

Vacant Lot Greening & Crime

Ajji Narayanan

12/6/2018

Introduction

Many American cities, particularly former industrial hubs like Detroit and Philadelphia, face the burgeoning problem of vacant lots in cities. Philadelphia has adopted an interesting solution in the form of vacant lot greening, which turns neglected vacant lots into mini public parks for local residents. The positive ecological and mental health benefits of greened vacant lots have been well documented, but one second order effect is the effect of greening on neighborhood safety. This paper seeks to analyze the impact of vacant lot greening on surrounding crime rates using data on greened vacant lots and crime in Philadelphia. We will use a Difference in Difference approach to compare crime rates around greened vacant lots to crime rates around ungreened control vacant lots. The specific question we will answer is what is the additive effect of vacant lot greening on surrounding crime rates? And how does it change after we take into account crime changes in control vacant lots and neighborhood level crime changes?

Literature Review

Some scholars believe that greened lots would increase the crime rates of the neighborhoods they're in as green spaces can provide cover for criminal activity (Kuo and Sullivan 2001b) and impede visibility/limit one's ability to easily escape (Jackson, Dannenberg, and Frumkin 2013). Finally, if green spaces bring in more people and vehicles into an area, there are more opportunities for criminals to strike in the first place. Wolfe et. al examined the the association of vegetation with crime in Philadelphia and found that the higher the rates of vegetation in the neighborhood, the higher the likelihood of theft (Wolfe and Mennis 2012). Garvin and co-authors conducted a similar analysis in 2012 on greened vacant lots in Philadelphia. They found that vandalism, burglaries, and robberies all significantly increased after a lot was greened (Garvin, Cannuscio, and Branas 2013). Another study by Branas and co-authors in 2011 found that the greened lots in Philadelphia were also associated with statistically insignificant (but still present) increases in drug use and distribution, as well as disorderly conduct (Branas et al. 2011).

On the other side of the argument are scholars who argue that green space should deter and lower crime rates. The main idea is that green spaces encourage more 'eyes on the street' and deters criminals from initiating a crime in the first place. Greenery and vegetation could also have a mentally restorative effect on potential criminals (Kaplan and Kaplan 1989). Wolfe and his coauthors (in the same study cited above) find that rates of vegetation are negatively correlated with rates of assaults and burglaries. A similar study by Kuo in and co-authors in 2012 find that vegetation is negatively correlated with aggravated assaults, and total violent crimes (Kuo and Sullivan 2001a). Branas and coauthors (in the same article cited above) find that vacant lot greening in Philadelphia led to a significant decrease in gun assaults and vandalism. A similar study of vacant lot greening in Austin by Snelgrove and coauthors found that greening was associated with a significant reduction in the total amount of crimes and an even sharper reduction in the amount of violent crimes (Snelgrove et al. 2004). Finally, another study by Kuo and co-authors in 2001 found that vacant lot greening led to a statistically significant reduction in violent crimes (Kuo and Sullivan 2001b). All in all, vacant lot greening seem to have a negative impact on gun crimes and violent crime, while other less serious place based crimes like robberies, burglaries and disorderly conducts seem to go up. The one limitation of most of these studies is that they are relatively old.

Data

For this study, we combined data from 3 sources:

- 1) LandCare greened vacant lot data (from the Pennsylvania Horticultural Society)
- 2) Vacant lot Violation data (from the Philadelphia Licenses and Inspections office)
- 3) Crime data (From the Philadelphia Police Department)

Greened Vacant Lot Data

The first LandCare dataset comes from the Pennsylvania Horticultural Society (PHS), which administers Philadelphia's vacant lot greening program. This data reported all lots in the city that have been greened through PHS's LandCare program. In total, there were 10,132 lots that were greened in the time frame between 1993 and 2017. The data gives the address of the lot, the season it was greened in (Fall or Spring) and the square footage of the lot. After filtering out lots where the date of greening wasn't recorded, we were left with 6,822 greened vacant lots.

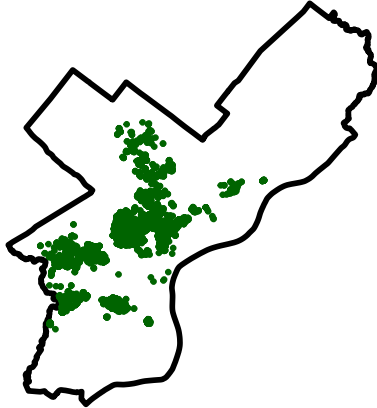
Ungreened Vacant Lot Data

The second dataset comes from the Philadelphia Licenses and Inspections office and lists all code violations in the city. The L&I office monitors building conditions in Philadelphia and issues code violations to properties that are in violation of the building code. One sub type of code violations is given to vacant and abandoned lots in the city. By filtering the data to only lots with these sub types of code violations, we obtain data on all vacant (and ungreened) lots in Philadelphia. Overall, this data gave us 15,612 ungreened vacant lots from the years 2007 to 2017.

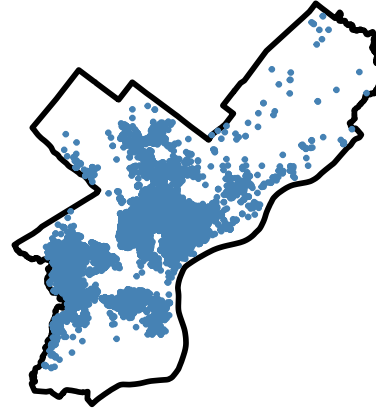
There are a couple of important notes about this dataset. First, there are many lots with multiple vacant violations and for consistency we only record the first violation. Second, a few of the lots appeared in both the LandCare greened lot data and the L&I lot violation data. We deleted these overlapping lots from the lot violation data and kept them in the LandCare greened lot data. Finally many of the violations have now been closed, implying that the lots are no longer vacant. Some violations are not closed, implying that the lots are still vacant. This allows us to construct a vacancy time frame for each lot where the vacancy start date is the date of the first issued violation and the vacancy end date is either the date the violation has been closed or the current date if the violation hasn't been closed. We limit the dataset to lots that have been vacant for at least 2 years so that we can compare crime counts a year before and after and be certain the lot was vacant for the entire time frame.

Below is a map of where all the greened and ungreened lots are located in Philadelphia.

