

Eviction Filings

Part 2: Follow-Up to Time Series

Bliss Cohen

December 14, 2020

Background

This document continues the work started in the [Time Series](#) project (see [Time Series Exploration with Eviction Filings](#)). The goal of the project is to identify the best way to present data to help answer the following questions:

- How are eviction filings trending for a particular region?
- Is one region's filings higher or lower than another region?

Once the answers to the above questions are obtained, the next step is to identify suitable thresholds for initiating specific intervention.

After Guy and I talked on November 25th, 2020, I revised my approach to explore aggregated daily data, rather than monthly aggregations, without a time series lens. In other words, back to basics.

Objective

Obtain feedback on data wrangling and initial smoothing approaches

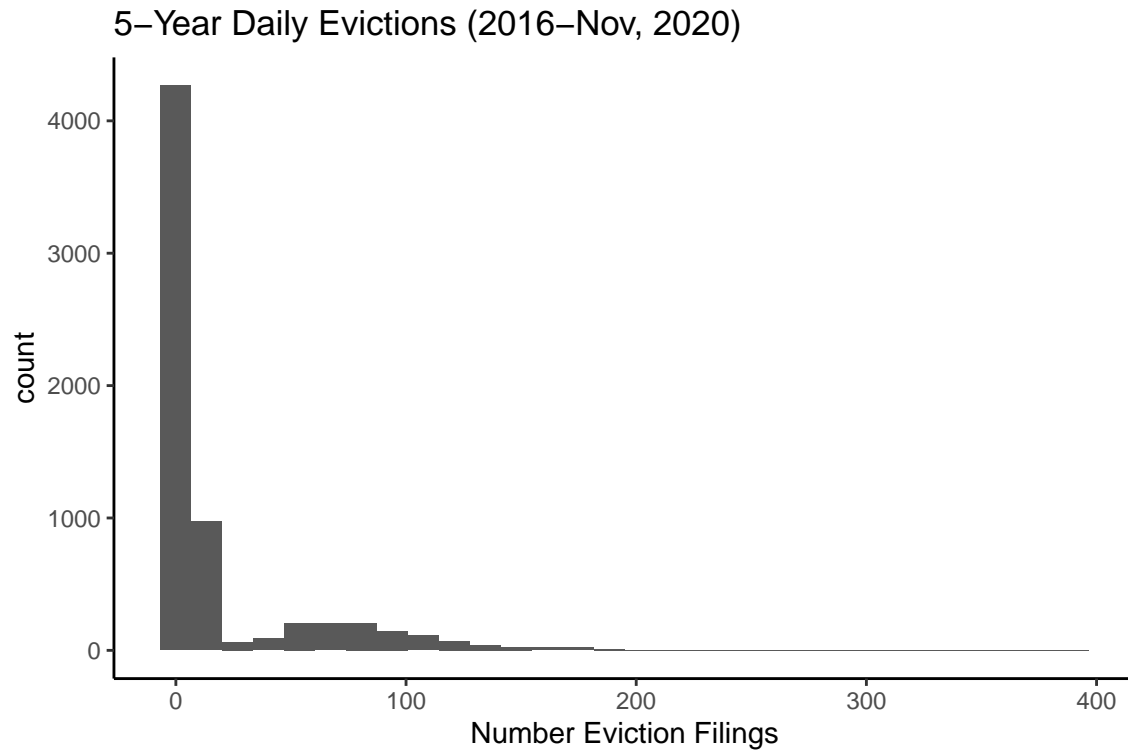
Read in Data

The following file aggregates the number of daily eviction filings from January, 2016 - November, 2020 across 12 regions of interest. The 'Est_Occ_Units' is an estimate of the number of occupied rental units for the region, as obtained from the U.S. Census American Community Survey.

```
## # A tibble: 6 x 7
##   New_Date   Month  Year   Day Region Num_evictions Est_Occ_Units
##   <date>     <dbl> <dbl> <int> <chr>         <int>         <dbl>
## 1 2016-01-04     1  2016     4  WI             77         775089
## 2 2016-01-05     1  2016     5  WI             59         775089
## 3 2016-01-06     1  2016     6  WI             52         775089
## 4 2016-01-07     1  2016     7  WI             54         775089
## 5 2016-01-08     1  2016     8  WI             50         775089
## 6 2016-01-11     1  2016    11  WI             61         775089
```

