

Change-point Analysis : Little Rock 2015 Crime

Jyotishka Datta

February 28, 2018

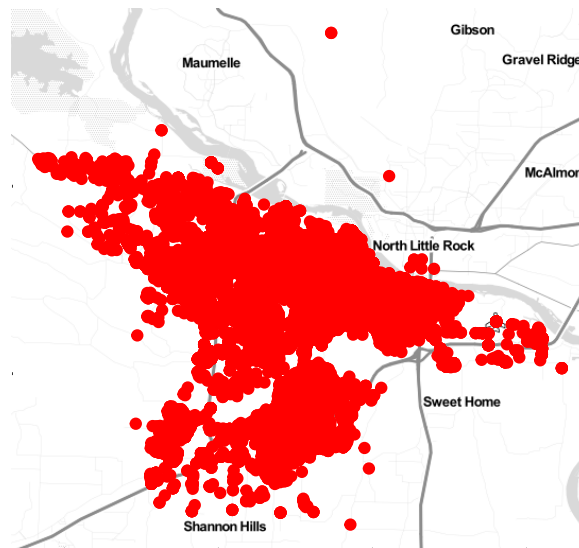
1 Little Rock 2017 Incidents Data

1.1 Background

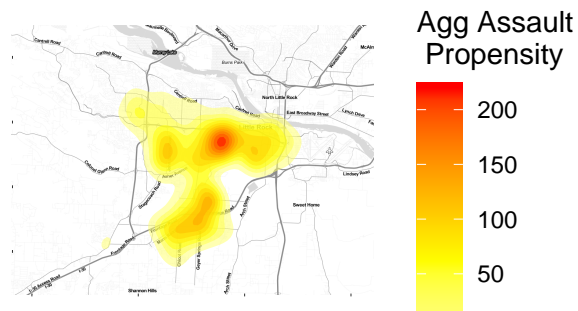
This is an exploratory analysis of a 2015 data-set that reports criminal incidents in Little rock restricted to aggravated assaults and homicides, with a few descriptive variables, such as incident type and address.

1.2 Map of Incidents

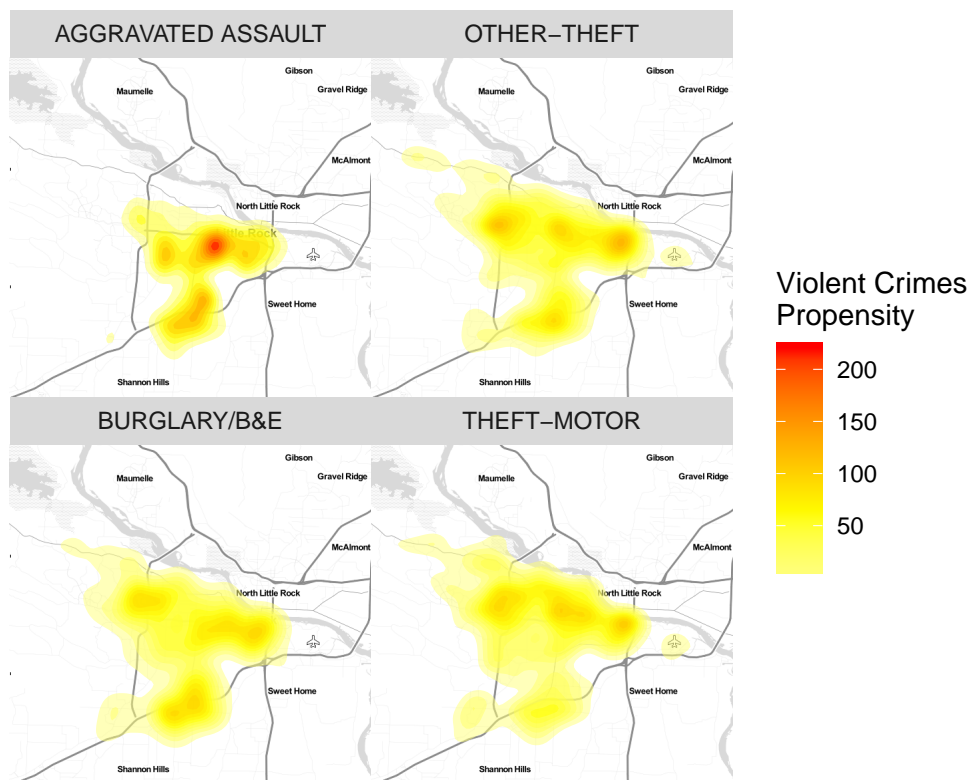
First we plot the incidents on a map like a point process using the latitude and longitudes extracted by `ggmap`.



