

Eviction Filings

Part 4

Bliss Cohen

February 03, 2021

Background

This project analyzes historical and current eviction filings for the state of Wisconsin (WI) and 10 WI counties (collectively called “BNGP Counties”). The general term “region” will be used to describe an aggregated or non-aggregated geographical area.

The next phase of the project is to share a proposal for a Shiny app. Stakeholders have the following key questions:

- How are eviction filings trending for a particular region?
- Can you present some visuals to help us brainstorm how to identify thresholds for initiating intervention programs?
- Is one region’s filings higher or lower than another?

This file contains these follow-ups from our January 13, 2021 discussion:

- Wrap Up Old Stuff and Set Aside
 - Examine WI seasonal trends by month-year
 - Compare regional time series trends via line plots
- Improve Old Stuff to Maybe Keep
 - Calculate month-on-month percent change
 - Rank yearly filing rates
 - Explore margin of error
 - Rank quarterly filing rates
- Add New Stuff and Assess Value
 - Identify insights through linear regression

Objective

Discuss which analyses to present in a Shiny app proposal

Data Snapshot

This part is the same as before - just showing as a refresher, filling in missing days and all:).

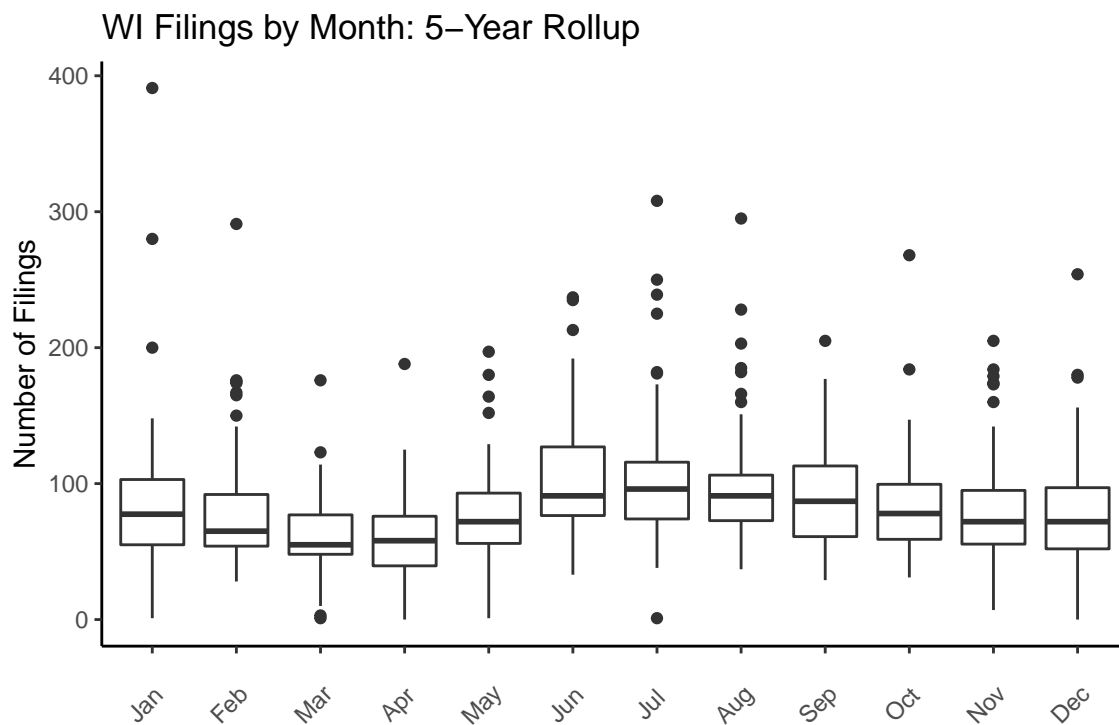
The base file aggregates the number of daily eviction filings from January, 2016 - December, 2020 across 12 regions of interest. Missing days were filled in with a filing value = 0 to ensure each region had a record for every day of the year. Since filings aren’t entered on weekends and certain holidays, each record was labeled with a “Yes/No” valid filing day label.

```
## Rows: 21,667
## Columns: 11
## $ New_Date      <date> 2016-01-01, 2016-01-01, 2016-01-01, 2016-01-01, 2016...
## $ Region        <fct> BNGP Counties, Brown, Calumet, Green Lake, Kewaunee, ...
## $ Month         <ord> Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan...
## $ Year          <dbl> 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2016,...
## $ Day           <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2,...
## $ Num_evictions <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ Weekday       <fct> Friday, Friday, Friday, Friday, Friday, Friday, Friday, Frida...
## $ Week          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ Quarter       <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ Holiday       <chr> "New Years Day", "New Years Day", "New Years Day", "N...
## $ Valid         <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No",...
```

