Daniel Addokwei Tetteh

SMALL CLAIMS LEGAL CASES

FINAL REPORT

2023-02-06

The exercise has two major sub-parts, Exercise 1 and Exercise 2, both of which are duly addressed below.

EXERCISE 1

1-1 [R] Subset of Eviction Cases

Please find the code for obtaining the subset of Eviction Cases below from R-Markdown. Also, refer to the R-code attached.

```
Cases_data$file_year=as.character(Cases_data$file_year)

Eviction_cases=Cases_data %>% filter(grepl('FORCIBLE ENTRY & DETAINER',
iss_desc))
glimpse(Eviction_cases)

## Rows: 173,476
## Columns: 9
```

1-2 [R] Create a line graph comparing the total number of evictions filed in Tulsa and Oklahoma countries each month.

Please find the codes for plotting the lines below. The complete files are in the R-codes attached.

```
# Now the Line Graph will be plotted
x_df_new %>%ggplot( aes(x=Month, y=Number_of_Cases, group=Court_County,
color=Court_County)) +
  geom_line(aes(colour=Court_County),size=0.8) +
  geom_point(aes(colour=Court_County)) +
  scale_color_manual(values=c('#FFFF00',"#663399")) +
  ggtitle("Number of Evictions filed for Tulsa and Oklahoma County") +
  ylab("Number of Eviction Cases")
```

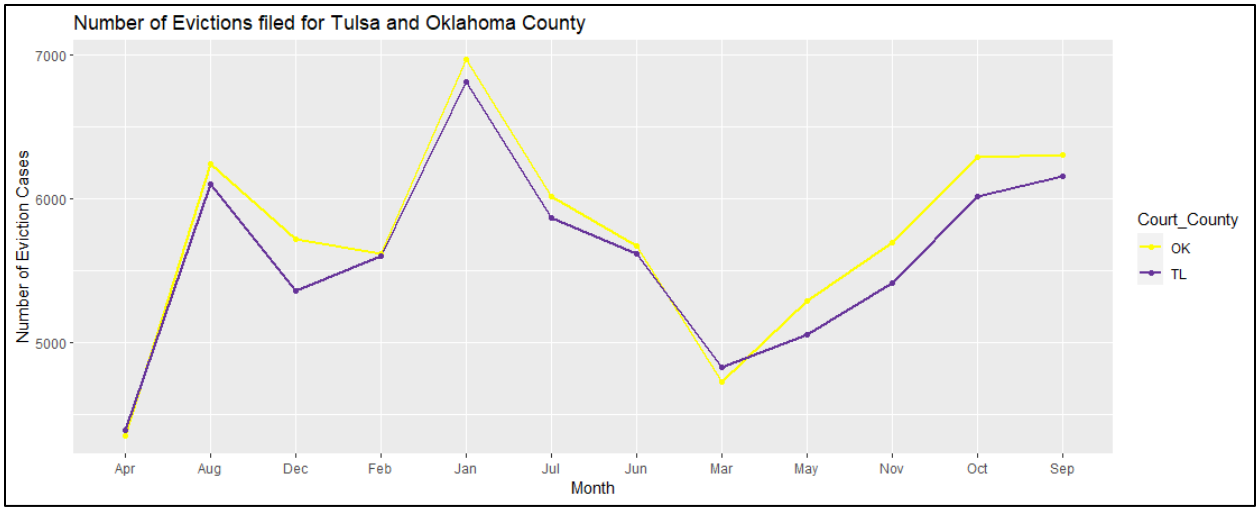


Fig-1 Eviction Cases filed in Tulsa and Oklahoma.