Treated Greened Lots



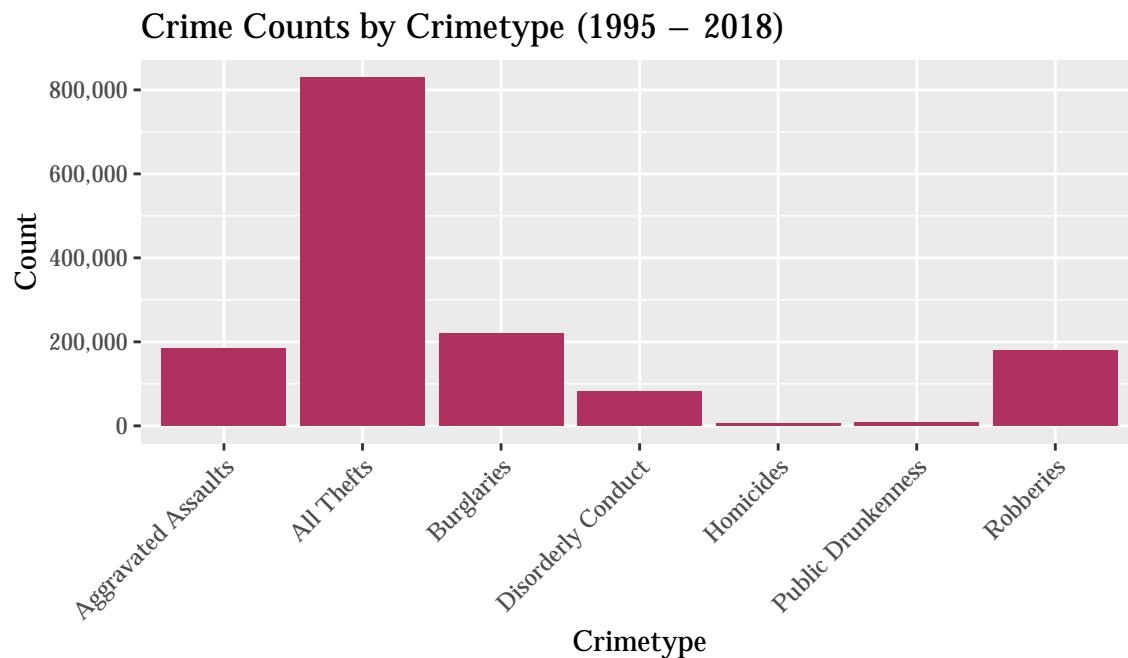
Control Vacant Lots



Notes: Locations of vacant lots in Philadelphia

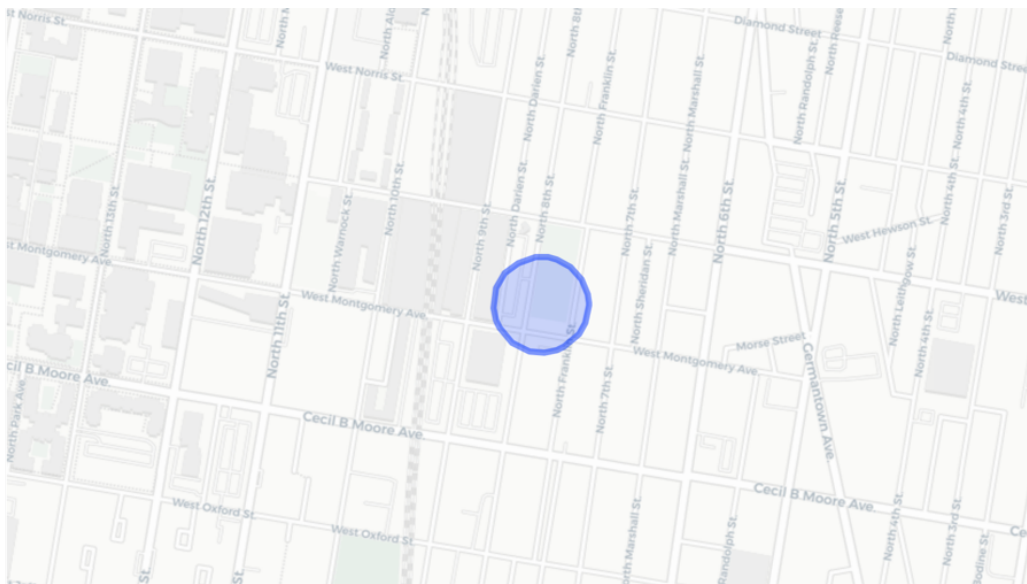
Crime Data

We obtained crime data from two separate sources. First, crime data from 1995 to 2005 was given to us by the Penn Criminology Department. And newer crime data from 2005 till October 2018 was obtained from the Philadelphia Police Department and is publicly available at opendataphilly.org. The crime data reports crimetype, latitude/longitude coordinates, and date of crime. One note about this dataset is that the types of crime that were reported in the old crime dataset were more limited than the types of crimes reported in the newer crime dataset. To maintain consistency, we only keep crimetypes which were recorded in both datasets, leaving us with 1,510,789 crime records from 1995 to 2018. Below is a histogram of the total number of crimes by crimetype across all years



Notes: These limited crimetypes are those reported in both the old and new crime data.

Finally, we combined all the above data to create one unified dataset on crime change around vacant lots in Philadelphia (both greened and ungreened). Specifically, we created a 70 meter buffer around each vacant lot and counted the number of crimes that occurred 11 months before and after the treatment date. The average block in Philadelphia has a radius of 140 meters, so a 70 meter buffer was chosen to cover the area of a full block around the lot. The 11 month before and after time frame was chosen to cancel out any seasonality in the crime data. The below chart shows what a 70 meter radius radius around an example lot in the data looks like.



We also assigned an intervention date to all the lots as follows. For greened lots, the intervention date was the date the lot was greened. For ungreened lots, the intervention date was the midpoint of the vacancy interval. So if a lot was vacant from January 2011 to January 2013, the treatment date was assigned as January 2012. This can be thought of as the date the lot would have been greened, had it been chosen to be a part of the LandCare program. Our final dataset was a listing of addresses, latitude/longitude pairs, a binary greening treatment variable, and crime statistics for each of our 22,434 lots. Below is a short excerpt of 3 random lots in our dataset.

Address	Intervention Date	Greened	Num Crimes before (70m)	Num Crimes after (70m)
3006 Broad St	2008-12-28	FALSE	2	9
1701 S 22nd St	2004-03-01	TRUE	6	6
1239 N 19th St	2015-04-28	FALSE	4	2

We first conduct 2 aggregate comparisons of control and greened lots using Wilcoxon rank-sum tests and then perform DID regressions.

I. Aggregate Comparisons (Unstandardized)

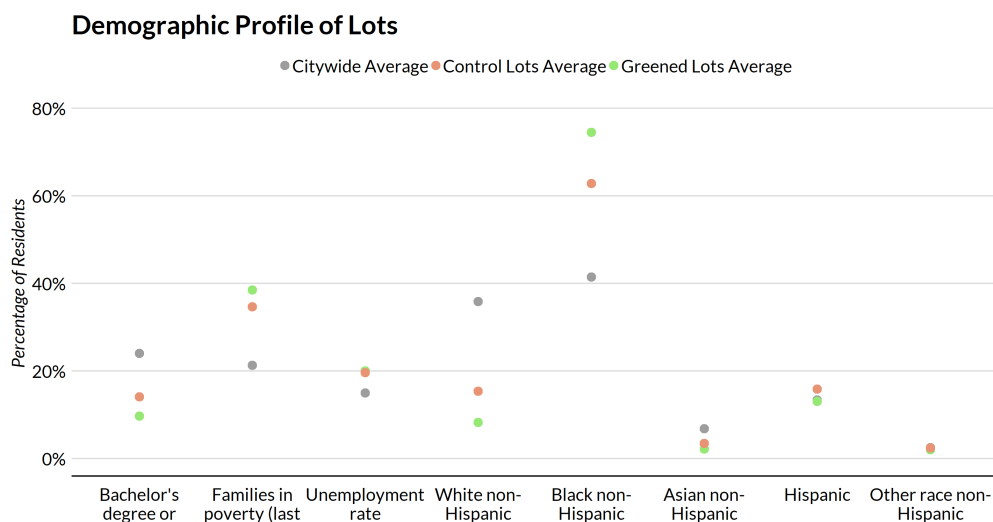
For the first part of our data analysis, we analyzed how crime rates had changed in aggregate at a 70 meter radius around greened lots and around ungreened control lots. The below table shows the average difference in number of crimes in the 11 months pre and post intervention date for greened vacant lots and control vacant lots broken down by crimetype. The Difference in Difference column is simply the average difference in number of crimes around greened lots minus the average difference in number of crimes around control lots. It represents the additive effect of greening on crime once you take into account crime trends around control neighborhoods. We also conduct a Wilcoxon rank-sum test to see if there is a statistically significant difference between crime change around greened lots and crime change around control lots. The Wilcoxon rank-sum test is similar to a t-test and assesses whether two means are significantly different from each other. The key advantage is that the Wilcoxon test allows for non-normal distributions around the means.

