

NY Taxi Rides

Summary of NYC Taxi data

```
In [5]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [7]: data_filename = "nyc_data.csv"
```

```
In [9]: data = pd.read_csv(data_filename, parse_dates=['pickup_datetime', 'dropoff_datetime'])
```

```
In [10]: # see documentation string
pd.read_csv?
```

```
In [11]: data.head()
```

```
Out[11]:
```

	medallion	hack_license	vendor_id	rate_
0	76942C3205E17D7E7FE5A9F709D16434	25BA06A87905667AA1FE5990E33F0E2E	VT	
1	517C6B330DBB3F055D007B07512628B3	2C19FBEE1A6E05612EFE4C958C14BC7F	VT	
2	ED15611F168E41B33619C83D900FE266	754AEBD7C80DA17BA1D81D89FB6F4D1D	CM	
3	B33E704CC189E80C9671230C16527BBC	6789C77E1CBDC850C450D72204702976	VT	
4	BD5CC6A22D05EB2D5C8235526A2A4276	5E8F2C93B5220A922699FEB AFC2F7A54	VT	

In [12]: `data.describe()`

Out[12]:

	rate_code	passenger_count	trip_time_in_secs	trip_distance	pickup_longitude	pickup_
count	846945.000000	846945.000000	8.469450e+05	8.469450e+05	846945.000000	846945
mean	1.026123	1.710272	8.125239e+02	9.958211e+00	-73.975155	40
std	0.223480	1.375266	1.609831e+04	6.525205e+03	0.035142	0
min	0.000000	0.000000	-1.000000e+01	0.000000e+00	-74.098305	40
25%	1.000000	1.000000	3.610000e+02	1.050000e+00	-73.992371	40
50%	1.000000	1.000000	6.000000e+02	1.800000e+00	-73.982094	40
75%	1.000000	2.000000	9.600000e+02	3.200000e+00	-73.968048	40
max	6.000000	6.000000	4.294796e+06	6.005123e+06	-73.028473	40

Coordinates

In [13]: `p_lng = data.pickup_longitude`
`p_lat = data.pickup_latitude`

In [14]: `p_lng.head()`

Out[14]:

```
0    -73.955925
1    -74.005501
2    -73.969955
3    -73.991432
4    -73.966225
Name: pickup_longitude, dtype: float64
```

In [15]:

```
def lat_lng_to_pixels(lat, lng):
    lat_rad = lat * np.pi / 180.0
    lat_rad = np.log(np.tan((lat_rad + np.pi / 2.0) / 2.0))
    x = 100 * (lng + 180.0) / 360.0
    y = 100 * (lat_rad - np.pi) / (2.0 * np.pi)
    return (x, y)
```

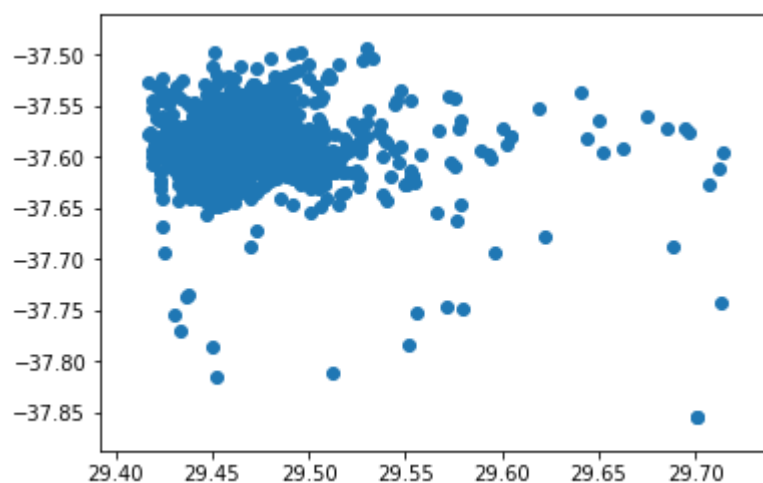
In [16]: `px, py = lat_lng_to_pixels(p_lat, p_lng)`
`#py.head()`
`type(py)`

