Water Utility

Wen Ye, Gautam Agarwal, Bryan Jin

October 2020

1 Introduction

Much of Madison's water main infrastructure dates back to World War II and the immediate postwar era, when pipes were made of a less durable material called spun cast iron.(reference needed) As the pipes start to age, repair work for these pipes has increased drastically for the City of Madison Water Utility Crew, especially in the winter. Meanwhile, PILOT (Payment in Lieu of Taxes) made by the Madison Water Utility to the city of Madison is increasing over time because its amount is determined by the value of the utility's infrastructure and area property tax rates. As a result of water main repair work, PILOT amount would get greater. We care because PILOT is part of the water bill and every resident in Madison who uses water pays it.

Hence, we want to address the question: **what are the factors that influence the number of water main breakages and how do they affect it?** By answering this question, we hope to build a predictive model for the number of water main breakages in the city of Madison.

As a separate topic, we also looked at water usage over the last two years. Specifically, we wanted to know: how has water usage changed from 2019 to 2020, possibly due to tumultuous events such as the pandemic and the subsequent lockdown?. Ideally, we would create a model that explains what factors affect water usage and use this model to predict water usage for the upcoming year.

2 How Does Water Usage Change from 2019 to 2020?

Figure 1 compares total water usage between 2019 and 2020. The total water usage between 2019 and 2020 is very similar until around June 24. In particular, Wisconsin's stay-at-home order, which was announced around March 24, did not seem to have a noticeable effect on total water usage.

On June 24, 2020, the total water usage plunged, and afterwards, the water usage in 2020 seems to exhibit a very different behavior from both the water usage in 2019 and the water usage in 2020 prior to June 24.

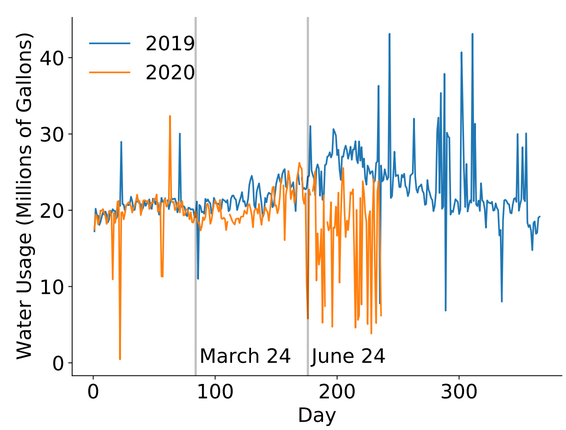

Figure 1: Total Water Usage, 2019 vs. 2020

Figure 1 poses two questions: why does the water usage level become erratic starting from late June 2020? Does water usage for specific classes (rather than just total water usage) show the same patterns?

Figure 2 attempts to address the second question. The plots show that a high variance pattern appears in each water usage category after June 24, 2020.

By plotting each category separately and with its own scale for the y axis, we highlight the fact that the pattern of increased variance after June 24 occurs in each category. The drawback of this approach is that visually comparing the line graphs with each other can lead to misleading results. For example, the industrial water usage graph seems to have greater variance across the year compared to the other categories, but this is because the y axis range for Industrial is much smaller than for categories such as Total or Residential.

We also observe patterns in the Figure 2 plots that were not discernible when we only looked at total water usage. For example, commercial water usage does show a decrease around March 24, which makes sense: the rise of Covid correlates with a decrease in commercial water usage. Interestingly, public authority water usage increases right after March 24.