We can look at the kernel density estimates of only aggravated assaults on the map as follows:



A natural question could be, does different incident types have different spatial concentration? It appears that Agg-assaults and homicides have a little different concentration zones.



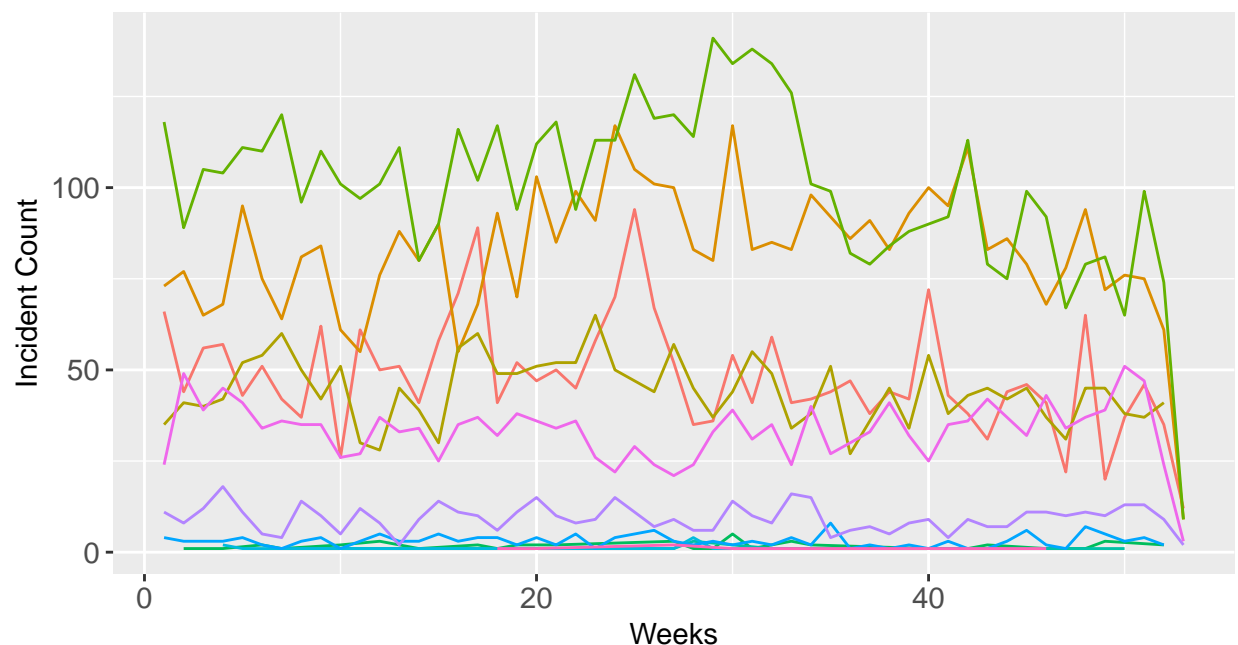
2 Trend Analysis for Weekly Counts

We count the number of incidents for each Offense Description category for each week. The goal is to see if there is any trend in the weekly counts of aggravated assaults / homicides. As

expected, there is no over-all increasing or decreasing trend but there might be small **local** change-points in the data.

```
library(lubridate)
# str(crime.gc)
crime.gc$INCIDENT_DATE <- as.character(crime.gc$INCIDENT_DATE)
crime.gc$INCIDENT_DATE <- as.Date(crime.gc$INCIDENT_DATE,format="%m/%d/%Y")
crime.gc$IncidentWeek <- lubridate::week(ymd(crime.gc$INCIDENT_DATE))
library(dplyr)

crime.ct <- crime.gc %>% group_by(OFFENSE_DESCRIPTION, IncidentWeek) %>%
  dplyr::summarise(count = n())
```