Crimetype	Average difference (greened)	Average difference (control)	Difference in Difference	Wilcox test p value
Aggravated Assaults	0.036	-0.062	0.099	0.0031
Burglaries	-0.064	-0.131	0.066	0.0531
Disorderly Conduct	-0.057	-0.043	-0.013	0.0027
Homicides	0.010	0.000	0.010	0.6598
Public Drinking	0.005	-0.002	0.008	0.0195
Robberies	-0.116	-0.045	-0.071	0.0006
Thefts	-0.049	-0.009	-0.039	0.0548
Violent Crime	-0.069	-0.107	0.038	0.1641
Non Violent Crime	-0.165	-0.186	0.021	0.3493
All Crimes	-0.234	-0.294	0.059	0.2109

Notes: Average difference is the number of crimes 11 month after - number of crimes 11 months before within 70 meter buffer zone.

There are a couple of notable results. First, the number of total crimes decreases more around control lots (-0.29) than around greened lots (-0.23), and the difference in differences is +0.06. This means even though crime had reduced around greened lots, greened lots had on average 0.06 more crimes than the control lots. In other words, greening seems to have slowed down the crime reduction that would have taken place otherwise. The crimetypes with significant positive difference in differences (at a 5% confidence level) were the number of aggravated assaults (0.10) and public drinking (0.01). The crimetypes with significant negative difference in differences (at a 5% confidence level) were disorderly conduct and robberies. This initial analysis seems to suggest that greening actually increases the total number of crimes, particularly aggravated assaults and public drinking, relative to control vacant lots.

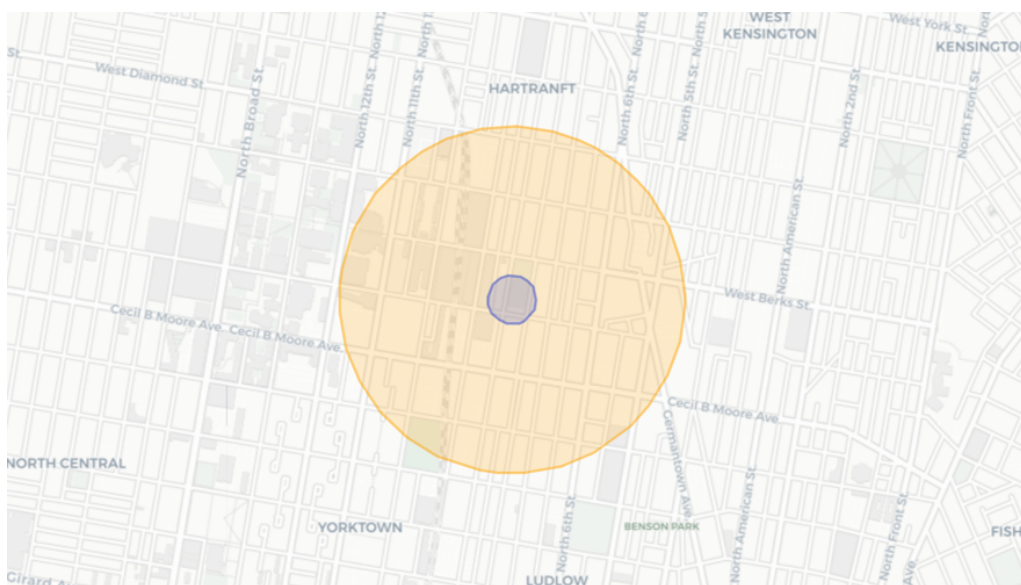
This analysis has one big shortcoming, Namely, we've assumed that control lots and vacant lots are directly comparable and the only difference between them is that some of them were greened. However, greened lots and control lots as an aggregate tend to be located in different kinds of neighborhoods. Below is a comparison of the socioeconomic characteristics of the average neighborhood that control lots and greened lots are located in. Control vacant lots tend to be in neighborhoods with higher poverty rates, higher share of black residents, lower share of white residents and lower college education rates than the average neighborhood in Philadelphia. Greened lots tend to be in neighborhoods with even higher poverty rates, even higher share of black residents, even lower share of white residents and even lower college education rates. We needed a way of measuring crime change while also taking into account the different surrounding neighborhoods of greened and control vacant lots.



Notes: These statistics were calculated by using a prototype Data Bias Assessment tool from the Urban Institute which I helped create over the summer. It uses tract level Census data to calculate citywide demographic averages, and then uses the proportion of the datapoints falling into each tract to ‘weight’ the Census variables and return the demographic averages implied by the data.

II. Aggregate Comparisons (Standardized)

So, we decided to create a crime metric for each lot that takes into account the crime change in the surrounding neighborhoods. We created a standardized crime change variable that was the difference in crimes at a 70 meter buffer minus the number of crimes at a 500 meter ‘doughnut’ buffer. In terms of the below chart, the standardized crime change is the difference in crimes within the blue circle minus the difference in crimes within the orange ‘doughnut’.



Notes: 70 meter and 500 meter buffers zones for 1803 N 8th Street

The crime change at a 500 meter ‘doughnut’ buffer tells us about the neighborhood level crime trends. And the crime change at a 70 meter buffer tells us about the localized crime trends around a vacant lot. The

difference between these crime changes (ie the Difference in Difference) will tell us what the additive effect of greening is relative to the neighborhood level crime trends. The crime change was also standardized per square kilometer - that is divided by the area of the buffer in square kilometers - so that the differences in crimes could be directly comparable across different buffer sizes. Below are the mean standardized differences in number of crimes per square kilometer for greened vacant lots and control vacant lots broken down by crimetype, along with the Wilcoxon rank-sum test's p-value for the differences in differences.

Crimetype	Standardized difference (greened)	Standardized difference (control)	Difference in Difference	Wilcox test p value
Aggravated Assaults	4.506	-2.364	6.870	0.000
Burglaries	-1.850	-2.446	0.596	0.514
Disorderly Conduct	-2.350	0.329	-2.678	0.996
Homicides	0.826	0.058	0.769	0.012
Public Drinking	0.241	-0.297	0.537	0.600
Robberies	-5.029	-0.512	-4.517	0.036
Thefts	-2.577	0.302	-2.878	0.060
Violent Crimes	0.304	-2.818	3.122	0.057
Non Violent Crimes	-6.536	-2.113	-4.423	0.477
All	-6.232	-4.931	-1.302	0.991

Notes: Standardized crime counts are the number of crimes per square Km in the 70 meter buffer - number of crimes per square Km in the 500 meter doughnut buffer. The standardized differences is the standardized crime counts 11 months after - the standardized crime counts 11 months before.

For greened lots, all crimes except for aggravated assaults, homicides and public drinking went down in relation to the neighborhood level crime trends. And for control lots, all crimes except for homicides, disorderly conduct and thefts went down in relation to the neighborhood level crime trends. Since homicides are a low frequency and unpredictable crime type, this result should not be read into too much. Finally the difference in differences - which can be thought of as the additive effect of greening on crime in relation to neighborhood level crime trends and control vacant lots - show similar results to the unstandardized results above. Namely, the number of disorderly conduct, robberies and thefts went down while the number of burglaries, homicides, and public drinking went slightly up. Interestingly, the total number of crimes and non violent crimes go down (thought not at a statistically significant level), in contrast to the unstandardized analysis above. This is because while the signs are the same between the unstandardized and standardized analyses, the magnitudes of the negative results are larger in the standardized analysis.