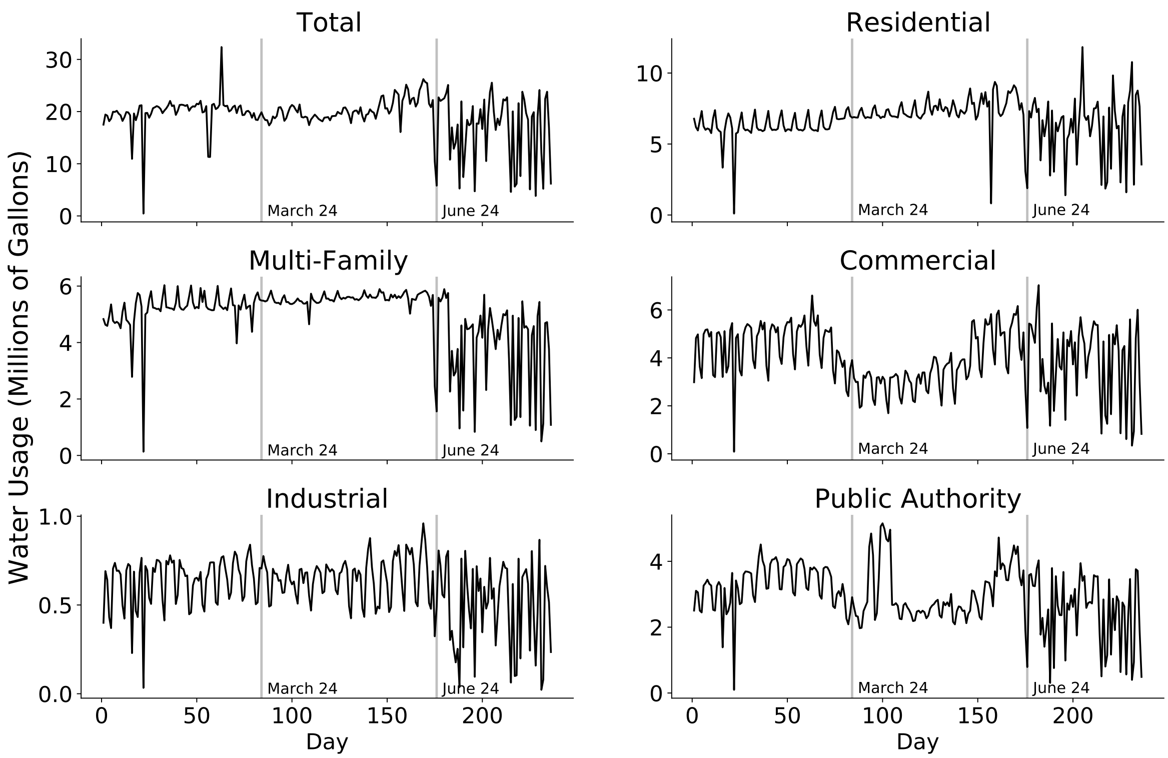

Figure 2: 2020 Water Usage by Category

3 What Factors Influence Water Main Breakages?

3.1 Temperature

Figure 3 shows the total number of water main breaks by month from 1980 to 2020. The total number of water main breaks substantially increases in the months of December, January and February which is the winter season. Low temperatures and other winter conditions may contribute to this apparent difference in the number of breaks between different months of the year. From this plot, we decide to examine more closely the relationship between temperature and the number of breaks.

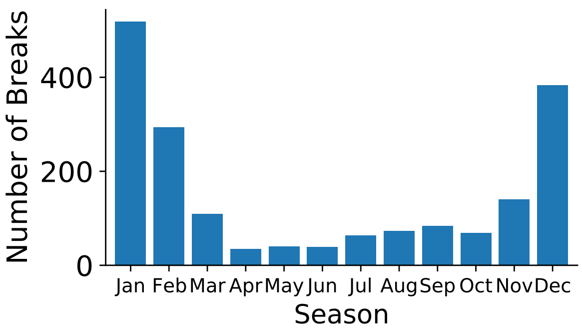

Figure 3: Number of Water Main Breaks by Month

Figure 4 compares the minimum temperature and the number of water main breaks for each year (again from 1980 to 2020). The minimum temperature of each year is scaled by 5 to make the peaks and dips more visible. Generally, peaks in the number of breakages correlate with dips in the minimum temperature.