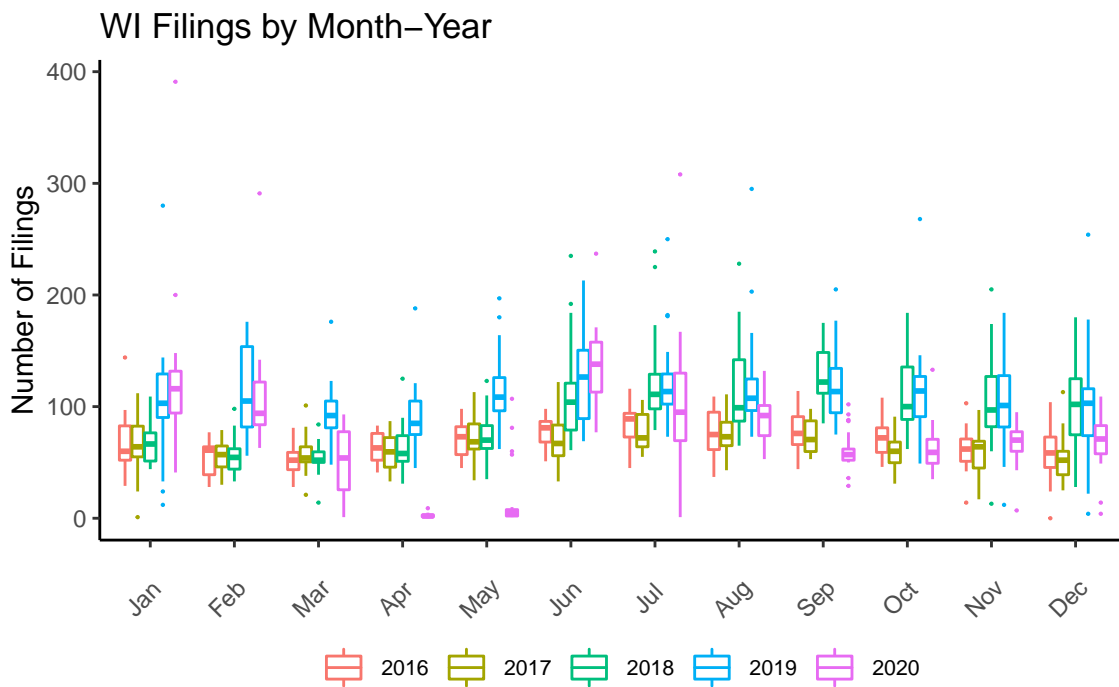
Wrap Up Old Stuff and Set Aside

1. WI Seasonal Trends

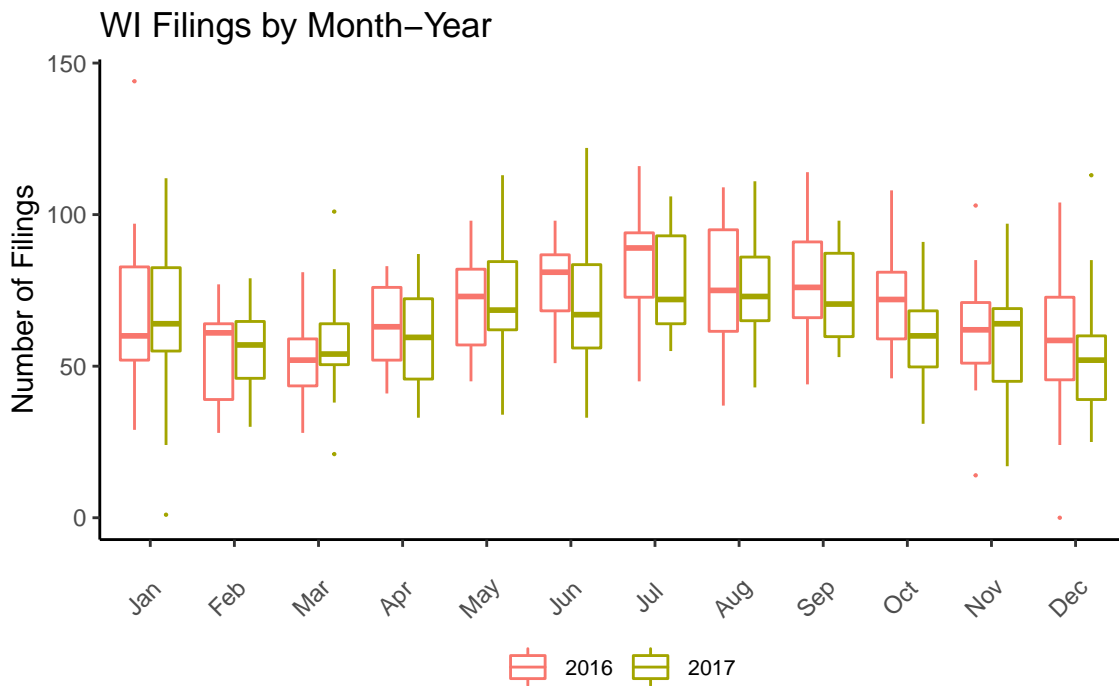
Below is the original graph showing how filings trend by season, aggregating all 5 years. Filings dip in the spring, peak in the summer, and then slowly taper off towards the end of the year, eventually bottoming out again in the spring of the following year.



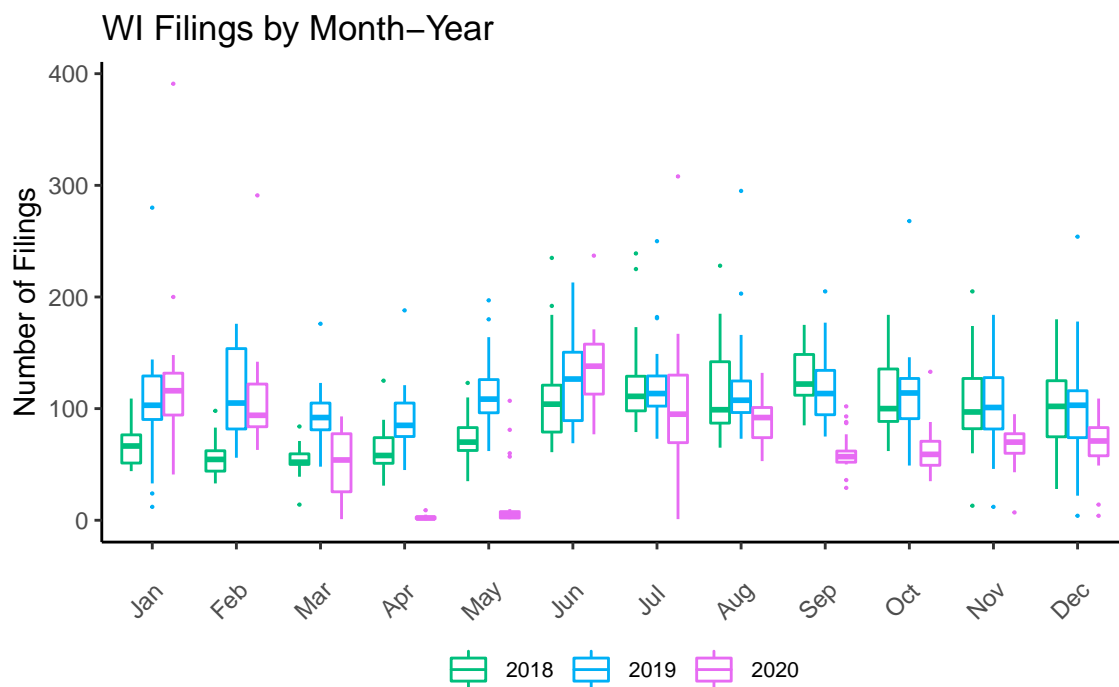
Each boxplot was split out by year to decompose five years of data to one year per boxplot. I have to admit, though, I have a hard time comparing one year's boxplots to another.



I felt like I could see these trends better if I filtered the years. 2016 and 2017 show similar seasonal trends: Filings dip in the spring, peak in the summer, and then taper off towards the end of the year.

