Water Utility

Wen Ye, Gautam Agarwal,Bryan Jin

October 2020

1 Introduction

TODO

2 Water Usage

TODO Add previous plots

3 Exploring Factors of Water Main Breaks

3.1 Temperature

[**Week05 Wen**]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.

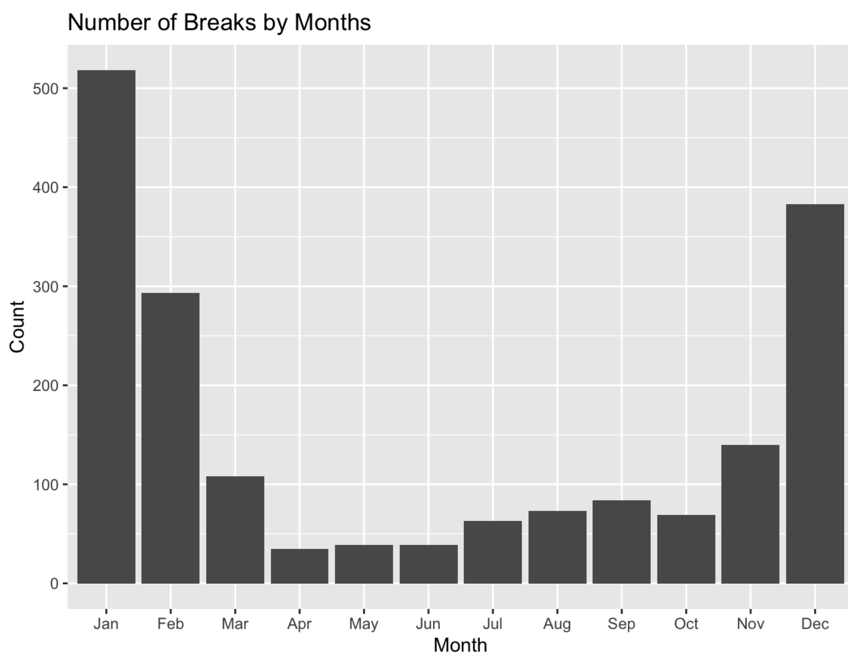

Figure 1: Number of Water Main Breaks by Months

[**Week05 Wen**]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.

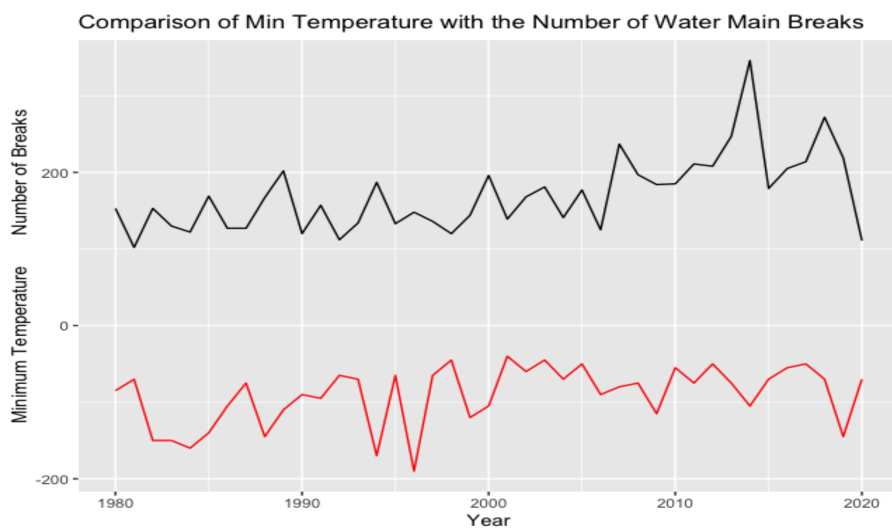

Figure 2: comparison between min temp and counts

[**Week06 Wen**]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.

