

# Time Series Exploration with Eviction Filings

Bliss Cohen

August 17, 2020

## Background

I am in the process of developing an [Eviction Tracker](#) for a client. Evictions are filed on a daily basis, 5 days/week excluding holidays. I aggregate these filings each month across some Wisconsin counties and for the State of Wisconsin.

I would like to add an ‘Overview’ first page to the Tracker. One of the things I’d like to include in the ‘Overview’ is trending information. If desired, people could then fish out specific information on the subsequent pages which are currently in place.

After exploring distributions of eviction filings, I have come to the conclusion that the distributions are definitely not normal. I believe there is a seasonal component. Besides being affected by seasonal fluctuations, filing patterns altered in response to a state-wide eviction moratorium in effect between March 27 - May 26, 2020.

I am not trying to develop a predictive model! I am, however, trying to figure out the best way to visualize ‘big picture’ trends by region.

## Objective

Develop the appropriate analysis and visual to show ‘big picture’ trends in eviction filings

## Read in Data

Below is a snapshot of eviction filing aggregations up to my last update of July 31, 2020. The ‘Date\_for\_Filter’ field just allows me to avoid plotting data past the last update where no data exists. I have also scaled the filings to the number of rental units to find eviction rates.

```
Summary_Year_Month <- read_rds("Summary_Year_Month.rds")
```

```
last_update <- read_lines("last_update")
```

```
head(Summary_Year_Month)
```

```
## # A tibble: 6 x 7
##   Region Year Month Evictions_Filed Date_for_Filter Renter_Occ_Units
##   <chr>   <dbl> <fct>           <dbl> <date>           <dbl>
## 1 Brown   2016 Jan           66 2016-01-01       36228
## 2 Brown   2016 Feb           53 2016-02-01       36228
## 3 Brown   2016 Mar           53 2016-03-01       36228
## 4 Brown   2016 Apr           61 2016-04-01       36228
```

```
## 5 Brown    2016 May           68 2016-05-01           36228
## 6 Brown    2016 Jun           98 2016-06-01           36228
## # ... with 1 more variable: Filings_Per_1000_Units <dbl>
```

## Graph Distributions

I examined eviction filing distributions via histograms and qq-plots. The distributions were not normal. The graphs below are specific to Wisconsin.

