# Analyze Boston

*Michael Rose*

```r
knitr::opts_chunk$set(echo = FALSE)
```

## Abstract

'Analyze Boston' is an open data initiative maintained by the city of Boston containing facts, figures, and maps related to the city. In this project we will look at some of the city's 133+ data sets and analyze them with descriptive and inferential statistics.The focus of the project is the discovery of interesting patterns.

## Intro to the Data

In this analysis, I will be using 4 different data sets.

### Employee Earnings Report

- Each year, the City of Boston publishes payroll data for employees. This dataset contains employee names, job details, and earnings information including base salary, overtime, and total compensation for employees of the City.
- You can see more at https://data.boston.gov/dataset/employee-earnings-report

```
## # A tibble: 22,235 x 12
##    NAME    `DEPARTMENT NAME` TITLE   REGULAR RETRO  OTHER OVERTIME INJURED
##    <chr>   <chr>             <chr>     <dbl> <dbl>  <dbl>    <dbl>   <dbl>
##  1 Miller~ Boston Police De~ Police~ 129531.    NA 13694.    8150.    NA
##  2 Sulliv~ Boston Police De~ Office~  56922.    NA  3595.    1548.    NA
##  3 O'Hara~ Boston Police De~ Police~ 124057.    NA  6432.   29044.    NA
##  4 Whalen~ Boston Police De~ Police~  94956. 4985. 13592.   85419.    58.0
##  5 Kelly,~ Boston Police De~ Tape L~  69995.    NA   300     7961.    NA
##  6 Carrol~ Boston Police De~ Police~  12757. 2390. 41612.     912.    NA
##  7 Connol~ Boston Police De~ Police~  93180. 2028. 13338.   19882.    NA
##  8 Ivens,~ Boston Police De~ Police~     NA    NA  60777.      NA   2659.
##  9 Kelly,~ Boston Police De~ Police~  13827.    NA 62393.     868.    NA
## 10 Klokma~ Boston Police De~ Police~ 107599.    NA 14482.   12825.    NA
## # ... with 22,225 more rows, and 4 more variables: DETAIL <dbl>,
## #   `QUINN/EDUCATION INCENTIVE` <dbl>, `TOTAL EARNINGS` <dbl>,
## #   POSTAL <int>
```

### Crime Incident Reports

- Crime incident reports are provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred. Records in the new system begin in June of 2015.
- You can see more at https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system

```
## # A tibble: 6 x 17
##   INCIDENT_NUMBER OFFENSE_CODE OFFENSE_CODE_GRO~ OFFENSE_DESCRIP~ DISTRICT
##   <chr>           <chr>        <chr>             <chr>            <chr>
## 1 I182017604      03115        Investigate Pers~ INVESTIGATE PER~ B3
## 2 I182017601      00520        Residential Burg~ BURGLARY - RESI~ B2
## 3 I182017596      03831        Motor Vehicle Ac~ M/V - LEAVING S~ A1
## 4 I182017595      03802        Motor Vehicle Ac~ "M/V ACCIDENT -~ <NA>
## 5 I182017594      01830        Drug Violation    DRUGS - SICK AS~ D14
## 6 I182017593      00361        Robbery           ROBBERY - OTHER  D4
## # ... with 12 more variables: REPORTING_AREA <int>, SHOOTING <chr>,
## #   OCCURRED_ON_DATE <dttm>, YEAR <int>, MONTH <int>, DAY_OF_WEEK <chr>,
## #   HOUR <int>, UCR_PART <chr>, STREET <chr>, Lat <dbl>, Long <dbl>,
## #   Location <chr>
```

## BPD Firearm Recovery Counts

- This dataset provides daily counts of firearms recovered by Boston Police Department since August 20, 2014. Recovery totals are provided for three distinct channels: crime, voluntary surrender, and gun buyback programs.
- You can see more at https://data.boston.gov/dataset/boston-police-department-firearms-recovery-counts

```
## # A tibble: 6 x 4
##   CollectionDate CrimeGunsRecovered GunsSurrenderedSafe~ BuybackGunsRecov~
##   <chr>                       <int>                <int>             <int>
## 1 8/20/2014                       2                    3                 1
## 2 8/21/2014                       2                    0                 4
## 3 8/22/2014                       0                    0                 2
## 4 8/25/2014                       8                    3                 0
## 5 8/26/2014                       9                    0                 0
## 6 8/27/2014                       1                    0                 0
```

## Economic Indicators

- The Boston Planning and Redevelopment Authority (BPDA), formerly known as the Boston Redevelopment Authority (BRA), is tasked with planning for and guiding inclusive growth within the City of Boston. As part of this mission, BPDA collects and analyzes a variety of economic data relating to topics such as the employment, housing, travel, and real estate development. This is a legacy dataset of economic idicators tracked monthly between January2013 and January 2015.
- You can see more at https://data.boston.gov/dataset/economic-indicators-legacy-portal

```
## # A tibble: 6 x 19
##    Year Month logan_passengers logan_intl_flights hotel_occup_rate
##   <int> <int>            <int>              <int>            <dbl>
## 1  2013     1          2019662               2986            0.572
## 2  2013     2          1878731               2587            0.645
## 3  2013     3          2469155               3250            0.819
## 4  2013     4          2551246               3408            0.855
## 5  2013     5          2676291               3240            0.858
## 6  2013     6          2824862               3402            0.911
## # ... with 14 more variables: hotel_avg_daily_rate <dbl>,
## #   total_jobs <int>, unemp_rate <dbl>, labor_force_part_rate <dbl>,
## #   pipeline_unit <int>, pipeline_total_dev_cost <dbl>,
```

```
## #   pipeline_sqft <int>, pipeline_const_jobs <dbl>, foreclosure_pet <int>,
## #   foreclosure_deeds <int>, med_housing_price <int>,
## #   housing_sales_vol <int>, new_housing_const_permits <int>,
## #   `new-affordable_housing_permits` <int>
```

**Data Cleaning**

The first thing that needs to be done is tidying up the data. We can start by removing any numeric NAs and turning them into 0s.

```
## [1] 0
```

```
## [1] 0
```

```
## [1] 0
```

Since minimum wage is $11/hr we can filter full time from part time. I will be removing those who make under 11 * 40 * 52 = $22880 / year. I will round down to $20,000.

```
## # A tibble: 6 x 8
##   NAME  `DEPARTMENT NAM~ TITLE REGULAR OVERTIME EXTRA_PAY `TOTAL EARNINGS`
##   <chr> <chr>            <chr>   <dbl>    <dbl>     <dbl>            <dbl>
## 1 Mill~ Boston Police D~ Poli~ 129531.    8150.    37981.          175663.
## 2 Sull~ Boston Police D~ Offi~  56922.    1548.     3595.           62065.
## 3 O'Ha~ Boston Police D~ Poli~ 124057.   29044.    52078.          205178.
## 4 Whal~ Boston Police D~ Poli~  94956.   85419.    54878.          235312.
## 5 Kell~ Boston Police D~ Tape~  69995.    7961.      300            78256.
## 6 Carr~ Boston Police D~ Poli~  12757.     912.    45566.           59234.
## # ... with 1 more variable: INJURED <dbl>
```

Now we can begin to clean up those department name factors to get a clearer view of the groups as a whole. It seems like the school system takes up the majority of the factors, so lets compress them all into one factor - "Education"

```
## [1] Boston Police Department       Workers Compensation Service
## [3] BPS East Boston High           BPS School Safety Service
## [5] Dpt of Innovation & Technology BPS Ohrenberger Elementary
## 224 Levels: Accountability Achievement Gap ... Youth Engagement & Employment
```

```
## [1] "BPS East Boston High"       "BPS School Safety Service"
## [3] "BPS Ohrenberger Elementary" "BPS Transportation"
## [5] "BPS Quincy Elementary"      "BPS McCormack Middle"
```

```
## [1] "Roosevelt K-8"       "Edison K-8"          "Higginson/Lewis K-8"
## [4] "Jackson/Mann K-8"    "Greenwood, S K-8"    "Lyon K-8"
```

```
## [1] "Tech Boston Academy"         "BPS Latin Academy"
## [3] "West Roxbury Academy"        "BPS MPH\\Crafts Academy"
## [5] "Kennedy, EM Health Academy"  "WREC: Urban Science Academy"
```

```
## [1] "Boston Police Department"       "Workers Compensation Service"
## [3] "Education"                      "Dpt of Innovation & Technology"
## [5] "Registry Division"              "Boston Fire Department"
```

```
## [1] "Boston Police Department"       "Workers Compensation Service"
## [3] "Education"                      "Dpt of Innovation & Technology"
## [5] "Registry Division"              "Boston Fire Department"
```

```
## [1] "Boston Police Department"       "Workers Compensation Service"
## [3] "Education"                      "Dpt of Innovation & Technology"
## [5] "Registry Division"              "Boston Fire Department"
```

```
## [1] "Boston Police Department"       "Workers Compensation Service"
## [3] "Education"                      "Dpt of Innovation & Technology"
## [5] "Registry Division"              "Boston Fire Department"
```

# Education Pay

```
## # A tibble: 6 x 8
##   NAME  `DEPARTMENT NAM~ TITLE REGULAR OVERTIME EXTRA_PAY `TOTAL EARNINGS`
##   <chr> <chr>            <chr>   <dbl>    <dbl>     <dbl>            <dbl>
## 1 Bott~ Education        BPS ~      0         0         0           285459.
## 2 Chan~ Education        Supe~ 264661.        0      6000.          270661.
## 3 McCa~ Education        Teac~  46981.        0    182189.          229170.
## 4 Jord~ Education        Unit~ 106762.        0     81789.          188551.
## 5 Estr~ Education        Depu~ 177625.        0         0           177625.
## 6 Wood~ Education        Asst~   4620         0    169047.          173667.
## # ... with 1 more variable: INJURED <dbl>
```

