Water Utility

Wen Ye, Gautam Agarwal, Bryan Jin

October 2020

1 Introduction

TODO: Add text

2 Water Usage

TODO: Add previous plots

3 Exploring Factors of Water Main Breaks

3.1 Temperature

Figure 1 shows the total number of water main breaks by months from 1980 to 2020. The total number of water main breaks has a substantial increase in the month of December, January and February which is in the Winter season. The low temperatures in the Winter could contribute to the number of breakages significantly.

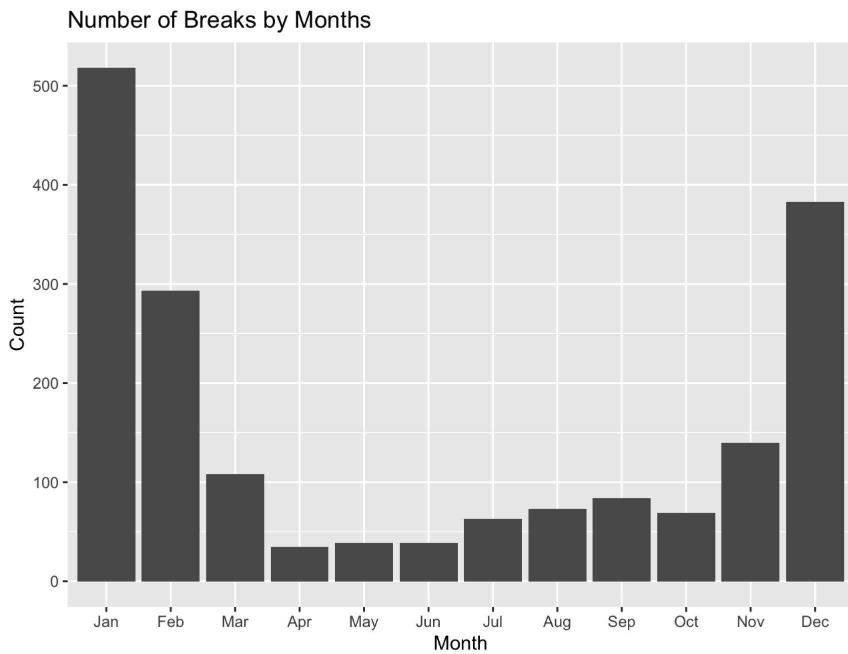

Figure 1: Number of Water Main Breaks by Months

Figure 2 shows the comparison between the minimum temperature of the year and the number of water main breakages in that same year. The minimum temperature of each year is scaled by 5 to make the peaks and dips more visible. In general peaks in the number of breakages correlate with dips in the minimum temperature.

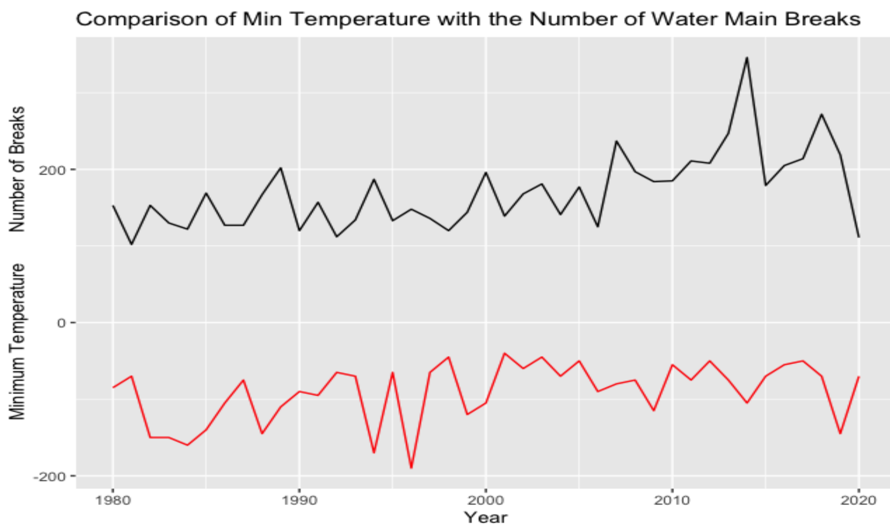

Figure 2: comparison between min temp and counts

Figure 3 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. Examining the January plot itself, at the point of Jan 7th, there's a significant sudden increase in temperature and the number of breakages went up alongside. This could be a sign that the sudden change in temperature is also a contributing factor in determining the number of breakages.