For robustness, we replicated the above analysis at a buffer distance of 150 meters instead of 70 meters, and the results stayed mostly the same.

III. Difference in Differences Regression

The above aggregate comparisons are very similar in methodology to what a Difference in Differences regression does. To check the validity of our results, we also conduct a formal DID regression. First our dataset had to be transformed to look like the following:

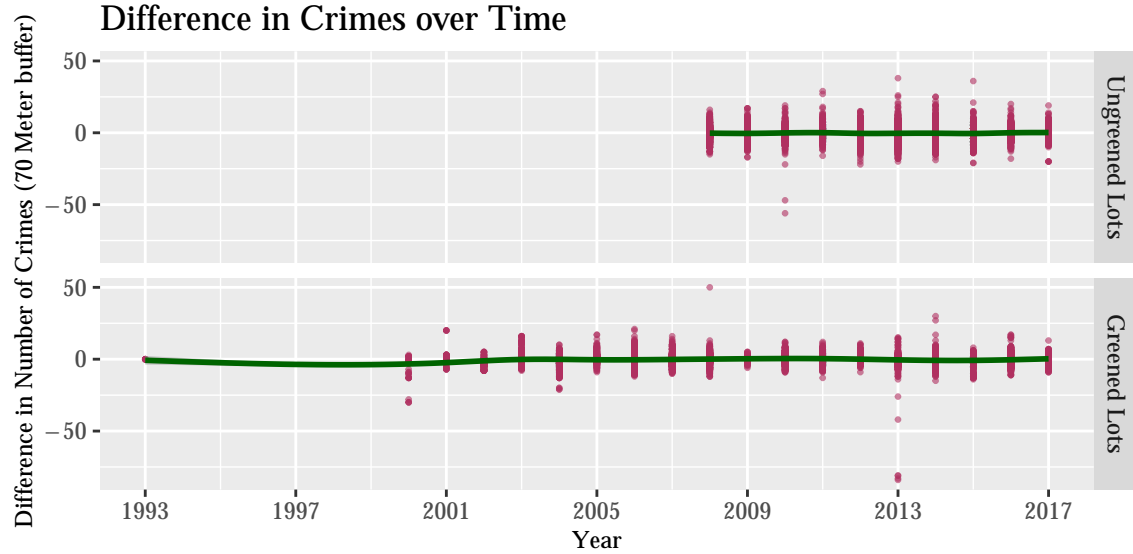
Address	Intervention Date	Greened	Time	Num Crimes
3006 Broad St	2008-12-28	0	0	2

Address	Intervention Date	Greened	Time	Num Crimes
3006 E Thompson St	2008-12-28	0	1	2
1701 S 22nd St	2004-03-01	1	0	6
1701 S 22nd St	2004-03-01	1	1	6

Then The regression setup is as follows

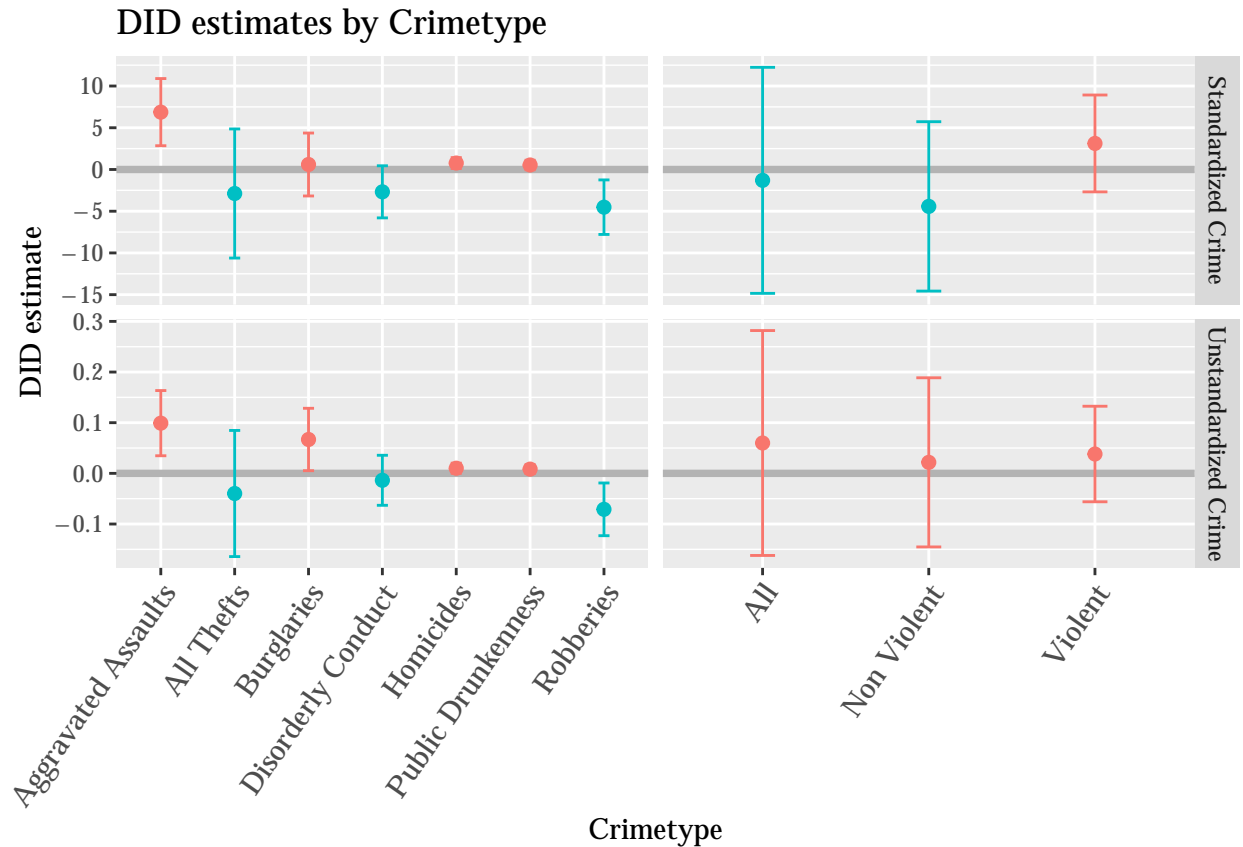
$$Y_i = \beta_0 + \beta_1 X_{Time} + \beta_2 X_{Greened} + \beta_3 X_{Greened:Time}$$

Where Y_i is the number of crimes. In this case the DID coefficient of interest is B_3 which tells us what the additive effect of a lot being greened is on crime rates. There are some important notes about our approach. First, this approach assumes there are no time specific fixed effects. In other words there should be no additional effect on crime whether a lot was greened in 2001 or 2018. This is an important assumption because our dataset spans the 21 years from 1997 to 2018. We verified this assumption was valid by plotting the difference in number of crimes for all the lots in the dataset. For easier viewing, we also overlaid a local polynomial regression line in green. There seem to be no clear patterns over time for either greened or ungreened control lots, so our assumption of no time specific fixed effects is valid.



Notes: Due to data limitations, the Ungreened control lots span fewer years than the greened lots. Each point represents a lot in the dataset. The x axis is the year of the intervention data and the y axis is the difference in total number of crimes 11 months after vs 11 months before in 70 meter buffer zone.

