In [50]:
```
# Add low points to the figure
```

In [ ]:

In [ ]:

Homework

The JFK international airport is located within longitude -73.82 to -73.78 and latitude 40.63 to 40.67. Extract all taxi trips who picked up passengers from this area.

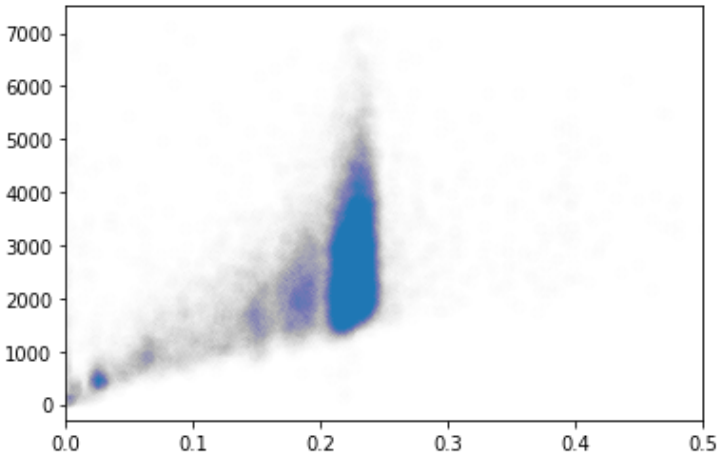
1. For this sub-dataset only, visualize the correlation between the aerial distance and the trip duration.
2. For this sub-dataset only, visualize the drop-off location using dropoff_longitude and dropoff_latitude.
3. There was a snow storm on Jan 23, 2016. Is the distribution of trip_duration on that day different from the rest of the year?

In [51]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

In [66]:
```
# Extract all taxi trips who picked up passengers from this area.

index = (taxidata["pickup_longitude"] > -73.82) & (taxidata["pickup_longitude"]
data = taxidata[index]
index2 = (data["pickup_latitude"] > 40.63) & (data["pickup_latitude"] < 40.67)
subdata = data[index2]
```

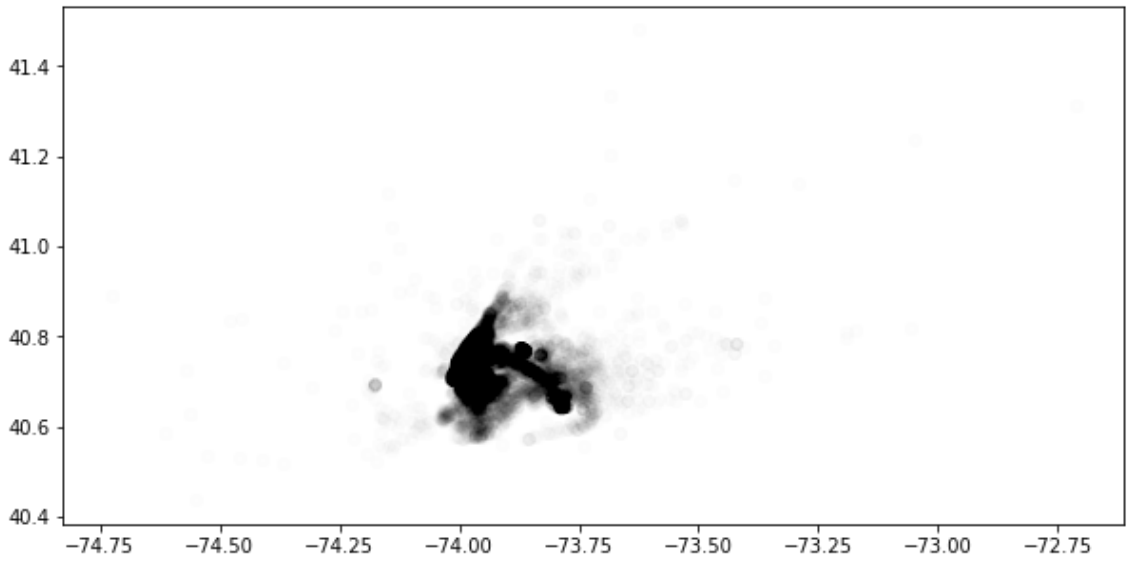
In [65]:
```
# 1. For this sub-dataset only, visualize the correlation between the aerial di

plt.scatter(subdata['aerial_dist'], subdata['trip_duration'], alpha=0.005)
plt.xlim(0, 0.5)
```

Out[65]:
```
(0, 0.5)
```



In [75]:
```
# 2. For this sub-dataset only, visualize the drop-off location using dropoff_l
plt.figure(figsize=(10, 5))
plt.plot(subdata['dropoff_longitude'],
         subdata['dropoff_latitude'],

         'ko',
         alpha=0.006)

# plt.xlim(-73.82, -73.78)
# plt.ylim(40.63, 40.67)
```

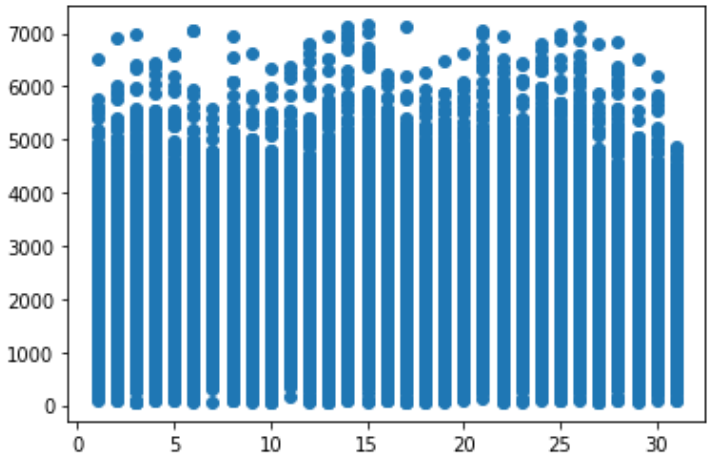
Out[75]:
```
[<matplotlib.lines.Line2D at 0x12639fc18>]
```



In [106]:
```
# fig = plt.figure()
# ax = fig.add_subplot(111)

# ax.plot(data['close'])

subdata['pickup_year'] = subdata['pickup_datetime'].dt.year
index = (subdata["pickup_year"] > 2015) & (subdata["pickup_year"] < 2017)
subdata_year = subdata[index]

subdata_year["pickup_month"] = subdata_year["pickup_datetime"].dt.month
subdata_year["pickup_day"] = subdata_year["pickup_datetime"].dt.day

plt.scatter(subdata_year['pickup_day'], subdata_year['trip_duration'])

# plt.figure(figsize=(10, 5))
# plt.plot(subdata_year["pickup_month"],
#          subdata_year["trip_duration"],
#          'bo',
#          alpha=0.006)


# subdata_year['trip_duration'].hist()
# subdata_year.head()
```

```
//anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:6: SettingWithCo
pyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/sta
ble/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas-d
ocs/stable/indexing.html#indexing-view-versus-copy)
```

Out[106]:
```
<matplotlib.collections.PathCollection at 0x1339b0208>
```



In [101]:
```
subdata_year.head()
```

Out[101]:

ickup_latitude	dropoff_longitude	dropoff_latitude	store_and_fwd_flag	trip_duration	pickup_hour	dayofweek
40.657879	-73.809189	40.690182	N	3528	11	1
40.644646	-73.981125	40.720886	N	1782	23	2
40.643559	-73.970665	40.687279	N	2824	16	2
40.643707	-73.978699	40.750343	N	2065	20	4
40.641460	-73.971771	40.749409	N	1884	20	0