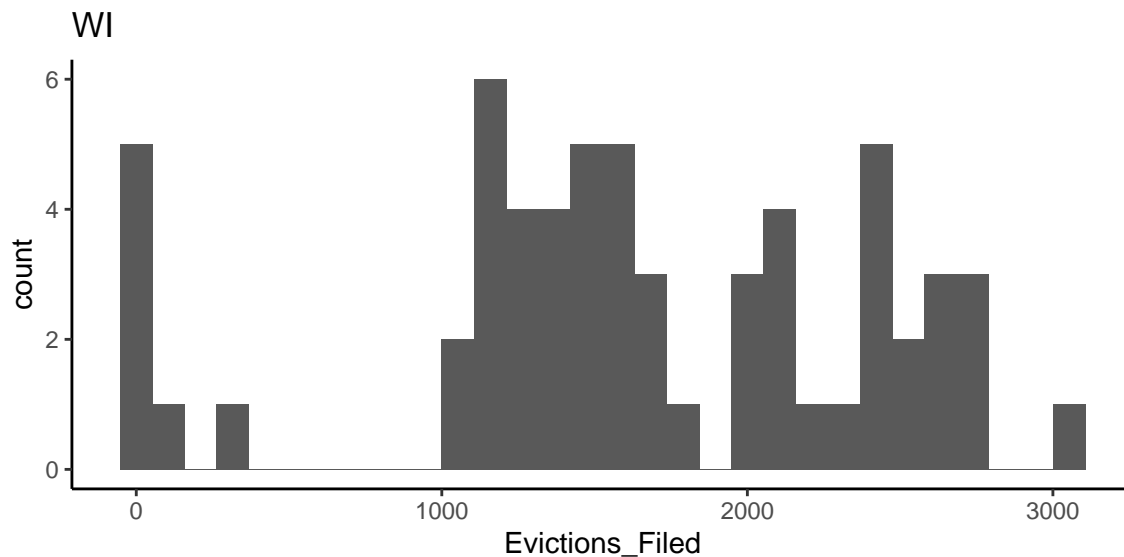
```
# histogram

v_hist <- function(my_region) {
  Summary_Year_Month %>%
    filter(Region == my_region) %>%
    ggplot(aes(x = Evictions_Filed)) +
    geom_histogram() +
    labs(title = my_region) +
    theme_classic()
}

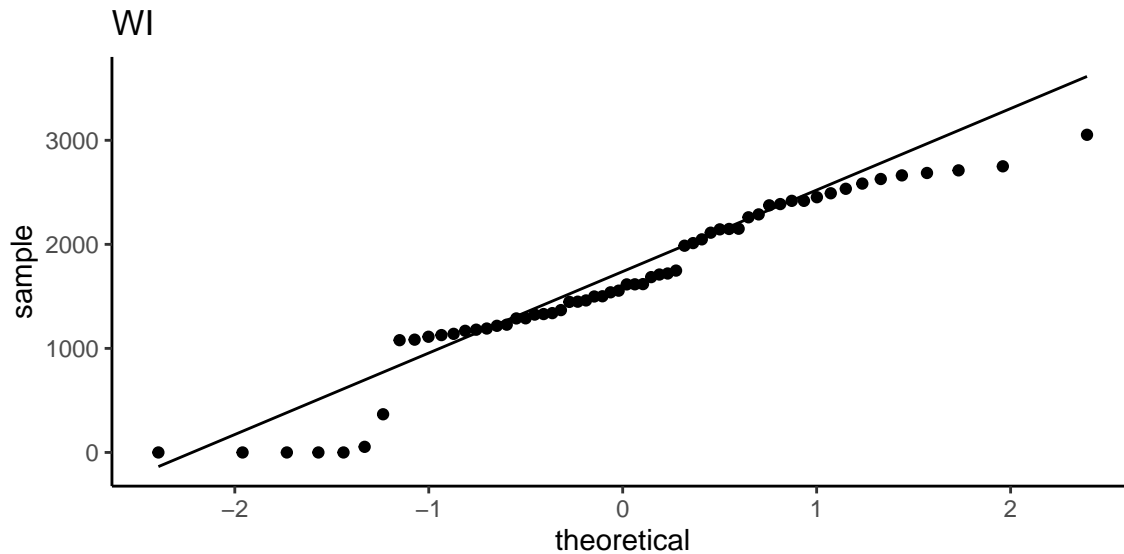
# qq plot

v_qq <- function(my_region) {
  Summary_Year_Month %>%
    filter(Region == my_region) %>%
    ggplot(aes(sample = Evictions_Filed)) +
    geom_qq() +
    geom_qq_line() +
    labs(title = my_region) +
    theme_classic()
}
```

```
v_hist("WI")
```



```
v_qq("WI")
```



### Assess Distribution Skew

I used this [guide](#) to figure out how to assess skew using the moments package. You can see from the skew numbers that none of the regions had a normal distribution. It just didn't make sense to try to transform the distribution since some had a positive skew and some had a negative skew.

```
Skew_evictions <- Summary_Year_Month %>%
  group_by(Region) %>%
  nest() %>%
  mutate("Skew" = map(data, ~skewness(.x$Evictions_Filed)))
```

