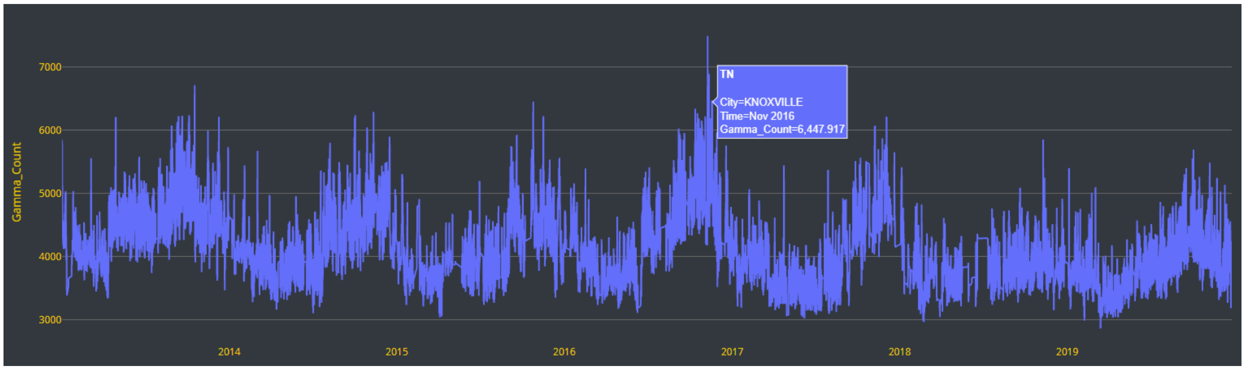

Figure 4. Radiation Time Series Plot. Shows gamma count information for Knoxville, TN. from January 2013 to December 2020. Highlighted field corresponds to November 2016.

It was noted from the data that most sensors behave differently, have sharp adjustments represented by large offsets in the rates which could be attributed to sensor maintenance or settings adjustments. To address the second objective, after searching for natural disasters in the US, the wildfires in the Great Smoky Mountains in Tennessee represent a good example of how this information can be quickly retrieved to make comparisons. In November of 2016, a large scale wildfire occurred in the Smoky Mountains, it can be seen in *Figure 4* that over a seven year timelapse there was a large spike in gamma counts per minute during the November timeframe in 2016. It is known that ashes serve a transport mechanism of radioactive particles.

The third objective looked at identifying a correlation between the elevation of where the sensor is located and the amount of radiation registered. The correlation plot from *Figure 5* shows elevation plotted against a radiation metric, either dose rate or gamma counts. It can be seen that the distribution of the scattered points does have an overall incremental pattern, that is, as elevation increases gamma count rates also increase. The same situation for the Gamma Dose Rate metric occurs but it isn't as easily identified as this metric also incorporates the energy of the radiation which is not considered in this visualization.

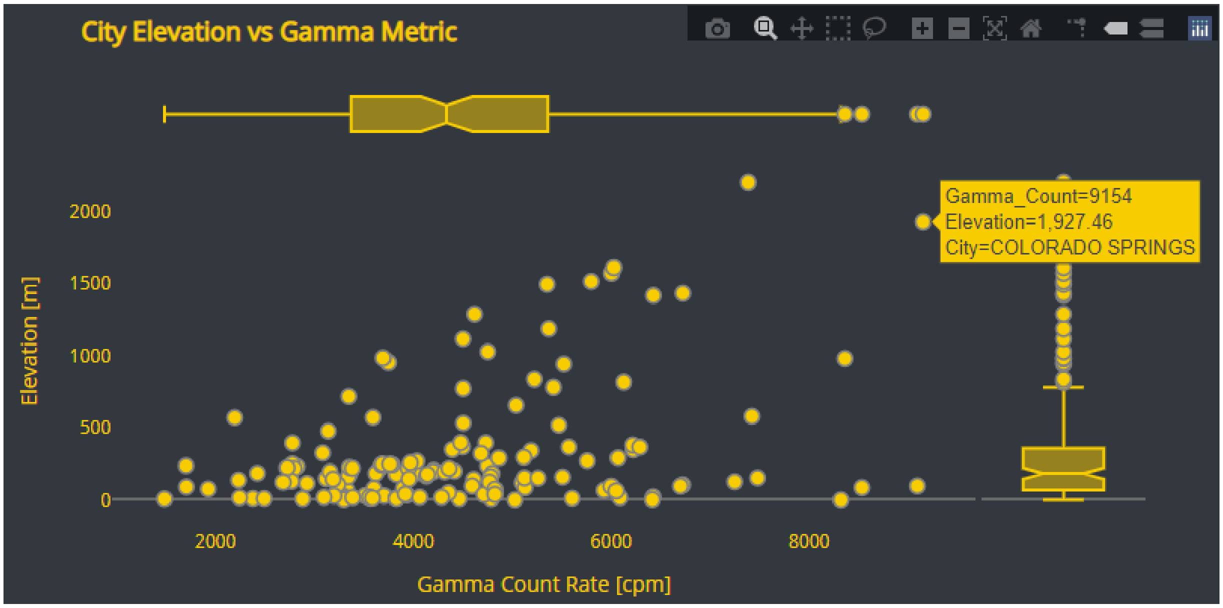

Figure 5. City Elevation vs Gamma Metric. Scatter plot using elevation and gamma count metric to explore correlation between these two features.

Overall the visualization accomplishes one of the main objectives that is to facilitate viewing of the information from RadNet and to quickly search and compare values by location and time. One of the current drawbacks from the dataset is that the exact locations of the sensors are not provided. The lack of positioning will affect the elevation estimation since the coordinates are only assumed to be inside of the cities boundaries which could be low or high in elevation. It is also questionable how good the data has been normalized across their different sensors, the efficiency of detecting radiation may vay between the sensors and therefore some normalization should be made in order to compare values between them.

References

- EPA RadNet Datasets
  - https://www.epa.gov/radnet/radnet-csv-file-downloads
- Latitude, Longitude, and Timezone Information
  - US Cities Database | Simplemaps.com
- City Elevation Data
  - https://nationalmap.gov