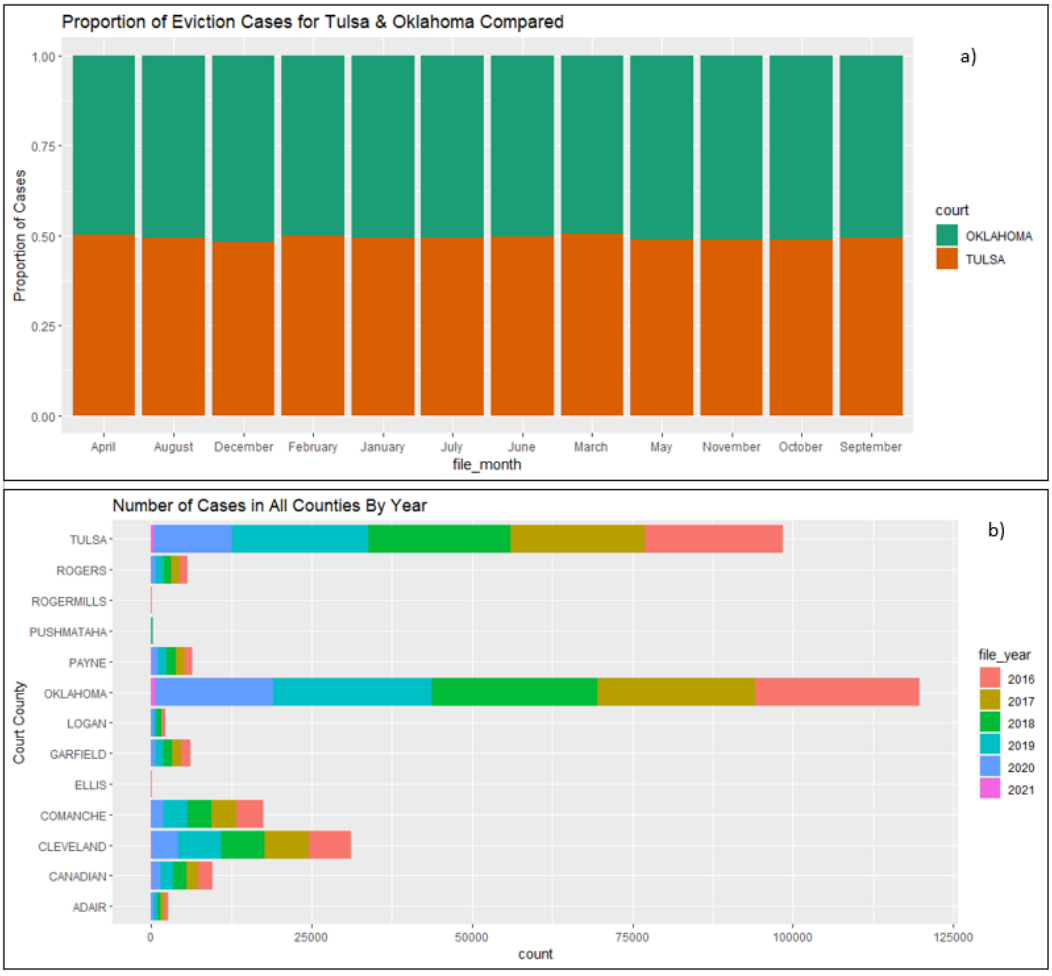


Fig 2. (a) Proportion of eviction cases in Tulsa and Oklahoma (b) Total number of cases filed for all counties and all years

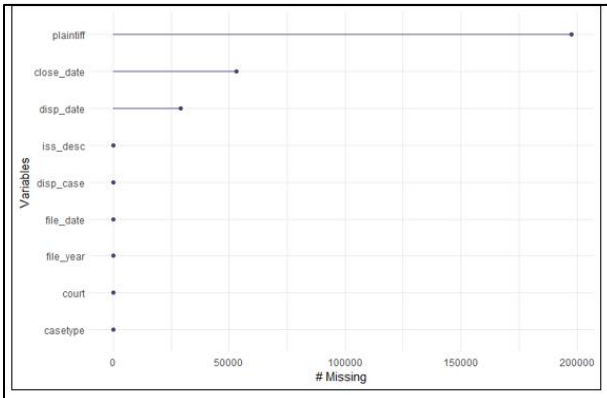
1-3 [R] (Narrative, 200 words) Describe at least two notable trends you can see in the data and offer possible explanations for what is causing them

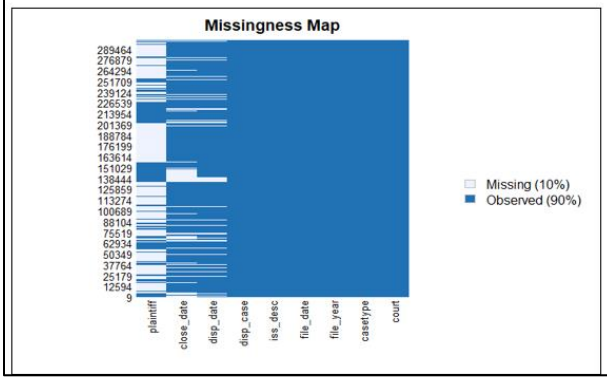
From the bar graph in Fig 2(a), the eviction cases filed for both Tulsa and Oklahoma are reasonably balanced in proportions for all months; thus, good grounds exist for comparison. Fig.1 indicates peak values for filed eviction cases in January for both Tulsa and Oklahoma, although Oklahoma is slightly higher. A possible cause for peak values in January is that it is the beginning of the year. Most cases may have been carried forward from the previous year and officially filed in January. Also, from Fig. 2(b), the county with the highest number of cases is Oklahoma, which explains its higher value than Tulsa. There is also a steady drop in filed cases from January to March for both counties until the least numbers were recorded in April. The number then began to rise steadily from May to September. For both Tulsa and Oklahoma, the number of filed cases was fairly close for all the months except for May, October, November and December, where filed cases were significantly more in Oklahoma than in Tulsa. March and April are the only months where filed eviction cases were more in Tulsa than in Oklahoma. Fig 1 compares monthly filed cases for all years.