*# Positive or negative skew...not like we can do a consistent transformation*

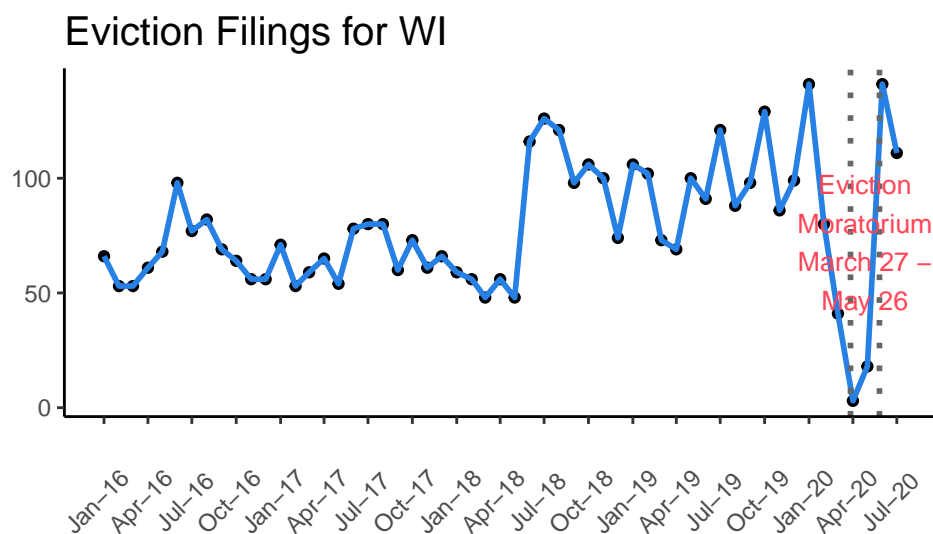
```
Skew_evictions[[3]]
```

```
## [[1]]
## [1] -0.3231594
##
## [[2]]
## [1] 1.133455
##
## [[3]]
## [1] 0.701935
##
## [[4]]
## [1] 0.9573817
##
## [[5]]
## [1] 0.5698529
##
## [[6]]
```

```
## [1] 1.021063
##
## [[7]]
## [1] 0.7461008
##
## [[8]]
## [1] -0.1644577
##
## [[9]]
## [1] 0.7488471
##
## [[10]]
## [1] -0.4836251
##
## [[11]]
## [1] -0.5543952
```

### Plot Eviction Filings vs. Time

The plot below shows how WI's raw data fluctuates over time. It seems like there might be some type of repeating pattern, and certainly the moratorium is introducing an anomaly.



### Create a Time Series Object and Decompose

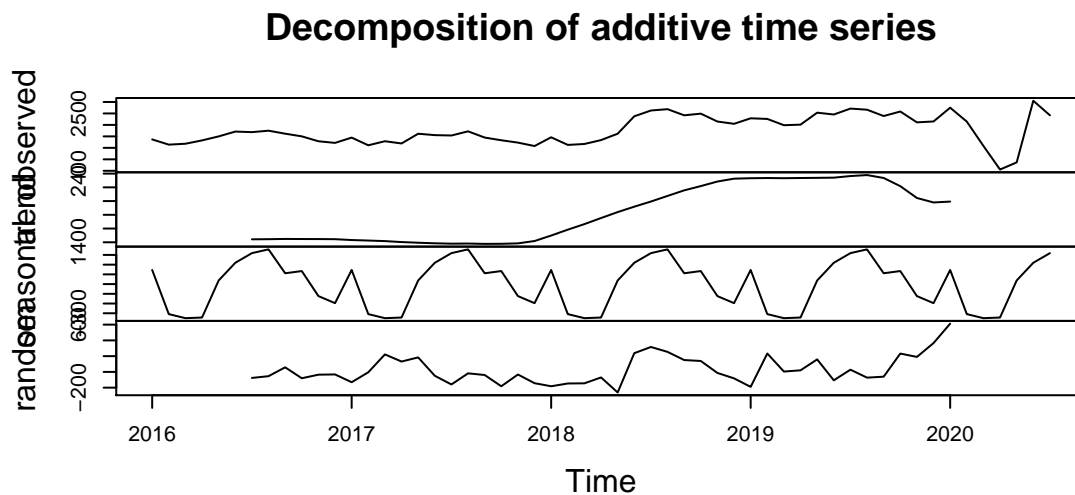
Since I suspected some kind of seasonality component, I started exploring time series objects. Specifically, I extracted the State data and converted WI to a time series object. I then used the `decompose()` function to separate the components into seasonal, trend, and random (residuals).