Out[16]: `pandas.core.series.Series`

Scatter plot of pickup locations

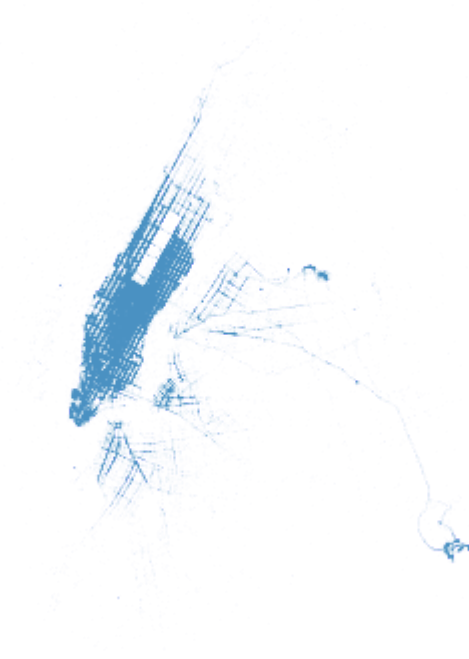
```
In [17]: plt.scatter(px, py)
```

```
Out[17]: <matplotlib.collections.PathCollection at 0x229d525b8d0>
```



```
In [18]: plt.figure(figsize=(8, 6))  
plt.axis('equal')  
plt.xlim(29.40, 29.55)  
plt.ylim(-37.63, -37.54)  
plt.axis('off')  
plt.scatter(px, py, s=.1, alpha=0.03)
```

```
Out[18]: <matplotlib.collections.PathCollection at 0x229d7a5b0b8>
```



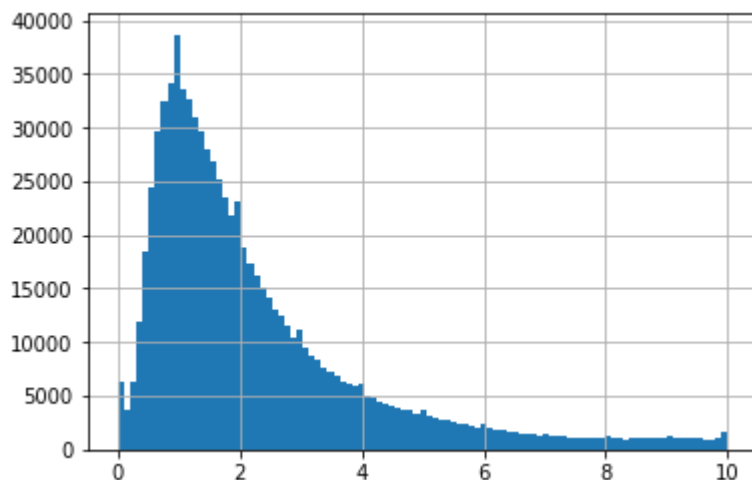
Histogram of the trip distance

```
In [19]: bin_array = np.linspace(start=0., stop=10., num=100)
         bin_array
```

```
Out[19]: array([ 0.          ,  0.1010101 ,  0.2020202 ,  0.3030303 ,  0.4040404 ,
                0.50505051,  0.60606061,  0.70707071,  0.80808081,  0.90909091,
                1.01010101,  1.11111111,  1.21212121,  1.31313131,  1.41414141,
                1.51515152,  1.61616162,  1.71717172,  1.81818182,  1.91919192,
                2.02020202,  2.12121212,  2.22222222,  2.32323232,  2.42424242,
                2.52525253,  2.62626263,  2.72727273,  2.82828283,  2.92929293,
                3.03030303,  3.13131313,  3.23232323,  3.33333333,  3.43434343,
                3.53535354,  3.63636364,  3.73737374,  3.83838384,  3.93939394,
                4.04040404,  4.14141414,  4.24242424,  4.34343434,  4.44444444,
                4.54545455,  4.64646465,  4.74747475,  4.84848485,  4.94949495,
                5.05050505,  5.15151515,  5.25252525,  5.35353535,  5.45454545,
                5.55555556,  5.65656566,  5.75757576,  5.85858586,  5.95959596,
                6.06060606,  6.16161616,  6.26262626,  6.36363636,  6.46464646,
                6.56565657,  6.66666667,  6.76767677,  6.86868687,  6.96969697,
                7.07070707,  7.17171717,  7.27272727,  7.37373737,  7.47474747,
                7.57575758,  7.67676768,  7.77777778,  7.87878788,  7.97979798,
                8.08080808,  8.18181818,  8.28282828,  8.38383838,  8.48484848,
                8.58585859,  8.68686869,  8.78787879,  8.88888889,  8.98989899,
                9.09090909,  9.19191919,  9.29292929,  9.39393939,  9.49494949,
                9.5959596 ,  9.6969697 ,  9.7979798 ,  9.8989899 , 10.          ])
```


```
In [20]: data.trip_distance.hist(bins=bin_array)
```

```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x229d91eafd0>
```



Widgets

```
In [21]: from ipywidgets import interact
@interact
def show_nrows(distance_threshold=(0, 200)):
    return len(data.loc[data.trip_distance > distance_threshold])
```

distance_th...  54

17

In []: