Disaster at St. Himark

An analysis of Radiation Data

Author: Andrew J. Otis

Background

St. Himark advertises various attractions that promotes the raising of a family accommodating various economic classes and the “beach go-er’s” lifestyle. One of the largest local employers is the nuclear plant “Always Safe Nuclear Power Plant”, considered to be the pride of the city[1]. Always Safe generates more than enough energy to make the Ocean city of St. Himark self-sustainable and surplus energy is exported to the mainland, providing even further value to the St. Himark community at large.

During the dates of April 6, 2020:12:00 am - April 10, 2020: 11:59 pm (*i.e. 120-hour timelapse*) St. Himark was struck by an earthquake. The nuclear plant suffered damage causing radioactive contamination by leaks. In addition, coolant containing radioactive waste sprayed the cars of employees.

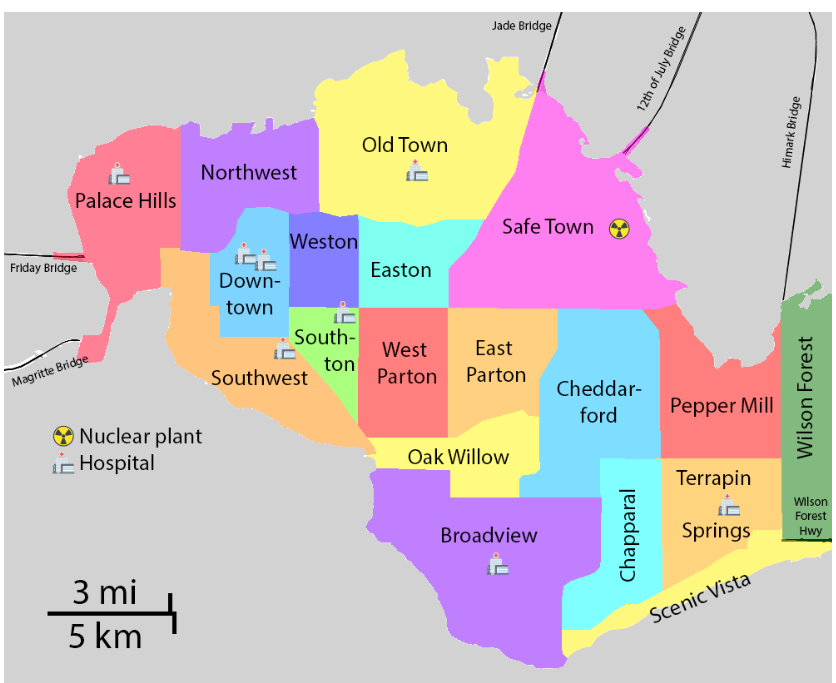

Figure 1: Neighborhood map of St. Himark

Research Question

To support disaster mitigation efforts, the task is to determine the extent of the radioactive contamination in the city of St. Himark. What are likely locations that will require further monitoring, cleanup or even evacuation[1]? By providing emergency services and city officials a holistic understanding of the leak's risk to public safety[1], city resources can be managed efficiently and effectively.

Additionally, determining if data uploaded citizen scientists assists in clarifying the situation or makes it more confusing and harder to respond to.

Description of Metadata

It's important to note the data analyzed is fictional, modeled and inspired by the real-world radiation map data created by the scientific organization "Safe Cast".[5]

Data collection precedes the day of the earthquake and subsequent radioactive contamination. Due to not following international regulations during construction, fears that plant operations increasing background radiation arose. Always Safe covered the expense to have static sensors professionally calibrated and installed at various locations throughout St. Himark. Additionally, as part of broader nuclear education local citizen scientists fitted their cars with homemade radiation sensors, making them mobile.