I wasn't sure if the type of decomposition should be "additive" or "multiplicative" - maybe the amplitude is changing around July 2018 in the graph above? In any case, I ultimately ran the decomposition both ways.

The decomposition plot below is for the additive version. I realize that the plots are shorter in "trend" and "random" because you need 6 months prior and post to find moving averages.

```
Just_WI <- Summary_Year_Month %>%
  filter(Region == 'WI',
         Date_for_Filter <= mdy(last_update)) %>%
  pull(Evictions_Filed) %>%
  ts(frequency = 12, start = 2016)
```

```
decomp_WI_add <- (decompose(Just_WI, type = "additive"))
plot(decomp_WI_add)
```



I can access the values from the decomposed time series. Here I'm just showing the values used for a seasonal adjustment in the additive decomposition.

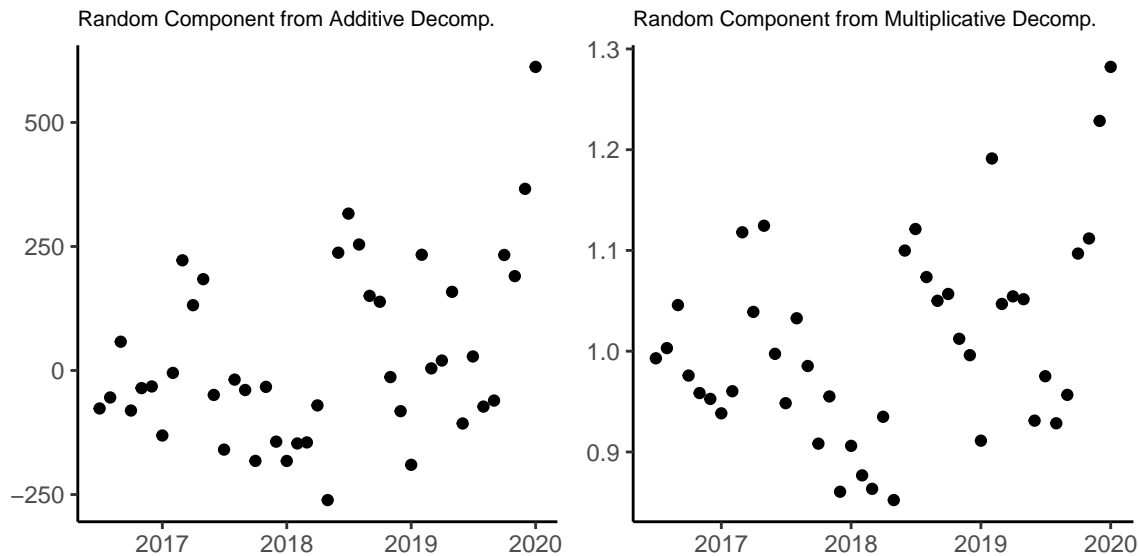
```
# decomp_WI_add$seasonal
#
# decomp_WI_add$trend
#
# decomp_WI_add$random

# Seasonal adjustment

decomp_WI_add$figure
```

```
## [1] 145.25550 -308.67159 -350.39381 -343.78270 37.39786 218.99508
## [7] 318.37008 357.01591 111.30758 134.04716 -123.42159 -196.11950
```

The graphs below show the random component for additive and multiplicative. Not sure if one is better or worse than the other...



### STL Confusion... I Like These Graphs But...

Conceptually, the decompose stuff made some sense to me. But now things get really fuzzy.

Although I am not trying to develop a predictive model, I realized that I needed some way to create estimates for the front and back ends of the decomposed trend line. After googling until I couldn't take it anymore, it seemed like the `stl` function might be useful (Seasonal Decomposition of Time Series by Loess). `stl` assumes an additive model (although you can transform the data if it is multiplicative).

I don't know what this means, but I'll spit it out: "stl is a lowess smoother that replaces values with a 'locally weighted' robust regression estimate of the value." I will just plow forward without fully understanding the next steps.

