Homework 5- Arianna Burford

```
import pandas as pd
print("import package libraries")
```

```
import package libraries
```

```
print("Loading dataset, this may take some time...")
tree_census = pd.read_csv('trees.csv')
```

```
Loading dataset, this may take some time...
```

```
tree_census.head()
```

	created_at	tree_id	block_id	the_geom	tree_dbh	stump_diam	curb_loc	status	health	spc_latin	...	st_assem	st_senate	nta	nta_name	boro
0	08/27/2015	180,683	348,711	POINT (-73.84421521958048 40.723091773924274)	3	0	OnCurb	Alive	Fair	Acer rubrum	...	28	16	QN17	Forest Hills	4073
1	09/03/2015	200,540	315,986	POINT (-73.81867945834878 40.79411066708779)	21	0	OnCurb	Alive	Fair	Quercus palustris	...	27	11	QN49	Whitestone	4097
2	09/05/2015	204,026	218,365	POINT (-73.93660770459083 40.7175807400991116)	3	0	OnCurb	Alive	Good	Gleditsia triacanthos var. inermis	...	50	18	BK90	East Williamsburg	3044
3	09/05/2015	204,337	217,969	POINT (-73.93445615919741 40.713537494833226)	10	0	OnCurb	Alive	Good	Gleditsia triacanthos var. inermis	...	53	18	BK90	East Williamsburg	3044
4	08/30/2015	189,565	223,043	POINT (-73.97597938483258 40.66677775537875)	21	0	OnCurb	Alive	Good	Tilia americana	...	44	21	BK37	Park Slope-Gowanus	3016

5 rows × 42 columns

```
tree_census.tail()
```

	created_at	tree_id	block_id	the_geom	tree_dbh	stump_diam	curb_loc	status	health	spc_latin	...	st_assem	st_senate	nta	nta_name	b
683783	08/18/2015	155,433	217,978	POINT (-73.95494401022562 40.713210783145)	25	0	OnCurb	Alive	Good	Quercus palustris	...	53	18	BK73	North Side-South Side	3(
683784	08/29/2015	183,795	348,185	POINT (-73.85665019989099 40.71519444267162)	7	0	OnCurb	Alive	Good	Cladrastis kentukea	...	28	15	QN17	Forest Hills	4(
683785	08/22/2015	166,161	401,670	POINT (-74.13651724205825 40.62076152739799)	12	0	OnCurb	Alive	Good	Acer rubrum	...	63	24	SI07	Westerleigh	5(
683786	08/29/2015	184,028	504,204	POINT (-73.90311472453581 40.850828186655754)	9	0	OnCurb	Alive	Good	Acer rubrum	...	86	33	BX41	Mount Hope	2(
683787	09/03/2015	200,607	306,527	POINT (-73.78752645502483 40.73216525220126)	23	0	OnCurb	Alive	Fair	Acer rubrum	...	25	11	QN41	Fresh Meadows-Utopia	41