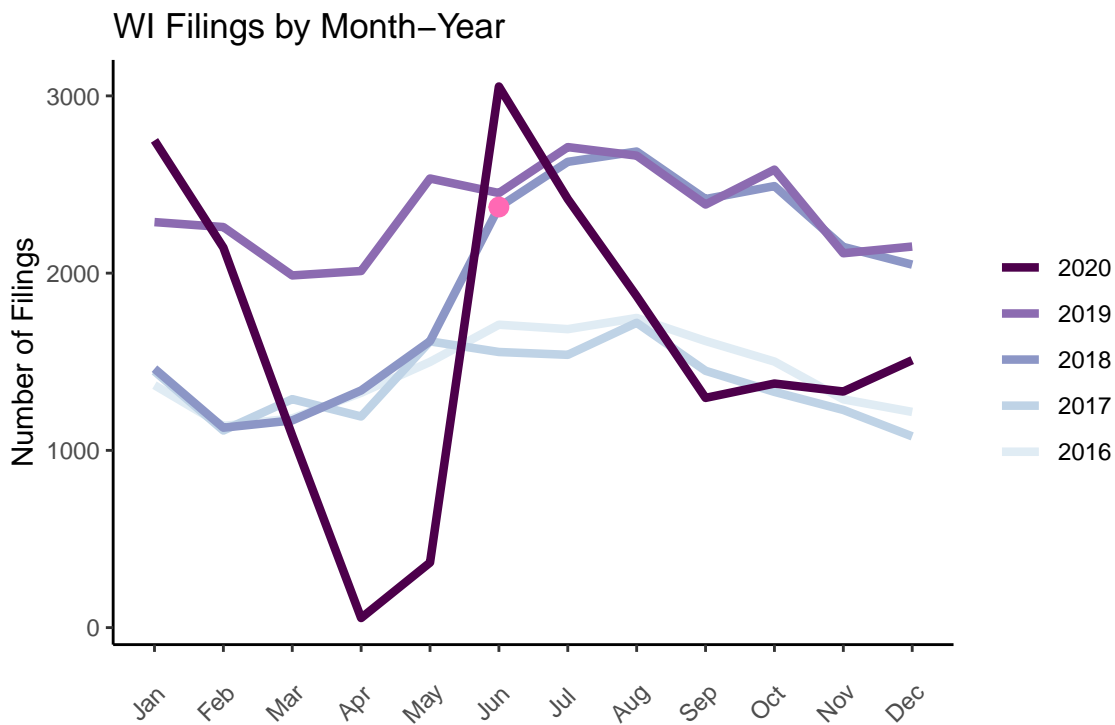


But seasonal trends are different in 2018, 2019, and 2020 vs 2016-2017, in addition to being different from each other.



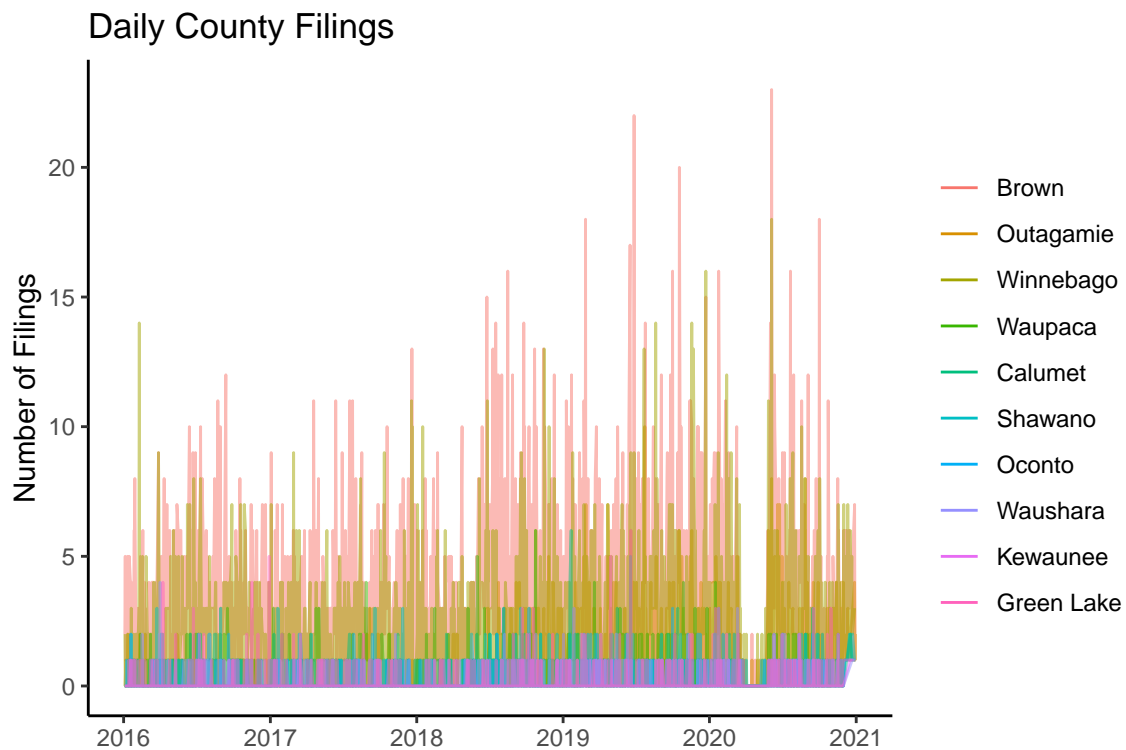
When all is said and done, I prefer aggregating monthly filings for a seasonal graph, as shown below:

- Something changes in May-June 2018 where filings spike up and remain high
- Filings are relatively flat across the 2019 seasons, although there is a slight bump in the summer which plateaus into the fall
- 2020 starts out with high filings, but these become erratic over the year due to the pandemic
 - It is unclear if Dec. 2020 has returned to pre-2018 levels or if filings are still suppressed by the pandemic