From the above the pays seem about normal. The first woman Torii Bottomley won a lawsuit against her employer for workplace bullying:

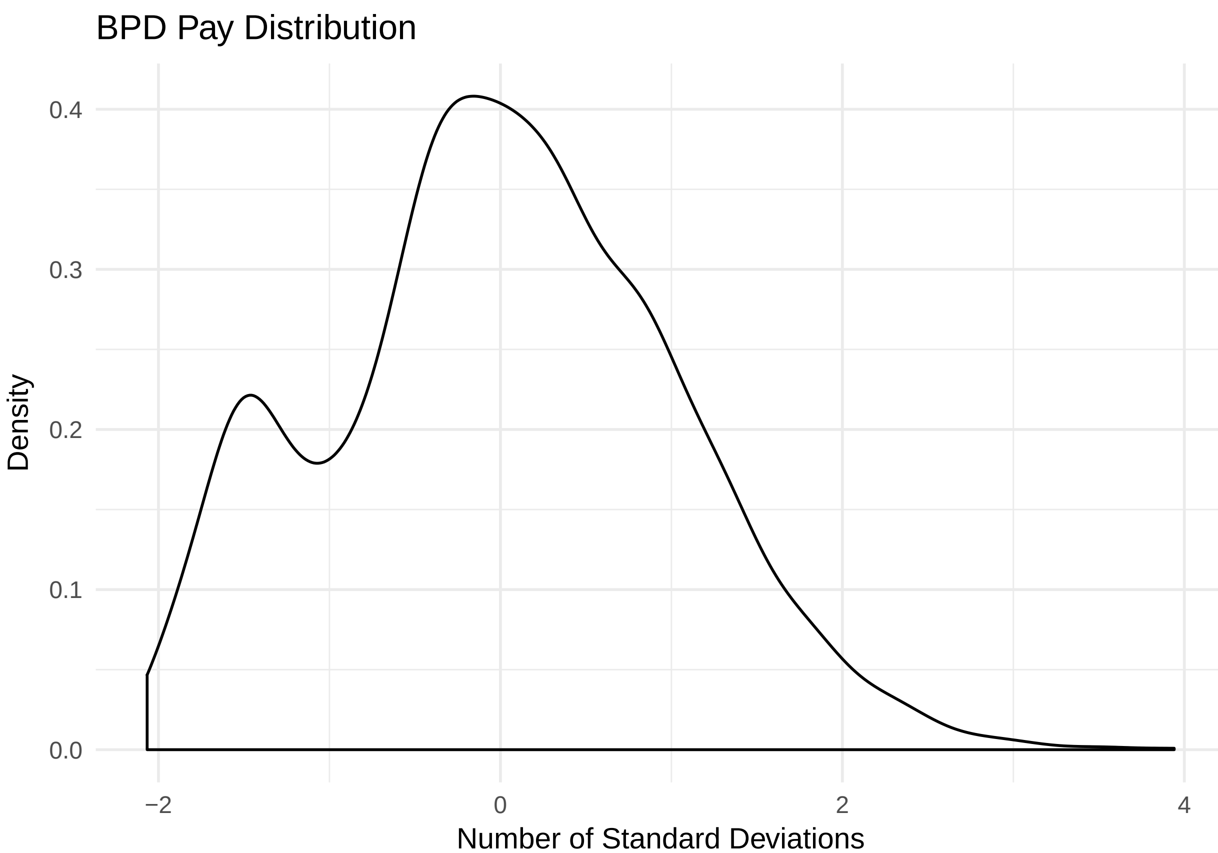
https://www.pacermonitor.com/public/case/22846991/Bottomley_v_Boston_Public_Schools_et_al

The only other outlier that it shown is Elaine M McCabe with a base salary of $46,981 and $182,189.26 in extra pay (not overtime). I was unable to find a reasoning for this.

## Police Pay

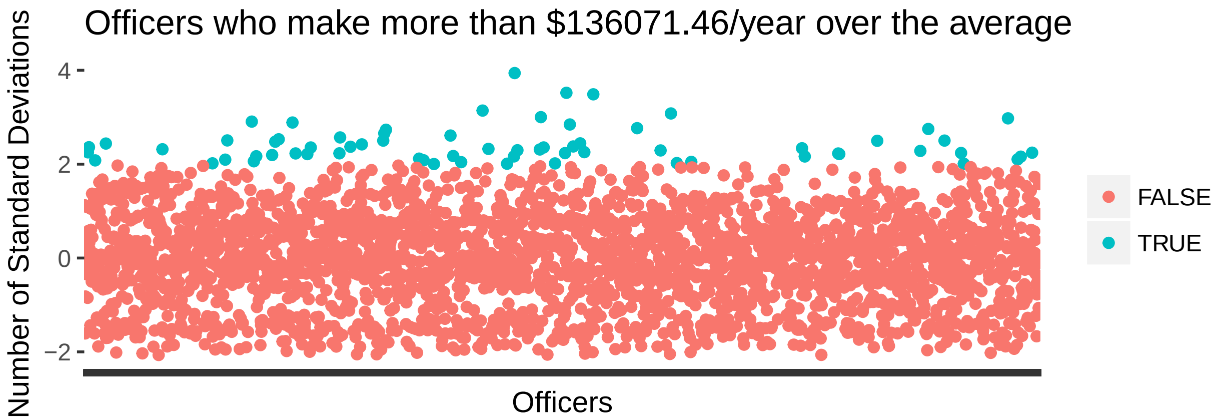
First lets look at the how the pay is distributed across the police department.

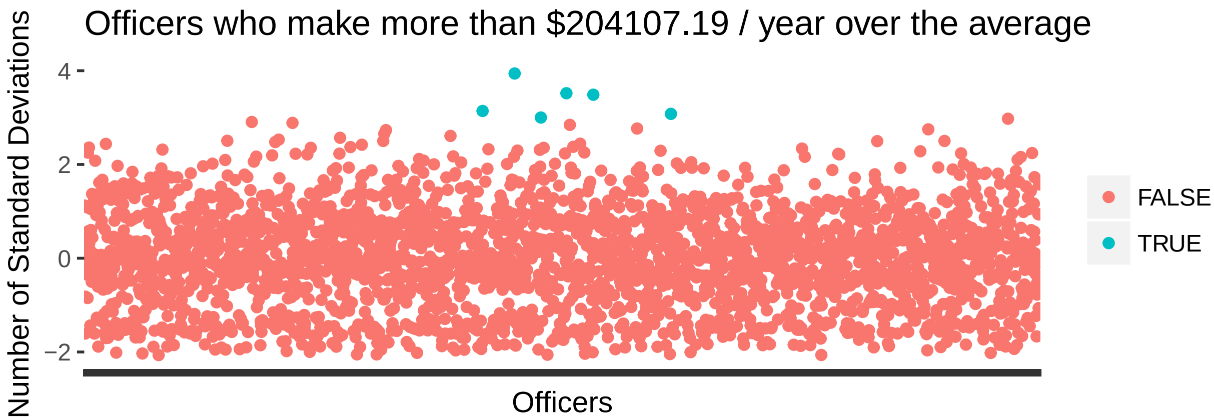
```
## # A tibble: 6 x 9
##   NAME  `DEPARTMENT NAM~ TITLE REGULAR OVERTIME EXTRA_PAY `TOTAL EARNINGS`
##   <chr> <chr>            <chr>   <dbl>    <dbl>     <dbl>            <dbl>
## 1 Hose~ Boston Police D~ Poli~ 146894.   62696.   156642.          366233.
## 2 Kerv~ Boston Police D~ Poli~ 125715.   66067.   150210.          341992.
## 3 Lee,~ Boston Police D~ Poli~  97414.   71669.   171093.          340176.
## 4 Hass~ Boston Police D~ Poli~ 137104.   72158.    99536.          320224.
## 5 McCo~ Boston Police D~ Poli~ 146894.   63708.   106072.          316674.
## 6 Jose~ Boston Police D~ Poli~  97414.   87746.   126997.          312156.
## # ... with 2 more variables: INJURED <dbl>, num_std_devs <dbl>
```




```
## [1] 57576.17
```

We can see from the plot above that our Boston PD officers have pay that is roughly normally distributed around our mean. The graph about shows how many standard deviations officers are away from the mean. The mean itself is $139345.4 per year and 1 standard deviation is $57576.17 per year. Lets now look at our distribution again with an emphasis on our superearners:



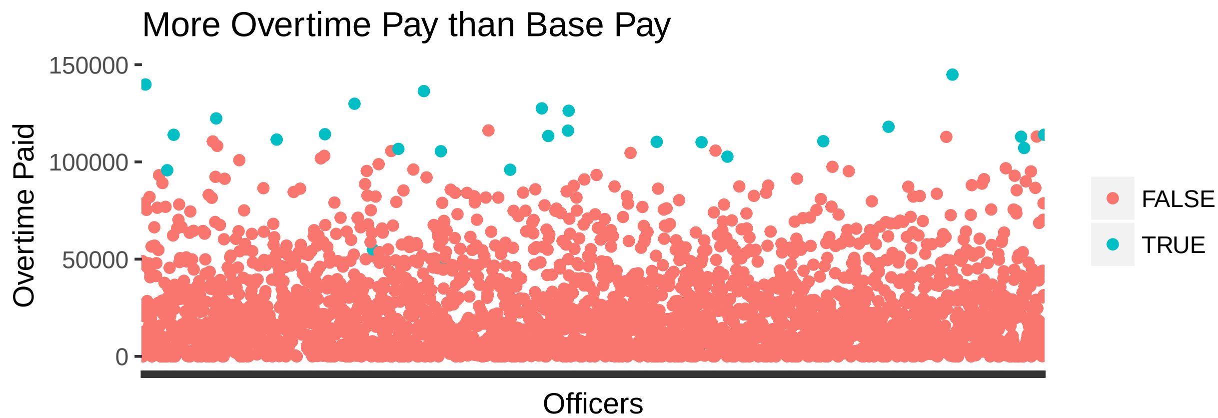



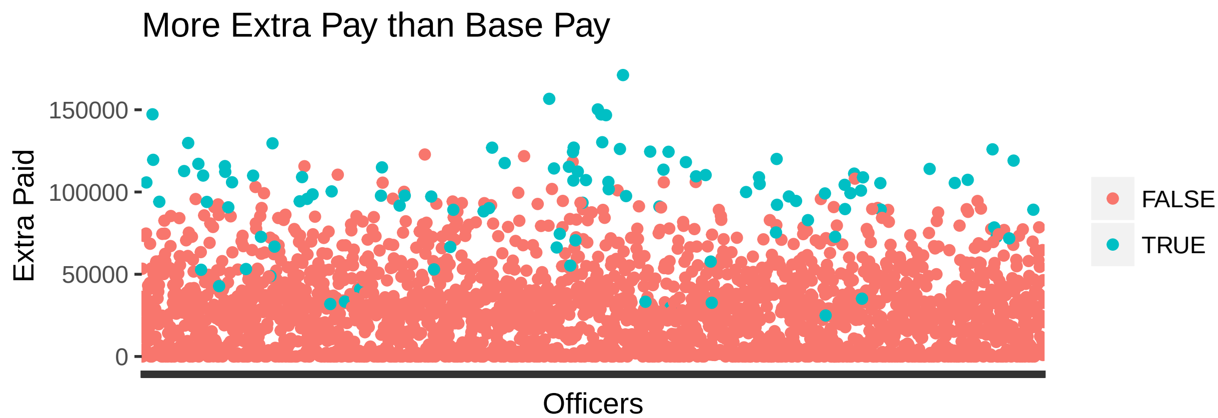

We can see from above that there are a few people who earn exorbitant sums. What could cause this? The first thing that comes to mind is a high base salary.

```
## # A tibble: 6 x 9
##   NAME  `DEPARTMENT NAM~ TITLE REGULAR OVERTIME EXTRA_PAY `TOTAL EARNINGS`
##   <chr> <chr>            <chr>   <dbl>    <dbl>     <dbl>            <dbl>
## 1 Evan~ Boston Police D~ Comm~ 230000.        0      8846.          238846.
## 2 Gros~ Boston Police D~ Supn~ 199244.        0     26094.          225338.
## 3 Buck~ Boston Police D~ Supn~ 181983.        0     57978.          239961.
## 4 Manc~ Boston Police D~ Supn~ 181983.        0     57978.          239961.
## 5 Holm~ Boston Police D~ Supn~ 181983.        0     47087.          229069.
## 6 Ridg~ Boston Police D~ Supn~ 180369.        0     57726.          238094.
## # ... with 2 more variables: INJURED <dbl>, num_std_devs <dbl>
```

We can see from the table above that the highest paid person in the Boston Police Department is the Commissioner with a salary of about $230,000 and a total earnings of about $240,000. Even with this high base salary, he is only 1.68 standard deviations above the average officer pay of $124254.60. In fact, the entire table of top base pay people have standard deviations less than 2, so we can see that base pay isn't what contributes to such high pay.

The next thing that comes to mind is a lot of overtime mixed with extra pay. This extra pay includes things like road detail and testifying in court. We could also check an assumption that those with a higher paygrade (e.g. captains and lieutenants) are likely to get more overtime money since their time and a half is generally a lot higher.




From the plots above we can see that there are quite a few people with more of their annual salary coming from overtime or extra pay. This is quite surprising.

If we can assume time and a half, then there are officers working their regular hours and then much more. There are a few who make over $100,000 per year extra through just overtime.

We can also see that there are even more officers who make more than their base salary in extra pay. This extra pay was defined as things like court appearances, detail, retrograde pay and the Quinn education incentive which gives a small salary bump for having a criminal justice degree. Why could this be? After looking into it, I came across some Boston Globe reports. Here is a quote from https://www.bostonglobe.com/metro/2017/06/20/for-some-boston-police-officers-extra-money-comes-easy/oS47lc7OuYyVZbTPBv1zQL/story.html :

**Quote: "**

In what critics call an extreme example of a systemic problem, Lee Waiman bolstered his wages thanks to police union contracts that require that officers who work detail shifts or testify in court be paid a minimum of four hours, even if the assignment lasts only 30 minutes.

Last year, Lee earned $58,600 by working more than 1,100 hours of overtime, according to a Globe review of police payroll records. Records show that Lee did not work 674 of those hours — more than 16 40-hour weeks — yet received time-and-a-half pay.

Most of Lee's overtime pay stemmed from court appearances that typically lasted no more than an hour, the Globe found. He was also paid for 2,771 hours for detail shifts, including 861 unworked hours. That allowed him to make close to $130,000, a sum that did not include his overtime pay.

"It's a generous system," said Sam Tyler, president of the Boston Municipal Research Bureau, a fiscal watchdog group. "You're paid for hours you don't work. It isn't a new issue, but it's one that really does need stricter focus and management to control those costs."

"

Clearly this is a known pattern and has been looked at before. For example, last year the BPD overtime alone hit $66.9 million.

https://www.bostonglobe.com/metro/2018/02/16/bpd-captain-was-city-top-earner/iI4G1pnC7MOUODxo0XR4aO/story.html

Looking back even further, I came across this boston.com article from 2007:

http://archive.boston.com/news/local/articles/2007/08/23/3_police_lieutenants_are_cited_for_alleged_detail_abuses/

**Quote:** "

The internal audit of shifts worked in 2005 concluded that Lieutenants Haseeb Hosein, Timothy Kervin, and Ghassoub Frangie engaged in untruthful reporting of hours, performed details that conflicted with a scheduled tour of duty, and received details through unauthorized means. Hosein and Kervin were also cited with breaking the law, but officials did not provide details on the alleged infractions.

The investigators accused Hosein, a 19-year veteran, of 203 violations that include 80 counts of inaccurate reporting on a detail card, 16 counts of receiving details outside the system, 24 counts of accepting a detail scheduled during his regular patrol shifts, and one count of breaking the law and conduct unbecoming an officer.

Kervin, a 20-year veteran, was charged with 191 violations that include 68 counts of inaccurate reporting on a detail card, 46 counts of receiving details outside the system, six counts of accepting a detail scheduled during his regular patrol shifts, and one count each of breaking the law and conduct unbecoming of an officer.

Frangie, a 29-year veteran, was charged with 84 violations that include 34 counts of inaccurate reporting on a detail card, 10 counts of accepting a detail scheduled during his regular patrol shifts, three counts of receiving details outside the system, and two counts of conduct unbecoming an officer.

"

```
## # A tibble: 5 x 9
##   NAME  `DEPARTMENT NAM~ TITLE REGULAR OVERTIME EXTRA_PAY `TOTAL EARNINGS`
##   <chr> <chr>            <chr>   <dbl>    <dbl>     <dbl>            <dbl>
## 1 Sull~ Boston Police D~ Poli~ 110100.  144874.    28440.          283414.
## 2 Acos~ Boston Police D~ Poli~ 101681.  139806.    27517.          269004.
## 3 Fitz~ Boston Police D~ Poli~ 112121.  136404.    34853.          283378.
## 4 Deva~ Boston Police D~ Poli~ 116109.  129912.    28925.          274946.
## 5 Hold~ Boston Police D~ Poli~  95326.  127517.    32208.          255051.
## # ... with 2 more variables: INJURED <dbl>, num_std_devs <dbl>
```