EXERCISE 2

2-1 [R] Exploratory Data Analysis (EDA)

Data Visualization





	variable	n_miss	pct_miss
1	plaintiff	197519	65.877216680
2	close_date	53173	17.734441965
3	disp_date	29158	9.724876513
4	iss_desc	36	0.012006844
5	disp_case	36	0.012006844
6	file_date	4	0.001334094
7	court	0	0.000000000
8	casetype	0	0.000000000
9	file_year	0	0.000000000

Fig.3 – Missing Data in Original Data

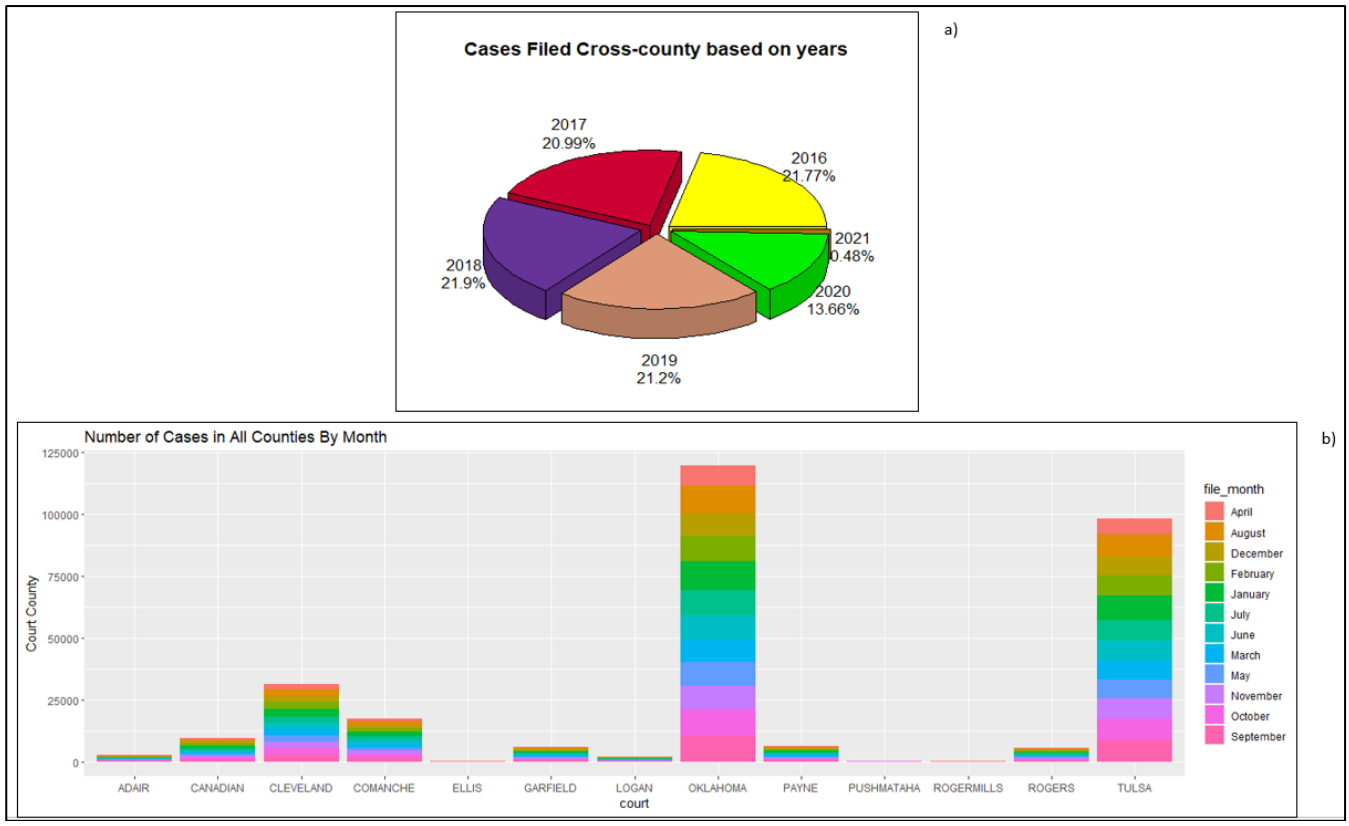


Fig4. (a) Percentage of cases filed cross-county based on years; (b) Number of cases filed in all counties by month for all years