Examine Distribution of Eviction Filings

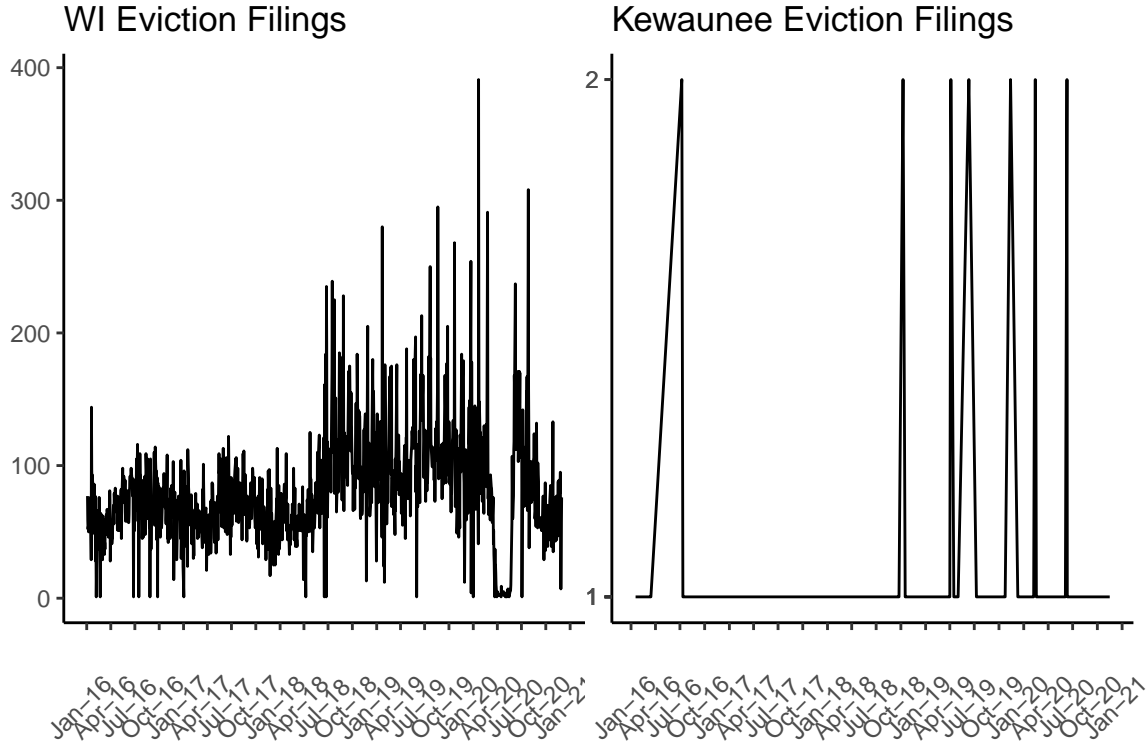
The clump on the left side of the histogram is due to smaller County-related filings while larger values belong to the State (WI). There are no '0' values or NA's.



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.00	1.00	3.00	19.09	10.00	391.00

Plot Filings by Day (Gut Check)

The plots below compare the most populated region - the entire state of Wisconsin - to the smallest region, Kewaunee County.



Fill in Daily Sequence with '0' Filings and Replot

The Kewaunee graph above seemed to flat line at '1' for months on end, leading me to suspect that there were many dates for which no evictions were filed. There are 365 days in a year. The tables below show the number of days in each year associated with a filing. It makes sense that the least populous region Kewaunee had very few filing days per year while the most populous region WI had filings on most days of the year.