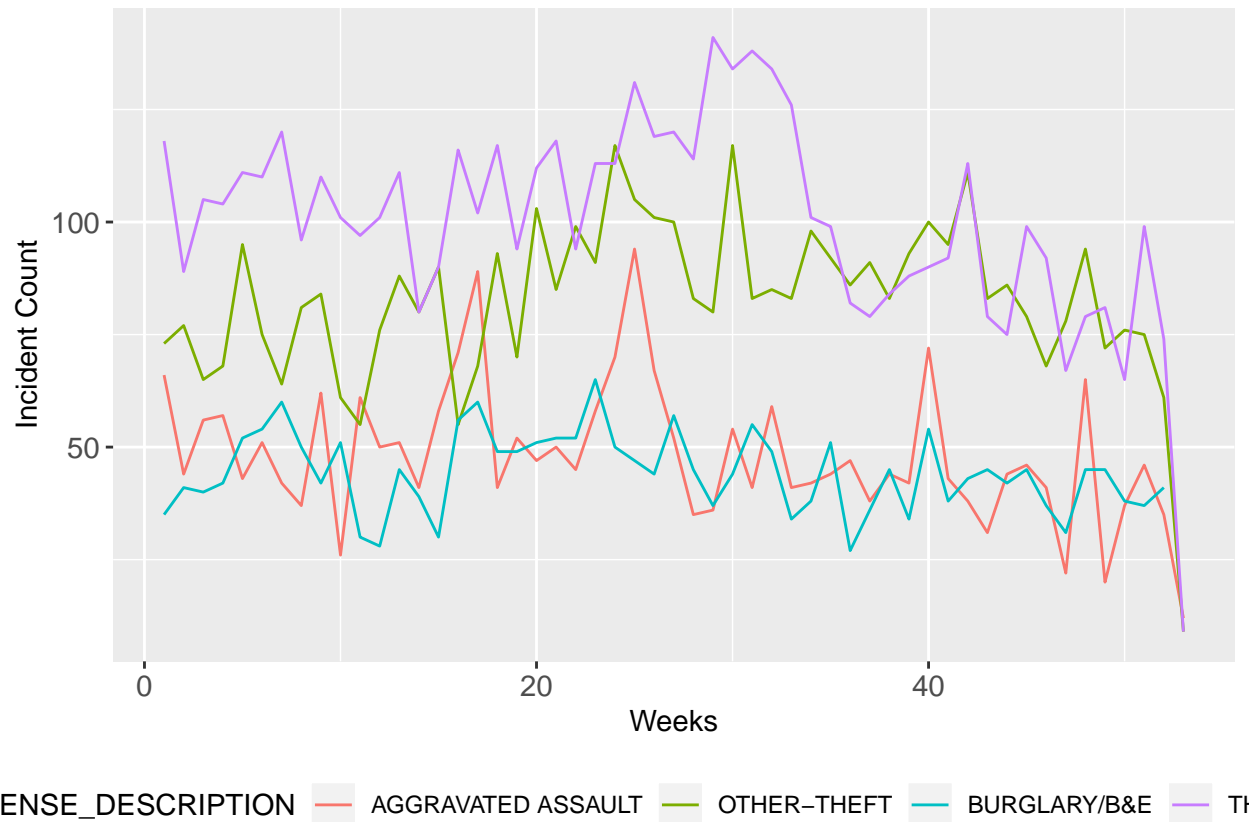


AGGRAVATED ASSAULT	THEFT-MOTOR	PURSE-SNATCHING
OTHER-THEFT	MURDER & NONNEGLIGENT MANSLAUGHTER	RAPE
BURGLARY/B&E	POCKET-PICKING	ROBBERY