We run a series of DID regressions using different values for the dependent variable Y_i . The dependent variables we used were total number of crimes broken down by crimetype (ie total number of Aggravated Assaults, Burglaries, Thefts etc) as well as number of non violent crimes, violent crimes, and all crimes. We also used standardized crime counts (which took into account the number of crimes in the surrounding neighborhood using the ‘doughnut’ approach described above) as dependent variables. The DID estimates from each of those regression is shown in the below graph



Notes: Standardized crime counts were computed exactly as in the Aggregate Comparisons. It is the number of crimes per square Km in the 70 meter buffer - number of crimes per square Km in the 500 meter doughnut buffer.

These chart mirrors the results found in the aggregate comparison section. When looking at the unstandardized crime counts, the DID estimate is negative only for thefts, disorderly conducts, and robberies. This remains true for the standardized crime counts, but now the DID estimates for total number of crimes and non violent crimes flips and becomes negative. In short, greening reduced the total number of non violent crimes, mostly through reductions in thefts, robberies, and disorderly conduct citations. There also seems to be an increase in violent crimes driven mostly by increases in aggravated assaults. The only strongly significant effects (at the 5% level) are the decrease in robberies and the increase in aggravated assaults.

The big takeaway is that greening reduces the number of non violent crimes - mainly through reductions in disorderly conduct citations, thefts and robberies - after taking into account crime trends in the surrounding neighborhood and crime trends around control vacant lots. These results are for the most part statistically significant at the 5% level. There are also statistically insignificant increases in the number of burglaries, homicides, public drinking citations, total violent crimes and a large and statistically significant increase in the number of aggravated assaults. This interestingly confirms some of the previous literature by showing reductions in disorderly conduct citations and thefts but also contradict the literature by showing an increase in violent crimes like aggravated assaults.

Limitations

This analysis has plenty of limitations and should be taken with a grain of salt. To list a few:

- 1) There is large potential for cross contamination in our results as many vacant and control lots are clustered closely together. In other words crime change could be due to the presence of multiple greened

lots within the study radius. In this paper we don't make any attempts to adjust for the number of surrounding greened/control vacant lots,

- 2) The parallel trend assumption of the DID regression is likely not met as the control and treatment lots are not exactly the same. Due to data limitations, the control lots only span the years 2001 to 2018 while the greened lot data span the years 1993 - 2018. In addition the control and treatment lots were in slightly different parts of Philadelphia and in socioeconomically different neighborhoods. Overall, it is highly unlikely our quasi experimental approach successfully
- 3) We tried to correct for the above problem by using doughnut 'standardized' crime counts. However our doughnut standardization increases the possibility of cross contamination by using an even larger radius of 500 meters for comparison.

Code Appendix

Below is all the code used in the R Markdown Document. To see the data cleaning/merging code, check out the other submitted R file - vacant_lots_data_readin_and_merge.R.

```
knitr::opts_chunk$set(echo = FALSE, warning = F, message = F, fig.align="left", cache = T)
knitr::opts_knit$set(root.dir = "C:/Users/ajjit/Google Drive/Documents/vacant_lots_final/vacant-lots-pr

library(png)
library(knitr)
library(grid)
library(data.table)
library(lubridate)
library(rgdal)
library(stringr)
library(scales)
library(mapttools)
library(sf)
library(broom)
library(tidyr)
library(lattice)
library(tidyverse)
# library(gridExtra)
# library(ggspatial)
library(mapview)
library(leaflet)
# library(htmltools)
library(patchwork)
library(extrafont)
select = dplyr::select
filter = dplyr::filter
melt = reshape2::melt
expand = tidyr::expand

loadfonts()

### Read in final cleaned dataset for use in Rmd plots
lots = readRDS("C:/Users/ajjit/Google Drive/Documents/vacant_lots_final/vacant-lots-project/final_data/

### Seperate lots into greened and ungreened, also read in Philly shapefile from web
lots_greened = lots %>% filter(greened) %>% st_transform(4326)
lots_ungreened = lots %>% filter(!greened) %>% st_transform(4326)
phl_city_bounds = st_read("http://data.phl.opendata.arcgis.com/datasets/405ec3da942d4e20869d4e1449a2be4

greened_lots_map = ggplot()+
  geom_sf(data = phl_city_bounds, col = "black", fill = "transparent",
    size = 1, show.legend = "point") +
  geom_sf(data = lots_greened, col = "darkgreen", fill = "darkgreen", size = 0.5, show.legend = F) +
  theme( axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid.major = element_line(colour = 'transparent'),
```

```

    panel.grid.minor = element_line(colour = 'transparent'),
    text = element_text(size = 9, family = 'CM Roman'),
    plot.title = element_text(size=11,family = 'CM Roman' ),
    plot.background = element_rect(fill = "transparent"),
    panel.grid = element_blank(),
    line = element_blank(),
    rect = element_blank())+
ggtitle("Treated Greened Lots")

ungreened_lots_map = ggplot()+
geom_sf(data = phl_city_bounds, col = "black", fill = "transparent",
        size = 1, show.legend = "point") +
geom_sf(data = lots_ungreened, col = "steelblue", fill = "steelblue", size = 0.5, show.legend = F) +
theme( axis.text = element_blank(),
        axis.ticks = element_blank(),
        panel.grid.major = element_line(colour = 'transparent'),
        panel.grid.minor = element_line(colour = 'transparent'),
        text = element_text(size = 9, family = 'CM Roman'),
        plot.title = element_text(size=11,family = 'CM Roman' ),
        plot.background = element_rect(fill = "transparent"),
        panel.grid = element_blank(),
        line = element_blank(),
        rect = element_blank())+
ggtitle("Control Vacant Lots")

### Use patchwork to show two ggplots side by side
greened_lots_map + ungreened_lots_map

crimes = readRDS('intermediate_data/crimes_agg_dt.rds')

crimetypes_over_time = ggplot(crimes, aes(x = crimetype_05)) +
  geom_bar(fill = "maroon")+
  xlab("Crimetype") +
  ylab("Count") +
  ggtitle("Crime Counts by Crimetype (1995 - 2018)") +
  scale_y_continuous(label = comma) +
  theme(plot.title =element_text(size=13, family="CM Roman"),
        axis.text.x = element_text(size=9, family="CM Roman", angle = 45, hjust = 1),
        axis.text.y = element_text(size=9, family="CM Roman"),
        axis.title = element_text(size=11, family="CM Roman"),
        text = element_text(family = "CM Roman"))

crimetypes_over_time

#Display 70 meter radius plot (generated in kepler.gl)
img <- readPNG("C:/Users/ajjit/Google Drive/Documents/vacant_lots_final/vacant-lots-project/writeup/gray
grid.raster(img)