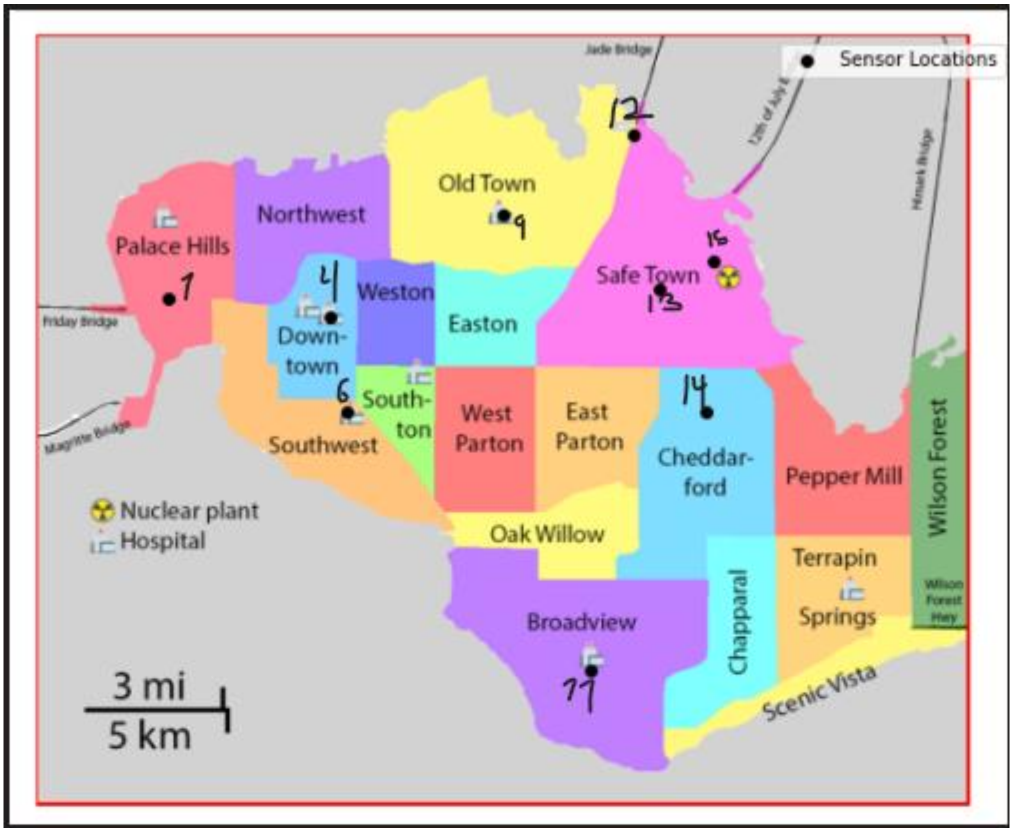

Figure 2: Location of Static Sensors paid for by Always Safe

Description of Dataset

The data consists of 3 csv files, 1 file for static sensor data(*radiation values only*), 1 file for static sensor location data, and 1 file for mobile sensor data(*location data already included*),

Including the initial powering of a sensor, radiation readings were collected in 5 second intervals. As a result, there are hundreds of thousands of datapoints per sensor, static and mobile, not all valid datapoints.

```
----Stationary Sensor Locations----
Dimensions: (9, 3)
```

Sensor-id	Lat	Long
12	0.20764	-119.81556
15	0.16849	-119.79033
13	0.15979	-119.80715
11	0.04147	-119.82861
6	0.12180	-119.90430

Table 1: Preview of spatial data for Static Sensors

```
----Stationary Sensor Data----
Dimensions: (744000, 4)
```

Timestamp	Sensor-id	Value	Units
4/6/2020 0:00	15	11.926	cpm
4/6/2020 0:00	6	14.816	cpm
4/6/2020 0:00	1	17.055	cpm
4/6/2020 0:00	12	16.787	cpm
4/6/2020 0:00	13	15.009	cpm

Table 2: Preview of temporal and radiation data for Static Sensors

```
----Mobile Sensor Data----
Dimensions: (3315711, 7)
```

Timestamp	Sensor-id	Long	Lat	Value	Units	User-id
2020-04-06 00:00:00	15	-119.83035	0.14007	0.0	cpm	CitizenScientist
2020-04-06 00:00:00	22	-119.76075	0.04205	0.0	cpm	HSS
2020-04-06 00:00:00	40	-119.89067	0.11658	0.0	cpm	MutantX
2020-04-06 00:00:00	1	-119.96665	0.18792	0.0	cpm	MySensor
2020-04-06 00:00:00	27	-119.80570	0.01711	0.0	cpm	CitizenScientist

Table 3: Preview of temporal, spatial, and radiation data for Mobile Sensors

Data Wrangling

Radiation sensor data was originally retrieved from GitHub profile[1] and files were provided, not much wrangling was required of the data. The load of the work was in visual analysis steps.

Visual Exploration

Difference in sensor readings have the same shape in visuals when transformed, so I saved plots of non-scaled values and also have my code plotting the scaled values. Ultimately, differences in radiation readings among mobiles become more apparent when filtering out the warm-up spikes that give inaccurately high readings. I assume this is just the nature of this particular electronic equipment.