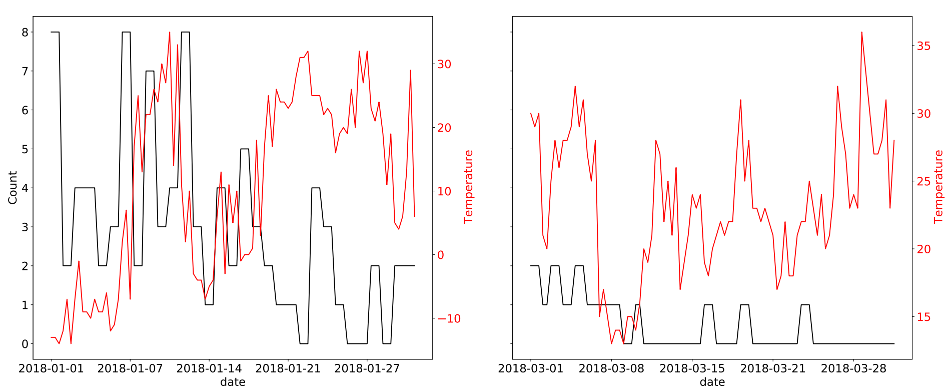

Figure 3: Jan and March comparison

3.2 Pipe Depth

Figure 4 describes the distribution of water main breaks with respect to the depth the main's pipe has been laid. It can be inferred that most pipes that break are laid 5-7 feet deep.

This may be because:

1. It is most common to use pipes at this depth
2. These pipes are laid in particular areas not suitable for them
3. The soil strata at this depth is not suitable

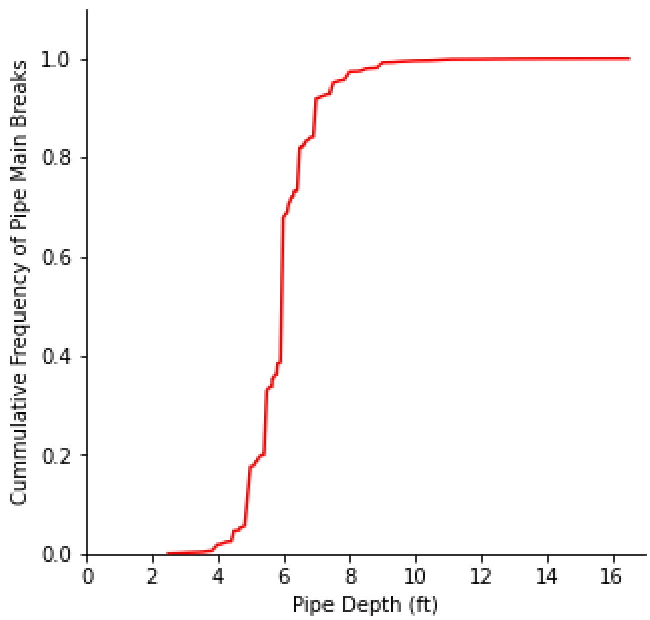

Figure 4: Cumulative Frequency Distribution of Main Breaks by Pipe Depth

3.3 Pipe Size

Figure 5 describes the distribution of water main breaks with respect to the depth the main's pipe has been laid. It can be inferred that 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 length are laid in particular areas not suitable for them
3. Pipes of this length are laid in between unsuitable soil

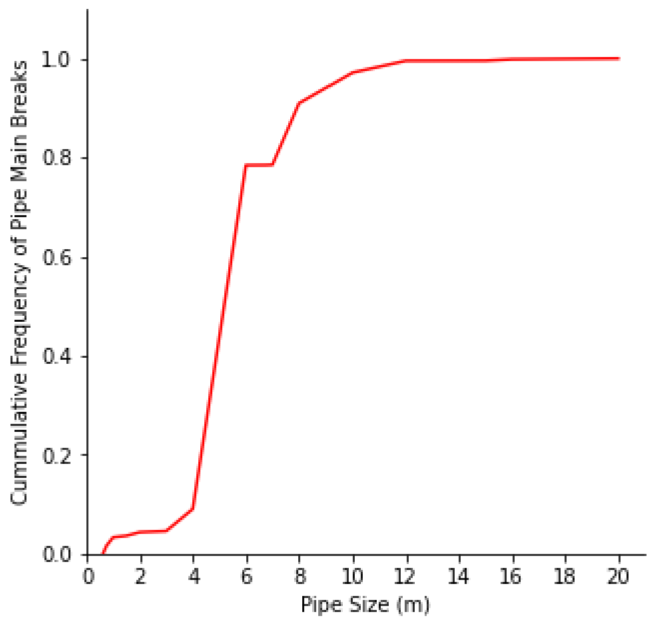

Figure 5: Cumulative Frequency Distribution of Main Breaks By Pipe Size

3.4 Location

Figure 6 describes the distribution of water main breaks on map for every decade. From the The number of breaks away from the inner city/downtown area have started to increase between 2000-2009 and 2010-2020. One possible parallel to this is the increasing population in the past 20 years. We also noticed that the number of breaks increase every decade.