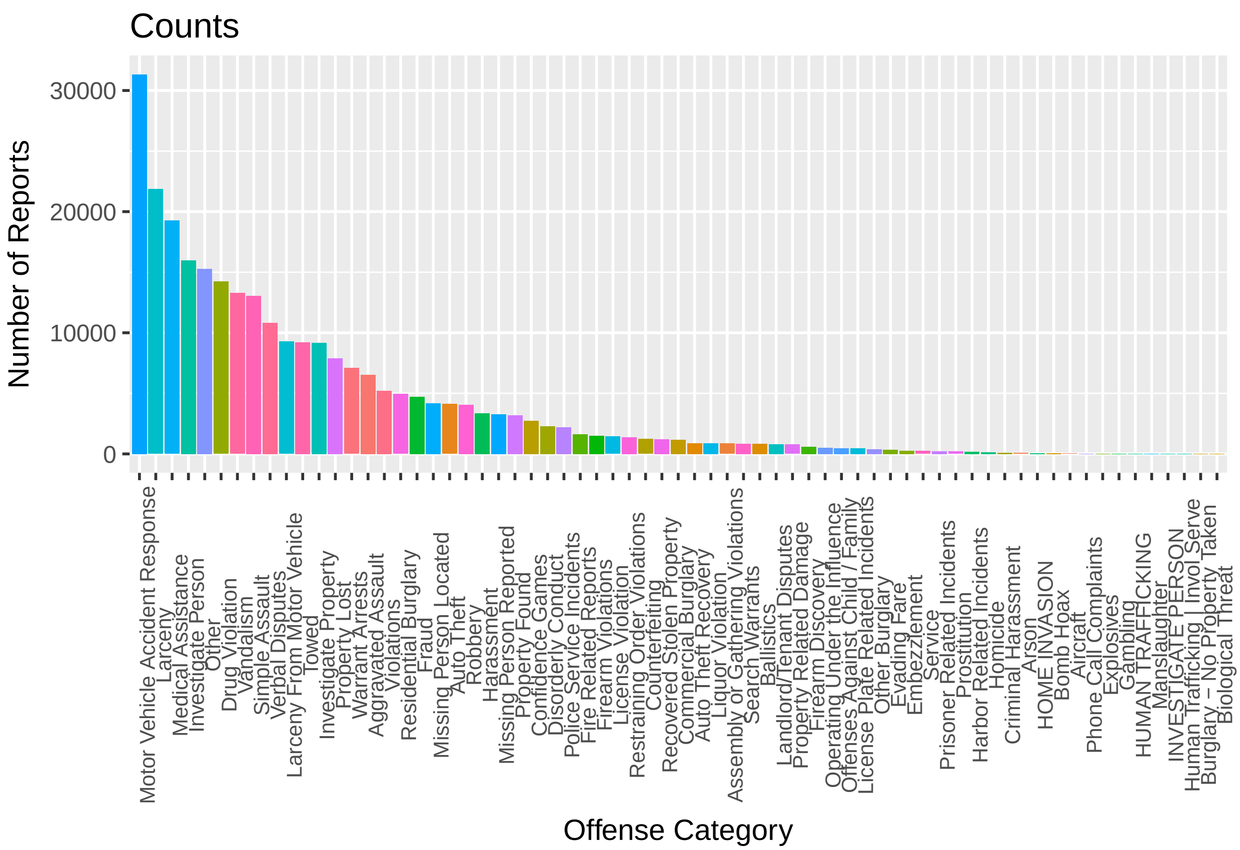
```
## # A tibble: 5 x 9
##   NAME  `DEPARTMENT NAM~ TITLE REGULAR OVERTIME EXTRA_PAY `TOTAL EARNINGS`
##   <chr> <chr>            <chr>   <dbl>    <dbl>     <dbl>            <dbl>
## 1 Lee,~ Boston Police D~ Poli~  97414.   71669.   171093.          340176.
## 2 Hose~ Boston Police D~ Poli~ 146894.   62696.   156642.          366233.
## 3 Kerv~ Boston Police D~ Poli~ 125715.   66067.   150210.          341992.
## 4 Alme~ Boston Police D~ Poli~  86918.   24289.   147234.          259096.
## 5 King~ Boston Police D~ Poli~ 129531.   26444.   147170.          303145.
## # ... with 2 more variables: INJURED <dbl>, num_std_devs <dbl>
```

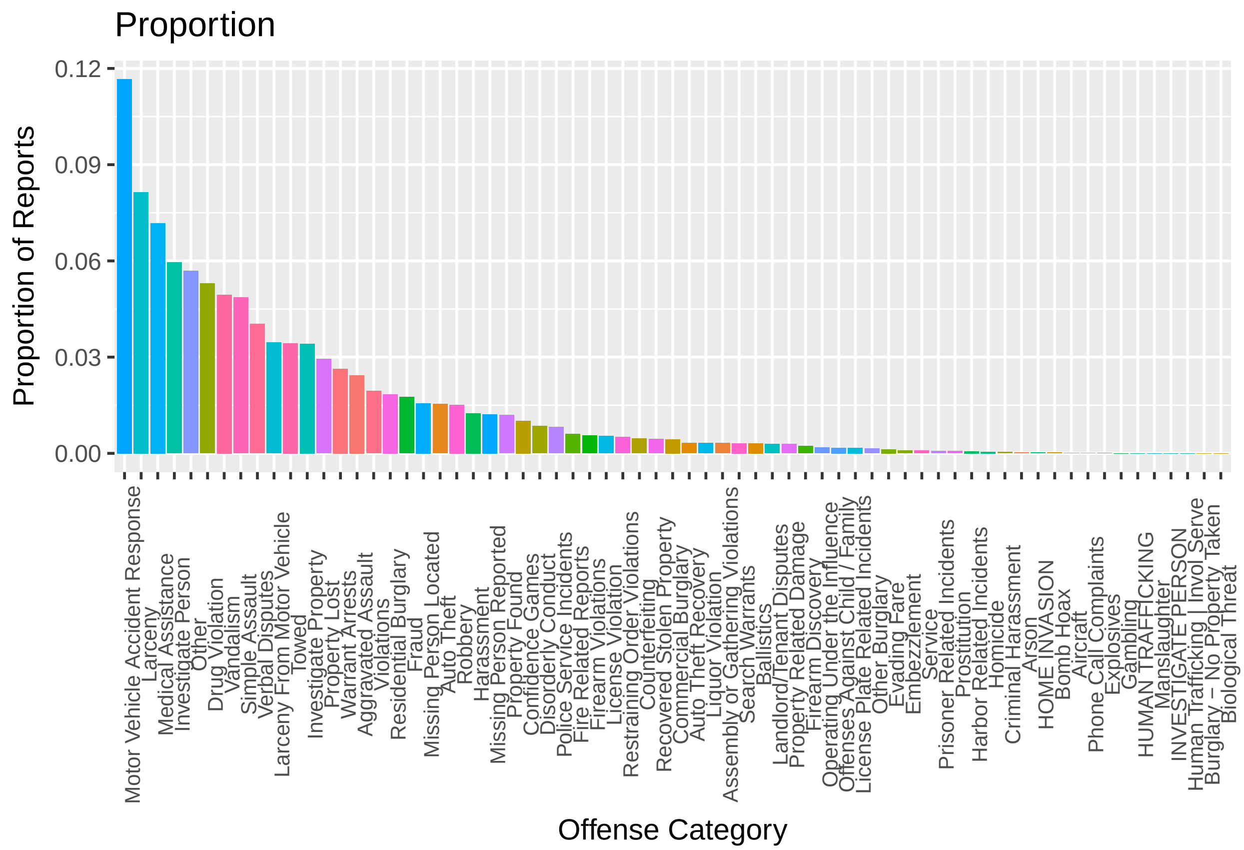
From the tables above we see some familiar names - including Lee Waiman from our first Boston Globe Article. We also see the names Haseeb Hosein and Timothy Kervin from our 2nd and 3rd articles. Clearly the city government is aware of the problem, but has not stopped it in at least 12 years.

That being said though, police are an important and vital part of society. Lets explore some of their valiant work.

## Crime

```
## # A tibble: 6 x 17
##   INCIDENT_NUMBER OFFENSE_CODE OFFENSE_CODE_GRO~ OFFENSE_DESCRIP~ DISTRICT
##   <chr>           <chr>        <chr>             <chr>            <chr>
## 1 I182017604      03115        Investigate Pers~ INVESTIGATE PER~ B3
## 2 I182017601      00520        Residential Burg~ BURGLARY - RESI~ B2
## 3 I182017596      03831        Motor Vehicle Ac~ M/V - LEAVING S~ A1
## 4 I182017595      03802        Motor Vehicle Ac~ "M/V ACCIDENT -~ <NA>
## 5 I182017594      01830        Drug Violation    DRUGS - SICK AS~ D14
## 6 I182017593      00361        Robbery           ROBBERY - OTHER  D4
## # ... with 12 more variables: REPORTING_AREA <int>, SHOOTING <chr>,
## #   OCCURRED_ON_DATE <dttm>, YEAR <int>, MONTH <int>, DAY_OF_WEEK <chr>,
## #   HOUR <int>, UCR_PART <chr>, STREET <chr>, Lat <dbl>, Long <dbl>,
## #   Location <chr>
```






As we see from above there are quite a few offense categories! Clearly Motor Vehicle Accident Response is the largest, followed by larceny (theft). This data set is quite interesting, so lets look at some more patterns.

## Most Common Crimes by Time of Day

The tables below show the top 3 crimes that occur by hour. Understandably, the number 1 is consistently motor vehicle accident response, with consistently around ~10% of all crimes committed within that hour frame over 3 years of data (2015 - 2017). Other interesting bits:

- Simple assaults are more common during hours 1,2,3. This could be due to bars and nightlife.
- Vandalism commonly occurs during the hours of 4, 5, and 6 am. This makes sense, as vandals would be likely to strike at night.
- People got towed more often during the hours of 7, 8, and 9 am. This is likely cars that were left overnight.
- Drug Violations were most common during the hours of 16, 17, 18 and 19 (or 4,5,6,7 pm).

```
## Hour: 1
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response   758 0.0962
## 2 Simple Assault                    657 0.0834
## 3 Medical Assistance                600 0.0762
## ---------------------------------------
##
## ---------------------------------------
## Hour: 2
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response   831 0.128
## 2 Simple Assault                    604 0.0931
## 3 Medical Assistance                486 0.0749
## ---------------------------------------
##
## ---------------------------------------
## Hour: 3
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response   529 0.136
## 2 Medical Assistance                336 0.0861
## 3 Simple Assault                    241 0.0618
## ---------------------------------------
##
## ---------------------------------------
## Hour: 4
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response   340 0.120
## 2 Medical Assistance                284 0.0999
## 3 Vandalism                         182 0.0640
## ---------------------------------------
##
## ---------------------------------------
## Hour: 5
## # A tibble: 3 x 3
```

```
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response   437 0.158
## 2 Medical Assistance                299 0.108
## 3 Vandalism                         190 0.0688
## ---------------------------------------
##
## ---------------------------------------
## Hour: 6
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response   739 0.175
## 2 Medical Assistance                349 0.0826
## 3 Vandalism                         265 0.0627
## ---------------------------------------
##
## ---------------------------------------
## Hour: 7
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1226 0.164
## 2 Towed                             954 0.128
## 3 Medical Assistance                477 0.0639
## ---------------------------------------
##
## ---------------------------------------
## Hour: 8
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1566 0.143
## 2 Towed                            1177 0.107
## 3 Larceny                           746 0.0681
## ---------------------------------------
##
## ---------------------------------------
## Hour: 9
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1462 0.118
## 2 Towed                            1068 0.0859
## 3 Larceny                           874 0.0703
## ---------------------------------------
##
## ---------------------------------------
## Hour: 10
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1435 0.105
## 2 Larceny                          1219 0.0892
```