Table 1: Kewaunee: Number of Days Evictions Were Filed by Year

Year	n
2016	11
2017	9
2018	16
2019	27
2020	28

Table 2: Wisconsin: Number of Days Evictions Were Filed by Year

Year	n
2016	261
2017	255
2018	258
2019	256
2020	232

I believe that missing days mean no evictions were filed on that day. In other words, number of filings = 0

rather than NULL or NA. I think it will be easier to compare regions if they share the same dates. As a result, I will fill in the daily sequence with 0 filings for missing days.

The tables show updates specific to WI and Kewaunee after completing the daily sequence.

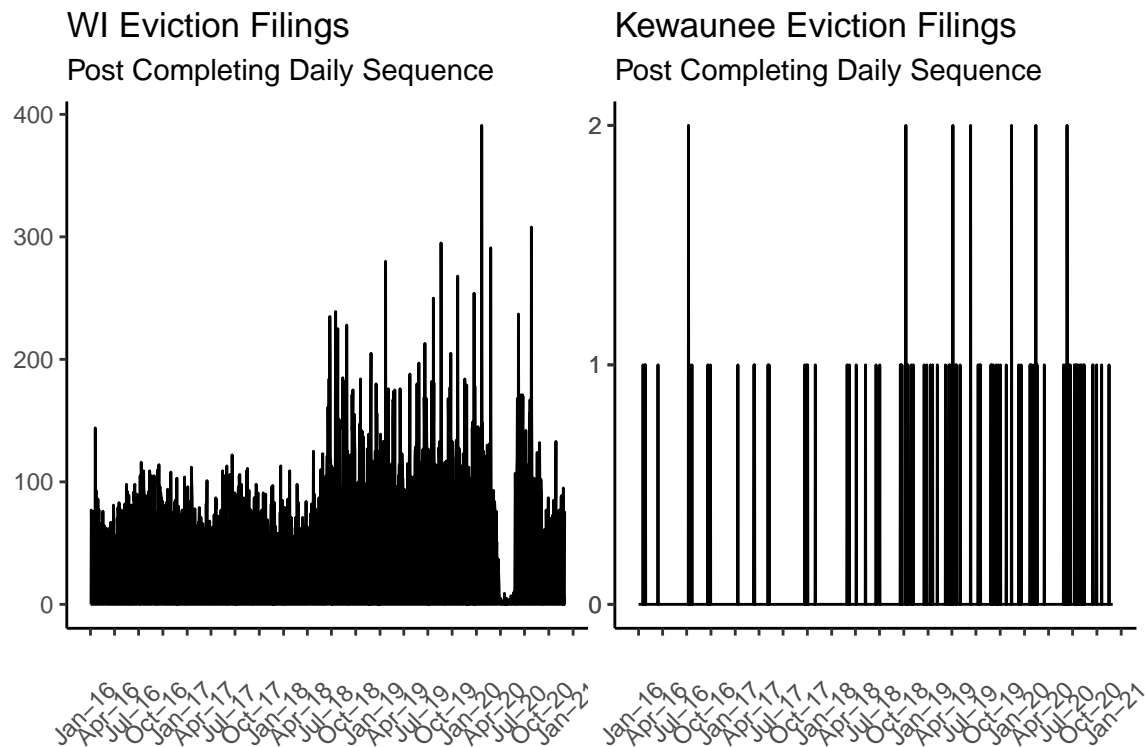
Table 3: Kewaunee: Each Day Now Has a Record

Year	n
2016	366
2017	365
2018	365
2019	365
2020	335

Table 4: Wisconsin: Each Day Now Has a Record

Year	n
2016	366
2017	365
2018	365
2019	365
2020	335

The WI and Kewaunee graphs are re plotted below after filling in the daily sequence. Now Kewaunee doesn't flat line at 1; if no filings were present, the baseline is 0.



Label Weekdays/Holiday Days

In order to identify trends, it makes sense to filter out days in which entering an eviction filing wasn't even possible in the first place. In other words, most likely, eviction filings are not entered on weekends and holidays. If true, including weekends and holidays - at least to some extent - may mask patterns.