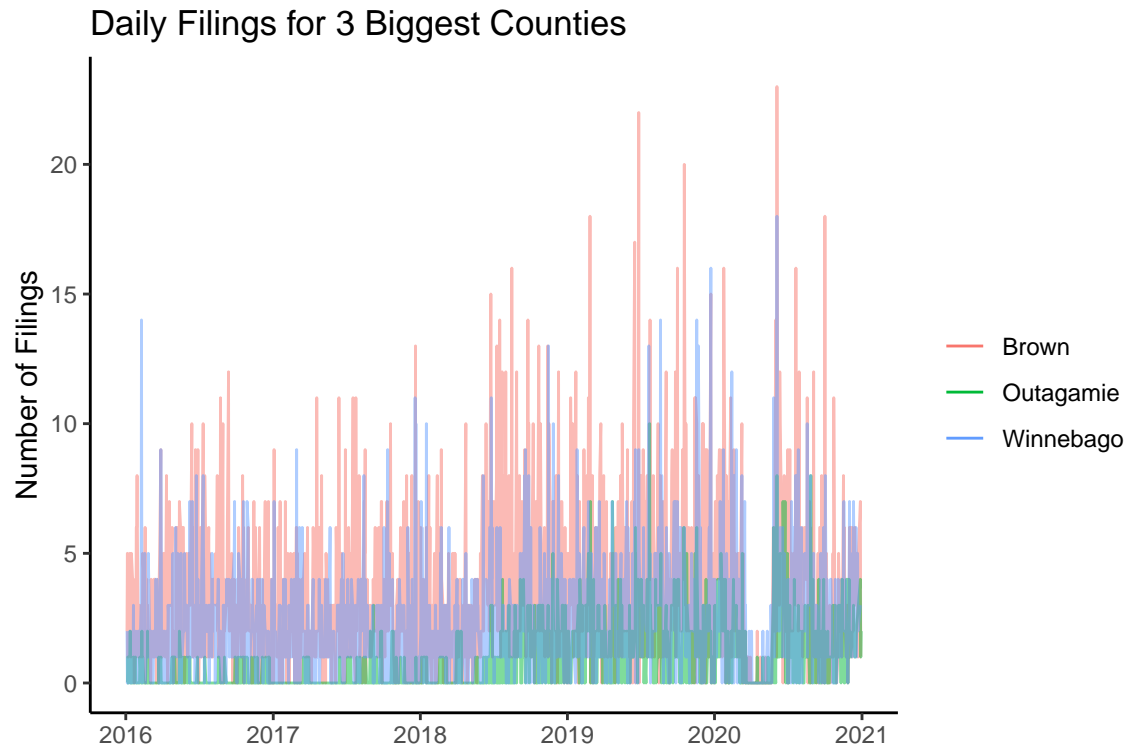


2. Time Series Trends

The graph below shows filings over time for non-aggregated regions. It is difficult to distinguish individual county trends, even with a low $\alpha = 0.5$. Nonetheless, all regions clearly dip in early 2020 due to the eviction moratorium.

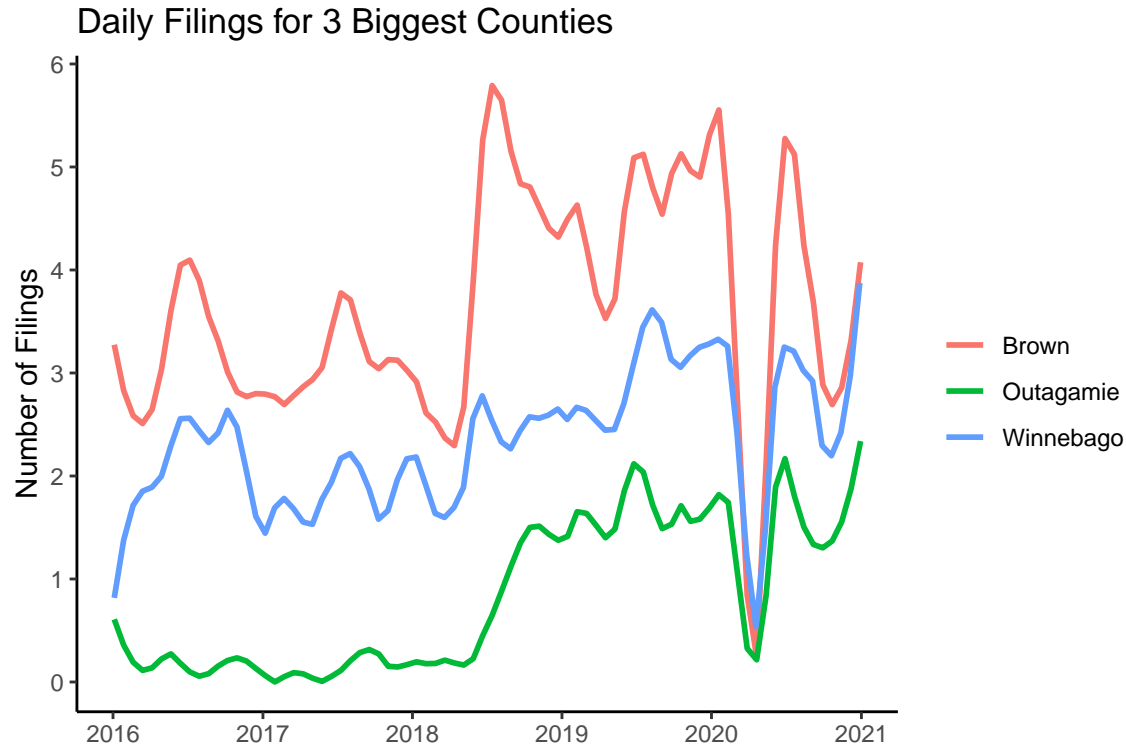


The utility of these plots doesn't really increase that much even when data is isolated to the three biggest counties.



Smoothing the trends in the graph below makes it easier to compare regions:

- To varying degrees, all regions show an uptick like WI about half-way through 2018 where filings stay high until the 2020 whiplash
- That said, trends prior to mid-2018 are different across the counties



I concluded from the time series investigation that line plots might not be the best way to compare trends across the regions.

Improve Old Stuff to Maybe Keep

1. Month-on-Month Percent Change

I'm still trying to think of a quick, dashboard-like way for users to identify trends.

In the previous iteration *Evictions_Part3*, I calculated year-on-year % filing changes. I like the idea of keeping year-on-year functionality, but I agreed with your point that month-on-month calculations would be more relevant for current activity assessments.

The graphs below show 'value boxes' for all the regions in a given time frame.

Table 1: Timeframe

Period	Start Day	Through this End Day
Current	2020-12-02	2020-12-31
Prior	2020-11-02	2020-12-01

A green box means percent change has dropped, red means it's gone up, and yellow means it's steady. In a real app, I would add icons so the user didn't have to rely on color.

The first graph shows the aggregated regions, and the second graph shows the individual counties. I could imagine a similar set of graphs for year-on-year % changes.

Percent Change in Monthly Filings for Aggregated Regions

Most Recent 30 Days to Prior 30 Days



Percent Change in Monthly Filings for Counties

Most Recent 30 Days to Prior 30 Days

-83 Green Lake	-67 Oconto
-51 Outagamie	-27 Winnebago
+11 Brown	+100 Calumet
+120 Shawano	+200 Waushara
+200 Kewaunee	+250 Waupaca

2. Rank Yearly Filing Rates

An annual filing rate was calculated for each region by dividing the number of filings by the estimated number of occupied rental units. The rate was ultimately expressed as the number of filings per 1000 rental units (filings/unit*1000).