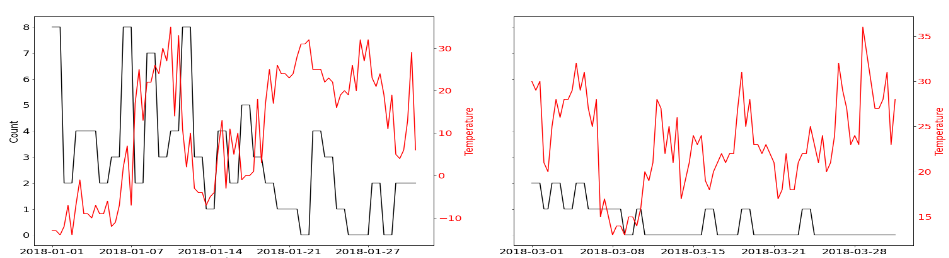

Figure 3: Jan and March comparison

3.2 Pipe Depth

[**Week06 Gautam**]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

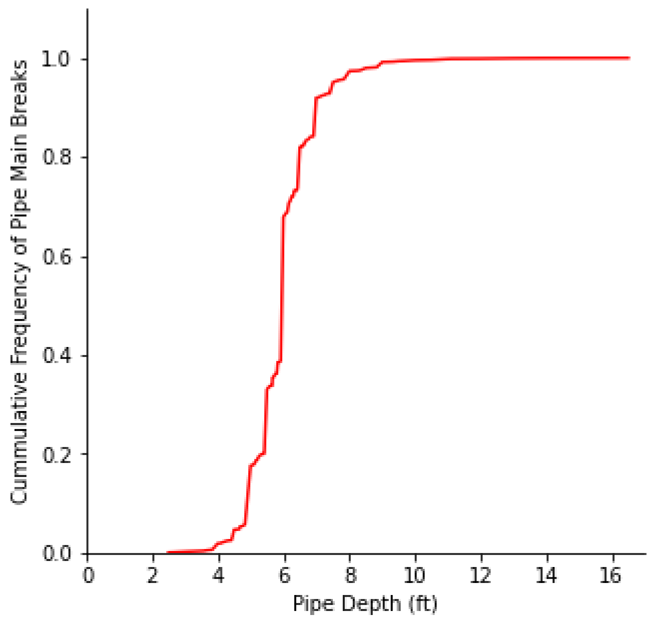

Figure 4: Cumulative Frequency Distribution of Main Breaks by Pipe Depth

3.3 Pipe Size

[**Week06 Gautam**]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

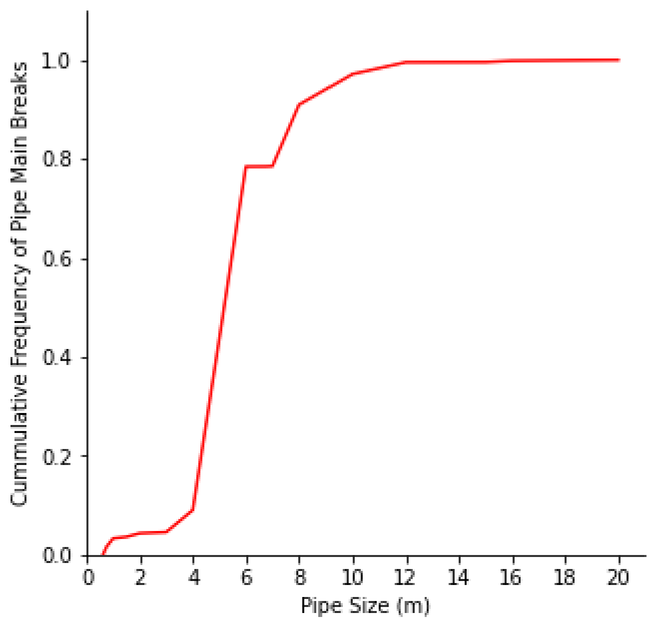

Figure 5: Cumulative Frequency Distribution of Main Breaks By Pipe Size

3.4 Location

[**Week05 Gautam**]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.