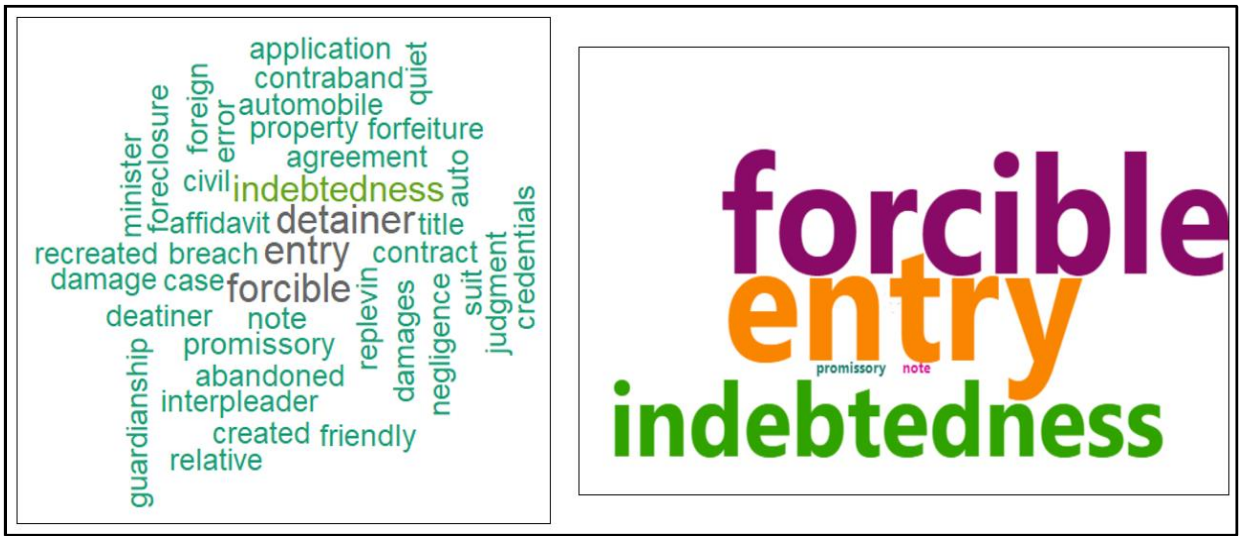


Fig4. Word cloud for issue types (description) for all small claims cases filed

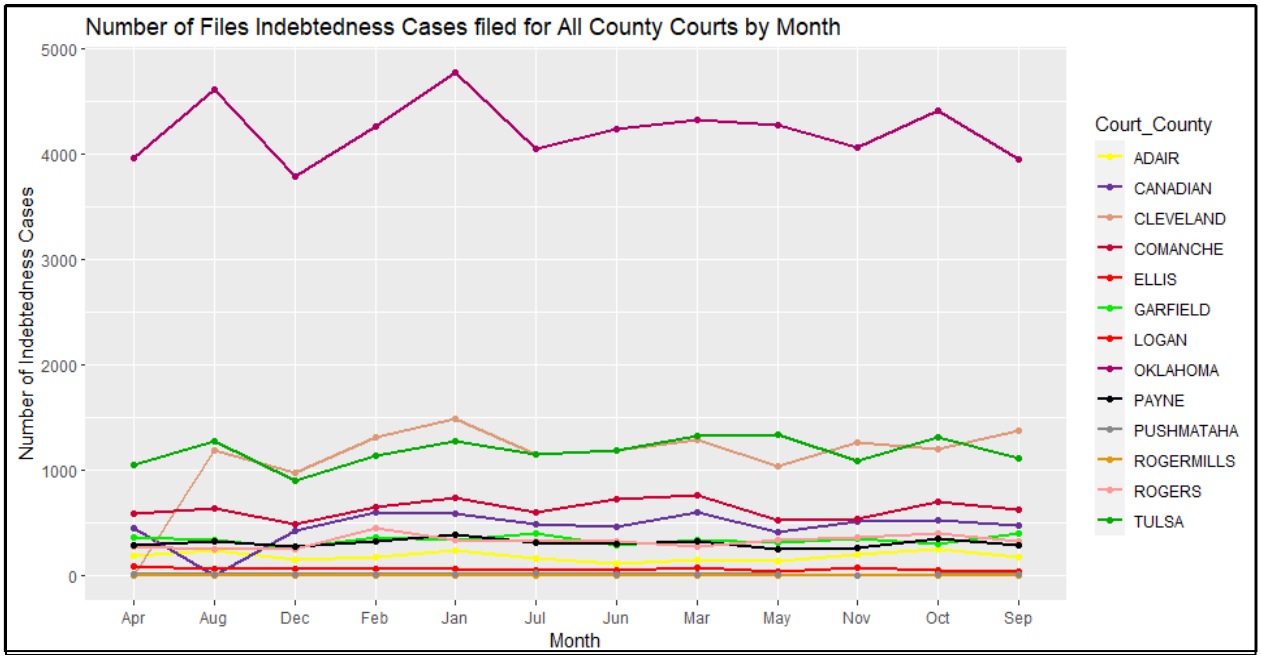


Fig5. Number of indebtedness cases filed for all county courts by month