From visual exploration, characteristics of the data can be identified. Relevant to this analysis is the change in radiation level overtime through out the various the community of St. Himark.

The following are characteristics for changes in radiation identified when readings are visualized by day or by hour.

When initially powered on , there is a large spike in readings before returning to actual surrounding values.

When powered down, sensors remain at current reading until powering back up, where the warm-up spike occurs. The amount of time a sensor has been off does not appear to be correlated with whether a warmup spike occurs (*i.e. these happen all the time*)

There are data where higher readings are not a warm-up spike. Some warmup spikes are lower than actual high readings, so it was not ideal filtering these out, as it would also filter out valid readings.

The challenge being to identify whether a sensor is in cool down, or if a mobile sensor is driving into or out of a contaminated area. Comparing the graphs by date and hours can assist with this.

Data Pre-Processing

Pre-processing steps require the data to be log-scaled, it's also important to note any 0's or large numbers will result in a -inf or inf value, respectively which were replaced with 0.

For the purposes of geographic plotting, the following transformations were performed. New coordinates in the form of pixels in order to do plot by time and location, converting the "Timestamp" column to pandas datetime dtype, and adding columns for hours for another perspective of radiation overtime.

To account for the largest of warmup spikes, the data was filtered down to only include values shown in CSU's radiation exposure severity categorization.[3] Spikes from sensor warmup are inescapable, but the most egregious offenders can be filtered out.

cpm	mSv/yr
Small fraction of background < 1.00 cpm Artificial Source (medical equipment) = 1.20 cpm Normal background from natural source < 40.0-60.0 cpm Safe < 150 cpm Max dose of uranium miner < 200 cpm Low cancer risk > 500.3 cpm Occupational regulation = 1000.68 cpm	Small fraction of background < 0.05 mSv/yr Artificial Source (medical equipment) = 0.3-0.6 mSv/yr Normal background from natural source < 40.0-3.0 mSv/yr "Safe < 7.50 mSv/yr \n" Max dose of uranium miner < 10 mSv/yr Low cancer risk > 25.0 mSv/yr Occupational regulation = 50 mSv/yr

Table 4: Conversion of industry standard unit of milli-Sievert to cpm (i.e. the units of the data analyzed).

It's important to note that the sensors are not Geiger counters, but with a provided conversion factor, I used the conversion factor for a Geiger counter to allow for the conversion of milli Sieverts per year(*mSv/yr*) to Counts Per Minute(*cpm*) and vice versa.

Data Analysis

Sensors Too Uncertain to Trust:

Regarding sensors too uncertain to trust. I believe any mobile sensors that were contaminated during the coolant spray are un-reliable when providing an accurate reading for an area of lower radiation. However, they can be used for further verification of entering a contaminated area since they're readings would increase just like the uncontaminated cars, just to a higher degree.

Mobile sensor 44 is a great example of this effect on uncertainty.