```

```

### Defining a crimetype name translation df for easier plotting
crime_trans_df = data.table(long_name = c('Aggravated Assaults', "All Thefts", "Burglaries",
                                           "Disorderly Conduct", "Homicides",
                                           "Public Drunkenness", "Robberies", "All", "Violent",
                                           "Non Violent"),
                             crimetype = c("agg_ass", "thefts", 'burg', 'disord', "homi",
                                           'pub_dr', "robb", "all",
                                           "vio", "nonvio"))

##### Helper fcn to create Agg. Comparison Table
make_eda_table = function(lots, radius){

  radius_m = paste0(radius, "m")
  lots1 = lots %>% st_set_geometry(NULL)
  lots_radius = lots1 %>%
    select(greened, contains(radius_m)) %>%
    select(greened, contains("diff"))

  ##### Doing T and Wilcox Tests #####

  std_cols_to_test = lots_radius %>% select(greened, contains("std")) %>%
    gather(key = "stat", value = "value", -greened) %>%
    mutate(greened = !greened)

  abs_diff_cols_regex = paste0("^diff\\w+", radius, "m$")
  ustd_cols_to_test = lots_radius %>% select(greened, matches(abs_diff_cols_regex)) %>%
    gather(key = "stat", value = "value", -greened) %>%
    mutate(greened = !greened)

  #helper do_test fcn for t tests and wilcox tests
  do_test = function(cols, test = "t", p_val_name){

    if (test == "t"){
      r = cols %>%
        group_by(stat) %>%
        do(tidy(t.test(value~greened, data=. ))) %>%
        ungroup() %>%
        mutate(crimetype = stringr::str_replace_all(stat, "std_diff_", ""),
               crimetype = stringr::str_replace_all(crimetype, paste0("_", radius_m,
                                                                       "_per_km2"), ""),
               crimetype = stringr::str_replace_all(crimetype, paste0("_", radius_m,
                                                                       ""),
               crimetype = stringr::str_replace_all(crimetype, "diff_", "")) %>%
        select(p.value, crimetype) %>%
        rename(!p_val_name := p.value)
    } else if (test == "w"){

      r = cols %>%
        group_by(stat) %>%
        do(glance(wilcox.test(value~greened, data=. ))) %>%

```

```

    ungroup() %>%
    mutate(crimetype = stringr::str_replace_all(stat, "std_diff_", ""),
           crimetype = stringr::str_replace_all(crimetype, paste0("_", radius_m,
                                                                    "_per_km2"), ""),
           crimetype = stringr::str_replace_all(crimetype, paste0("_", radius_m,
                                                                    ""),
           crimetype = stringr::str_replace_all(crimetype, "diff_", "")) %>%
    select(p.value, crimetype) %>%
    rename(!p_val_name := p.value)
  } else {
    stop("test parameter is not valid, must be 't' or 'w'")
  }
}

return(r)
}

std_diff_normed_area_t_test = do_test(std_cols_to_test, test="t", "std_t_test_p_value")
std_diff_normed_area_w_test = do_test(std_cols_to_test, test="w", "std_w_test_p_value")
ustd_diff_normed_area_t_test = do_test(ustd_cols_to_test, test="t", "ustd_t_test_p_value")
ustd_diff_normed_area_w_test = do_test(ustd_cols_to_test, test="w", "ustd_w_test_p_value")

tests = std_diff_normed_area_t_test %>%
  left_join(std_diff_normed_area_w_test) %>%
  left_join(ustd_diff_normed_area_t_test) %>%
  left_join(ustd_diff_normed_area_w_test)

#### Creatning aggregate means, differences, and p-val in final dataframe ####
s = lots_radius %>%
  gather(key = "stat", value = "value", -greened) %>%
  group_by(stat, greened) %>%
  summarize_all(mean) %>%
  spread(greened, value) %>% ungroup() %>%
  mutate(crimetype = stringr::str_replace_all(stat, "std_diff_", ""),
         crimetype = stringr::str_replace_all(crimetype, paste0("_", radius_m,
                                                                    "_per_km2"), ""),
         crimetype = stringr::str_replace_all(crimetype, paste0("_", radius_m,
                                                                    ""),
         crimetype = stringr::str_replace_all(crimetype, "diff_", ""))
  ) %>%
  rename(greened = "TRUE",
         ungreened = "FALSE") %>%
  arrange(stat) %>%
  group_by(crimetype) %>%
  summarize(
    abs_diff_g = dplyr::nth(greened, n = 1),
    abs_diff_g_per_km2 = dplyr::nth(greened, n = 2),
    std_abs_diff_g_per_km2 = dplyr::nth(greened, n = 3),
    abs_diff_c = dplyr::nth(ungreened, n = 1),
    abs_diff_c_per_km2 = dplyr::nth(ungreened, n = 2),

```

```

    std_abs_diff_c_per_km2 = dplyr::nth(ungreened, n = 3)) %>%
  ungroup() %>%
  mutate(diff_abs_diff = abs_diff_g - abs_diff_c,
         diff_std_diff_per_km2 = std_abs_diff_g_per_km2 - std_abs_diff_c_per_km2) %>%
  left_join(tests, by = "crimetype") %>%
  arrange(factor(crimetype,
                 levels = c("agg_ass", "burg", "disord", "homi", "pub_dr",
                           "robb", "thefts", "vio", "nonvio", "all")))

s = s %>% select(crimetype, abs_diff_g, abs_diff_c, diff_abs_diff, ustd_w_test_p_value,
                ustd_t_test_p_value, abs_diff_g_per_km2, abs_diff_c_per_km2,
                std_abs_diff_g_per_km2, std_abs_diff_c_per_km2, diff_std_diff_per_km2,
                std_w_test_p_value, std_t_test_p_value)

return(s)
}

lots_70m_eda = make_eda_table(lots, 70)
# Wrote table below in Markdown because alignment was very bad in stargazer/kable

#Display demographic profile of lots plot (generated in kepler.gl)
img <- readPNG("C:/Users/ajjit/Google Drive/Documents/vacant_lots_final/vacant-lots-project/writeup/graphs/
grid.raster(img)

#Display 70/500 meter radius plot (generated in kepler.gl)
img <- readPNG("C:/Users/ajjit/Google Drive/Documents/vacant_lots_final/vacant-lots-project/writeup/graphs/
grid.raster(img)

#add Year column for easier plotting
lots = lots %>%
  mutate(year_season_begin = year(date_season_begin))

# convert logical vec to human readable names
lot_greened_names = c(`TRUE` = "Greened Lots",
                      `FALSE` = "Ungreened Lots")

theme_lato = theme( title = element_text(size = 11, family = 'CM Roman'),
                    axis.title = element_text(size = 9, family = 'CM Roman'),
                    axis.text = element_text(size = 9, family = 'CM Roman'),
                    strip.text = element_text(size = 9, family = 'CM Roman'),
                    legend.text= element_text(size=11, family = 'CM Roman'))

ggplot(lots, aes (x = year_season_begin, y = diff_all_70m)) +
  geom_point(col = "maroon", alpha = 0.6, size = 0.5) +
  geom_smooth(formula = y~x, col = "darkgreen", alpha = 0.5) +
  labs(x = "Year", y = "Difference in Number of Crimes (70 Meter buffer)",
       title = "Difference in Crimes over Time" ) +