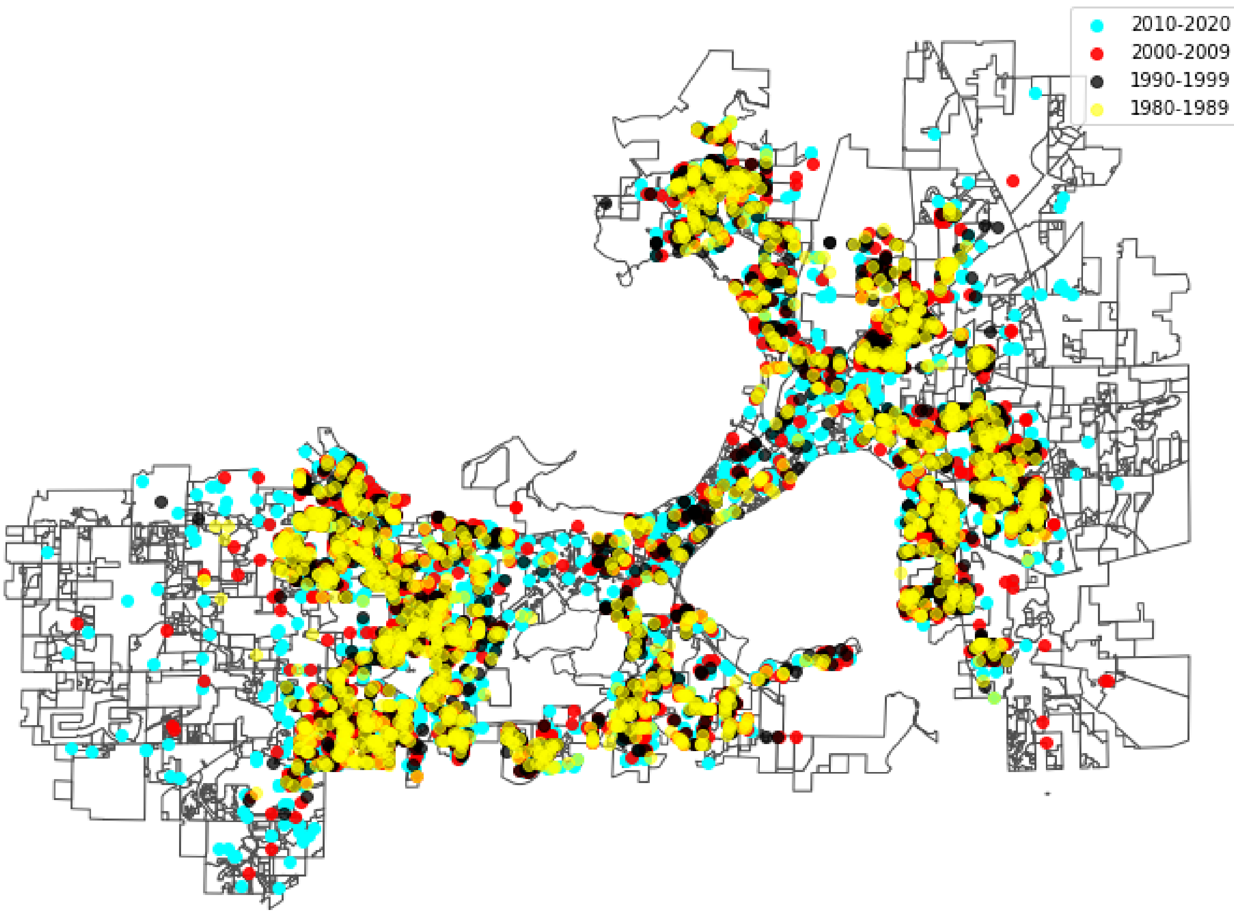

Figure 6: Pipe Main Breaks In City of Madison Every Decade

3.5 Soil Type

[**Week07 Bryan**]Figure 7 shows the distribution of the 5 most common atomic soil types that appear in the water main break dataset. In the dataset, some soil types involve one atomic type, like "clay" or "sand". Other soil types involve a combination of two or more atomic types, such as "clay, rock" or "rock, sand". For this visualization, we wanted to count how many times the most common atomic types appear in the dataset (so "clay, rock" adds one to the count of both "clay" and "rock"). This helps us better understand the soil type variable and whether soil type is a viable predictor variable for breaks. Looking at the distribution, we may consider boolean variables that each indicate presence of clay, sand, rock, or gravel.

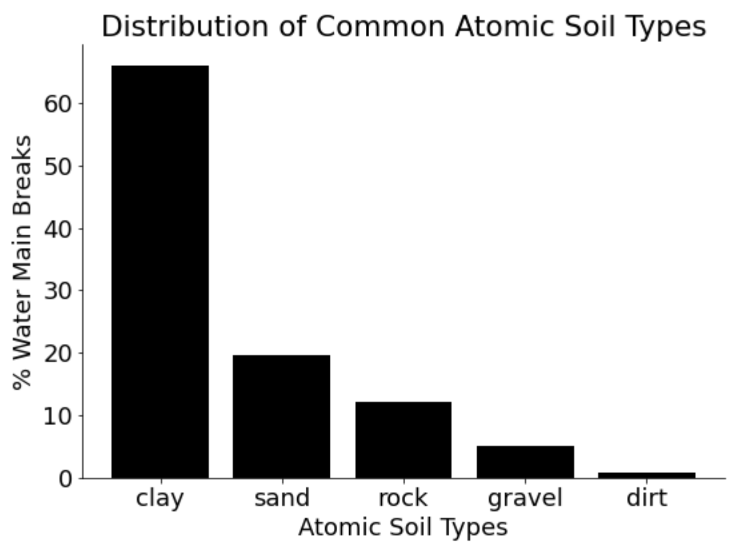

Figure 7: Soil Type and number of Breakages

4 Building the model

[**Week07 Wen**]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 8 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.