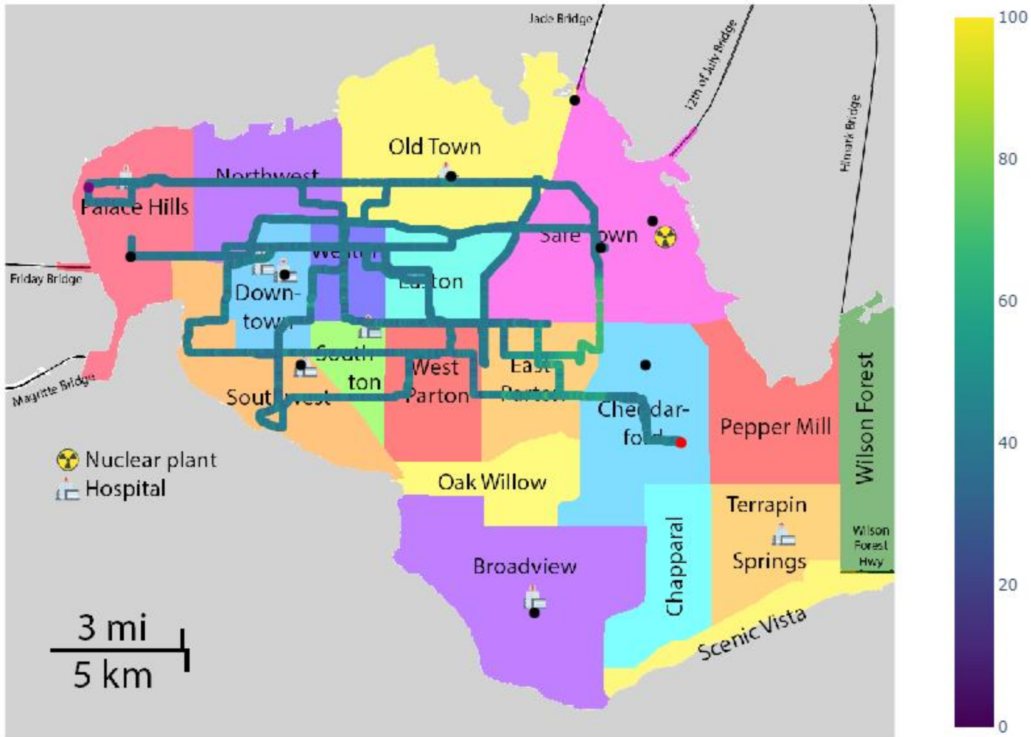

Figure 3: Path of mobile sensor 44 by radiation and location

Plotly puts this sensor at the nuclear factory at the expected time of contamination. Additionally, as the car drove by static sensor 13 and 9, it showed higher radiation readings, implying a likely contamination of the mobile sensor. The start of the timelapse is represented by a purple dot while the end of the timelapse is represented by a red dot.

Sensors Causing Uncertainty in Measurements:

The following are regions in St. Himark that intrinsically have greater uncertainty in radiation levels. Any neighborhoods a contaminated car drove, where no un-contaminated mobile or static sensors was also present and anywhere a mobile sensor did not drive or where a static sensor is located

Effects of the Earthquake on Measurements:

Main earthquake appears to be the Morning April 8th. Possible aftershocks occurring on April 9th due to unaffected static sensors beginning to rise. This is supported by the gradual increase of all average sensor readings.

The effect on static sensors is that some (e.g. 1) lost power and when powered on again, experienced the same warm-up spikes as mobile sensors.

Some mobile sensors, while located at nuclear plant and after April 8th increased to above normal background radiation(*i.e. coolant spillage*).

Regarding static sensors, some will show unusually high maximum radiation readings, far higher than measured even at potentially high contamination sights. These, like with the mobile sensors, are warm-up spikes. Which were accounted for by filtering. In a dataset of hundreds of thousand points, getting rid of these outliers shouldn't affect analysis greatly and provide a much more intuitive idea of how the sensors are behaving pre and post-earthquake. Referring to the "by hour" of a maximum plot assisting in determining whether or not the extremely high spikes in *Figure 4* are sensors warming up or an actual increase in radiation level.

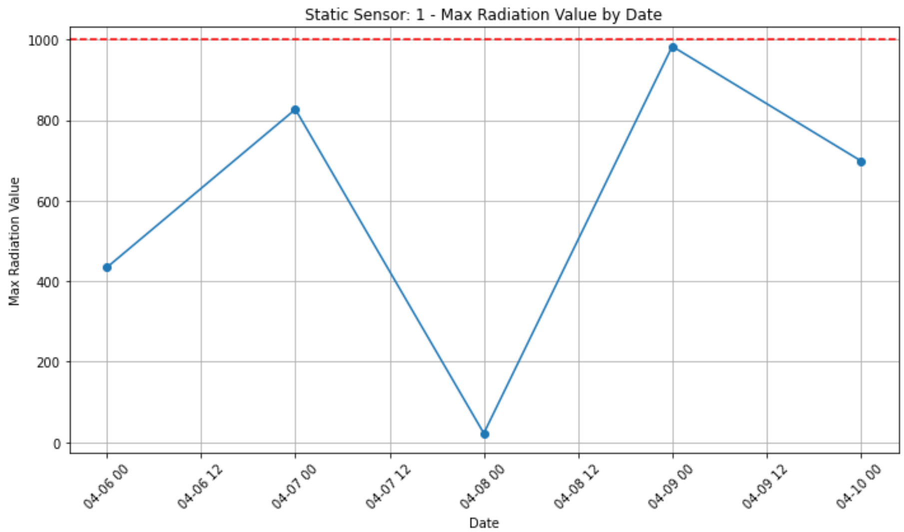

Figure 4: Unfiltered measurements show that static sensor 1, all points except for 1 were warm-up spikes. This is further supported by observing the max radiation by hour

.