There are 260 non-weekend days in a year (261 for leap years) and 17 official holidays. I need to label dates if they fall into these categories:

- Weekdays
- Holidays

A handful of records did have a filing date on the weekend. These weekend filings only had one eviction and all were at the State level (outside the Counties of interest). Thus, in general, evictions are not filed on the weekend.

Table 5: Number of Weekend Filings

New_Date	Region	Weekday	Num_evictions
2016-02-06	WI	Saturday	1
2016-02-21	WI	Sunday	1
2016-06-26	WI	Sunday	1
2016-07-16	WI	Saturday	1
2016-08-27	WI	Saturday	1
2016-09-25	WI	Sunday	1
2018-04-08	WI	Sunday	1
2018-06-17	WI	Sunday	1
2018-06-24	WI	Sunday	1
2019-06-01	WI	Saturday	1
2020-03-22	WI	Sunday	1

The next task is to determine which holidays are non-working days. The table below compiles eviction filings by holiday. Certainly the first 5 holidays are honored: Christmas Day, Independence Day, Labor Day, Memorial Day, and Thanksgiving.

Table 6: Holiday Filings

Holiday	Number_evictions
Christmas Day	0
Independence Day	0
Labor Day	0
Memorial Day	0
Thanksgiving Day	0
New Years Day	3
Good Friday	162
Veterans Day	178
MLK Birthday	273
Inauguration Day	284
Decoration Memorial Day	299
Lincoln Birthday	369
CPulaski's Birthday	402
Election Day	423
Columbus Day	473
Washingtons Birthday	492
Presidents Day	522

Only 3 filings were recorded on New Year's Day. Actually, there was only a single filing on New Year's Day in Outagamie County; BNGP Counties and WI are aggregated from this filing.

Table 7: New Years Day Filings

New_Date	Region	Num_evictions	Holiday	Weekday
2020-01-01	BNGP Counties	1	New Years Day	Wednesday
2020-01-01	Outagamie	1	New Years Day	Wednesday
2020-01-01	WI	1	New Years Day	Wednesday

Perhaps the single New Year's Day filing and the weekend filings were the result of dutiful, committed employees. Or perhaps they were data entry errors. In any case, the tables below show the days to remove when identifying trends.