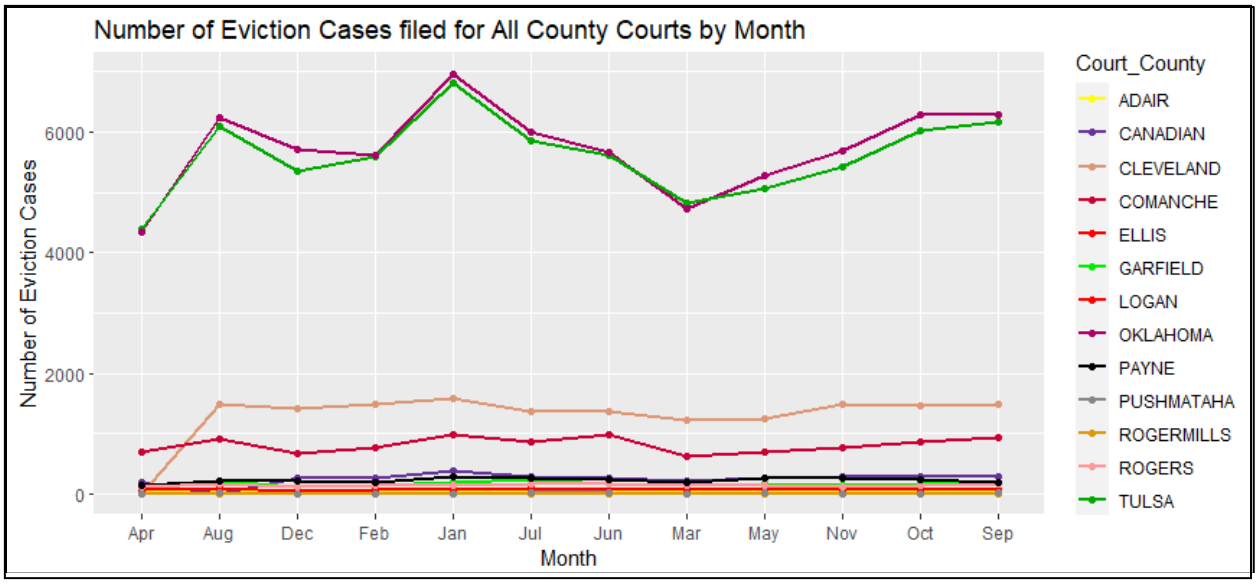


Fig6. Number of eviction cases filed for all county courts by month

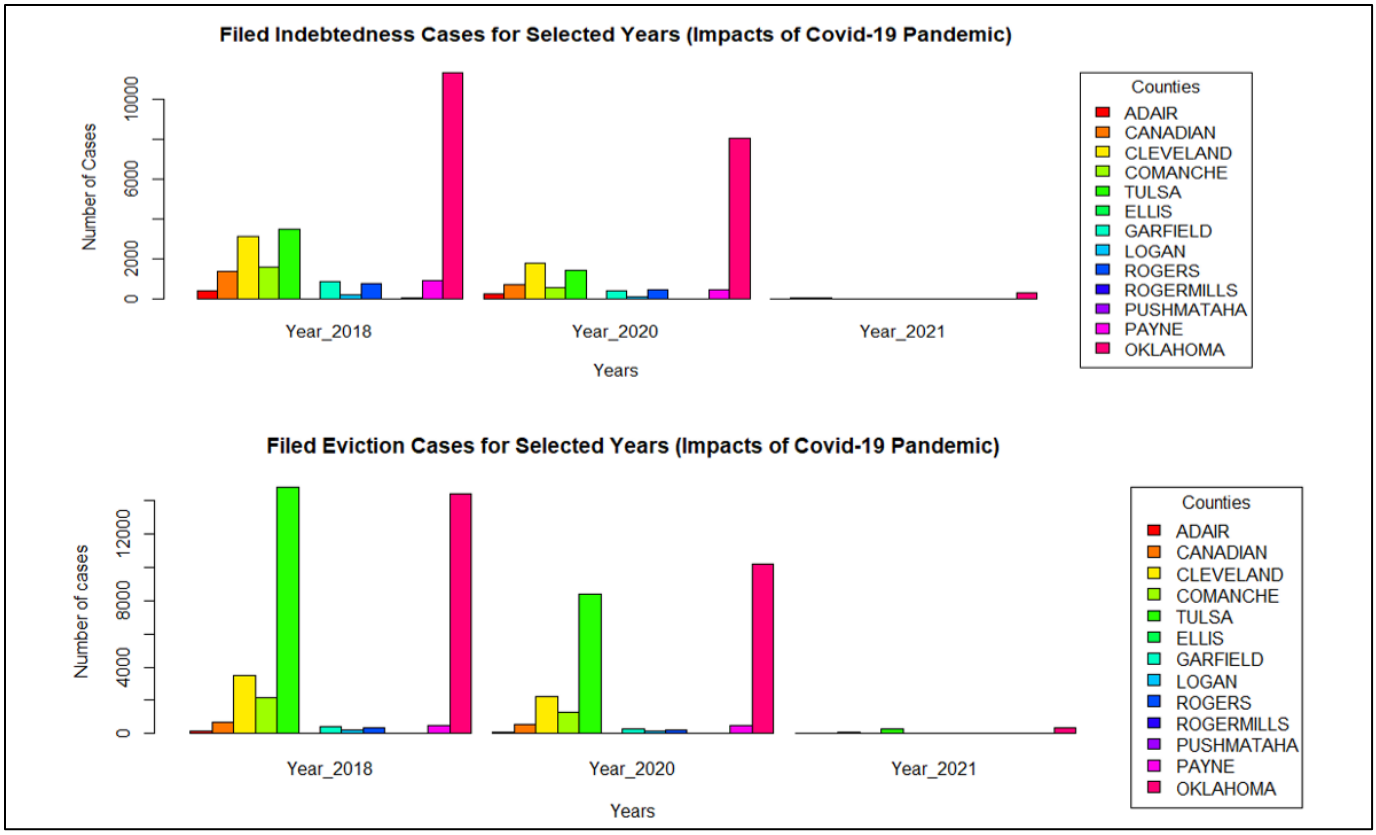


Fig7. Trends in filed Indebtedness and Eviction cases for all counties in selected years.