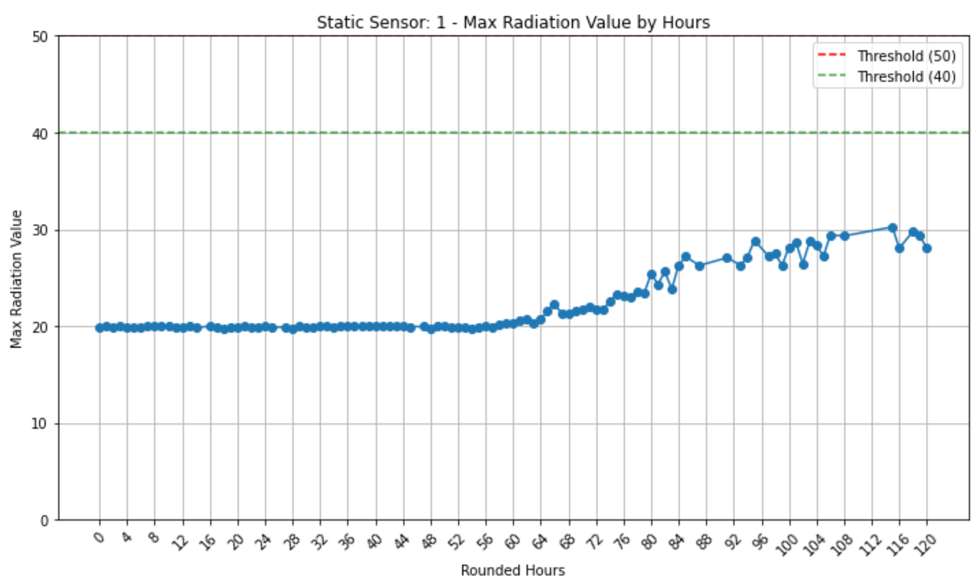

Figure 5: Average radiation by hours for static sensor.

Observe, at about after hour 64, the maximum sensor values steadily increased. However, it's also important to note that the highest values for this particular static sensor don't even go above background, but has the potential of getting worse if a contamination has occurred in that area(*determined in further data analysis*)

Potential Locations of Contamination and Uncertainty in Readings:

Using pandas "Cuts" functionality, values for x and y pixel of mobile and sensor data coordinates are selected to calculate the mean values per area as seen in the figure below.

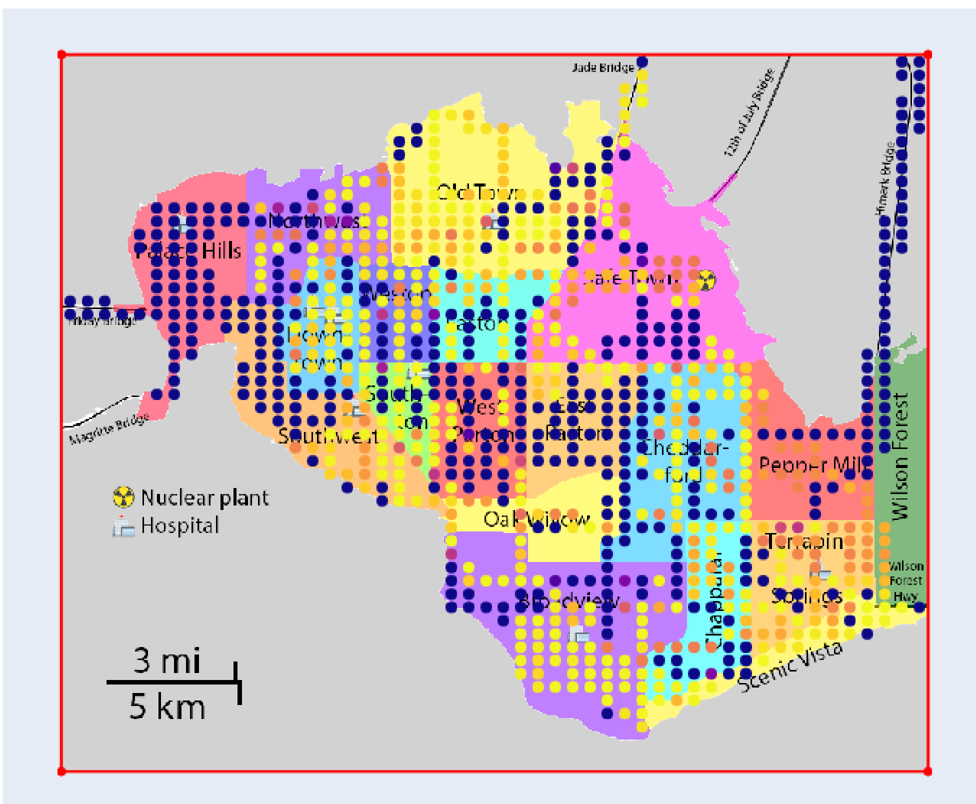

Figure 6: The color scale here is opposite that of the mobile sensors paths. Darker blue tone means higher radiation and lighter yellow tones means lower mean radiation values.