2.1 SELECTED TYPES

```
violent.ct <- crime.ct %>%
  filter(OFFENSE_DESCRIPTION %in% c("AGGRAVATED ASSAULT",
    "BURGLARY/B&E",
    "THEFT-MOTOR",
    "OTHER-THEFT"))
```

```
library(ggplot2)
crime.plot <- ggplot(violent.ct, aes(x = IncidentWeek, y = count, group = OFFENSE_DESCRIPTION)) +
  geom_line() + ylab("Incident Count") +
  xlab("Weeks") + theme_bw() +
  # theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(), axis.
  # facet_grid(OffenseDes~., scales="free_y") +
  theme(legend.position="bottom")
crime.plot <- crime.plot + theme(axis.title.y = element_text(size = rel(1), angle = 90)) +
  theme(axis.title.x = element_text(size = rel(1)))
crime.plot <- crime.plot + theme(axis.text = element_text(size = rel(1)))
crime.plot <- crime.plot + theme(strip.text.x = element_text(size=12, face="bold"), strip.t
print(crime.plot)
```



2.2 Change-point Analysis (without covariates)

Next, we perform a change-point test using the PELT (Pruned Exact Linear Timing) method by Killick et al. (2012). It is impossible to know if there should be any “true” change-points in 2015 data in the absence of any covariates like policy change or new law, but this can be treated as an preliminary exploratory analysis.

PELT is one instance of a series of popular changepoint detection methods that can identify points within a data-set where some statistical properties change. Specifically, if we have ordered data-points: y_1, y_2, \dots, y_n , and m change-points τ_1, \dots, τ_m that divide the data into $m + 1$ partitions. The changepoint detection methods then seek to minimize a function:

$$\sum_{i=1}^m C(y_{\tau_{i-1}+1:\tau_i}) + \text{pen}(n)f(m),$$

where C is a cost-function and $\text{pen}(n) \times f(m)$ is the penalty applied to prevent over-fitting. Suppose we have ‘IID’ data-points $y_1, \dots, y_n \sim f(y \mid \theta)$ for some unknown underlying parameter θ . The PELT method uses the negative log-likelihood as the cost function:

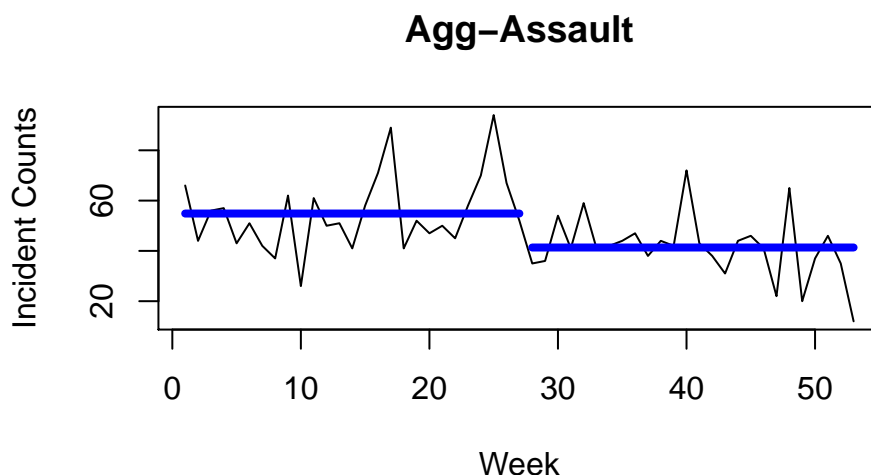
$$C(y_{(t+1):s}) = -\max_{\theta} \sum_{i=t+1}^s f(y_i \mid \theta)$$