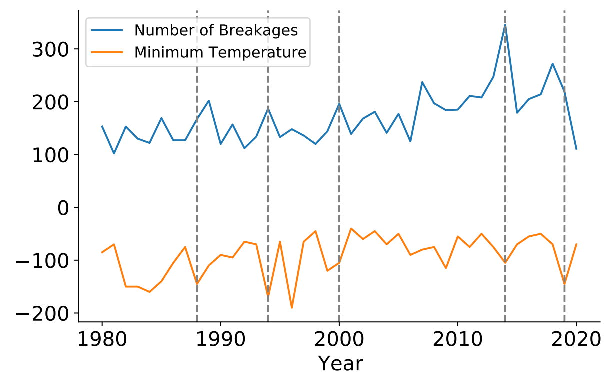

Figure 4: Comparison Between Min Temp and Breaks

Figure 5 gives a more detailed comparison to tease out the influence of temperature and season. If we look at 15 Fahrenheit for January and March 2018, the month of March has 1 break on that day while the month of January has 7 breaks on the day with the same temperature. Here, we can infer that season is more influential in determining the number of breaks than the minimum temperature of the day itself. Meanwhile, examining the day of Jan 7th and Mar 18th, there's a significant sudden increase in temperature and the number of breakages went up alongside. This suggests that a sudden increase in temperature which could cause the soil to thaw too quickly is a contributing factor to breakages.

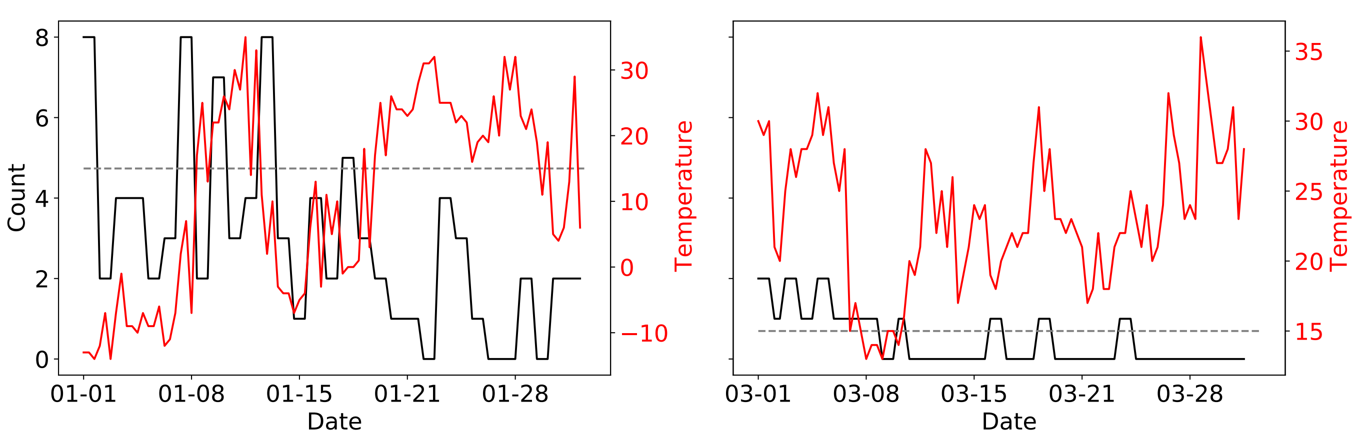

Figure 5: Jan and March comparison

3.2 Pipe Depth

Figure 6 describes the distribution of water main breaks with respect to the depth at which the main's pipe was laid. The cumulative frequency is 0.1 for pipes with depth at most 5 feet and 0.9 for pipes with depth at most 7 feet. This means 80 percent of water mains that break are laid 5-7 feet deep.

This may be because:

1. It is most common for pipes to be laid at this depth

2. Pipes that are laid at this depth are the most vulnerable to break

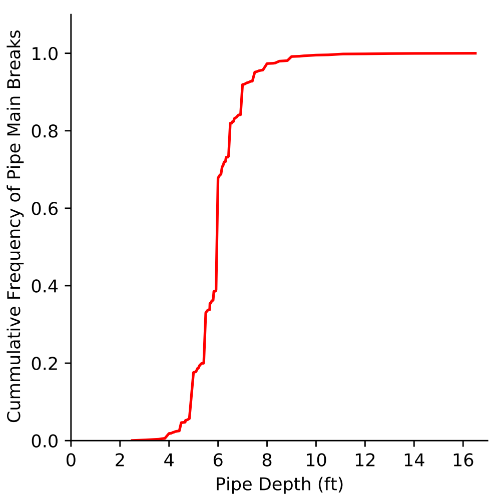

Figure 6: Cumulative Frequency Distribution of Main Breaks by Pipe Depth

3.3 Pipe Size

Figure 7 describes the distribution of water main breaks with respect to the size of the main's pipe. Most pipes that break are 5-6 metres long.

This may be because:

1. It is most common to use pipes of this length
2. Pipes of this size are the most vulnerable to break

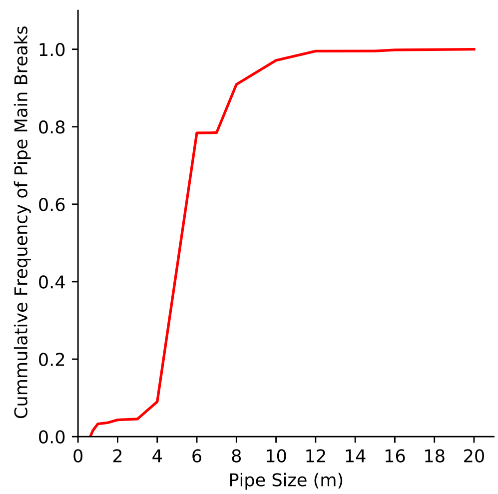

Figure 7: Cumulative Frequency Distribution of Breaks By Pipe Size

3.4 Location

Figure 8 shows the geographic distribution of water main breaks for every decade. The number of breaks away from the inner city/downtown area have started to increase between 2000-2009 and 2010-2020. One possible explanation is the increasing population outside of the downtown area in the past 20 years. We also noticed that the number of breaks increase every decade.

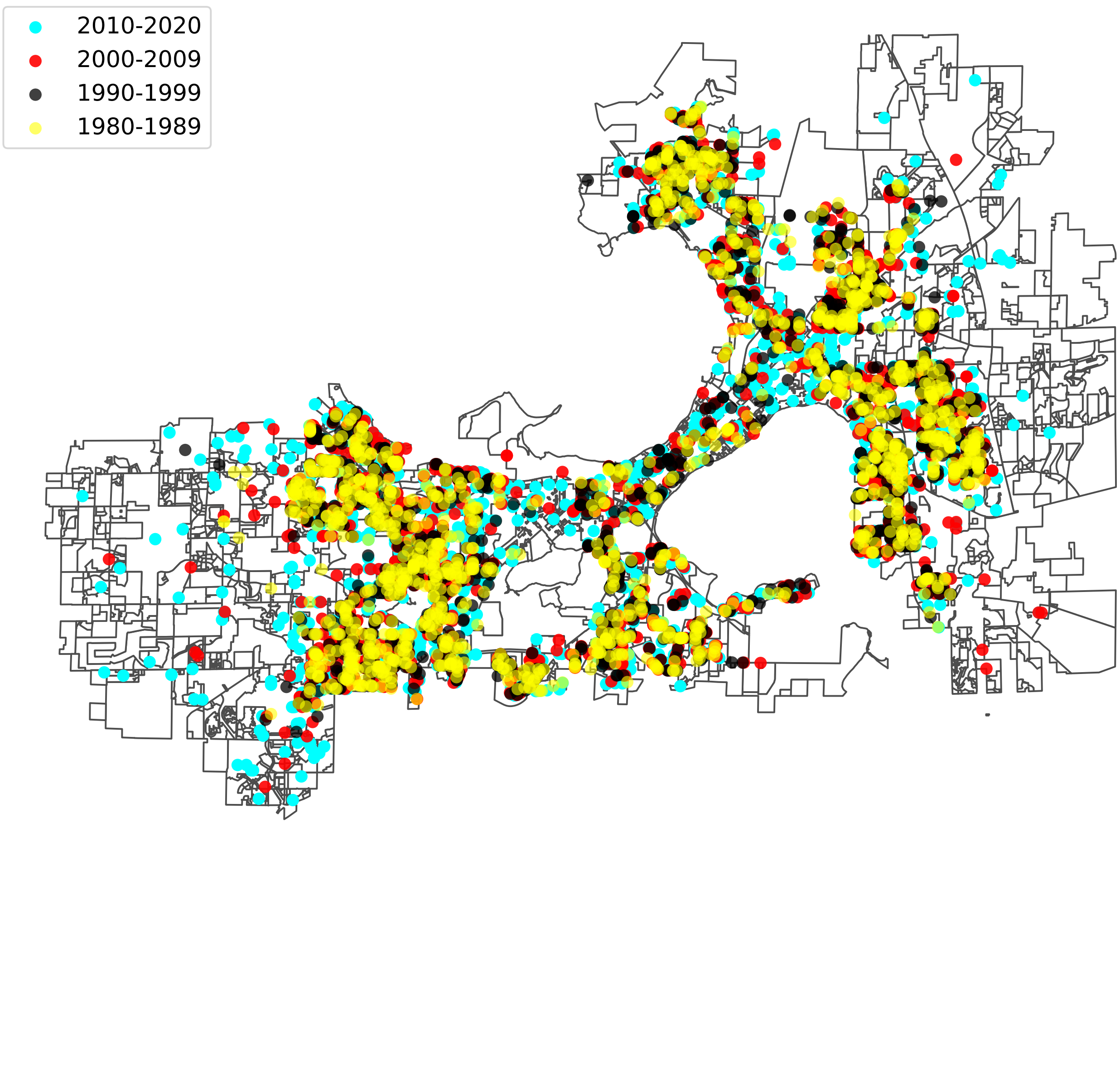

Figure 8: Pipe Main Breaks In City of Madison Every Decade

3.5 Soil Type

For the soil type variable, many water mains had a singular soil type: clay, sand, gravel, etc. Other water mains had two soil types: clay and sand, sand and gravel, etc. A few water main breaks even had three soil types: clay, sand, rock or clay, gravel, rock. The most common soil types were clay, sand, gravel, dirt, and rock.

To get a better understanding of the soil type variable, Figure 9 shows how often the most common soil types appeared on their own vs. in combination with other types. For example, the leftmost bar shows that clay often appeared as the only soil type, while the rightmost bar shows rock mostly appeared with other soil types (such as sand and rock or clay and rock).

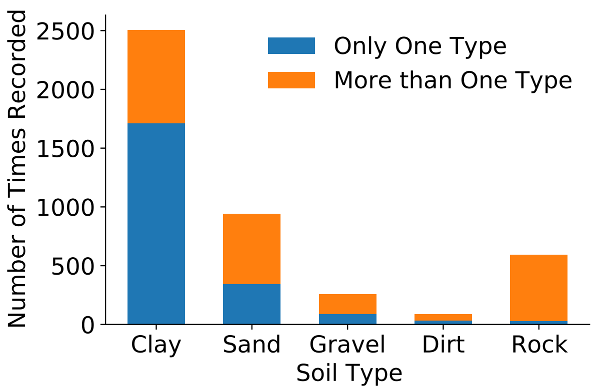

Figure 9: One Soil Type vs Multiple Soil Types

Figure 10 focuses on the instances of the soil type variable where two soil types are recorded. The leftmost bar shows that when clay appears as one of two types, the other type is most likely either sand or rock. Note that the three bars on the right indicate that when gravel, dirt, or rock appear in a pair, they are always paired with clay or sand.

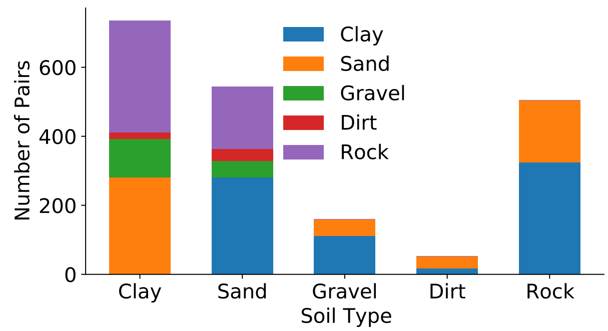

Figure 10: Common Pairs of Soil Types

Overall, clay and sand seem to be the main soil types. Gravel, dirt, and rock seem to function as secondary soil types that can be associated with either clay or sand. To better understand the soil type variable, we may investigate how soil type relates to other variables such as location or pipe dimensions.