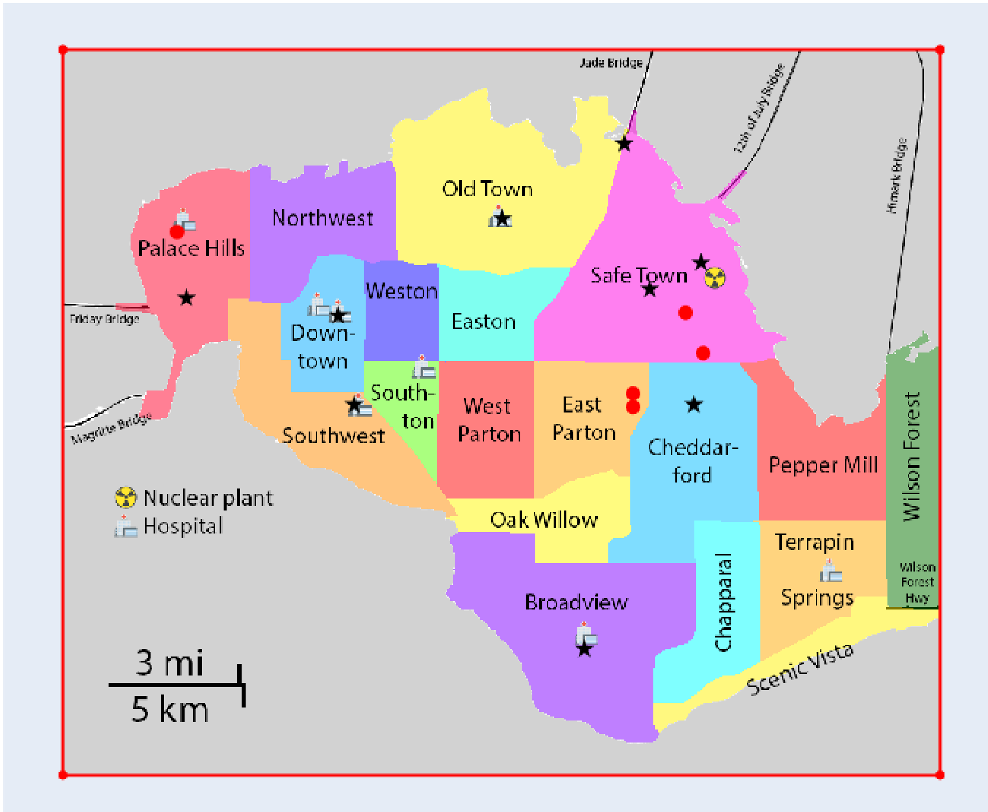

Figure 7: Plotting the top 5 highest readings to identify potential contamination areas by mean radiation level.

Top 5 Highest Mean Values Plotted
1. Mean Value: 66.35, Border of SafeTown and Cheddarford
2. Mean Value: 61.64, East Parton
3. Mean Value: 53.05, SafeTown
4. Mean Value: 51.77, East Parton
5. Mean Value: 49.36, Palace Hills

The area Palace Hills from *Figure 7* implies that there is a contamination spot near the hospital. However, mobile sensor 44 and other contaminated cars driving in the area contributed to the mean radiation level higher than the low threshold for safe exposure.

Thus, likely contamination spots are the 2 mean values above the higher threshold of safe radiation exposure located on the border of SafeTown/Cheddarford and in the Eastern part of East Parton, respectively.

Contaminated Cars:

I don't think officials need to be worried about contaminated cars spreading radiation. The concentrated radiation point in palace hills is the contaminated car 44. Additionally, sensor 2 was at the same hospital post contamination and did not show any higher-than-normal radiation measurements when leaving palace hills.

Recall, concentrated points of contamination appear to be concentrated around the tri-county area of SafeTown/Cheddarford, and East Parton.

Additionally, notice as mobile sensors drive by these concentrated contaminated sections, their readings will go up. Those sensors that remain above the normal background radiation and were located at the nuclear site the day of the earthquake, are potentially contaminated.

