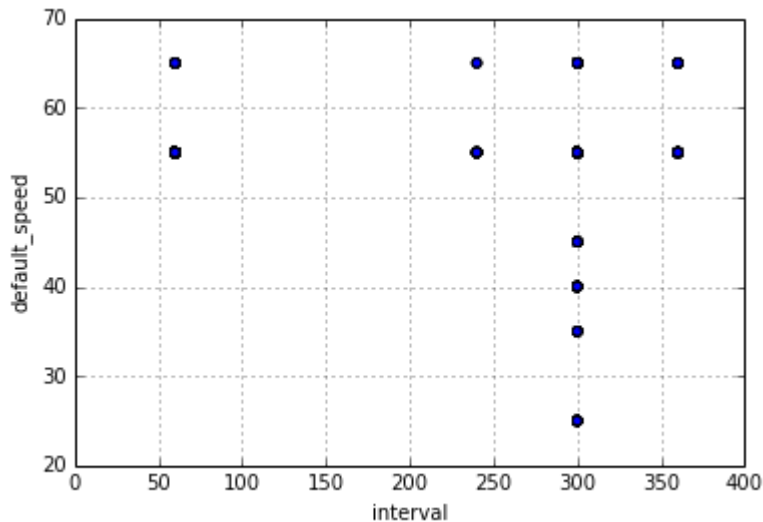


```
interval_df = df.groupby('interval')
%matplotlib inline
df.plot(kind='scatter', x='interval', y='default_speed')
```



This helps in seeing if the default\_speed is varying with the interval size.

This helps in seeing what values of default\_speed were taken in which interval. This might help us in the cleaning task.

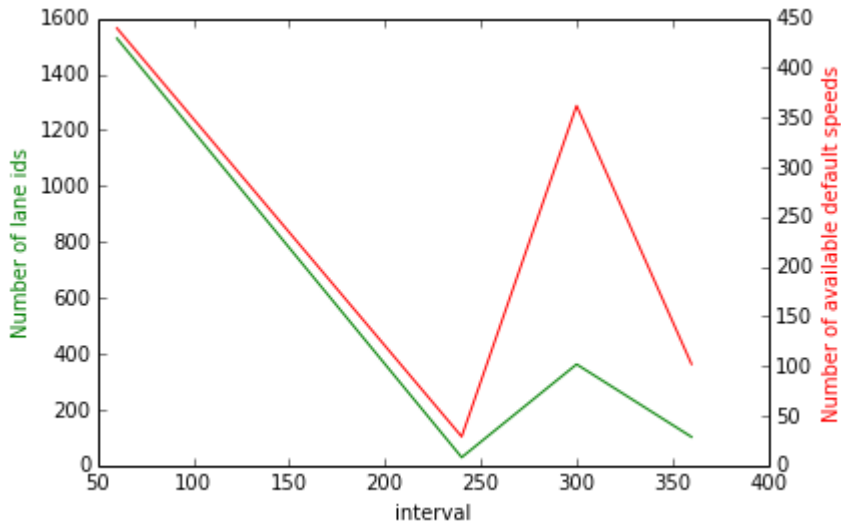
```
%matplotlib inline
```

```
import matplotlib.pyplot as plt
fig, ax1 = plt.subplots()
ax2 = ax1.twinx()
x = interval_df['interval'].nunique().index
ax1.plot(x, interval_df['lane_id'].nunique(), 'g-')
ax2.plot(x, interval_df['default_speed'].count(), 'r-')
```

```
ax1.set_xlabel('interval')
```

```
ax1.set_ylabel('Number of lane ids', color='g')
```

```
ax2.set_ylabel('Number of available default speeds', color='r')
```



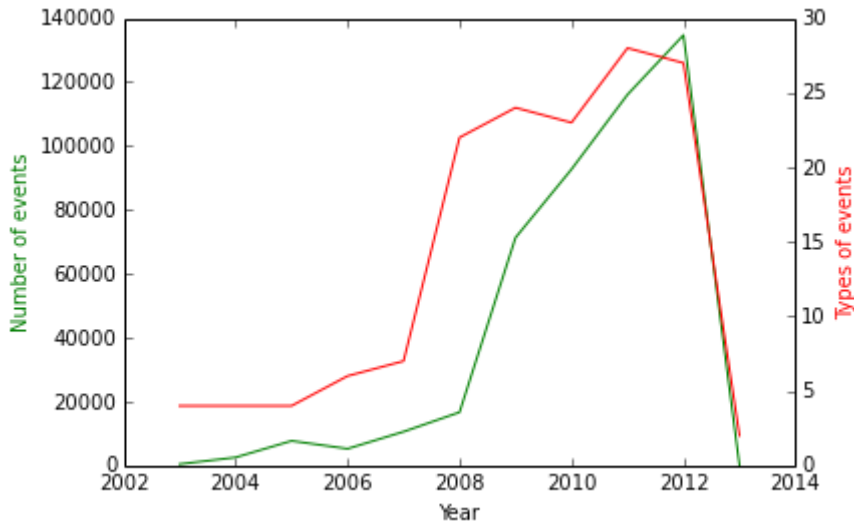
This helps in seeing the number of lane ids in an interval along with the number of available default\_speed values for the intervals. This helps in determining if the default\_speed values can be used for flow correction in the cleaning task. On analyzing it, it shows that most of the lane ids have an interval of value 60 and many of them don't have a default\_speed. So default\_speed cannot be used for flow correction where it is missing.

```
is_task_not_null = ~isNaN(temp_df_no_duplicates['event_type'])
temp_df_no_duplicates_event = temp_df_no_duplicates[is_task_not_null]
%matplotlib inline
```

```
import matplotlib.pyplot as plt
fig, ax1 = plt.subplots() ax2 =
ax1.twinx()
x = event_year_grouped['year'].nunique().index
```

```
ax1.plot(x, event_year_grouped['event_id'].nunique(), 'g-') ax2.plot(x,
event_year_grouped['event_type'].nunique(), 'r-')
```