4 How to use Temperature to Predict the Number of Breakages?

4.1 Building the model

Before building a model to predict the number of breakages, we first want to know if the four seasons actually have distinct patterns in terms of the number of breakages. And Figure 11 shows that they indeed do. The range of temperature varies as well as the maximum number of breakages reached. The number of breakages is always an integer, but to reduce overlapping and have a better sense of where all the points lie, we added a little noise with a normal distribution centered at 0 and standard deviation of 0.1.

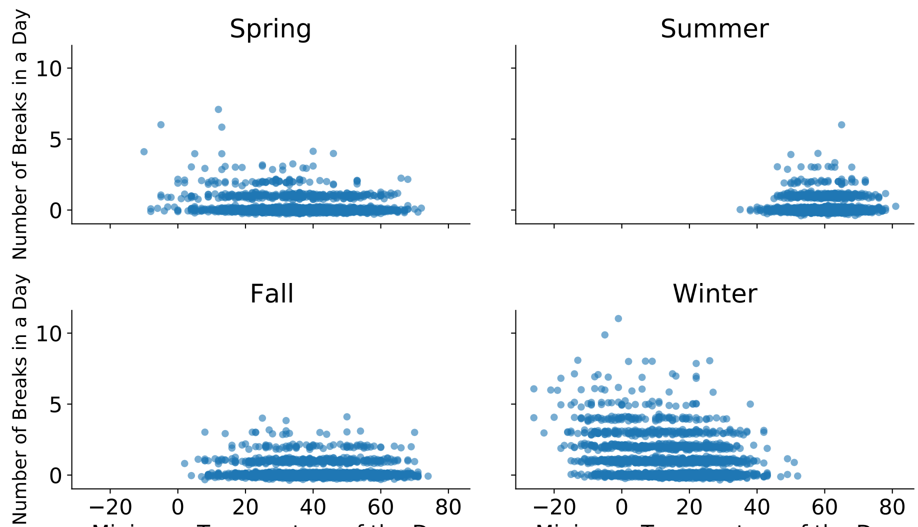

Figure 11: Number of Breakages by Soil Type categorized by Season

Fig 12 shows the number of breakages by soil type in each season. The soil types are stacked in the order of the most number of breakages to the least. Clay has the highest number of breakages and Mud/Dirt has the lowest. Relative to Summer, breaks in rock/stone are up by 8.56% during winter, whereas breaks in gravel are only up by 0.54%. For all the soil types, rock/stone has the greatest increase in number of breaks and gravel has the lowest. It could be inferred that gravel is more resistant to Winter than all the other soil types and implementing new water main pipes in gravel in the future could reduce the number of breakages.

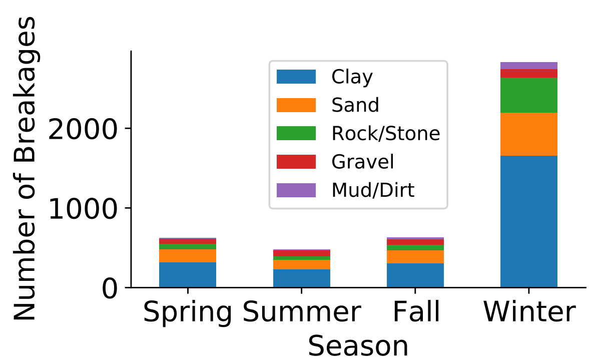

Figure 12: Percentage of Breaks by Season

4.2 Visualizing the model

Figure 13 gives the visualization of the prediction model given the minimum temperature of a day and what season we are in. This is a second degree polynomial regression. Season is fed in using One Hot Encoding.

The minimum temperature of the day is better at predicting the number of breaks than the maximum temperature of the day by 3% in terms of explained variance score. It makes sense because most of the breaks happen in the winter and min temperature is probably more representative of the weather that day than max temperature. We tried feeding in month of the year instead of season using One Hot Encoding and that changed the explained variance score by little. So we think splitting it into 12 months is probably too specific and season is a nice enough split. We also fed in the change in temperature from the previous day to the current day, and surprisingly, it also didn't really improve the model. In the end, it turned out that the combination of minimum temperature and season is the best at predicting the number of breaks on a given day explaining 26% variance.