For having consistently higher than background radiation after the date of the earthquake mobile sensors suspected of being contaminated are 9, 13, 22, 32, 39, 44. The cars suspected of low to no contamination are, 31, 45, 46, with an uncontaminated sensor 2 detecting a contaminated sensor 44 at the Palace Hills Hospital.

At the end of the timelapse contaminated mobile sensors 9, 13, 22, 32, 39, 44 are located in southern border of Cheddarford/PepperMill, northwest end of PepperMill, on the Scenic Vista Highway to the mainland, border of Cheddarford/PepperMill, southern part of SafeTown, and Cheddarford, respectively. I don't believe city officials should be worried about these parked cars. So long as they were in a non-contaminated area, readings demonstrated normal background radiation. Implying that though contaminated, do not affect their surroundings.

Statistical Analysis

The ANNOVA Test was performed in two ways. One comparing the mean radiation of static to mobile sensors. Another comparing mean radiation of sensors to self, before and after estimated earthquake date of April 8th (*e.g. static readings during first 60 hours, static readings during last 60 hours*).

Results indicate that there is a significant difference in means for both cases. In other words, some static and mobile sensors have significantly different values and sensors of the same type

Model Construction

Due to the task of the project being to better understand the radiation contamination's severity and current spread, any machine learning used in this analysis was unsupervised. Thus, splitting the data for training and test sets was not necessary for analysis.

Isolation forest was used, also the data does not need to be split. Isolation Forest does not require splitting the data into training and test sets as traditional supervised learning algorithms do. Isolation Forest is an unsupervised machine learning algorithm designed for anomaly detection. Further, it was calculated what percentage of the data consisted of these anomalies.

The purpose was to support that radiation spikes observed from sensor warm-up did not consist of more than 30% for each type of sensor. Essentially demonstrating the reliability of each type of sensor.

Static sensor ID	Percentage of Anomalies
1	4.972337
4	4.977025
6	4.993171
9	4.986227
11	4.988541

Table 5: Preview for percentage of static sensor readings that are considered anomalies by the Isolation Forest mode

Mobile Sensor ID	Percentage of Anomalies
1	21.414914
2	17.230706
3	10.778914
4	11.304348
5	15.369360

Table 5: Preview for percentage of mobile sensor readings that are considered anomalies by the Isolation Forest model

For both average radiation by date and by hours, the final 12 hours of static sensor readings are higher than the average from earlier timestamps. This implies that at the date or hour of the rise in sensor readings, it would be reliable to assume that radioactive contamination occurred. These trends are observable in *Figures 8 and 9*. These trends hold true for maximum values by date and hour as well, please refer to the folder “generated_pictures” as part of my submission to view extra visuals.