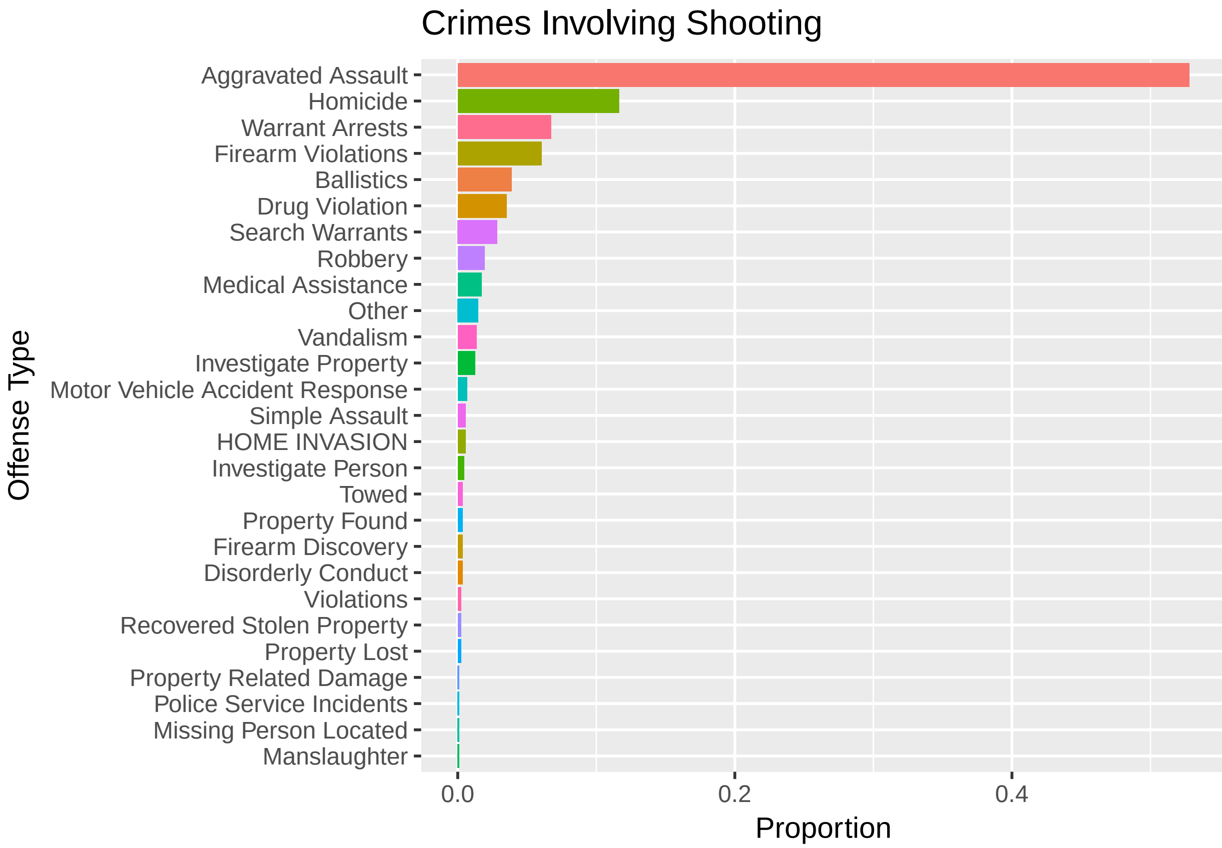
```
## 3 Medical Assistance              1078 0.0789
## ---------------------------------------
##
## ---------------------------------------
## Hour: 11
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1431 0.104
## 2 Larceny                          1275 0.0927
## 3 Medical Assistance               1067 0.0776
## ---------------------------------------
##
## ---------------------------------------
## Hour: 12
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Larceny                          1702 0.109
## 2 Motor Vehicle Accident Response  1504 0.0961
## 3 Medical Assistance               1116 0.0713
## ---------------------------------------
##
## ---------------------------------------
## Hour: 13
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1562 0.110
## 2 Larceny                          1449 0.102
## 3 Medical Assistance               1100 0.0777
## ---------------------------------------
##
## ---------------------------------------
## Hour: 14
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1635 0.114
## 2 Larceny                          1614 0.113
## 3 Medical Assistance               1073 0.0749
## ---------------------------------------
##
## ---------------------------------------
## Hour: 15
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1763 0.127
## 2 Larceny                          1542 0.111
## 3 Medical Assistance               1017 0.0730
## ---------------------------------------
##
## ---------------------------------------
```

```
## Hour: 16
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  2061 0.122
## 2 Larceny                          1666 0.0989
## 3 Drug Violation                   1471 0.0873
## ---------------------------------------
##
## ---------------------------------------
## Hour: 17
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  2204 0.126
## 2 Drug Violation                   1878 0.107
## 3 Larceny                          1648 0.0938
## ---------------------------------------
##
## ---------------------------------------
## Hour: 18
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1943 0.114
## 2 Drug Violation                   1752 0.102
## 3 Larceny                          1554 0.0908
## ---------------------------------------
##
## ---------------------------------------
## Hour: 19
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1655 0.111
## 2 Larceny                          1297 0.0873
## 3 Drug Violation                   1269 0.0854
## ---------------------------------------
##
## ---------------------------------------
## Hour: 20
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1451 0.109
## 2 Larceny                          1078 0.0810
## 3 Medical Assistance                996 0.0749
## ---------------------------------------
##
## ---------------------------------------
## Hour: 21
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
```

```
## 1 Motor Vehicle Accident Response  1341 0.113
## 2 Medical Assistance                 977 0.0825
## 3 Larceny                            784 0.0662
## ---------------------------------------
##
## ---------------------------------------
## Hour: 22
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1342 0.123
## 2 Medical Assistance                860 0.0790
## 3 Vandalism                         768 0.0705
## ---------------------------------------
##
## ---------------------------------------
## Hour: 23
## # A tibble: 3 x 3
##   `.$OFFENSE_CODE_GROUP`              n   prop
##   <chr>                           <int>  <dbl>
## 1 Motor Vehicle Accident Response  1075 0.121
## 2 Vandalism                         657 0.0737
## 3 Medical Assistance                641 0.0719
## ---------------------------------------
##
## ---------------------------------------
```

## Shooting

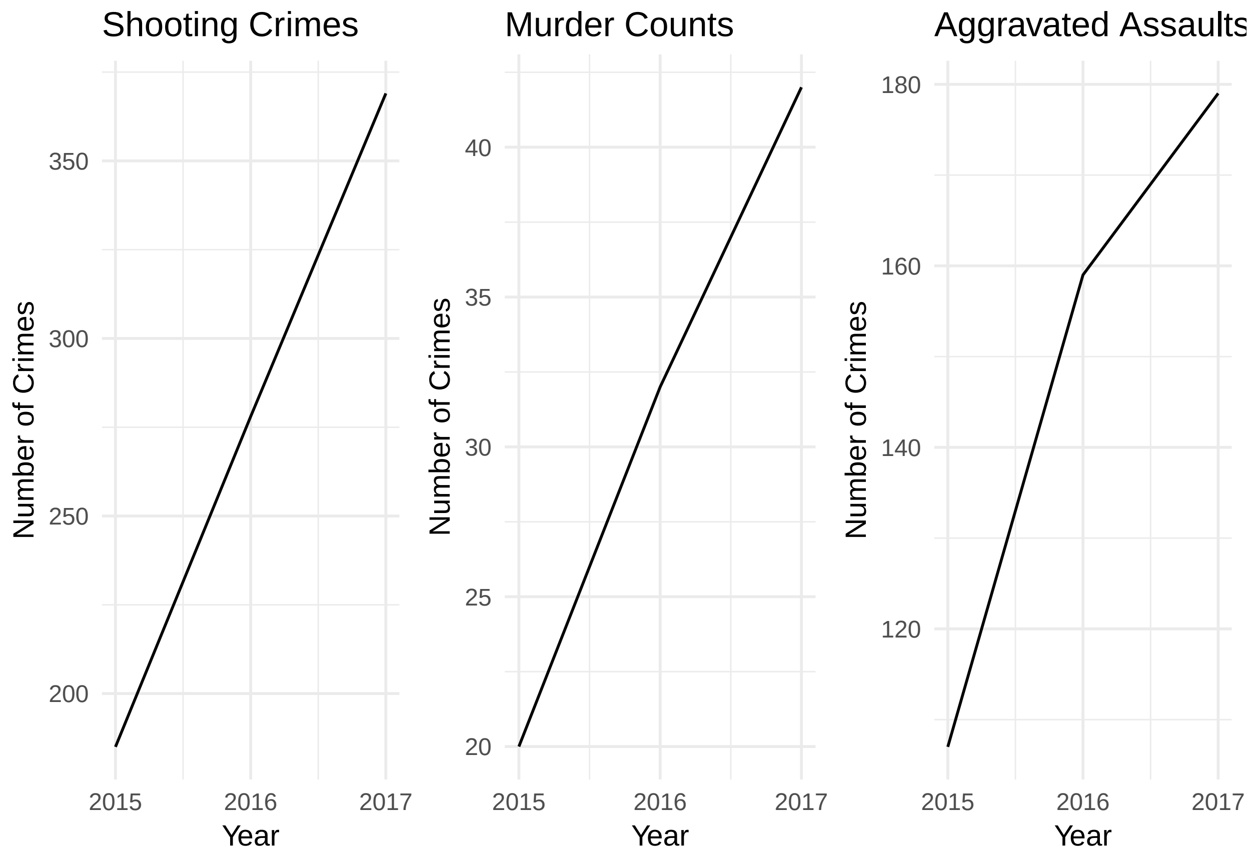
Lets take a look at the crimes involving shooting. The shooting column of the Crime Incidents Reports data indicates that a shooting took place. Lets take a look at the This data is from 2015 - 2018.