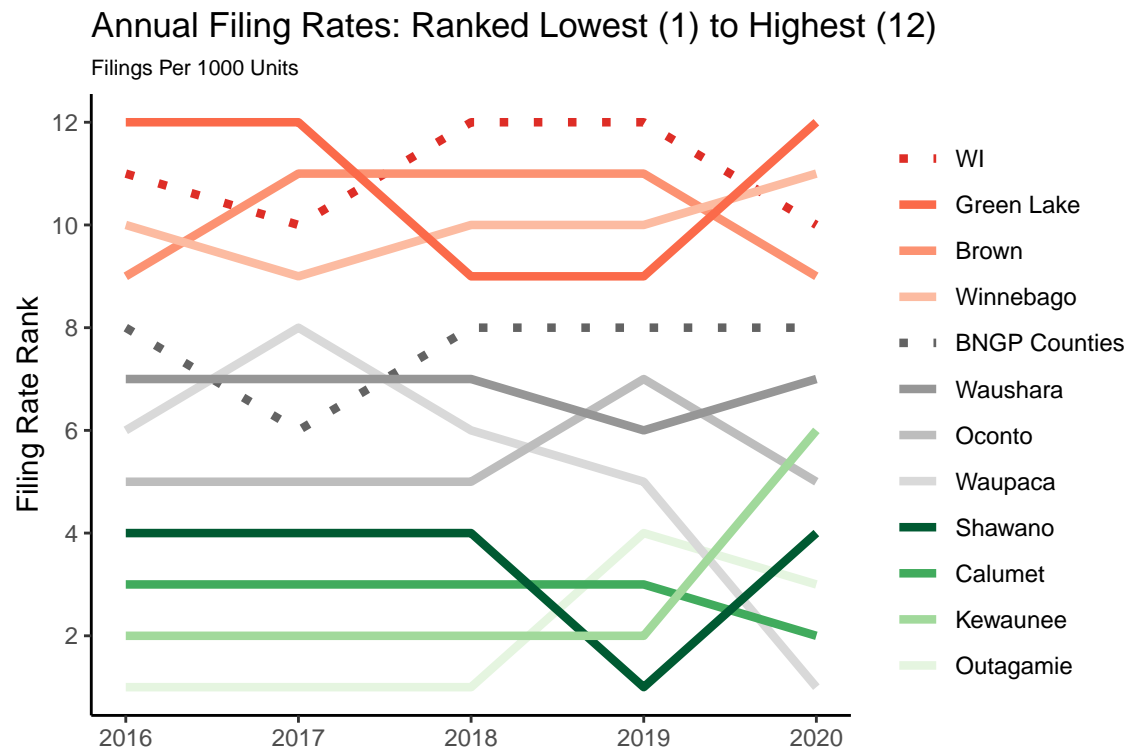
Filing rates were then ranked from 1 to 12 with 1 being the lowest and 12 being the highest. The low, medium, and high groups are visualized below across the years:

- Similar to WI, Green Lake, Brown, and Winnebago have the highest rates
- Outagamie, Kewaunee, Calumet, and Shawano have the lowest rates
- The remaining counties, including the BNGP County aggregation, are somewhere in between

That said, some trends may be changing:

- Kewaunee ticked up in 2020 to a “medium” rate
- Waupaca dipped in 2020 to a “low” rate

No doubt rates will continue to flux until the pandemic is controlled.



3. Margin of Error for Yearly Filing Rates

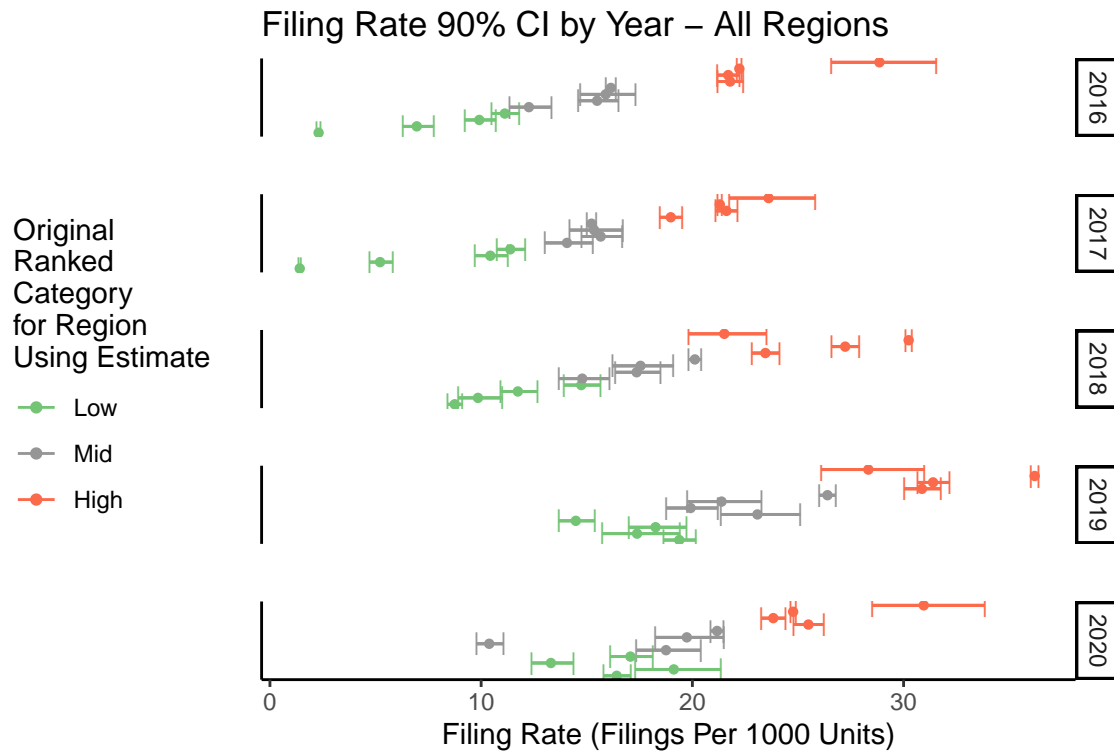
All this ranking is well and good, but one unknown lurks in the background. Rate is calculated by dividing the number of filings by the **estimated** number of rental units, as obtained from the U.S. Census (2015-2019 ACS 5-Yr sample). Each estimate is associated with a Margin of Error (MOE) at 90% confidence. So the question is this: do rankings fall apart when considering the confidence interval (CI) for the number of rental units? This question was investigated more fully for the yearly ranked plot.

Fortunately, the plot below tells a similar story as the original ranking:

- Even with the CI, rankings cluster into Low, Mid, and High groups

- 2020 shows the same trend as the ranked plot; there is more mixing of Low and Mid groups

Thus, it seems that MOE's are tight enough such that relying on estimates for ranking is reasonable.