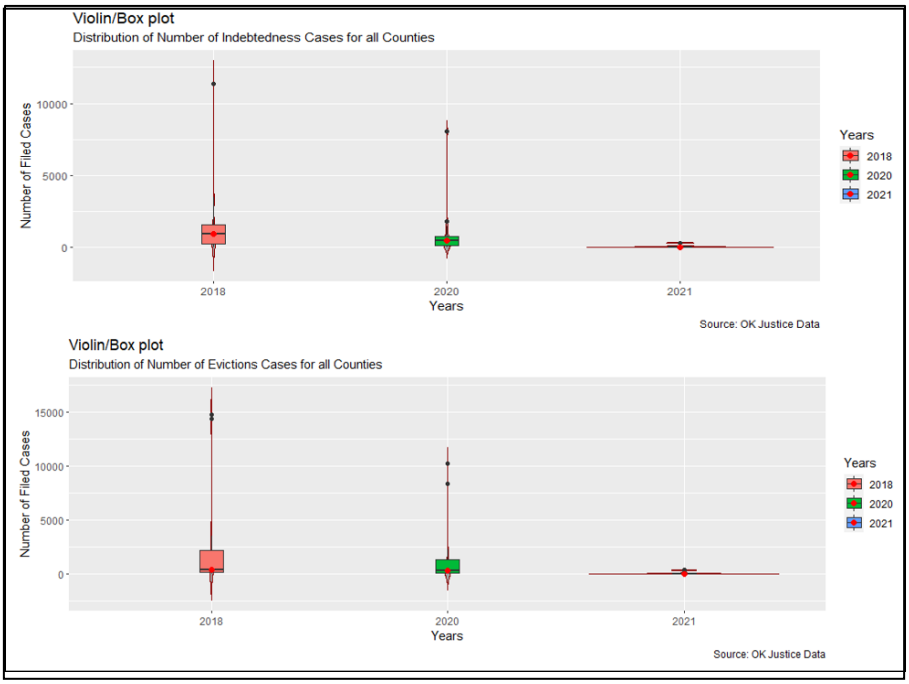


Fig8. Boxplot of Trends in filed Indebtedness and Eviction cases for all counties in selected years.

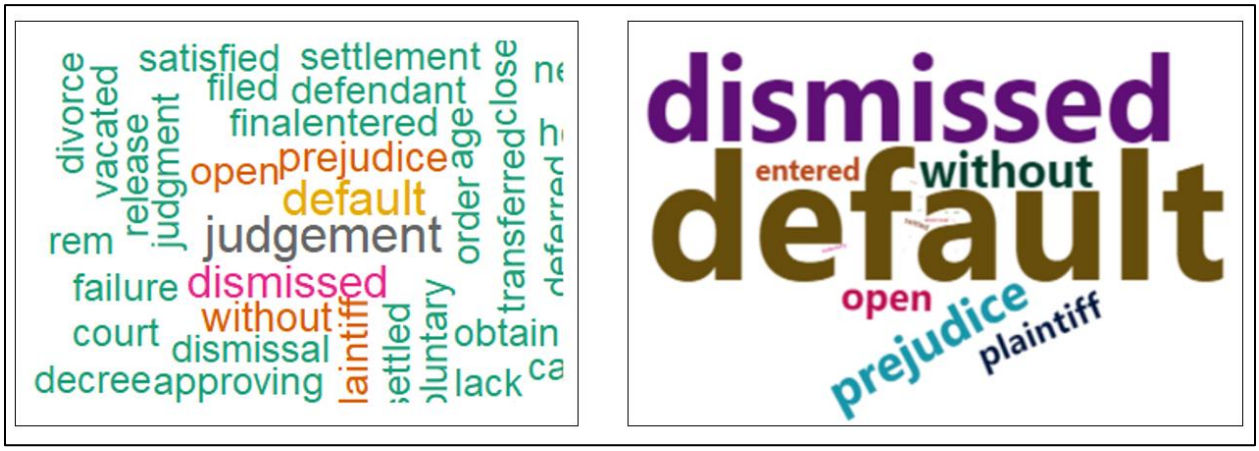


Fig9. Word cloud for dispositions for all small claims cases filed.