As we can see from the table above, over 50% of crimes involving shooting were aggravated assaults. At around 10% of shooting crimes, there were 102 homicides. This is a pretty low number for 3 years in a major city - good job BPD!

Lets take a look at how the number of murders by shooting has changed over the years 2015 - 2018.

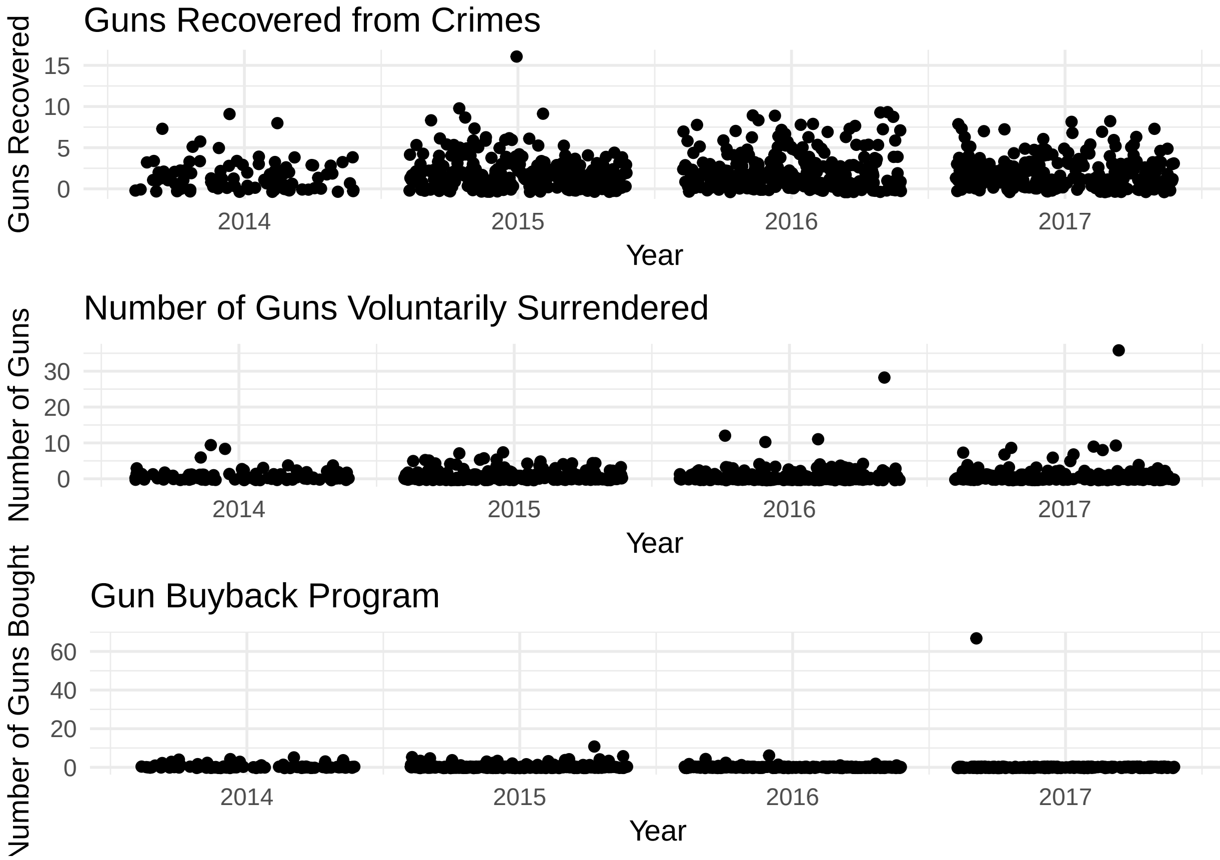
```
## # A tibble: 6 x 4
##   CollectionDate CrimeGunsRecovered GunsSurrenderedSafe~ BuybackGunsRecov~
##   <chr>                       <int>                <int>             <int>
## 1 8/20/2014                       2                    3                 1
## 2 8/21/2014                       2                    0                 4
## 3 8/22/2014                       0                    0                 2
## 4 8/25/2014                       8                    3                 0
## 5 8/26/2014                       9                    0                 0
## 6 8/27/2014                       1                    0                 0
```




```
## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_point).
```




From the first set of graphs we see the following:

- The number of shooting crimes jumps by about 80 each year above the previous year
- The number of murders jumps by about 10 each year more than the previous year
- The number of aggravated assaults rises, but the rate of increase has been slowed significantly between 2016 and 2017

From the second set of graphs related to gun recovery we see the following:

- The number of guns retrieved from crimes increased from 2014 to 2015, but there wasn't a large increase for the years 2015 - 2017
- The number of guns voluntarily surrendered seems to have been relatively consistent except for a few outliers (such as the 30+ guns surrendered in 2017)
- The gun buyback program has returned less guns than police work, but is still making a dent in the number of guns on the street. There is one remarkable data point in which 60+ guns were recovered.

## Economic Indicators

Lets take a look at the Economic Indicators Dataset. This set contains information on

- Tourism/Flights
- Hotel Market
- Labor Market
- Real Estate: Board Approved Development Projects (Pipeline)

- Real Estate Market: Housing

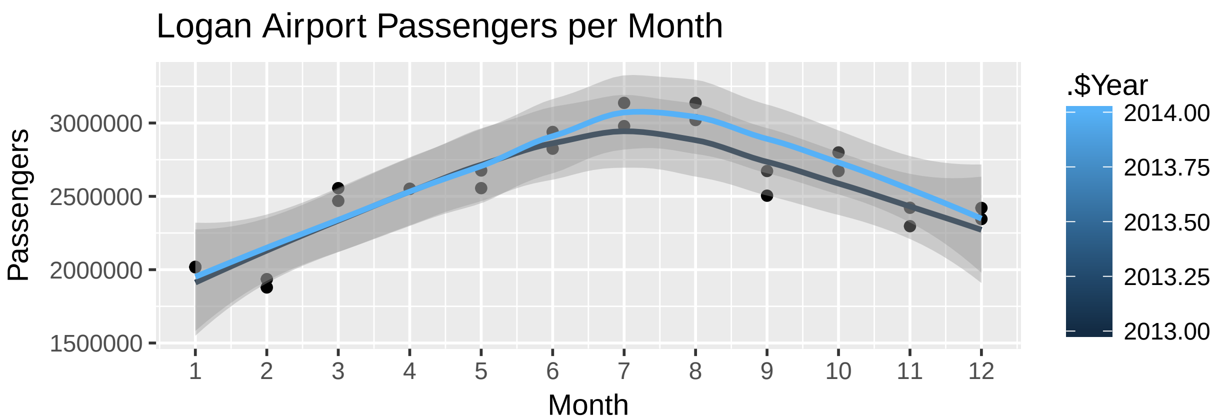
```
## # A tibble: 6 x 19
##    Year Month logan_passengers logan_intl_flights hotel_occup_rate
##   <int> <int>            <int>              <int>            <dbl>
## 1  2013     1          2019662               2986            0.572
## 2  2013     2          1878731               2587            0.645
## 3  2013     3          2469155               3250            0.819
## 4  2013     4          2551246               3408            0.855
## 5  2013     5          2676291               3240            0.858
## 6  2013     6          2824862               3402            0.911
## # ... with 14 more variables: hotel_avg_daily_rate <dbl>,
## #   total_jobs <int>, unemp_rate <dbl>, labor_force_part_rate <dbl>,
## #   pipeline_unit <int>, pipeline_total_dev_cost <dbl>,
## #   pipeline_sqft <int>, pipeline_const_jobs <dbl>, foreclosure_pet <int>,
## #   foreclosure_deeds <int>, med_housing_price <int>,
## #   housing_sales_vol <int>, new_housing_const_permits <int>,
## #   `new-affordable_housing_permits` <int>
```

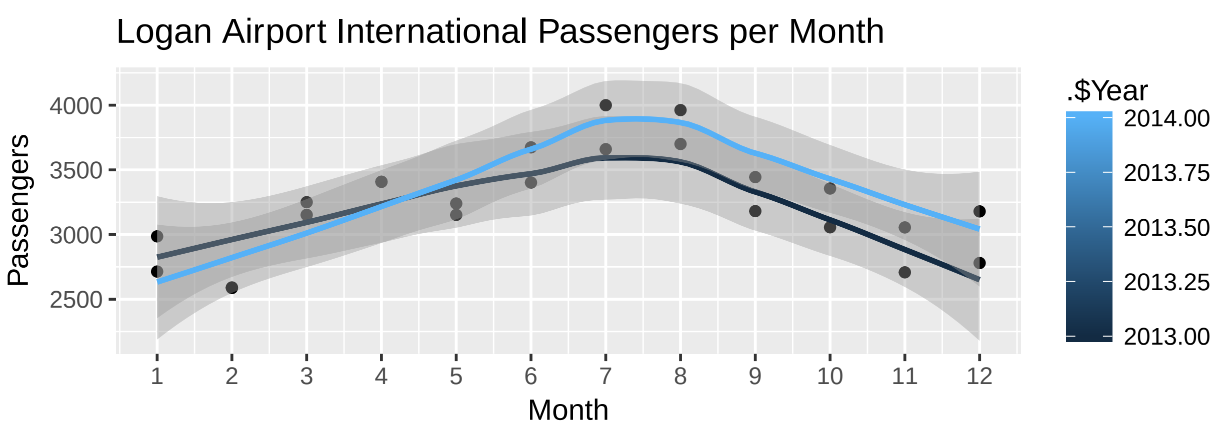
```
##  [1] "Year"                           "Month"
##  [3] "logan_passengers"               "logan_intl_flights"
##  [5] "hotel_occup_rate"               "hotel_avg_daily_rate"
##  [7] "total_jobs"                     "unemp_rate"
##  [9] "labor_force_part_rate"          "pipeline_unit"
## [11] "pipeline_total_dev_cost"        "pipeline_sqft"
## [13] "pipeline_const_jobs"            "foreclosure_pet"
## [15] "foreclosure_deeds"              "med_housing_price"
## [17] "housing_sales_vol"              "new_housing_const_permits"
## [19] "new-affordable_housing_permits"
```

```
## [1] 0
```