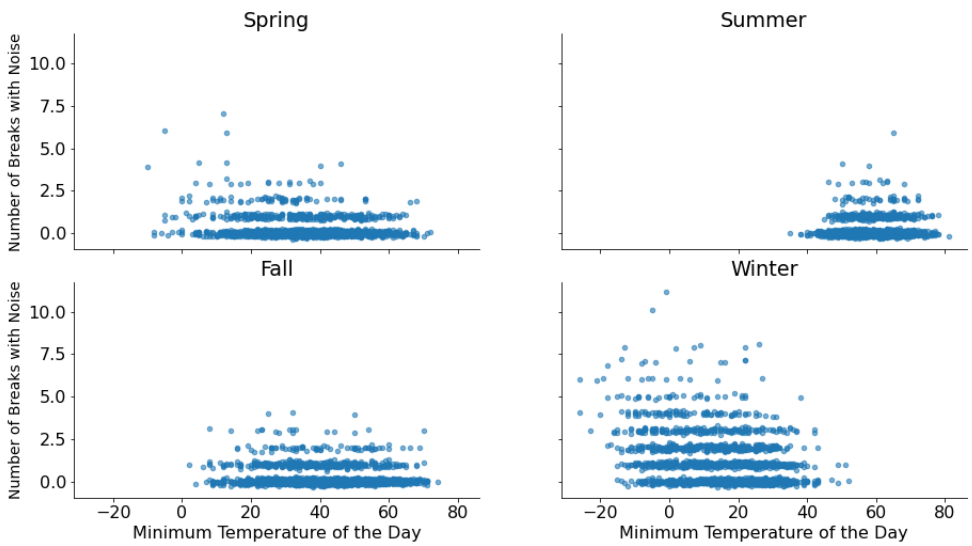

Figure 8: Break Pattern by Season

5 Visualizing the model

[**Week07 Wen**]Figure 9 gives the visualization of the prediction model given the minimum temperature of a day and what season we are in. This is a second order con-caved up parabola.

NOTE: This is incomplete. Only min temp and season is fed in. I think I also want to feed in the change in temp from the previous day. I want to try giving month of the year instead of season as input as well to see if the model does better in prediction.

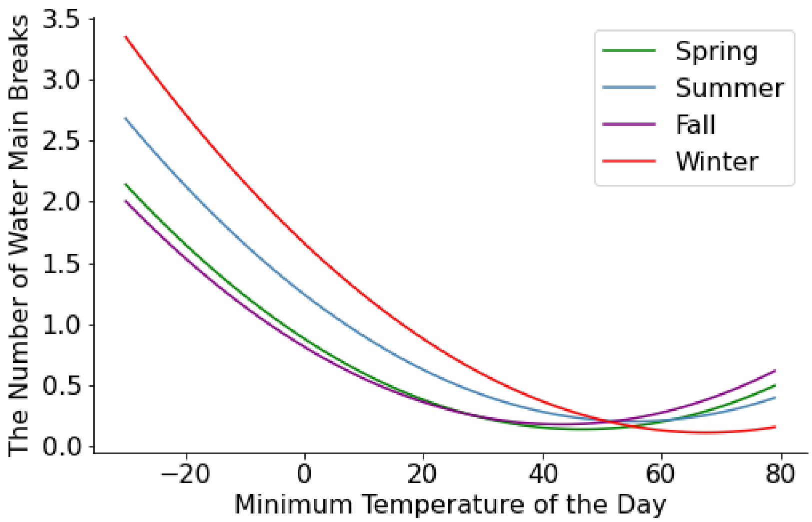

Figure 9: Prediction of the Number of Breakages given Season and Min Temp

[Week07 Gautam]Figure 10 gives the visualization of the prediction model for the number of breakages in the next decade, given the pipe depth in feet. The polynomial regression is a third order function.

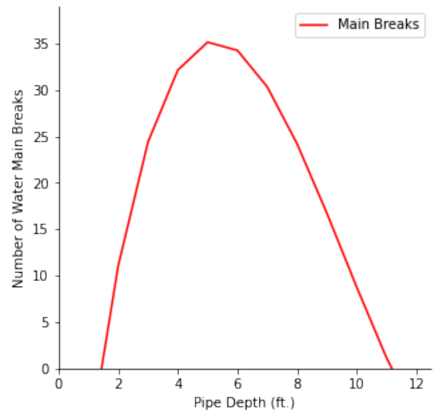

Figure 10: Prediction of Water Main Breaks by pipe depth, for the next decade

[Week07 Gautam]Figure 11 gives the visualization of the prediction model for the number of breakages in the next decade, given the pipe size in metres. The polynomial regression is a third order function.

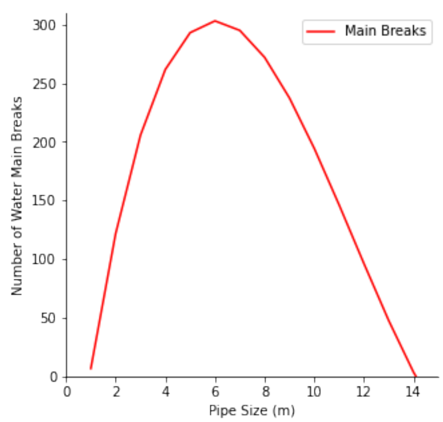

Figure 11: Prediction of Water Main Breaks by pipe size, for the next decade