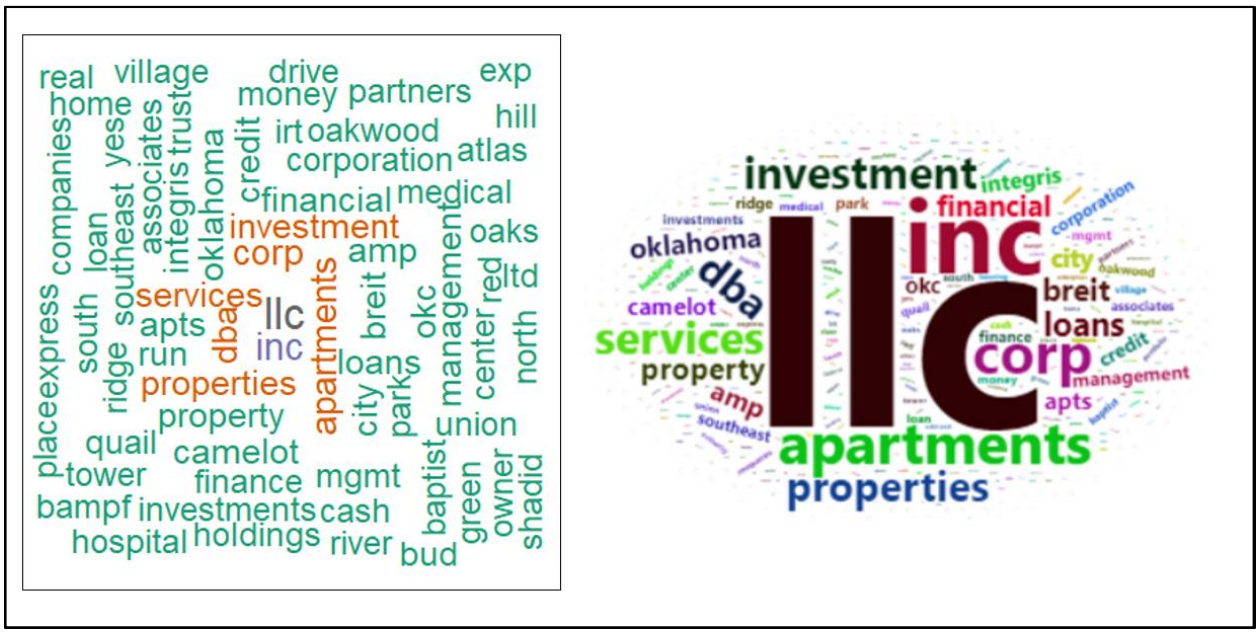


Fig10. Word cloud for plaintiffs for all small claims cases filed

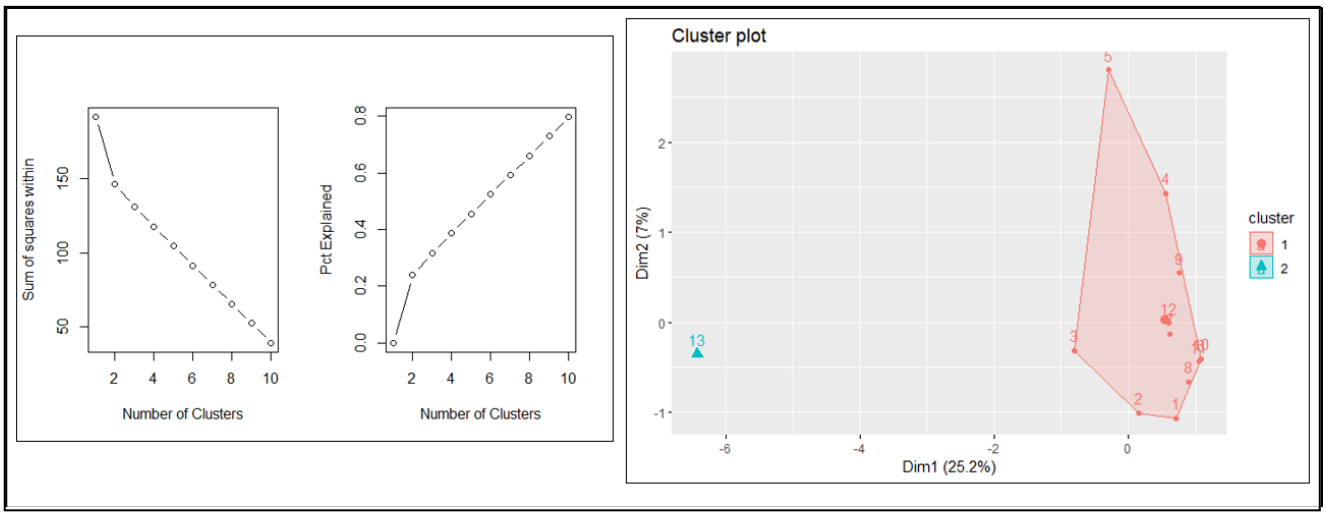


Fig11. Elbow plot and Clusters for Kmeans using sampled data for number of cases filed for all counties.

2-1 [R] EDA Narrative

First, the missing data within the data was explored. Fig.3 shows about 10% missingness in the data (Please find the data summary in the appendix). "file_date" was converted to date format and the case file year was extracted. Initial exploration showed that for all 6 years, 2018 had the highest number of filed cases across all counties with the first two county-highest being Oklahoma and Tulsa (Fig 4). The number of cases for each month was also fairly even for all the counties. The small claims issue types was analyzed to determine the most prevalent: “Indebtedness” and “Forcibly Entry and Detainer” were identified as the most prevalent. Other issue types observed were “promissory notes”, “damages,” and “random notes,” among others. The top two issue types were analyzed for all counties as shown in Figs 5 and 6. For the indebtedness cases, Oklahoma had the highest record for all months and the number exceeded the other counties by a count of about 2500 cases. Tulsa and Cleveland followed as the second and third counties with the highest number of indebtedness cases with Rogers having the least. The trend was a little different for Eviction cases. Here, Tulsa and Oklahoma had the highest and were significantly larger than the rest by about 2000 cases for all months as seen in Fig 6. In Figs 7 and 8, the effects of the Covid-19 pandemic were examined for both eviction and indebtedness cases across all counties by analyzing a pre-covid (2018), covid (2020), and post-covid (2021) years. The trends in Fig.7 showed a drastic reduction in the filed cases after 2020 and can be hypothesized to be due to the pandemic. In Fig 8, the boxplots showed a more even distribution for the indebtedness cases compared to the eviction cases, as seen in the interquartile range. From Fig 9 and 10, the word cloud showed “default” and “dismissed” cases as the most prevalent disposition on the filed cases and the plaintiffs for the filed cases (from a broad perspective) usually involved limited liability companies ((LLCs). An attempt was made at statistical modeling using the number of cases filed for the selected years in Fig7. KMeans clustering (unsupervised Learning) was used to cluster the data, named “df_stats” (Please refer to codes). The clustering was very poor and this was mainly because the data was predominantly categorical with encoded values for counties. KMODES will be a better modeling approach.