5 rows × 42 columns

```
tree_census.columns
```

```
Index(['created_at', 'tree_id', 'block_id', 'the_geom', 'tree_dbh',
       'stump_diam', 'curb_loc', 'status', 'health', 'spc_latin', 'spc_common',
       'steward', 'guards', 'sidewalk', 'user_type', 'problems', 'root_stone',
       'root_grate', 'root_other', 'trnk_wire', 'trnk_light', 'trnk_other',
       'brnch_ligh', 'brnch_shoe', 'brnch_othe', 'address', 'zipcode',
       'zip_city', 'cb_num', 'borocode', 'boroname', 'cncldist', 'st_assem',
       'st_senate', 'nta', 'nta_name', 'boro_ct', 'state', 'Latitude',
       'Longitude', 'x_sp', 'y_sp'],
      dtype='object')
```

```
tree_census.shape
```

```
(683788, 42)
```

```
tree_census.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 683788 entries, 0 to 683787
Data columns (total 42 columns):
 #   Column      Non-Null Count   Dtype  
---  ------      --------------   -----  
 0   created_at  683788 non-null  object 
 1   tree_id     683788 non-null  object 
 2   block_id    683788 non-null  object 
 3   the_geom    683788 non-null  object 
 4   tree_dbh    683788 non-null  int64  
 5   stump_diam  683788 non-null  int64  
 6   curb_loc    683788 non-null  object 
 7   status      683788 non-null  object 
 8   health      652172 non-null  object 
 9   spc_latin   652169 non-null  object 
 10  spc_common  652169 non-null  object 
 11  steward     164350 non-null  object 
 12  guards      79866 non-null   object 
 13  sidewalk    652172 non-null  object 
 14  user_type   683788 non-null  object 
 15  problems    225844 non-null  object 
 16  root_stone  683788 non-null  object 
 17  root_grate  683788 non-null  object 
 18  root_other  683788 non-null  object 
 19  trnk_wire   683788 non-null  object 
 20  trnk_light  683788 non-null  object 
 21  trnk_other  683788 non-null  object 
 22  brnch_ligh  683788 non-null  object 
 23  brnch_shoe  683788 non-null  object 
 24  brnch_othe  683788 non-null  object 
 25  address     683788 non-null  object 
 26  zipcode     683788 non-null  int64  
 27  zip_city    683788 non-null  object 
 28  cb_num      683788 non-null  int64  
 29  borocode    683788 non-null  int64  
 30  boroname    683788 non-null  object 
 31  cncldist    683788 non-null  int64  
 32  st_assem    683788 non-null  int64  
 33  st_senate   683788 non-null  int64  
 34  nta         683788 non-null  object 
 35  nta_name    683788 non-null  object 
 36  boro_ct     683788 non-null  int64  
 37  state       683788 non-null  object 
 38  Latitude    683788 non-null  float64
 39  longitude   683788 non-null  float64
 40  x_sp        683788 non-null  object 
 41  y_sp        683788 non-null  object 
dtypes: float64(2), int64(9), object(31)
memory usage: 219.1+ MB
```

```
tree_census.health.value_counts(dropna=False)
```

```
health
Good    528850
Fair     96504
NaN      31616
Poor     26818
Name: count, dtype: int64
```

```
tree_census.status.value_counts(dropna=False)
```

```
status
Alive    652173
Stump     17654
Dead      13961
Name: count, dtype: int64
```

```
trees_subset = tree_census[['tree_id', 'tree_dbh',
       'stump_diam', 'curb_loc', 'status', 'health', 'spc_latin', 'spc_common',
       'steward', 'guards', 'sidewalk', 'user_type', 'problems', 'root_stone',
       'root_grate', 'root_other', 'trnk_wire', 'trnk_light', 'trnk_other',
       'brnch_ligh', 'brnch_shoe', 'brnch_othe']]
trees_subset.head()
```

	tree_id	tree_dbh	stump_diam	curb_loc	status	health	spc_latin	spc_common	steward	guards	...	problems	root_stone	root_grate	root_other	trnk_wire	1
0	180,683	3	0	OnCurb	Alive	Fair	Acer rubrum	red maple	NaN	NaN	...	NaN	No	No	No	No	
1	200,540	21	0	OnCurb	Alive	Fair	Quercus palustris	pin oak	NaN	NaN	...	Stones	Yes	No	No	No	
2	204,026	3	0	OnCurb	Alive	Good	Gleditsia triacanthos var. inermis	honeylocust	1or2	NaN	...	NaN	No	No	No	No	
3	204,337	10	0	OnCurb	Alive	Good	Gleditsia triacanthos var. inermis	honeylocust	NaN	NaN	...	Stones	Yes	No	No	No	
4	189,565	21	0	OnCurb	Alive	Good	Tilia americana	American linden	NaN	NaN	...	Stones	Yes	No	No	No	