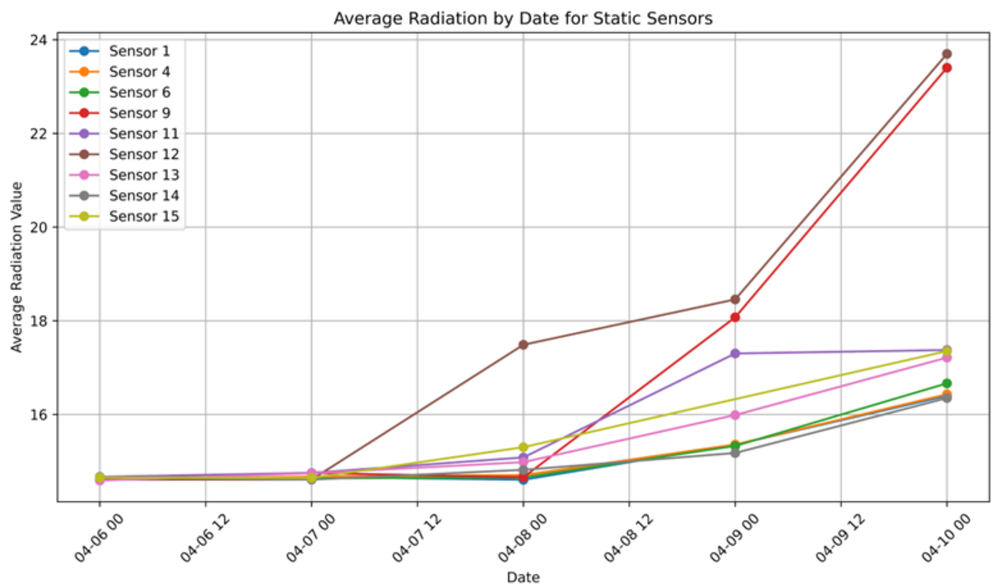

Figure 8: Plot of average static sensor reading over day

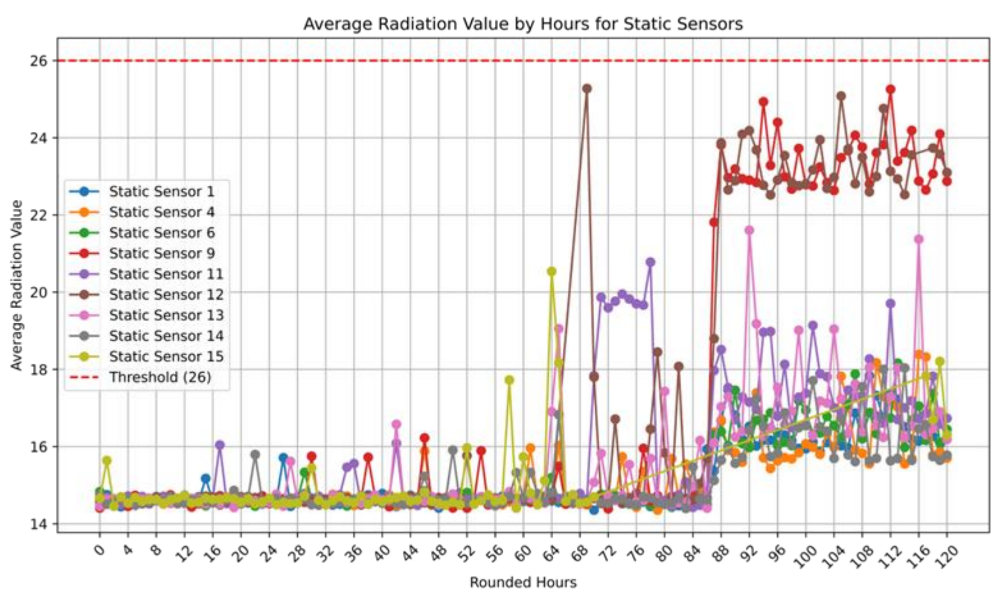

Figure 9: Plot of average static sensor reading over hours

Discussion

Regarding contamination zones, it's recommended that 2 additional static sensors be placed. Specifically, one at the border of East Parton and Cheddarford.

another at the border of Safetown and Cheddarford. These are the areas identified to be the most severely affected areas by the radiation leak.

Based on analysis, I think that mobile sensor data is a complement to the static sensor data. Without the mobile sensor data and just mean radiation readings, one could mistakenly assume that there was a contamination all the way in Palace Hills when it was actually a contaminated mobile sensor in that area artificially raising average radiation values.

Additionally, neither mobile nor static sensors consist of a lot of warmup spikes, a majority of the data is valid and surprisingly the mobile sensors can sometimes be just as reliable as the static ones. Quite remarkable, considering that they're homemade.

The data posed a temporal and locational challenge, it was quite difficult to keep all the various temporal and mapping plots organized in my head for optimized relation to one another.s

Both sensor types share the same weakness if warm-up spikes, but that is easily addressed by removing or filtering the data to exclude those warm-up values.

Conclusion

Analysis has demonstrated that there insn't a huge risk to public safety. Other than the hotspots show in *Figure 7*, background radiation throughout St. Himark appears to remain unchanged. Even the, the hotspots are not significantly higher than the upper threshold of what is considered to be a safe amount of exposure to ionizing radiation.

I don’t see it necessary to evacuate any of the neighborhoods, I think the best course of action would be to make a public announcement to avoid the contaminated zones until readings show normal background radiation levels again.

I believe it would be necessary to warn the owners of these contaminated cars so that they may be responsibly cleaned or disposed of.

References

1. Russi, E. (n.d.). *Emmanueliarussi/datasciencecapstone*. GitHub.
2. *How to use a geiger counter in the USA and APAC in 2020:* . ECOTEST. (2020, March 30).
3. Murray, A. (2017, January 23). *How much ionising radiation is dangerous?* research.csu.edu.au.
4. Parlante, N. (2012, July). *Nick's Geiger Counter Page*. cs.stanford.edu.
5. Safecast. (n.d.). *Radiation Levels Map*. map.safecast.org. https://map.safecast.org/
6. Thomas Hainke, Dipl. Inform. Univ. (2005). *Convert Nanosievert per second to nanosievert per hour (NSV/s to NSV/h):*. Convert Measurement Units.com.
7. Dalton, D. (2023). Data Science Captsone [COMP 4449]. University of Denver.