## Flights / Tourism

```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```



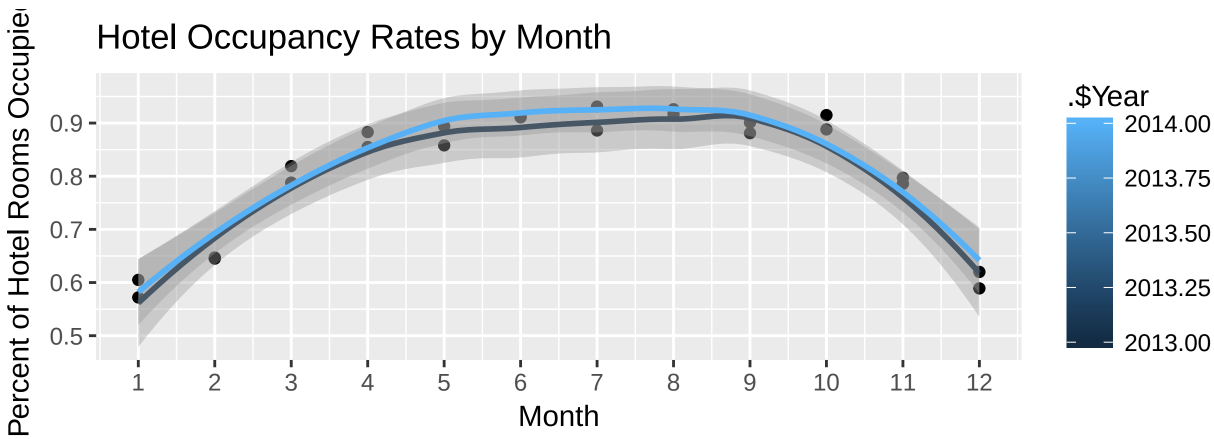



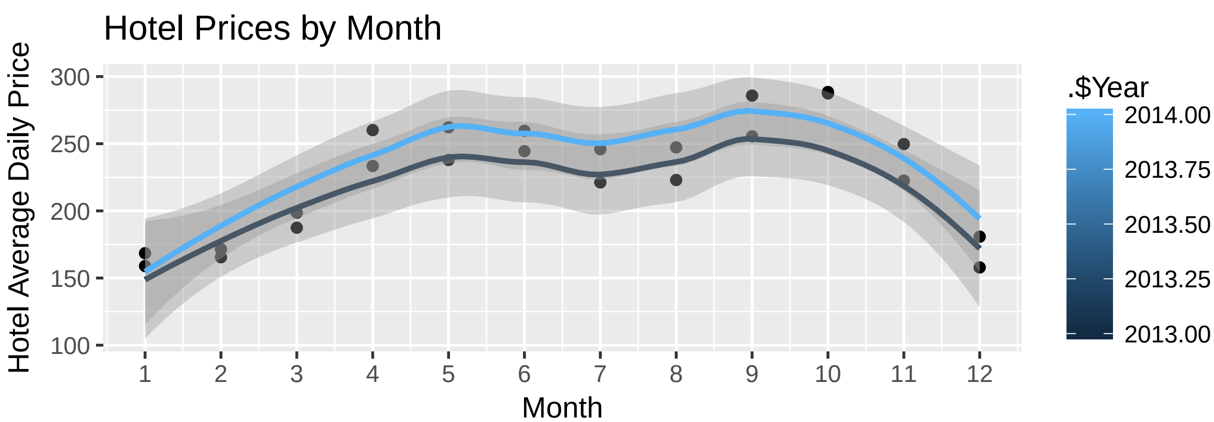

From the graphs above we see the following trends:

- 2014 was a slightly better year for tourism. This makes sense from an economic perspective as we will see in the graphs that follow.
- Flights jump about 50% from the beginning of the year to the summer and then level off as it gets colder
- There is roughly a thousand times as many domestic passengers as there are international passengers

## Hotels

```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```



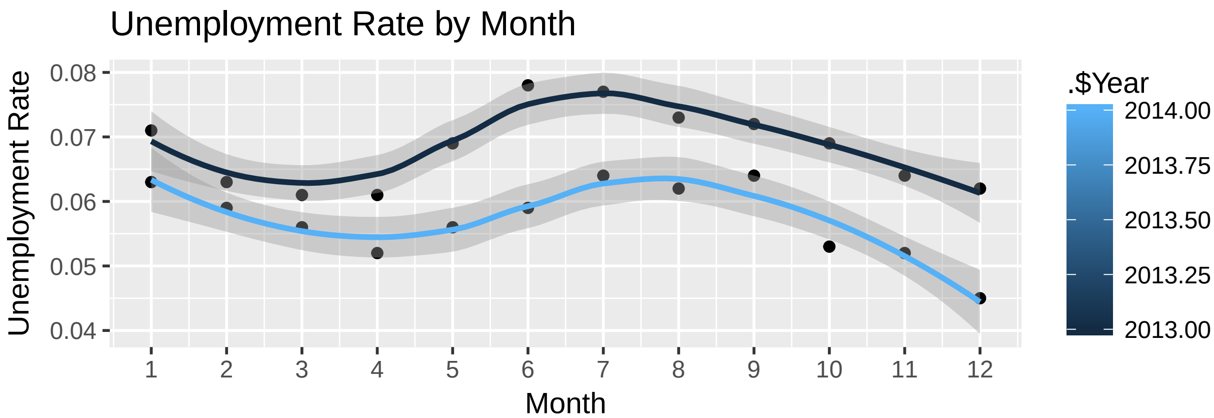



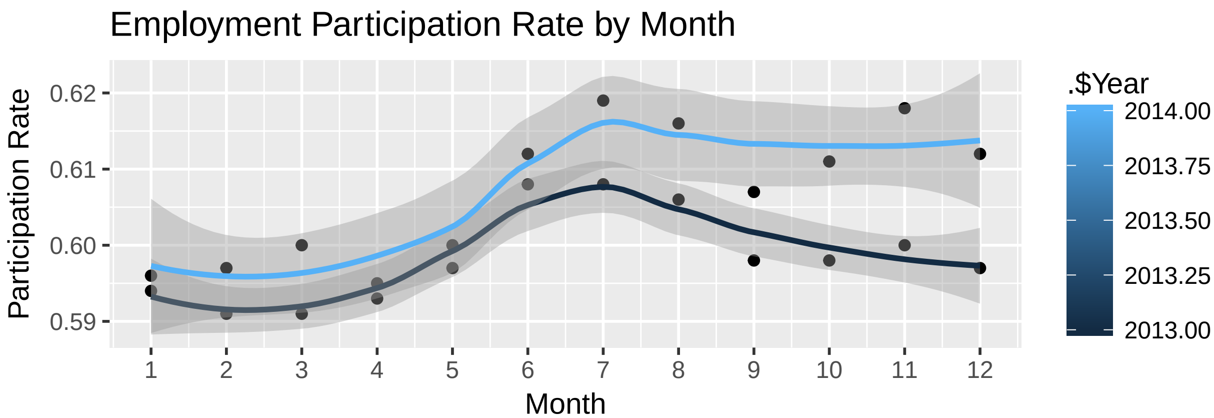

From the graphs above we see the following trends:

- 2014 had similar occupation rates to 2013
- The average price of hotels increased from 2013 to 2014
- The hotels are around half full during the winter and almost completely full during the summer
- Christmas is not that popular of a time to book a hotel in Boston
- The price of hotels increases as occupation increases