The seasonal, trend and remainder values are shown below after applying the `stl` function to WI data.

```
stl_WI <- stl(Just_WI, "periodic")
head(stl_WI$time.series)
```

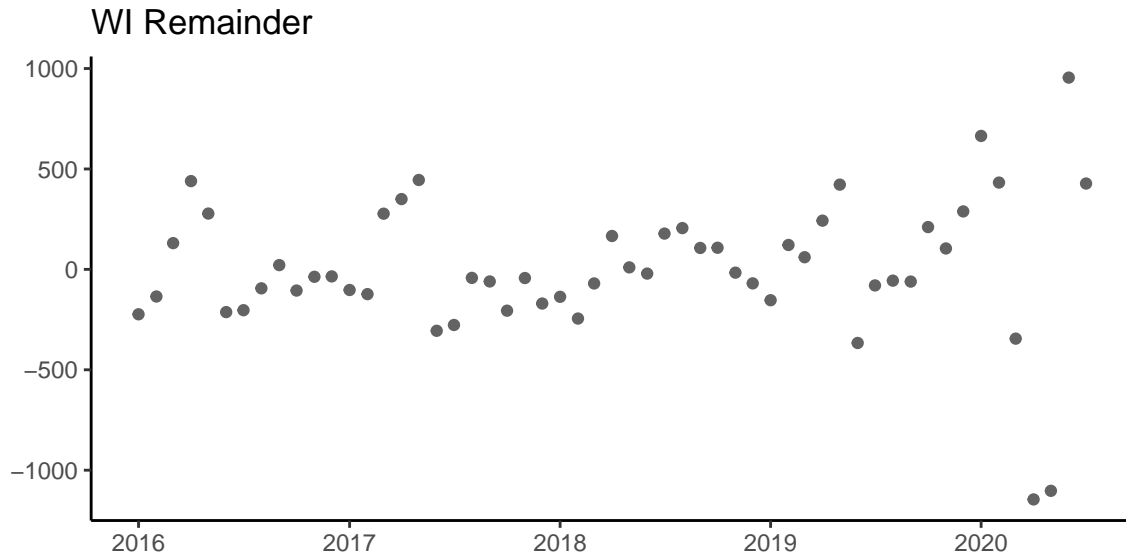
```
##      seasonal      trend remainder
## [1,]  109.3001 1482.337 -223.6375
## [2,] -200.1731 1475.207 -135.0341
## [3,] -418.6461 1468.077  130.5692
## [4,] -577.0465 1462.548  439.4989
## [5,] -235.8472 1457.018  277.8288
## [6,]  468.3061 1454.478 -212.7846
```

I combined the `stl` output with the original data so I could make some graphs using one data frame.

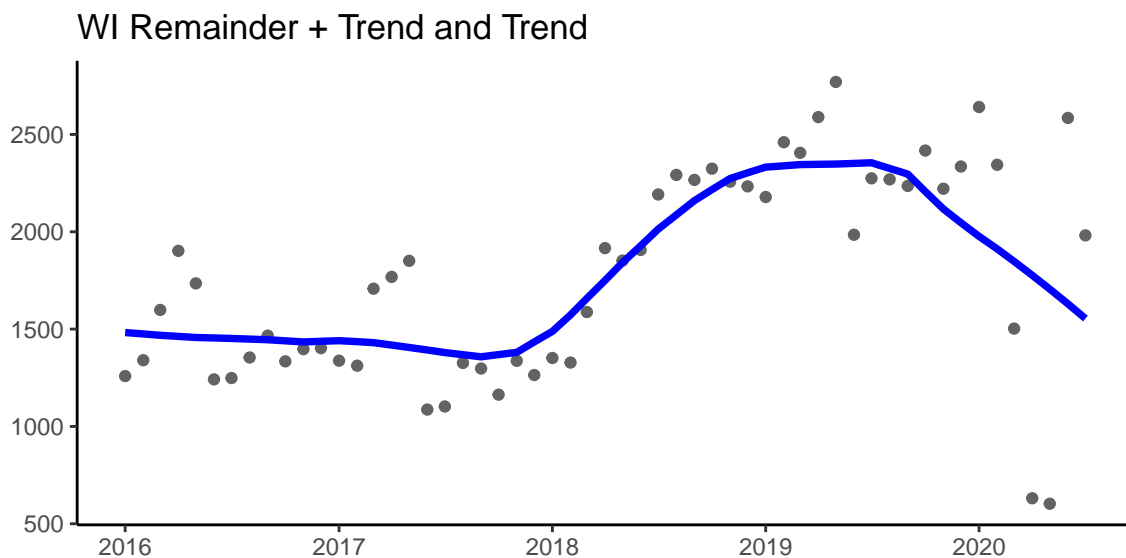
```
stl_WI_df <- stl_WI %>%
  pluck(1) %>%
  data.frame() %>%
  mutate("New_Date" = seq(from = ymd('2016-01-01'),
                           to = floor_date(mdy(last_update), 'month'), by = 'months')) %>%
  mutate("trend_remainder" = trend + remainder) %>%
```

```
select(New_Date, seasonal, trend, remainder, trend_remainder) %>%
  cbind(Just_WI)
```