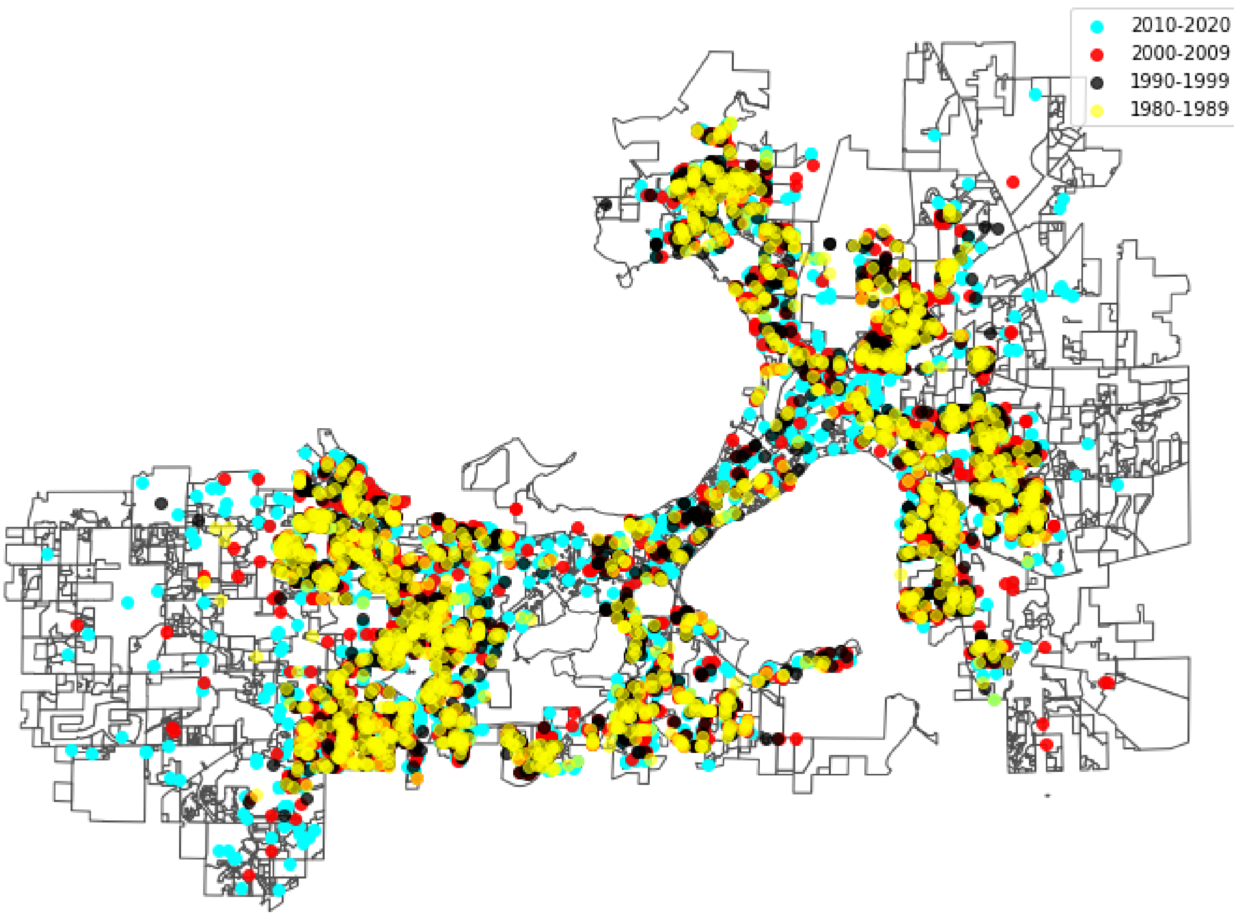

Figure 6: Pipe Main Breaks In City of Madison Every Decade

3.5 Soil Type

[**BEGIN Bryan's 10/19 results**] For the soil type variable, some water main breaks had a singular soil type: clay, sand, gravel, etc. Other water main breaks had two soil types: clay and sand, sand and gravel, etc. Some 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 7 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).