```

```

scale_x_continuous(breaks = c(1993, 1997, 2001,
                             2005, 2009, 2013, 2017)) +
facet_grid(greened ~ ., labeller = as_labeller(lot_greened_names)) +
theme_lato

#####
##### Setup for ID regressions #####
#####

lots = readRDS("final_data/lots.rds")

### Defining a crimetype name translation df for easier plotting
crime_trans_df = data.table(crimetype = c('Aggravated Assaults', "All Thefts", "Burglaries",
                                           "Disorderly Conduct", "Homicides",
                                           "Public Drunkenness", "Robberies", "All", "Violent",
                                           "Non Violent"),
                             short_name = c("agg_ass", "thefts", 'burg', 'disord', "homi",
                                             'pub_dr', "robb", "all",
                                             "vio", "nonvio"))

### Setting up dataframe for use in DID - need to melt and reshape data

lots_70m_did = lots %>% select(contains("70m"), contains("500m"),
                             greened, date_season_begin, full_address) %>%
  select(-contains("diff")) %>%
  mutate(greened = as.numeric(greened)) %>%
  st_set_geometry(NULL)

lots_70m_expanded = lots_70m_did %>%
  #expand and left join by a dummy time variable
  tidyr::expand(after = c(0,1), full_address) %>%
  left_join(lots_70m_did, by = "full_address") %>%
  arrange(full_address) %>%
  mutate(all_70m = dplyr::if_else(after==0, num_all_70m_before, num_all_70m_after),
         agg_ass_70m = dplyr::if_else(after==0, num_agg_ass_70m_before, num_agg_ass_70m_after),
         thefts_70m = dplyr::if_else(after==0, num_thefts_70m_before, num_thefts_70m_after),
         burg_70m = dplyr::if_else(after==0, num_burg_70m_before, num_burg_70m_after),
         disord_70m = dplyr::if_else(after==0, num_disord_70m_before, num_disord_70m_after),
         homi_70m = dplyr::if_else(after==0, num_homi_70m_before, num_homi_70m_after),
         pub_dr_70m = dplyr::if_else(after==0, num_pub_dr_70m_before, num_pub_dr_70m_after),
         robb_70m = dplyr::if_else(after==0, num_robb_70m_before, num_robb_70m_after),
         vio_70m = dplyr::if_else(after==0, num_vio_70m_before, num_vio_70m_after),
         nonvio_70m = dplyr::if_else(after==0, num_nonvio_70m_before, num_nonvio_70m_after),

         agg_ass_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_agg_ass_70m_before,
                                                         num_agg_ass_70m_after)) / (pi*(70^2))),
         thefts_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_thefts_70m_before,
                                                         num_thefts_70m_after)) / (pi*(70^2))),

```



```

burg_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_burg_70m_before,
num_burg_70m_after)) / (pi*(70^2))),
disord_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_disord_70m_before,
num_disord_70m_after)) / (pi*(70^2))),
homi_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_homi_70m_before,
num_homi_70m_after)) / (pi*(70^2))),
pub_dr_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_pub_dr_70m_before,
num_pub_dr_70m_after)) / (pi*(70^2))),
robb_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_robb_70m_before,
num_robb_70m_after)) / (pi*(70^2))),
all_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_all_70m_before,
num_all_70m_after)) / (pi*(70^2))),
vio_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_vio_70m_before,
num_vio_70m_after)) / (pi*(70^2))),
nonvio_70m_per_km2_num = 1000000 * ((if_else(after ==0, num_nonvio_70m_before,
num_nonvio_70m_after)) / (pi*(70^2))),

agg_ass_500m_per_km2_num= 1000000 * ((if_else(after ==0, num_agg_ass_500m_before - num_agg_ass_500m_after - num_agg_ass_70m_after )) / (pi*(500^2))),
thefts_500m_per_km2_num = 1000000 * ((if_else(after ==0, num_thefts_500m_before - num_thefts_500m_after - num_thefts_70m_after )) / (pi*(500^2))),
burg_500m_per_km2_num = 1000000 * ((if_else(after ==0, num_burg_500m_before - num_burg_500m_after - num_burg_70m_after )) / (pi*(500^2))),
disord_500m_per_km2_num = 1000000 * ((if_else(after ==0, num_disord_500m_before - num_disord_500m_after - num_disord_70m_after )) / (pi*(500^2))),
homi_500m_per_km2_num = 1000000 * ((if_else(after ==0, num_homi_500m_before - num_homi_500m_after - num_homi_70m_after )) / (pi*(500^2))),
pub_dr_500m_per_km2_num = 1000000 * ((if_else(after ==0, num_pub_dr_500m_before - num_pub_dr_500m_after - num_pub_dr_70m_after )) / (pi*(500^2))),
robb_500m_per_km2_num = 1000000 * ((if_else(after ==0, num_robb_500m_before - num_robb_500m_after - num_robb_70m_after )) / (pi*(500^2))),
all_500m_per_km2_num = 1000000 * ((if_else(after ==0, num_all_500m_before - num_all_500m_after - num_all_70m_after )) / (pi*(500^2))),
vio_500m_per_km2_num = 1000000 * ((if_else(after ==0, num_vio_500m_before - num_vio_500m_after - num_vio_70m_after )) / (pi*(500^2))),
nonvio_500m_per_km2_num = 1000000 * ((if_else(after ==0, num_nonvio_500m_before - num_nonvio_500m_after - num_nonvio_70m_after )) / (pi*(500^2))),

agg_ass_70m_per_km2_std_num= agg_ass_70m_per_km2_num - agg_ass_500m_per_km2_num,
thefts_70m_per_km2_std_num = thefts_70m_per_km2_num - thefts_500m_per_km2_num,
burg_70m_per_km2_std_num = burg_70m_per_km2_num - burg_500m_per_km2_num,
disord_70m_per_km2_std_num = disord_70m_per_km2_num - disord_500m_per_km2_num,
homi_70m_per_km2_std_num = homi_70m_per_km2_num - homi_500m_per_km2_num,
pub_dr_70m_per_km2_std_num = pub_dr_70m_per_km2_num - pub_dr_500m_per_km2_num,
robb_70m_per_km2_std_num = robb_70m_per_km2_num - robb_500m_per_km2_num,
all_70m_per_km2_std_num = all_70m_per_km2_num - all_500m_per_km2_num,
vio_70m_per_km2_std_num = vio_70m_per_km2_num - vio_500m_per_km2_num,
nonvio_70m_per_km2_std_num = nonvio_70m_per_km2_num - nonvio_500m_per_km2_num

) %>%
select(full_address, greened, after, matches("^\\w+70m$"), contains("std"))

```

```
#####
##### DID regressions for 70 Meters Buffer @#####
#####

did_regressions_70m_unstd = map2(1:10, crime_trans_df$short_name,
  ~ do.call("lm", list(as.formula(paste0(.y, "_70m ~ greened + after + greened:after")),
    data=as.name("lots_70m_expanded"))))

did_regressions_70m_std = map2(1:10, crime_trans_df$short_name,
  ~ do.call("lm", list(as.formula(paste0(.y,
    "_70m_per_km2_std_num ~ greened + after + greened:after")),
    data=as.name("lots_70m_expanded"))))

did_regressions_70m_unstd_r2 = map(did_regressions_70m_unstd,
  ~ .x %>% glance() %>% pull(adj.r.squared)) %>% unlist()

did_regressions_70m_std_r2 = map(did_regressions_70m_std,
  ~ .x %>% glance() %>% pull(adj.r.squared)) %>% unlist()

#####
##### Plots of DID regressions (70m buffer) #####
#####

### Creating dataframe of DID beta estimators from regressions
interaction_effects_70m_unstd = map(did_regressions_70m_unstd, ~ .x %>% tidy() %>%
  filter(term == "greened:after") ) %>%

  bind_rows() %>%
  mutate(short_name = crime_trans_df$short_name,
    summary_crimetype = c(F,F,F,F,F,F,T,T,T),
    dist_radis = "70m",
    standardized = c("Unstandardized Crime")) %>%
  left_join(crime_trans_df) %>%
  mutate(crimetype = as.factor(crimetype),
    negative_effect = if_else(estimate <= 0, T, F),
    sig_.05 = if_else(p.value <= 0.05, T, F))

interaction_effects_70m_std = map(did_regressions_70m_std, ~ .x %>% tidy() %>%
  filter(term == "greened:after") ) %>%

  bind_rows() %>%
  mutate(short_name = crime_trans_df$short_name,
    summary_crimetype = c(F,F,F,F,F,F,T,T,T),
    dist_radis = "70m",
    standardized = c("Standardized Crime")) %>%
  left_join(crime_trans_df) %>%
  mutate(crimetype = as.factor(crimetype),
    negative_effect = if_else(estimate <= 0, T, F),
    sig_.05 = if_else(p.value <= 0.05, T, F))
```

```

coefficients = bind_rows(interaction_effects_70m_unstd, interaction_effects_70m_std)

# Creating combined plot of all DID estimators
did_estimates_plot = ggplot(coefficients, aes(x = crimetype, y = estimate,
                                              col = negative_effect)) +
  geom_hline(yintercept = 0, col = "gray70", size = 1.3) +
  geom_point(size = 2) +
  geom_errorbar(aes(ymin = estimate - 2*std.error,
                   ymax = estimate + 2*std.error),
               alpha = 1, width = 0.15) +
  facet_grid(rows = vars(standardized), cols = vars(summary_crimetype), scales = "free") +
  labs(y = "DID estimate", x = "Crimetype", title = "DID estimates by Crimetype") +
  theme(text = element_text(family = "CM Roman"),
        strip.text.x = element_blank(),
        axis.text.x = element_text(size = 11, angle = 55, hjust = 1),
        legend.text = element_text(size = 11)) +
  guides(col = F)

did_estimates_plot