4. Rank Quarterly Filing Rates

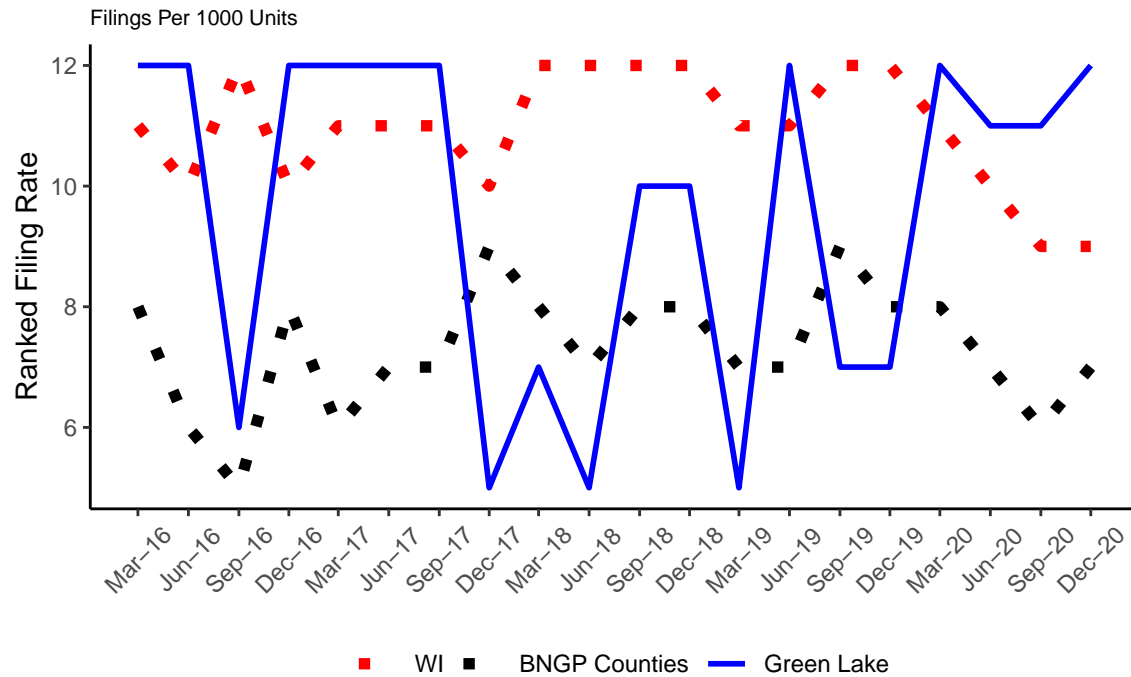
Year aggregations provide a 30,000 foot view. I examined the possibility of aggregating by month, but the line plots were just too messy. Quarterly aggregations appeared to be a decent compromise.

I created 2 functions based on a county of interest:

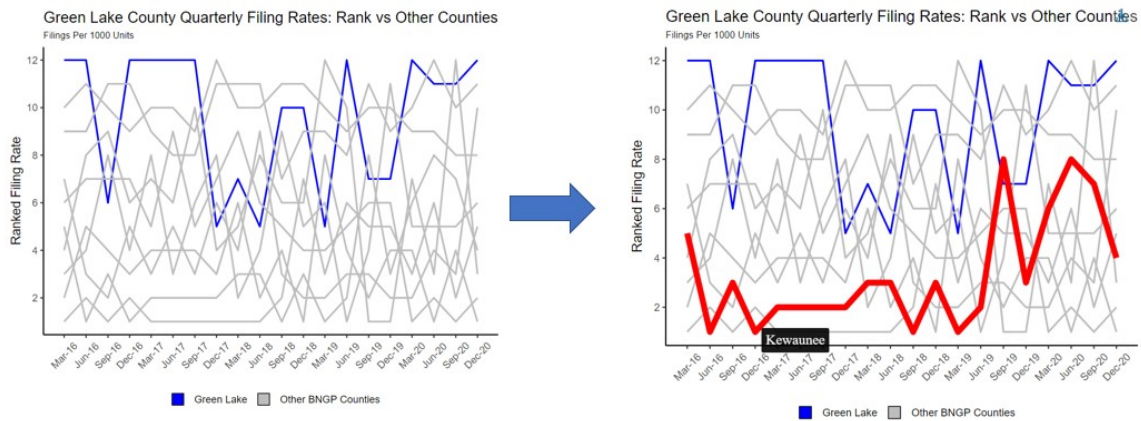
- Compare county to aggregated regions
- Compare county to other counties (non-aggregated)

The plot below shows how “Green Lake” county compares to WI and the 10 aggregated BNGP counties.

Green Lake Quarterly Filing Rates: Rank vs. Aggregated Regions



The following plots compare the same county of interest - Green Lake - to other counties. Interactive graphs can't be rendered in pdf, so I am resorting to low-resolution screen shots to give a feel for how they might look in a Shiny app, especially if the user hovered over a line.



Add New Stuff and Assess Value

1. Features

Additional features were joined to each region:

- Avg HH Income
 - Average Household Income

- **Pct Poverty HH**
 - Percent households below Official Poverty Level
- **Pct ALICE HH**
 - Percent households above the Official Poverty Level but unable to meet basic needs
- **Units Per 1000 People**
 - Number of occupied rental units per 1000 people
- **Avg Yearly Filing Rate**
 - Average yearly filing rate, as calculated from monthly rates

2. Number of Units Per 1000 People vs Features

What features, if any, correlate to the number of rental units?

For this analysis, each region only has one value per feature since ACS 5-Year estimates were used. The correlation plot below reveals the following:

- **Units Per 1000 People**
 - As **Avg HH Income** or **Pct Poverty HH** increase, the number of units increases slightly
 - As **Pct ALICE HH** increases, the number of units decreases slightly
- As **Avg HH Income** rises, **Pct Poverty HH** sharply decreases while **Pct ALICE HH** shows a more modest decrease
- As **Pct ALICE HH** increases, **Pct Poverty HH** increases slightly

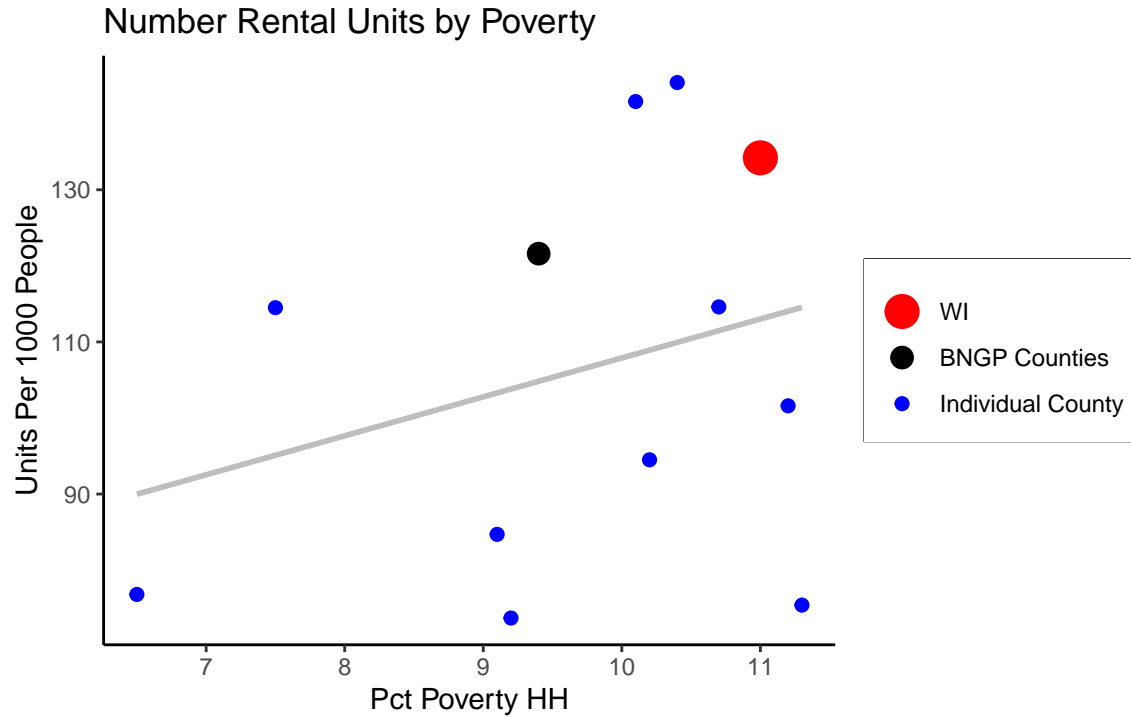


The positive association between `Units Per 1000 People` and both `Avg HH Income` and `Pct Poverty HH` is interesting. This positive association is surprising because of the strong negative association between `Avg HH Income` and `Pct Poverty HH`. Obviously, `Avg HH Income` and `Pct Poverty HH` are negatively correlated: as income goes up, poverty goes down.