In the first graph, I just plot the remainder. This is my interpretation: the remainder pretty much fluctuates around 0 up until 2020. Then the remainder goes out of whack, indicating that our model is not picking up other influencers like the moratorium.

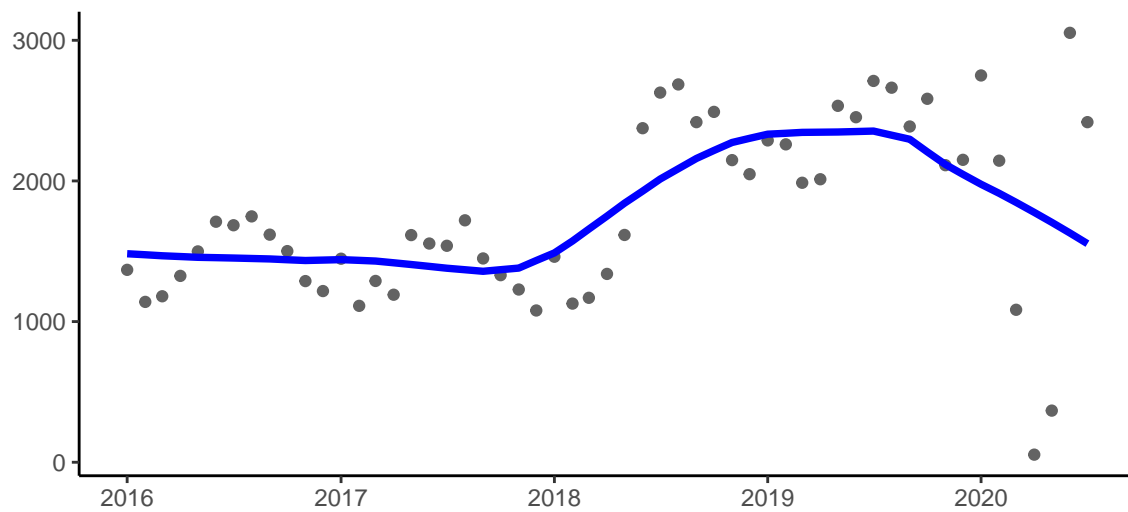


In the next graph, I added the trend and remainder values together for the dots and added a blue line for the trend.



And in the last graph, the dots reflect the raw data against the trend line.

WI Raw Data and Trend



### Questions

1. At a high level, is there something wrong or misleading in my approach to presenting trend information?
2. Can I confirm that seasonality does in fact exist at some confidence level/p-value, or is this an eyeball assessment?
3. How to check multiplicative vs additive?
4. What is stl doing that is different than decompose?

The graphs produced via the stl components seem like they could give people a basic overview of how eviction filings are trending for a particular region... but it feels a little bit like...