## Job Market

```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```




The graphs above show the following trends:

- 2013 had a higher unemployment rate than 2014
- Unemployment dropped on average 1%-2% between the years, dependent on the month
- Employees were more likely to participate in the employment market in 2014 than 2013.
- Unemployment rates are generally around 6-8%
- Market participation is roughly 60%

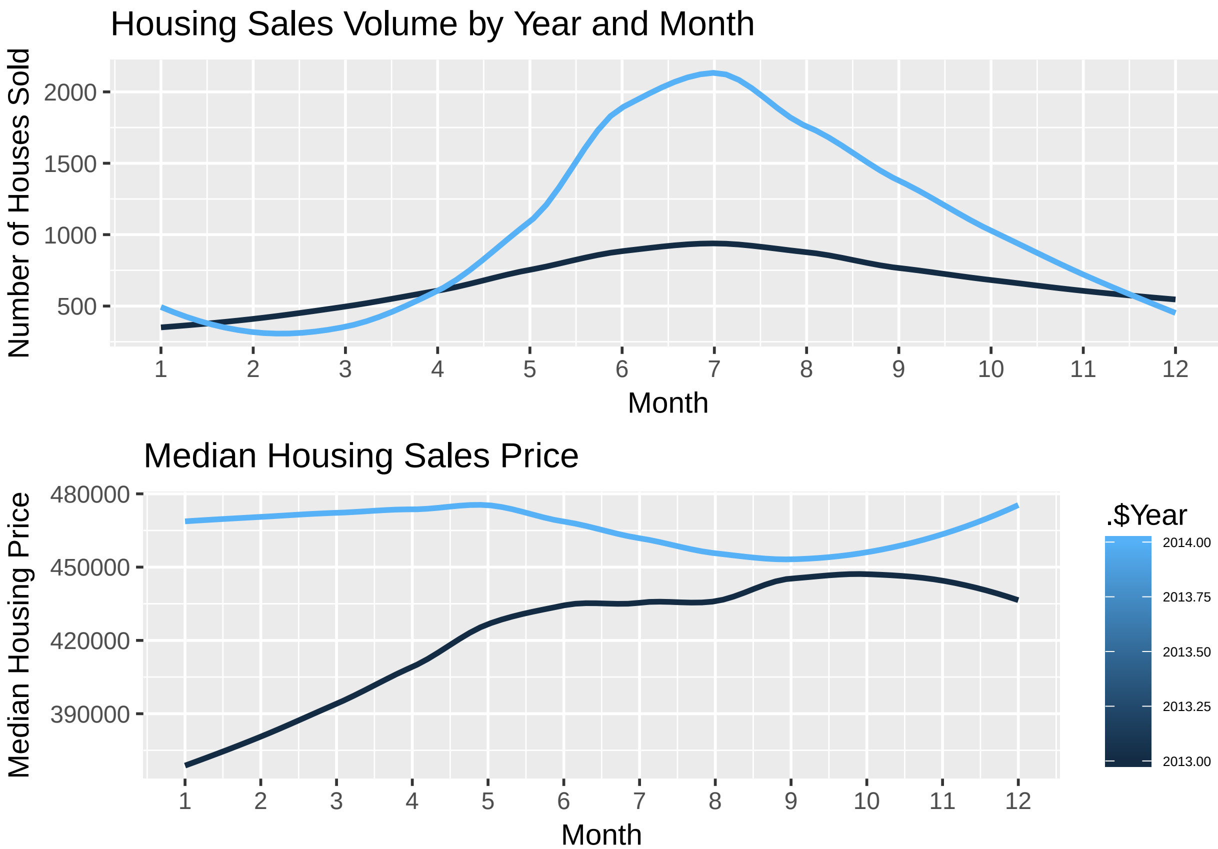
## Real Estate Board Approved Development Projects

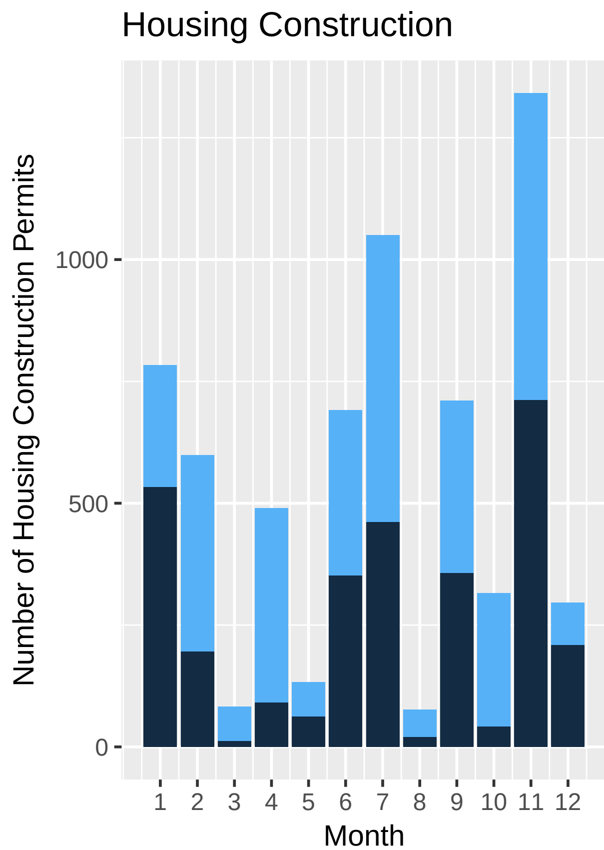

The graphs above show the following trends:

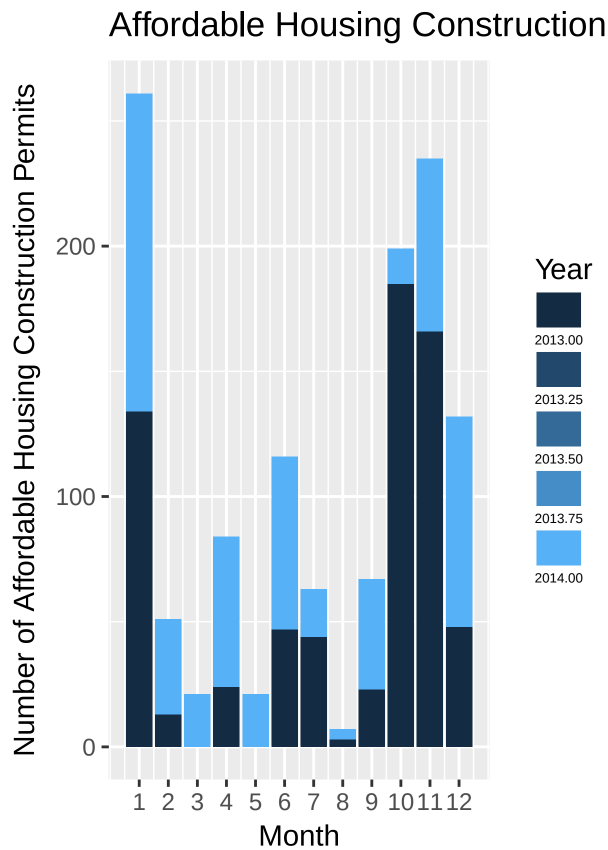
- 2014 had a significant amount more development than 2013
- The boom in development also lead to an increase in construction jobs
- The units that were developed also took up more space

# Housing

```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```

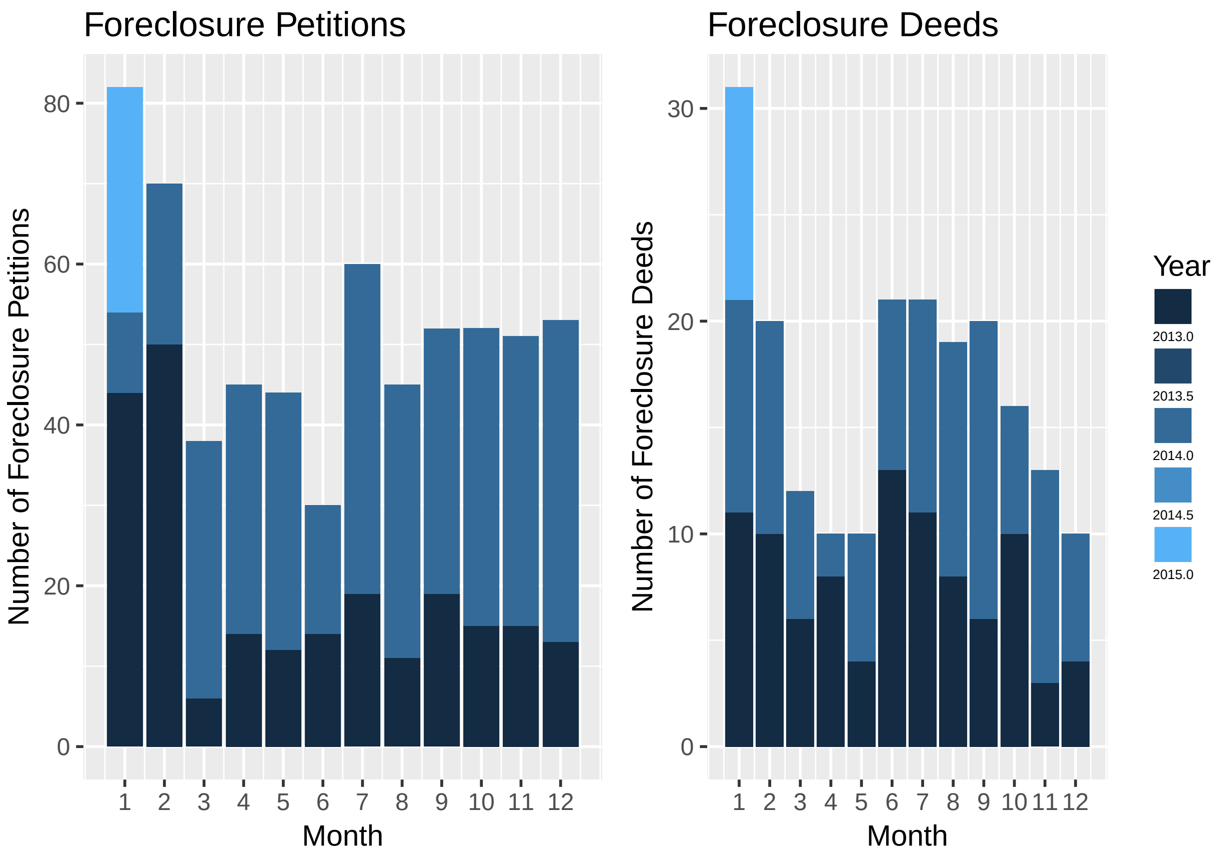





From the above graphs we can see the following trends:

- Housing sales increase during the summer
- The housing market had over twice the number of sales in 2014 and it did in 2013
- Houses sold in 2013 started off quite low at the beginning of the year and gradually increased in price
- Housing sales prices were roughly uniform over 2014
- Housing construction (affordable and regular) increased a good amount in 2014
- The number of construction permits for regular housing was quite high in November
- There was a lot more affordable housing being built in 2014 than in 2013




A foreclosure petition is when a lender (typically a bank) sues a delinquent tenant by filing a court document for foreclosure. This petition for foreclosure is then delivered to the homeowner along with a court summons.

A foreclosure deed is when a lender accepts the deed (document stating ownership) of a property instead of foreclosing on a house.

The graphs above show the following trends:

- The amount of foreclosures rose in 2014 from 2013
- Foreclosure notices in 2013 and 2014 were generally distributed between all the months of the year
- Foreclosure notices in 2015 were generally given in January
- 2015 seems to have had less foreclosures than 2013 and 2014 (perhaps a sign of the housing market crash of 08 recovering)