Multiple linear regression was used to investigate correlation significance. Both `Avg HH Income` and `Pct Poverty HH` exert significant, positive leverage on the number of rental units.

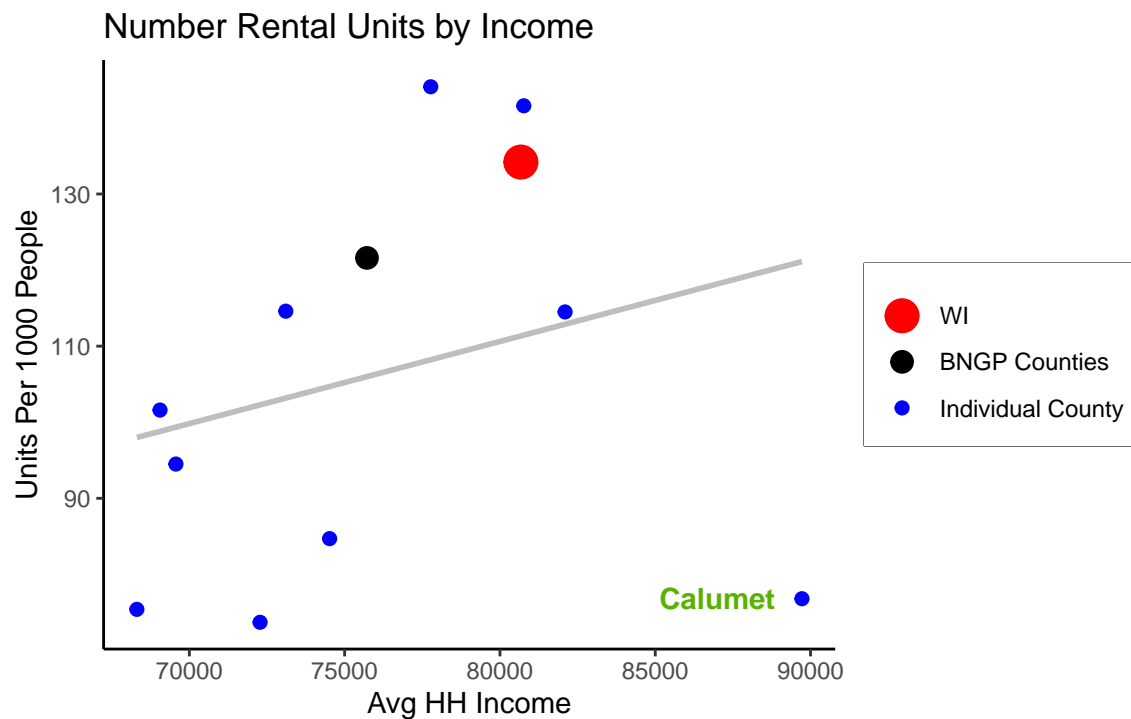
```
##
## Call:
## lm(formula = Units_per_1000p ~ Mean_HH_Income + Pct_ALICE + Pct_pov,
##     data = Prep_df_units_pop)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.188 -10.537  -2.446   13.453   24.558
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.359e+02  1.643e+02  -2.044   0.0752 .
## Mean_HH_Income  3.889e-03  1.371e-03   2.837   0.0219 *
## Pct_ALICE      -1.013e+00  2.822e+00  -0.359   0.7290
## Pct_pov         1.745e+01  5.915e+00   2.950   0.0184 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.2 on 8 degrees of freedom
## Multiple R-squared:  0.5551, Adjusted R-squared:  0.3882
## F-statistic: 3.327 on 3 and 8 DF,  p-value: 0.07728
```

As poverty increases, perhaps the availability of rental units also increases, since renting doesn't require the same level of financial security as home ownership. It would be interesting to repeat the analysis for home units.



2015–2019 American Community Survey 5–Year Estimates (Tables S1901, S2502, B01003)

Similarly, as income increases, the availability of rental units increases, perhaps because more people can support independent living. Curiously, Calumet county seems to defy this trend. Calumet is the wealthiest county, as measured by average income. Perhaps there is a threshold where people shift from renting to home ownership?



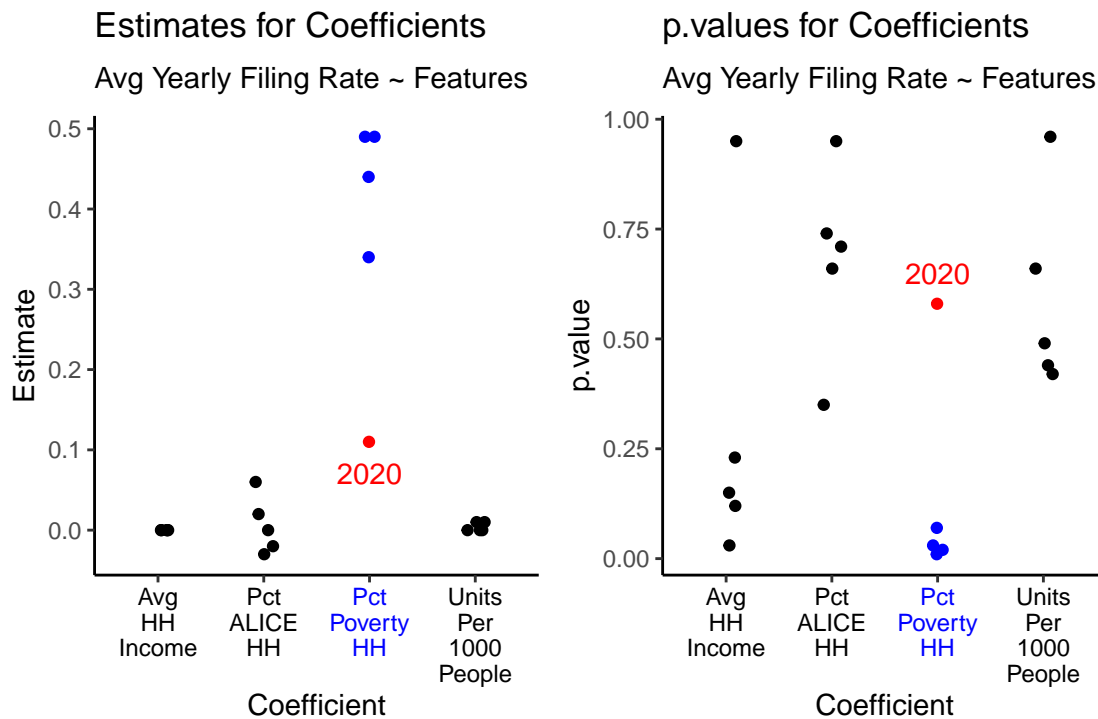
2015–2019 American Community Survey 5–Year Estimates (Tables S1901, S2502, B01003)

3. Average Yearly Filing Rates vs Features

What features, if any, correlate to the Avg Yearly Filing Rate ?