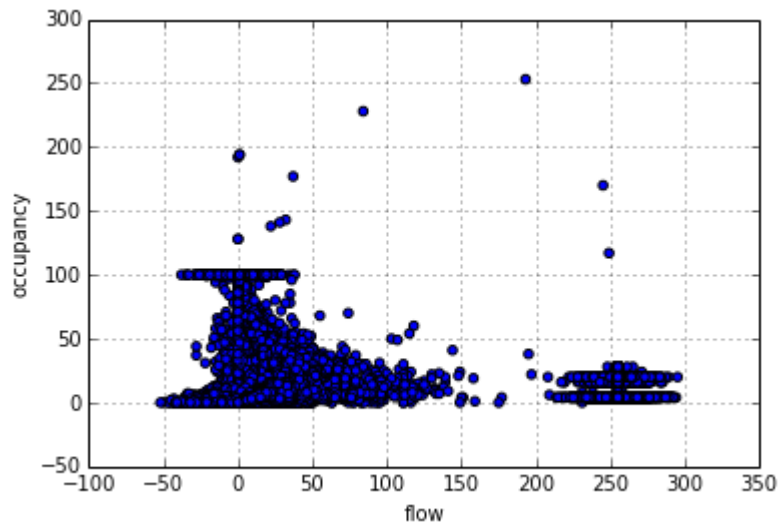
```
ax1.set_xlabel('Year') ax1.set_ylabel('Number
of events', color='g') ax2.set_ylabel('Types of
events', color='r')
```



This plot helps in showing the number of events which took place in each year and the number of the types of events which took place each year. This might help us in our prediction task and show which tuples might be more variable (when it comes to the type and number of events).

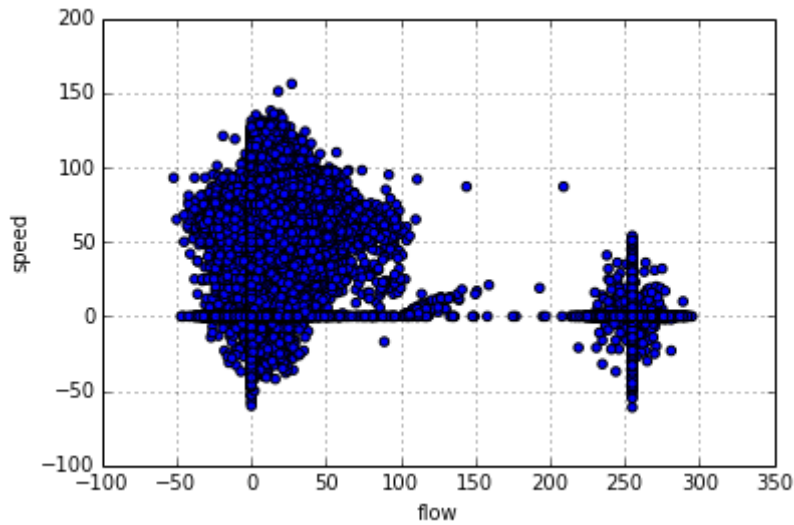
```
%matplotlib inline
```

```
events_ds.plot(kind='scatter', x='flow', y='occupancy')
```

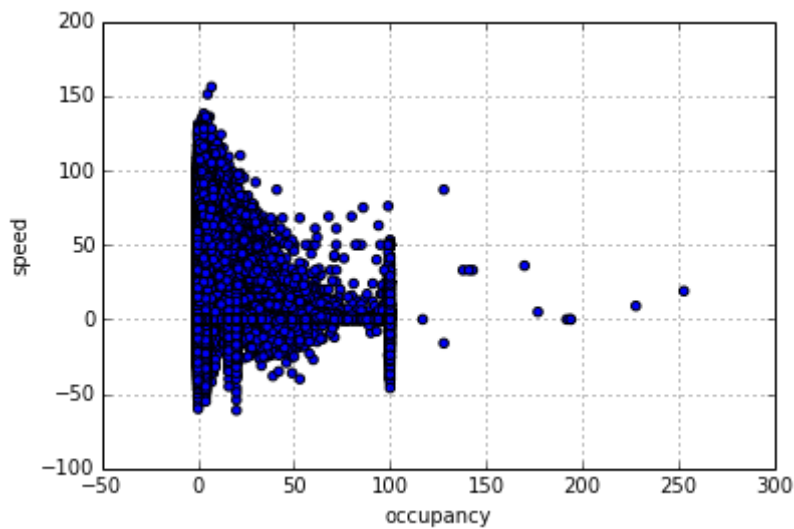


```
%matplotlib inline
```

```
events_ds.plot(kind='scatter', x='flow', y='speed')
```



```
%matplotlib inline
events_ds.plot(kind='scatter', x='occupancy', y='speed')
```



These plots show how the speed, flow and occupancy vary with each other. We thought that this might be useful in the cleaning task. Based on this we could see some erroneous data and outliers.