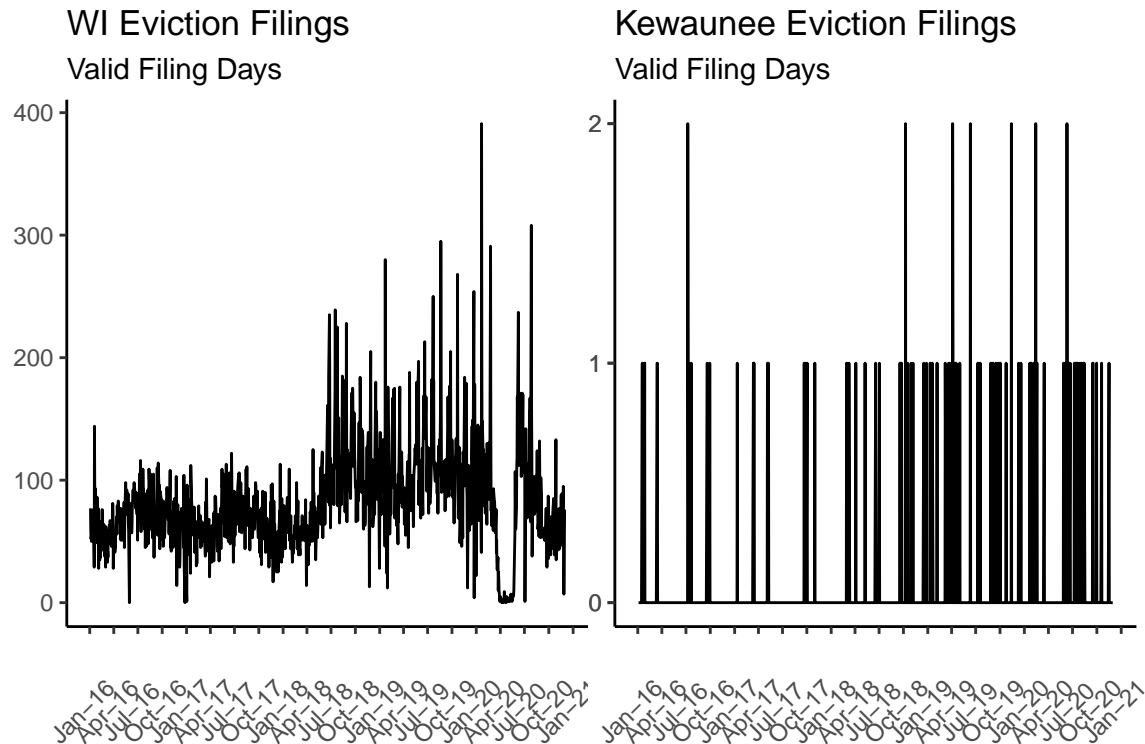
Table 8: Number Holiday Rows to Remove

Holiday	Valid	n
Christmas Day	No	36
Independence Day	No	48
Labor Day	No	60
Memorial Day	No	60
New Years Day	No	48
Thanksgiving Day	No	60

Table 9: Number Weekend Rows to Remove

Weekday	Valid	n
Saturday	No	3084
Sunday	No	3084

Back to good ole' WI and Kewaunee. The plots now show filings after filtering out non-valid days.



Smoothing (After Removing Weekends and Holidays)

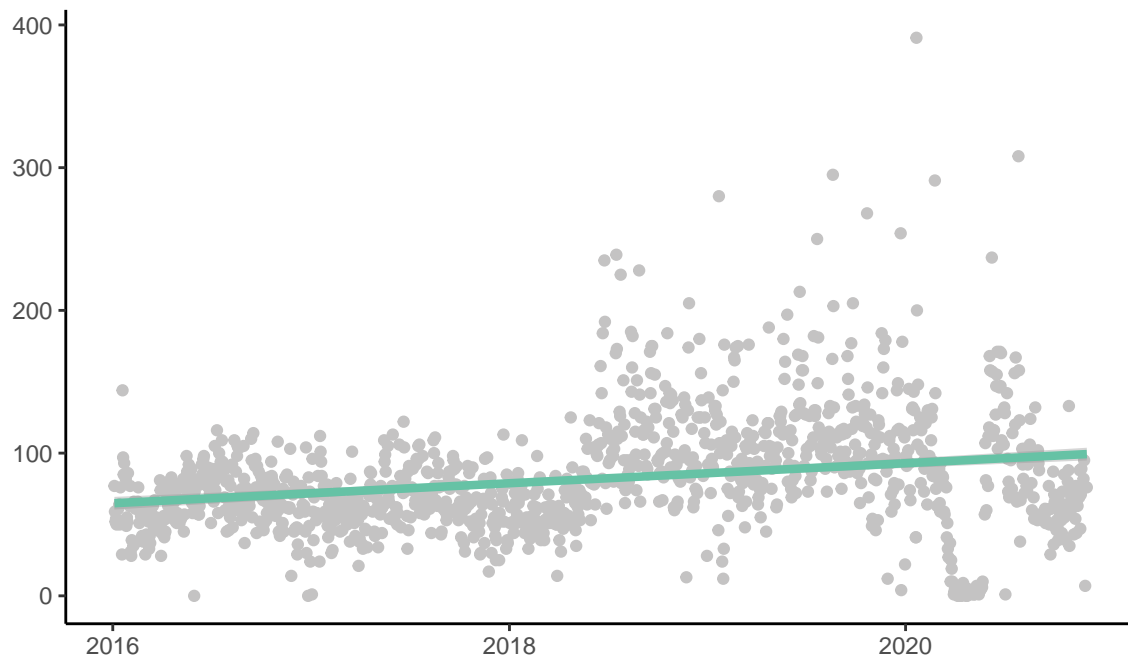
Obviously, the data is way too noisy to identify trends so I will experiment with different types of smoothing. To keep things simple, we'll focus on WI as a test drive.

Linear Regression Linear regression tells us that there is a significant association between the number of evictions over time, but the model only accounts for less than 6% of the variance.

```
##
## Call:
## lm(formula = Num_evictions ~ New_Date, data = Just_WI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -95.439 -21.713  -2.973  18.397 297.709
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -257.59922   38.77002  -6.644 4.53e-11 ***
## New_Date      0.01919    0.00219   8.765 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40.29 on 1256 degrees of freedom
## Multiple R-squared:  0.05764,    Adjusted R-squared:  0.05689
## F-statistic: 76.83 on 1 and 1256 DF,  p-value: < 2.2e-16
```

WI Eviction Filings

Linear Regression

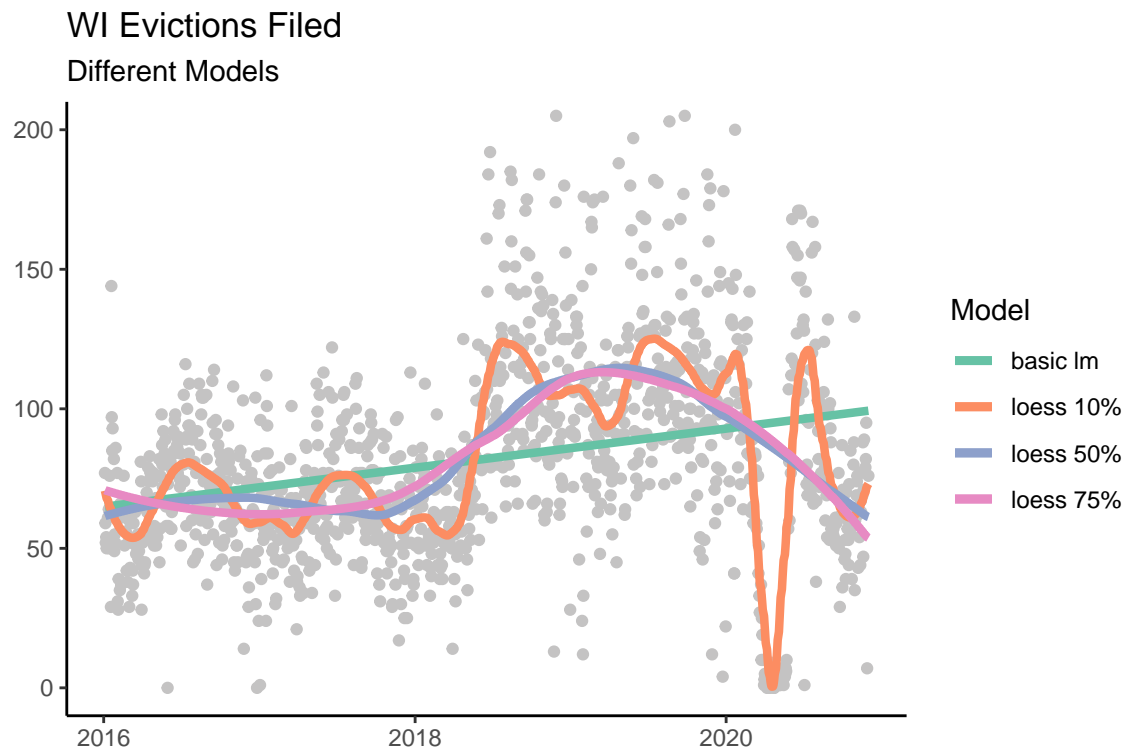


Loess The linear fit is not good because the data doesn't follow a linear trend. After googling some stuff, I decided that local regression via loess wouldn't be too much of a stretch for me.

I was getting an error when I tried using the dates as 'x'. I used this [guide](#) to learn that dates needed to be converted to a numeric sequence. Not quite sure why, but maybe the model can't do math on date objects as it tries to figure out which x values fit a span.