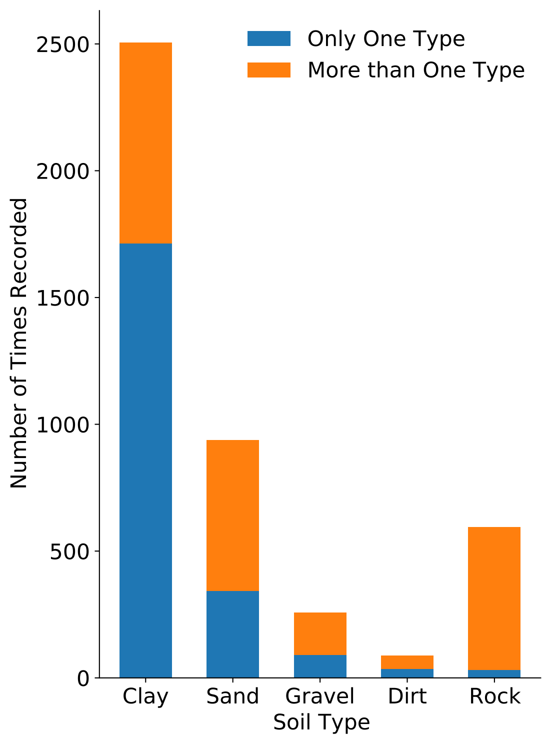

Figure 7: One Soil Type vs Multiple Soil Types

Figure 8 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.

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. [**END Bryan's 10/19 results**]

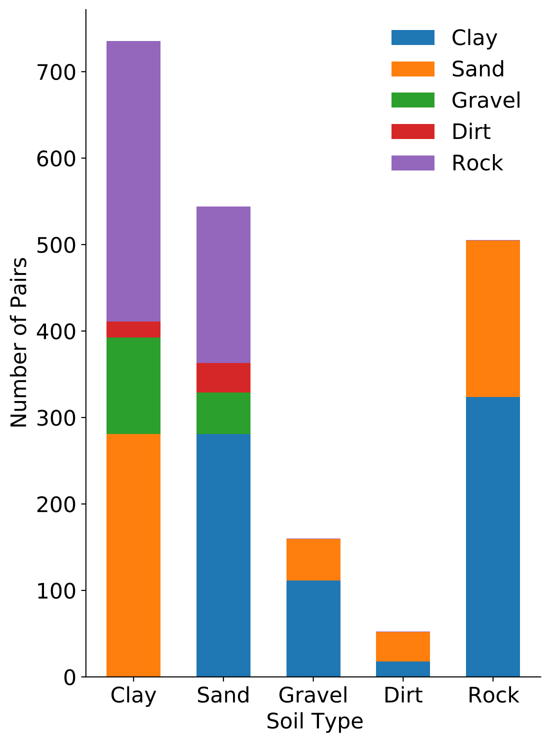

Figure 8: Common Pairs of Soil Types

4 Building the model

Before building a model to predict the number of breakages, I first want to know if the four seasons actually have distinct patterns in terms of the number of breakages. And Figure 9 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, I added a little noise with a normal distribution centered at 0 and standard deviation of 0.1.

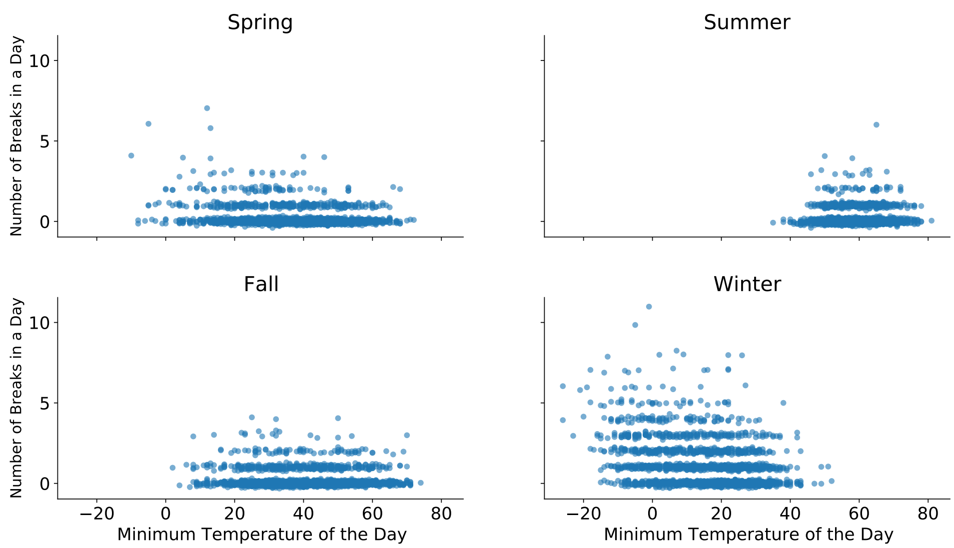

Figure 9: Break Pattern by Season

[**Begin Wendy's Oct 19th results**]Fig 10 shows the percentage of breaks by soil type in each season. It seemed that the order from the highest to the lowest percentage by soil type is the same across the seasons so I intentionally stacked them in order. Clay has the highest percentage of breaks and Mud/Dirt has the lowest percentage of breaks. From this plot, it can be inferred that when predicting a break, it will have a higher probability that the pipe broken is in Clay. (SideNote, I tried different positions of the legend and this position seems to make the most sense. I also thought about teasing out the difference between pipes in clay is more likely to break and there are simply more clay areas in Madison than other soil types. We'll need to find datasets on that if possible