DANIEL ADDOKWEI TETTEH

APPENDIX

1. Initial Report of Data

```
describe(Cases_data)

## Cases_data
##
##  9  Variables      299829  Observations
## ---------------------------------------------------------------------------
------
## court
##        n  missing distinct
##   299829        0       13
##
## lowest : ADAIR      CANADIAN   CLEVELAND  COMANCHE   ELLIS
## highest: PAYNE      PUSHMATAHA ROGERMILLS ROGERS     TULSA
##
## ADAIR (2687, 0.009), CANADIAN (9506, 0.032), CLEVELAND (31194, 0.104),
COMANCHE
## (17502, 0.058), ELLIS (49, 0.000), GARFIELD (6060, 0.020), LOGAN (2193,
0.007),
## OKLAHOMA (119728, 0.399), PAYNE (6481, 0.022), PUSHMATAHA (315, 0.001),
## ROGERMILLS (49, 0.000), ROGERS (5686, 0.019), TULSA (98379, 0.328)
## ---------------------------------------------------------------------------
------
## casetype
##        n  missing distinct    value
##   299829        0        1       SC
##
## Value          SC
## Frequency  299829
## Proportion      1
## ---------------------------------------------------------------------------
------
## file_year
##        n  missing distinct
##   299829        0        6
##
## lowest : 2016 2017 2018 2019 2020, highest: 2017 2018 2019 2020 2021
##
## Value       2016  2017  2018  2019  2020  2021
## Frequency  65285 62935 65652 63558 40946  1453
## Proportion 0.218 0.210 0.219 0.212 0.137 0.005
## ---------------------------------------------------------------------------
------
## file_date
##        n  missing distinct
##   299825        4     1271
##
## lowest : 01/01/2016 01/02/2018 01/02/2019 01/02/2020 01/03/2017
## highest: 9/29/2017  9/29/2020  9/30/2016  9/30/2019  9/30/2020
## ---------------------------------------------------------------------------
```

```
------
## plaintiff
##        n  missing distinct
##   102310   197519    18823
##
## lowest : , GILMORE SEAN              ;OYAL LOANS OF OKLA LLC     ;OYAL
LOANS OF OKLAHOMA LLC 1 PROPERTY GROUP LLC        1 PROPERTY GROUP, LLC
## highest: ZUBIK, JOAN                 ZUNIGA, FATIMA              ZUSPANN
RENTALS             ZYX LLC                    ZYX, LLC
## ----------------------------------------------------------------------
------
## iss_desc
##        n  missing distinct
##   299792       37     4288
##
## lowest : ABANDONED PROPERTY                      ABANDONED PROPERTY $537
ABANDONED PROPERTY >$5000.00            AFFIDAVIT FOR REPLEVIN
AFFIDAVIT FOR REPLEVIN  OR INDEBTEDNESS
## highest: PROMISSORY NOTE >$5000.00.              QUIET TITLE
RELATIVE GUARDIANSHIP                   REPLEVIN
REPLEVIN - CIVIL NO DAMAGES
## ----------------------------------------------------------------------
------
## disp_date
##        n  missing distinct
##   270671    29158     1293
##
## lowest : 01/01/2018 01/02/2018 01/02/2019 01/02/2020 01/03/2017
## highest: 9/29/2017  9/29/2020  9/30/2016  9/30/2019  9/30/2020
## ----------------------------------------------------------------------
------
## disp_case
##        n  missing distinct
##   299793       36       37
##
## lowest : ADJUDICATED IN NEED OF TREATMENT           BANKRUPTCY FILED
CLOSED JUVENILE AGE 18                      CONSOLIDATED
DEFAULT JUDGEMENT
## highest: RIGHTS OF MAJORITY GRANTED (ORDER OR OTHER) SUMMARY JUDGEMENT
ENTERED                  TRANSFERRED TO ANOTHER JURISDICTION
TRANSFERRED TO FEDERAL COURT               VACATED (ORDER OR JUDGMENT)
## close_date
##        n  missing distinct
##   246656    53173     1283
## lowest : 01/01/2018 01/02/2018 01/02/2019 01/02/2020 01/03/2017
## highest: 9/29/2017  9/29/2020  9/30/2016  9/30/2019  9/30/2020
## ----------------------------------------------------------------------
------
```