I explored different span sizes but did not change the 2 degree polynomial fitting default.

The graph below is zoomed in to show the different smooths. The loess with a 10% span best follows the curve.



The table below, which summarizes the Sum of Squared Error for each smooth, tells us the same thing as our eyeballs: the loess 10% span yields the smallest SSE.

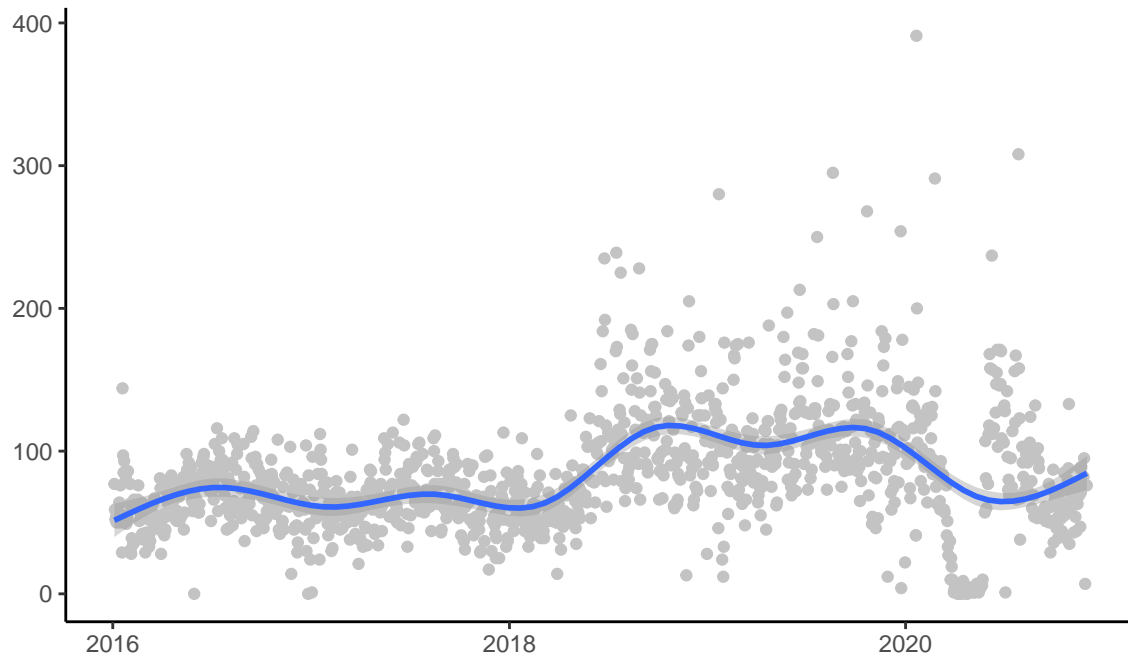
Table 10: Sum of Squared Error

Model	SSE
Basic lm	2038794
Loess 10%	1195210
Loess 50%	1701222
Loess 75%	1751367

GAM Default In the above graphs, I specified the type of smooth. Here is what the graph looks like if I let R loose to choose the default.

WI Eviction Filings

Method = Default GAM ($y \sim s(x, bs='cs')$)



I then started down the rabbit hole of trying to figure out what GAM and all that ‘flexible spline’ stuff meant and very soon was way, way out of my league... time for an intervention.

Questions

- Data Wrangling
 - Would you complete a daily sequence and fill in missing days with 0 filings?
 - Would you filter out weekends and holidays?
- Smoothing
 - I understand that loess is different than finding a moving average, but how do we know if it is “better”?
 - I basically understand `summary(lm)`, but `summary(loess)` baffles me. It seems that people don’t get too caught up in the model output since the focus is visual. Is this true, or should I investigate certain model outputs? How do I come to grips with not really understanding the model output?
 - Is there anything I should explore with GAM?
- Next Steps
 - Investigate loess with 10% span for all regions’ filings?
 - Repeat loess but with calculated rates so I can compare one region to another (rate = eviction filings per 1000 occupied rental units)