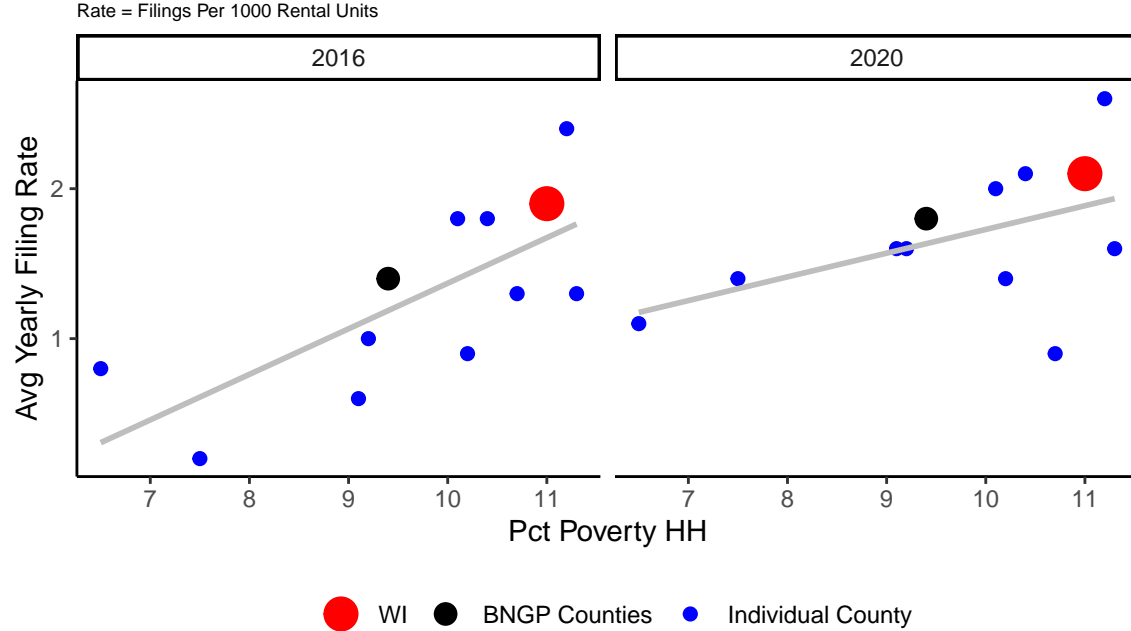
Monthly aggregations are about as granular as I can get while still extracting useful information from all the regions. Therefore, I aggregated monthly filings to find a monthly filing rate from which I calculated a yearly average.

Multiple linear regression was run for each year. The plots below show the range of Estimates and p.values associated with the coefficients. In general, Pct Poverty HH exerts significant, positive leverage vs other features, as evident by the higher Estimates and lower p-values. In 2020, however, there was no longer a correlation between Avg Yearly Filing Rate and Pct Poverty HH , or any other features for that matter.



The contrast of Avg Yearly Filing Rate vs Pct Poverty HH between historical (2016) and current (2020) is shown below.

Filing Rate vs Poverty



2015–2019 American Community Survey 5–Year Estimates (Tables S1901, S2502, B01003)

The table below shows output for the overall model by year. As already discussed, the model is no longer significant in 2020, as indicated by the largest p.value.

Year	r.squared	adj.r.squared	sigma	f_statistic	p.value
2016	0.68	0.50	0.44	3.80	0.06
2017	0.67	0.49	0.40	3.63	0.07
2018	0.87	0.79	0.26	11.52	0.00
2019	0.72	0.56	0.37	4.46	0.04
2020	0.41	0.07	0.45	1.21	0.39

In summary, it is not surprising that **Pct Poverty HH** consistently correlates with eviction rates. The Official Poverty Level is, by the own U.S. Census admission, simply a statistical yardstick; developed in the early 1960's, it has no connection to what is really needed to survive in today's society. The Official Poverty Level captures people in the most extreme financial distress. Eviction is also extreme. People under the Official Poverty Level cannot pay their rent.

Other measures, such as **Pct ALICE HH** developed by the United Way, offer a more realistic measure of poverty for working families. ALICE HH are above the Official Poverty Level yet unable to meet basic needs. **Pct ALICE HH** did not correlate with **Avg Yearly Filing Rate**. ALICE is less extreme than the Official Poverty Level. ALICE HH will prioritize rent or mortgage. But since rent/mortgage take up a huge chunk of their paychecks, they lack funds for other needs such as childcare, transportation, etc.

Questions

1. Any unresolved business in “Wrap Up Old Stuff and Set Aside”?
2. I am thinking of sharing all the plots from “Improve Old Stuff to Maybe Keep” in the proposal, except for the CI one (I did the CI check for my own peace of mind). I like these plots because they give a current snapshot relative to historical. I could imagine supplementing them a time series plot that allow users to investigate specific monthly numbers if they want. What do you think?

3. I'm leaning towards sharing a graph of Avg Yearly Filing Rate vs Pct Poverty HH , like one of those on page 16. But there are some things about the graph that bother me:

- Do you think multiple linear regression was appropriate? Would you recommend another approach?
- I really struggle talking about linear regression. It's been beaten into our heads that correlation doesn't mean causation. But how do you talk about the findings in a non-causation way???
- The graphs on page 16 are two dimensional, but multiple regression is multi dimensional. Is there a way to "see" the output from multiple regression? Is it misleading to make two-dimensional plots?

4. Nit-picky, less important questions:

- I find the term 'Residual standard error' from `summary(lm)` confusing (see page 13). I know it is the standard deviation of the residuals, but why isn't it called standard deviation or root mean squared error? This [link](#) kind of discusses this issue.
- Just a nagging question about labeling graphs. I prefer a very simple, clean look, so I try to remove as much clutter as possible. Sometimes I wonder if I really need to label axes. For example, do you think I need the "Coefficient" x-label on page 15 (Estimate and p-value)?