5 rows × 22 columns

```
trees_subset.isna().sum()
```

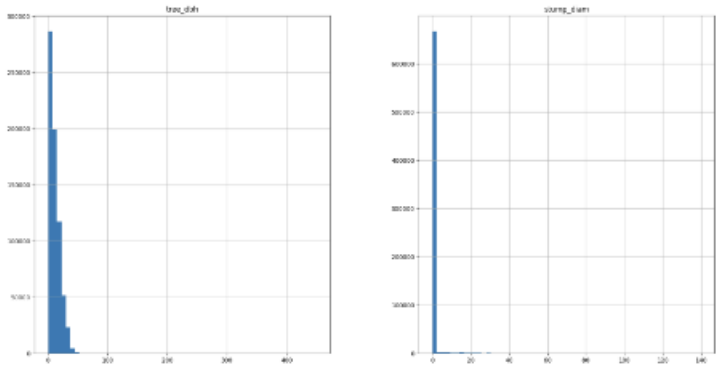
```
tree_id           0
tree_dbh          0
stump_diam        0
curb_loc          0
status            0
health        31616
spc_latin     31619
spc_common    31619
steward      519438
guards       603922
sidewalk      31616
user_type         0
problems     457944
root_stone        0
root_grate        0
root_other        0
trnk_wire         0
trnk_light        0
trnk_other        0
brnch_ligh        0
brnch_shoe        0
brnch_othe        0
dtype: int64
```

```
trees_subset.describe()
```

	tree_dbh	stump_diam
count	683788.000000	683788.000000
mean	11.279787	0.432463
std	8.723042	3.290241
min	0.000000	0.000000
25%	4.000000	0.000000
50%	9.000000	0.000000
75%	16.000000	0.000000
max	450.000000	140.000000

```
trees_subset.hist(bins=50, figsize=(20,15))
```

```
array([[<Axes: title={'center': 'tree_dbh'}>,
        <Axes: title={'center': 'stump_diam'}>]], dtype=object)
```



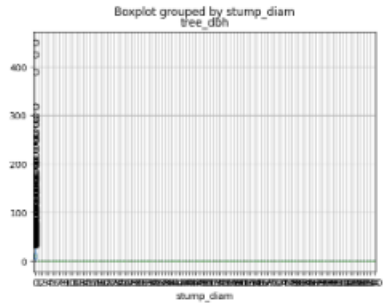

```
big_trees = trees_subset[trees_subset['tree_dbh']> 50]
big_trees.head()
```

	tree_id	tree_dbh	stump_diam		curb_loc	status	health	spc_latin	spc_common	steward	guards	...	problems	root_stone	root_grate	root_othe
2385	168,583	425	0		OnCurb	Alive	Good	Quercus bicolor	swamp white oak	1or2	NaN	...	NaN	No	No	N
2724	199,546	51	0		OnCurb	Alive	Good	Acer saccharinum	silver maple	NaN	NaN	...	Stones	Yes	No	N
4874	139,665	72	0	OffsetFromCurb	Alive	Good	Acer saccharinum	silver maple	NaN	NaN	...	NaN	No	No	No	N
6711	209,349	122	0		OnCurb	Alive	Good	Quercus palustris	pin oak	NaN	NaN	...	NaN	No	No	N
10053	215,075	169	0		OnCurb	Alive	Good	Gleditsia triacanthos var. inermis	honeylocust	NaN	NaN	...	NaN	No	No	N

5 rows × 22 columns

```
tree_census.boxplot(column='tree_dbh', by='stump_diam')
```

```
<Axes: title={'center': 'tree_dbh'}, xlabel='stump_diam'>
```




```
big_trees[['tree_id', 'tree_dbh']].plot(kind='scatter', x='tree_id', y='tree_dbh')
```

```
<Axes: xlabel='tree_id', ylabel='tree_dbh'>
```