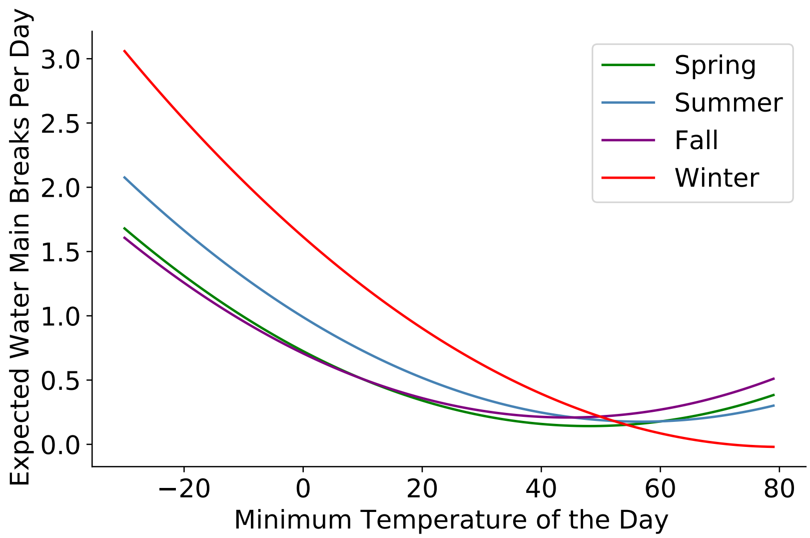

Figure 13: Prediction of the Number of Breakages given Season and Min Temp

5 How to Predict the Number of Breakages for Different Pipe Sizes and Depths?

Figure 14 predicts of the number of breakages for January 2021 with the pipe size in metres and pipe depth in feet. Using the data on the pipe mains broken in the month of January during the past 20 years, this model forecasts the number of pipe breaks in the future years with their sizes and depths. From the figure below, we can infer that during this month we can expect around 80 pipes to break and a majority of these pipes will be between 5 feet and 8 feet deep or between 4 metres and 8 metres long.

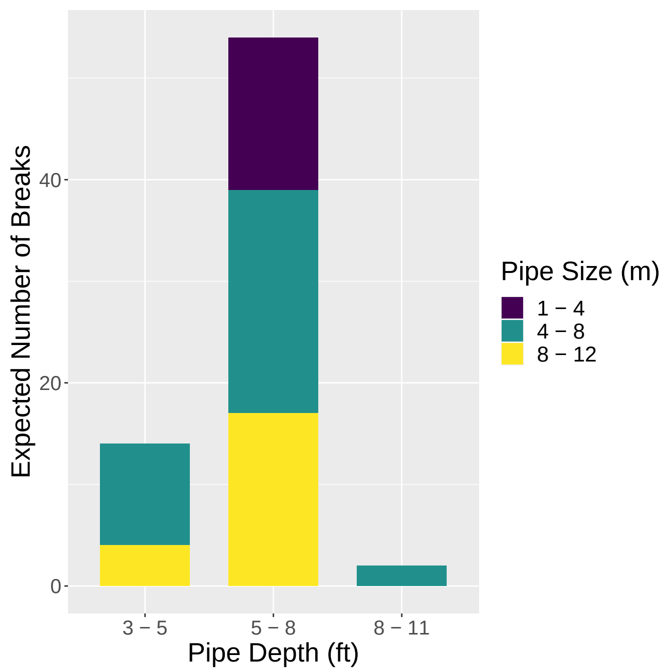

Figure 14: Predicting Water Main Breaks by pipe size and depth for January 2021

6 Highlight some Conclusions:

- low temperatures and sudden increases in temperature cause more pipes to break
- Season is a contributing factor to the number of breakages
- most pipes that break are laid 5-7 feet deep and are 5-6 meters long
- clay and sand are primary soil types while gravel, dirt and rock are secondary soil types
- pipes laid in gravel are more resistant to breaking during winter, and pipes laid in rock/stone are more vulnerable to breaking during winter

TODO: Full conclusion