The penalty is chosen based on the inferential goal, e.g. $\text{pen}(n) = n \log(n)$ is the popular BIC penalty and $f(m) = m$ assumes that penalization is linear with the number of change-points. When n is not too large, BIC favours a parsimonious model and can be shown to be model selection consistent. The number and location of change-points would depend on these choices as well.

```
library(changepoint)
agg.ct <- crime.ct %>%filter(OFFENSE_DESCRIPTION == "AGGRAVATED ASSAULT")

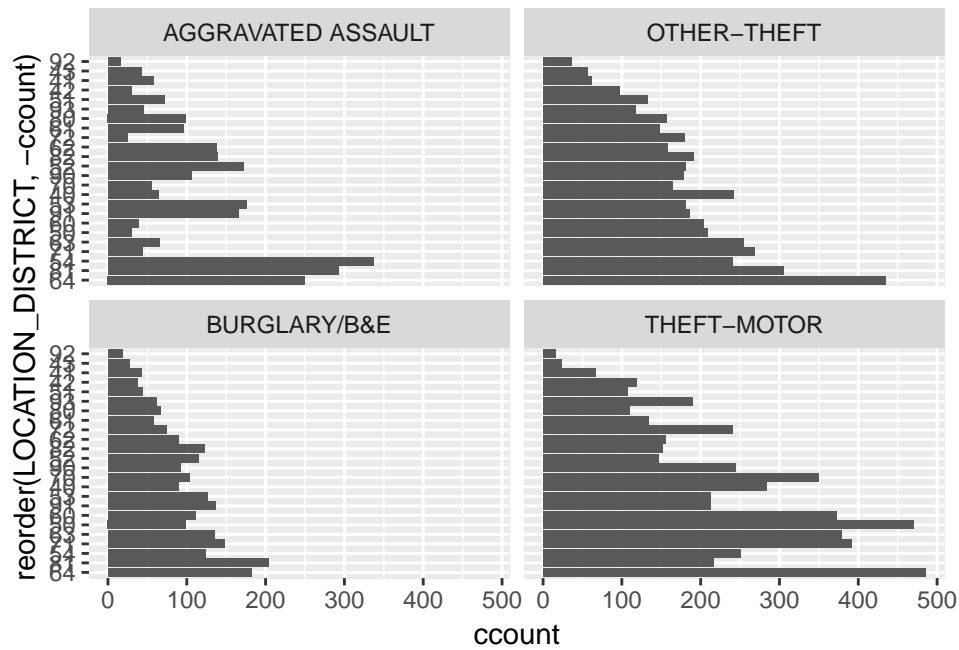
# results <- cpt.mean(agg.ct$count,penalty = "BIC", method="PELT")
results <- cpt.mean(agg.ct$count,penalty="None",method="AMOC")
cat("Location of change-points \n", cpts(results))

## Location of change-points
## 27
```

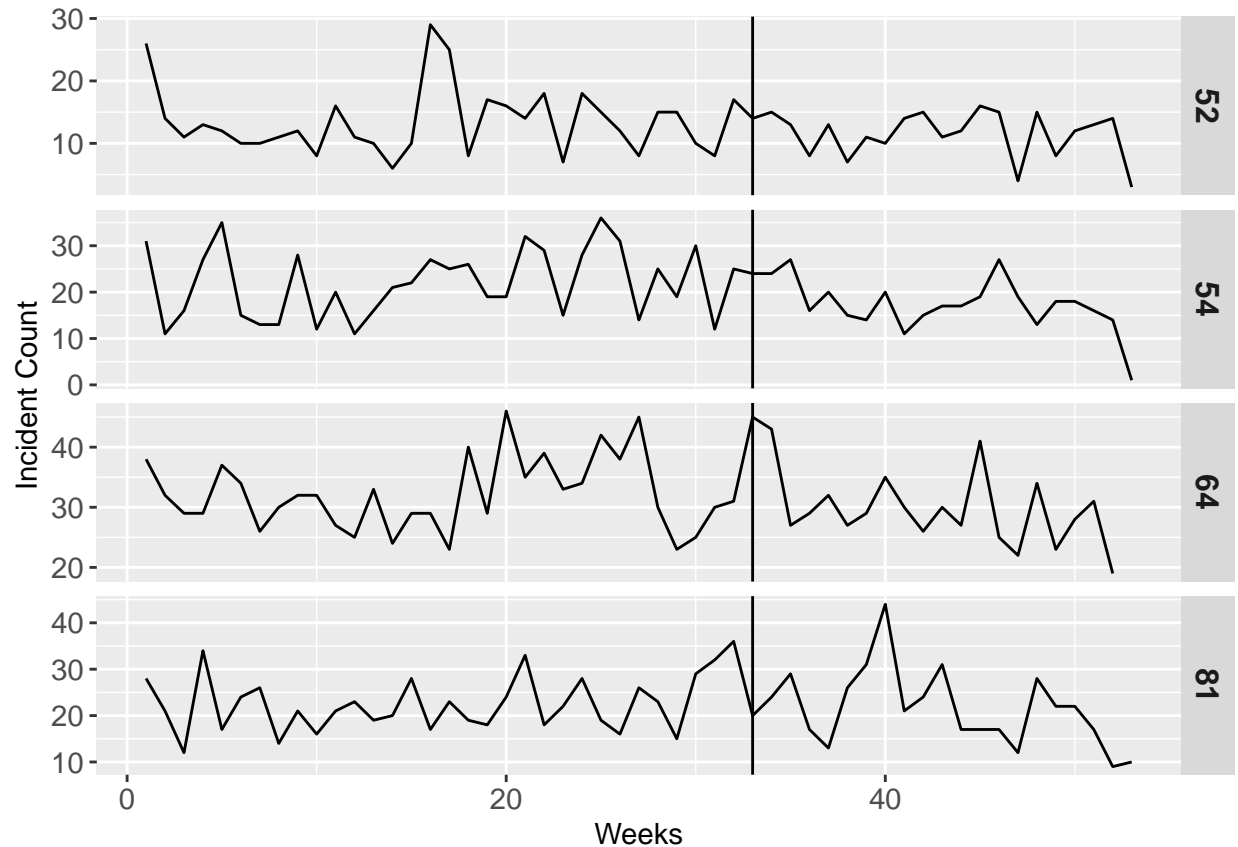


The difficult part is not yet done: which is investigating if these change-points are ‘real’, i.e. triggered by some external stimuli. For example, in May 2015, LRPD assigned an extra patrol [news link](#), which might be linked to temporary dip in agg-assault counts after week 20. Although, it is hard to make such associations without more detailed information / study.

2.2.1 District wise Analysis



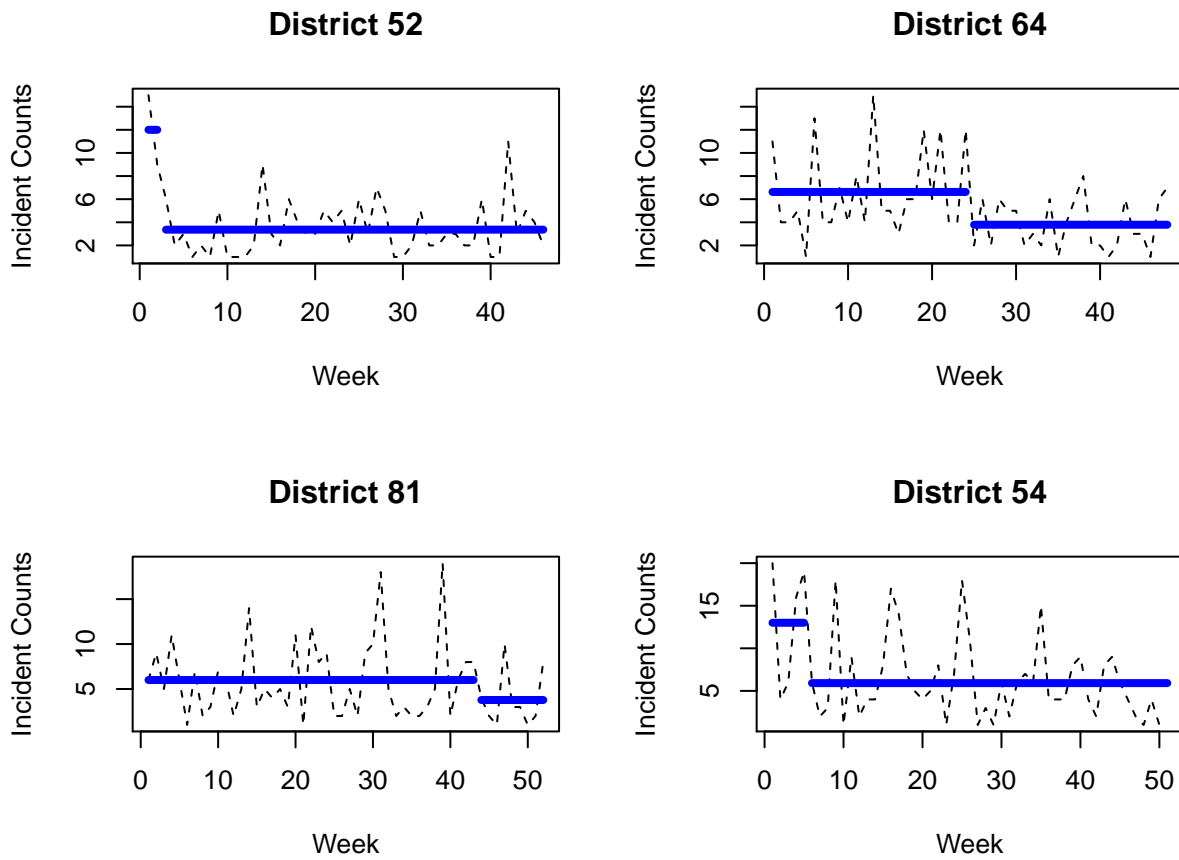
The figure above shows the distribution of district level incident counts. District 64 appears near the top. If we restrict ourselves to only a few districts with the highest number of agg-assault incidents (e.g. districts 52, 64, 81, and 54), do we see any temporal patterns?



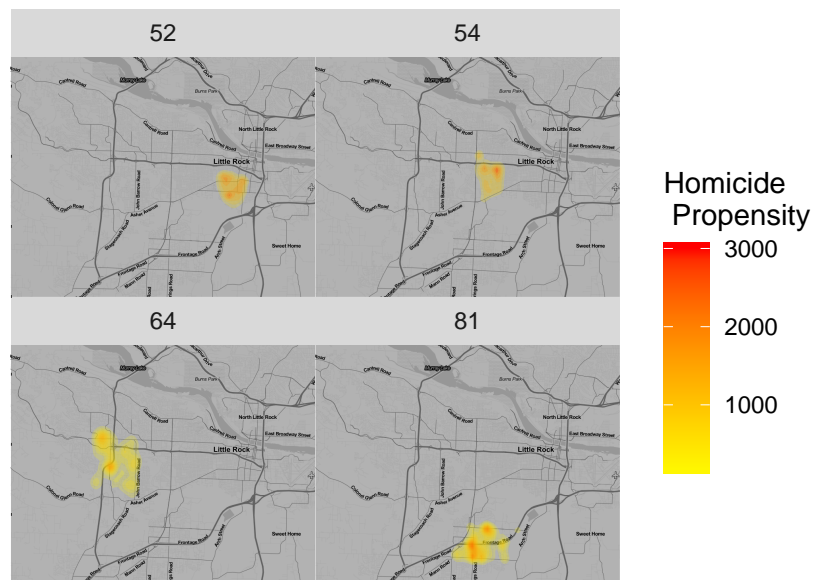
2.3 Most significant change-point

The figure below shows the change-point analysis for each of these top districts in 2015, with a minor change in the method: we restrict (artificially) the number of change-points detected to at most 1 - to detect the most significant change-point of crime frequencies in these districts. We should note here that LRPD is specially interested in District 64 because of its high volume of crimes.

```
## The Location of change-point for district 52 is week 2
## The Location of change-point for district 64 is week 24
## The Location of change-point for district 81 is week 43
## The Location of change-point for district 54 is week 5
```



Finally, where are these districts?



3 Reference:

1. PELT Algorithm: Killick R, Fearnhead P, Eckley IA (2012) **Optimal detection of changepoints with a linear computational cost**, **JASA** 107(500), 1590-1598.