```

References

- Anderson, James. 2018. “Much to Do About (Vacant) Lots.” *Next City*. Next City. nextcity.org/features/view/much-to-do-about-vacant-lots.
- Bivand, Roger, Tim Keitt, and Barry Rowlingson. 2017. *Rgdal: Bindings for the Geospatial Data Abstraction Library*. <https://CRAN.R-project.org/package=rgdal>.
- Bivand, Roger, and Nicholas Lewin-Koh. 2017. *Maptools: Tools for Reading and Handling Spatial Objects*. <https://CRAN.R-project.org/package=maptools>.
- Branas, Charles C, Rose A Cheney, John M MacDonald, Vicky W Tam, Tara D Jackson, and Thomas R Ten Have. 2011. “A Difference-in-Differences Analysis of Health, Safety, and Greening Vacant Urban Space.” *American Journal of Epidemiology* 174 (11). Oxford University Press: 1296–1306.
- Chang, Winston. 2014. *Extrafont: Tools for Using Fonts*. <https://CRAN.R-project.org/package=extrafont>.
- Cheng, Joe, Bhaskar Karambelkar, and Yihui Xie. 2018. *Leaflet: Create Interactive Web Maps with the Javascript ‘Leaflet’ Library*. <https://CRAN.R-project.org/package=leaflet>.
- Detroit Future City. 2017. “A Dfc Special Report: Green Culture Shift.”
- Donovan, Geoffrey H, and Jeffrey P Prestemon. 2012. “The Effect of Trees on Crime in Portland, Oregon.” *Environment and Behavior* 44 (1). Sage Publications Sage CA: Los Angeles, CA: 3–30.
- Dowle, Matt, and Arun Srinivasan. 2017. *Data.table: Extension of ‘Data.frame’*. <https://CRAN.R-project.org/package=data.table>.
- Dunnington, Dewey. 2018. *Ggspatial: Spatial Data Framework for Ggplot2*. <https://CRAN.R-project.org/package=ggspatial>.
- Econsult Corporation, Penn Institute for Urban Research, and May 8 Consulting. 2010. “Vacant Land Management in Philadelphia: The Costs of the Current System and the Need for Reform.”
- Garvin, Eugenia C, Carolyn C Cannuscio, and Charles C Branas. 2013. “Greening Vacant Lots to Reduce Violent Crime: A Randomised Controlled Trial.” *Injury Prevention* 19 (3). BMJ Publishing Group Ltd: 198–203.
- Goodyear, Sarah. 2013. “A 140-Acre Forest Is About to Materialize in the Middle of Detroit.” *Citylab*. The Atlantic Cities Blog. <https://www.citylab.com/design/2013/10/140-acre-forest-about-materialize-middle-detroit/7371/>.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <http://www.jstatsoft.org/v40/i03/>.
- Jackson, Richard J, Andrew L Dannenberg, and Howard Frumkin. 2013. “Health and the Built Environment: 10 Years After.” *American Journal of Public Health* 103 (9). American Public Health Association: 1542–4.
- Kaplan, Rachel, and Stephen Kaplan. 1989. *The Experience of Nature: A Psychological Perspective*. CUP Archive.
- Kuo, Frances E, and William C Sullivan. 2001a. “Aggression and Violence in the Inner City: Effects of Environment via Mental Fatigue.” *Environment and Behavior* 33 (4). Sage Publications Sage CA: Thousand Oaks, CA: 543–71.
- . 2001b. “Environment and Crime in the Inner City: Does Vegetation Reduce Crime?” *Environment and Behavior* 33 (3). Sage Publications Sage CA: Thousand Oaks, CA: 343–67.
- Mallach, Alan. 2018. “The Empty House Next Door: Understanding and Reducing Vacancy and Hypervacancy in the United States.” *Lincoln Institute of Land Policy*.
- Pebesma, Edzer. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R Journal*. <https://journal.r-project.org/archive/2018/RJ-2018-009/index.html>.

- Pedersen, Thomas Lin. 2017. *Patchwork: The Composer of Ggplots*. <https://github.com/thomasp85/patchwork>.
- Plerhoples, Christina, and Cleveland FRB Policy Summit. 2012. "The Effect of Vacant Building Demolitions on Crime Under Depopulation." *Economics Job Market Paper*. Lansing: Michigan State University, Department of Agricultural, Food, and Resources.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schilling, Joseph, and Jonathan Logan. 2008. "Greening the Rust Belt: A Green Infrastructure Model for Right Sizing America's Shrinking Cities." *Journal of the American Planning Association* 74 (4). Taylor & Francis: 451–66.
- Schmelzkopf, Karen. 1995. "Urban Community Gardens as Contested Space." *Geographical Review*. JSTOR, 364–81.
- Snelgrove, AG, JH Michael, TM Waliczek, and JM Zajicek. 2004. "Urban Greening and Criminal Behavior: A Geographic Information System Perspective." *HortTechnology* 14 (1). American Society for Horticultural Science: 48–51.
- South, Eugenia C, Michelle C Kondo, Rose A Cheney, and Charles C Branas. 2015. "Neighborhood Blight, Stress, and Health: A Walking Trial of Urban Greening and Ambulatory Heart Rate." *American Journal of Public Health* 105 (5). American Public Health Association: 909–13.
- Urbanek, Simon. 2013. *Png: Read and Write Png Images*. <https://CRAN.R-project.org/package=png>.
- Vacant Property Campaign'. 2005. "Vacant Properties: The True Cost to Communities."
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <http://ggplot2.org>.
- . 2018. *Stringr: Simple, Consistent Wrappers for Common String Operations*. <https://CRAN.R-project.org/package=stringr>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2018. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, and Lionel Henry. 2018. *Tidyr: Easily Tidy Data with 'Spread()' and 'Gather()' Functions*. <https://CRAN.R-project.org/package=tidyr>.
- Wickham, Hadley, and Dana Seidel. 2018. *Scales: Scale Functions for Visualization*.
- Wolch, Jennifer, John P Wilson, and Jed Fehrenbach. 2005. "Parks and Park Funding in Los Angeles: An Equity-Mapping Analysis." *Urban Geography* 26 (1). Taylor & Francis: 4–35.
- Wolfe, Mary K, and Jeremy Mennis. 2012. "Does Vegetation Encourage or Suppress Urban Crime? Evidence from Philadelphia, Pa." *Landscape and Urban Planning* 108 (2-4). Elsevier: 112–22.