

Analysis of NYC 311 Service Requests

John Perez

For this exercise I chose to focus on Housing Preservation & Development (HPD) Service Requests between 1/1/2018 and 12/31/2018. I initially downloaded the CSV as per instructions, loaded it into a dataframe in R, and performed various transformations and operations to answer the following questions:

1) How many total HPD Complaints in 2018?

Counting the unique values for the **Unique.Key** column, gives 607,278 complaints received for HPD.

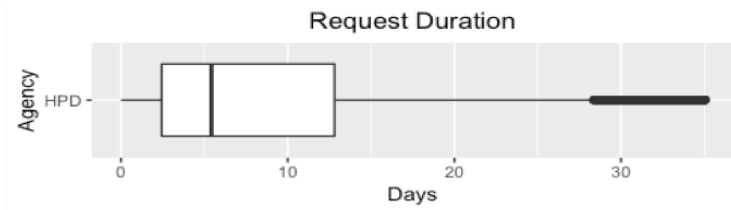
2) How many complaints were there for each borough?

Here I create a subset dataframe containing **Complaint.Type**, and **Borough** variables, then `group_by(borough)` and count number of complaints for each.

Borough	Complaints
BROOKLYN	201382
BRONX	192121
MANHATTAN	124412
QUEENS	74609
STATEN ISLAND	10798
Unspecified	3956

3) How long does it take to resolve HPD complaints?

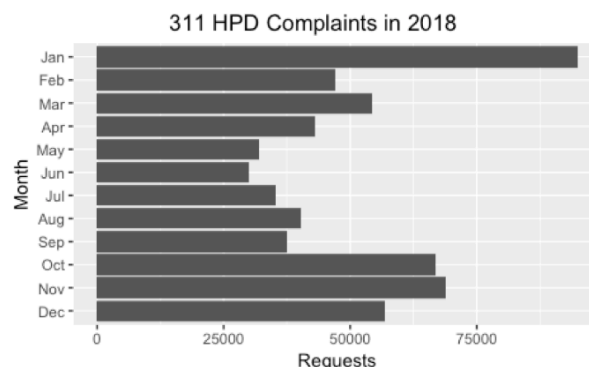
To answer this question, I create a vector **comp_diff** by subtracting the `Closed.Date` from `Created.Date`, using `lubridate` package to parse the date strings, then convert this into days unit. I then run `summary` function on that vector to conclude that the median turnaround time is ~6 days, with 75% of requests taking approximately 3-15 days. (median: 5.89, mean: 11.16, 1Q: 2.58, 3Q: 14.64, min: 0, max: 367.8)



4) What is the Monthly Breakdown of HPD complaints for 2018?

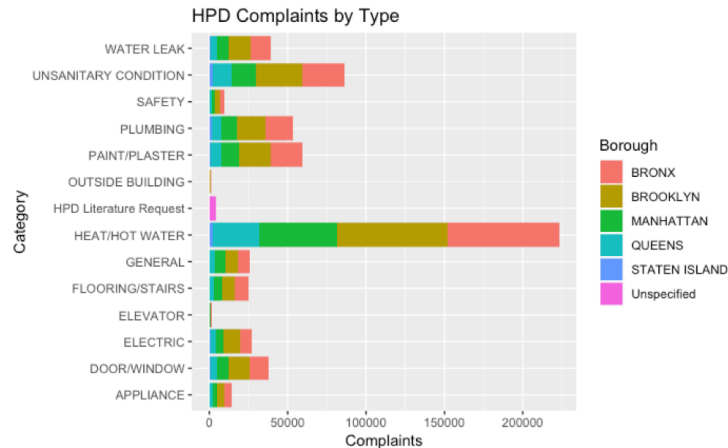
For this analysis, I created a subset dataframe containing the **Unique.Key**, **Created.Date**, and **Complaint.Type** columns, used the `mutate()` function (`dplyr`) to add a column specifying that respective claim's **Month**, then counted the total complaints for each group.

As shown below, the number of complaints increases significantly during the fall/winter months beginning in October, peaking in January, then reducing during the summer months.



5) How many complaints were there for each type for the different boroughs?

Here, I create a subset dataframe, **comp_types** which groups the complaints by type and borough, then use the summarize function to count each. There is an outlier complaint with type “AGENCY” so I’ve removed that for graphing purposes and later use ggplot to visualize the counts. As shown in the “HPD Complaints by Type” above, Heat/Hot water, Unsanitary Condition, and Paint/Plaster were the top 3 complaints in 2018.



6) What days and times of the week see the most complaints?

For this analysis I created an R function **heatmap()** which can be used for all agencies combined, or any individual agency and date range desired (given the data is available). This function takes agency abbreviation, start date, end date as parameters and outputs a heatmap visualization with each hour of each day having a color gradient depending on the total volume for that hour. As shown, for 2018, weekdays are busier than weekends, with Monday being the busiest day. The hours between 10AM and 3PM see the highest complaint volumes.