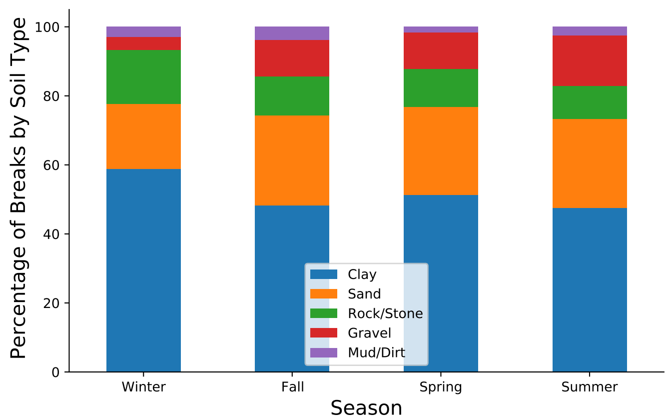

Figure 10: Percentage of Breaks by Season

[**END**]

5 Visualizing the model

Figure 11 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.

[**Begin Wendy's Oct 19th results**]The minimum temperature of the day is better at predicting the number of breaks than the maximum temperature of the day by 3%. 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. I tried feeding in month of the year instead of season using One Hot Encoding and that changed the explained variance score by little. But it gives 12 lines in the model which is very messy. So I think splitting it into 12 months is probably too specific and season is a nice enough split. I 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. [**END**]

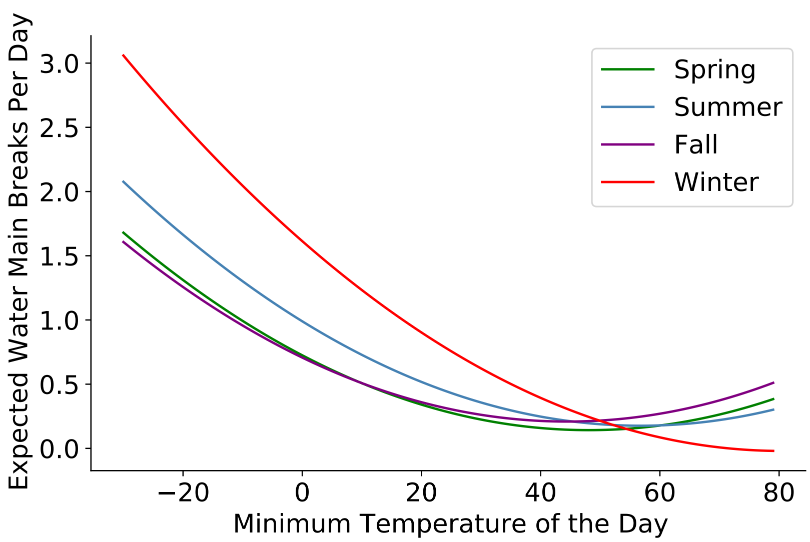

Figure 11: Prediction of the Number of Breakages given Season and Min Temp

[October 19th, Gautam Begin]

Figure 12 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.

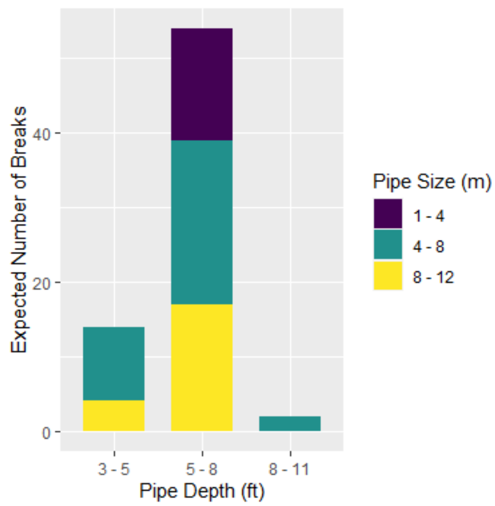

Figure 12: Predicting Water Main Breaks by pipe size and depth for January 2021

[October 19th, Gautam End]