

# Data Quality Report for NY Property Data

Bei Wang

The NY property data set contains the property valuation and assessment data, which were collected by employees from City of New York, including Property Assessors, Property Exemption specialists, ACRIS reporting, Department of Building reporting, etc. All these data are from the fiscal year 2001/2002 assessments. The data set is consisted of 32 fields and 1070994 records. Below is a table which contains summary statistics information about each field.

Field Name	Field Type (C = Categorical, N = Numeric)	# of Records That Have a Value	% of Populated	# of Unique Values	# of Records With Value Zero	Mean	Standard deviation	Min	Max	Most Common Field Value
RECORD	C	1,070,994	100	1,070,994	0					/
BBLE	C	1,070,994	100	1,070,994	0					/
B	C	1,070,994	100	5	0					4
Block	C	1,070,994	100	13,984	0					3944
LOT	N	1,070,994	100	6,366	0	364.7	853.2	1	9978	
EASEMENT	C	4,636	0.4	13	0					E
OWNER	C	1,039,249	97.0	863347	0					PARKCSTER PRESERVAT
BLDGCL	C	1,070,994	100	200	0					R4
TAXCLASS	C	1,070,994	100	11	0					1
LTFRONT	N	1,070,994	100	1297	169,108	36.6	74.0	0	9999	
LTDEPTH	N	1,070,994	100	1370	170,128	88.9	76.4	0	9999	
EXT	C	354,305	33.1	3	0					G
STORES	N	1,014,730	94.7	111	0					2.0
FULLVALAR	N	1,070,994	100	109,324	13,007	874,264.5	11,582,431	0	6,150,00,000	
AVLAND	N	1,070,994	100	70,921	13,009	850,67.92	4,057,260	0	2,668,500,000	
AVTOT	N	1,070,994	100	112,914	13,007	227,238.2	6,877,529	0	4,668,308,947	
EXLAND	N	1,070,994	100	33,419	491,699	36,423.89	3,981,576	0	2,668,500,000	
EXTOT	N	1,070,994	100	64,255	432,572	91,186.98	6,508,403	0	4,668,308,947	
EXCD1	C	638,488	59.6	130	0					1017
STADDR	C	1,070,318	99.9	839,281	0					501 SURF AVENUE
ZIP	C	1,041,104	97.2	196	0					10314

<b>EXMPT CL</b>	C	15,579	1.5	14	0					X1
<b>BLDFR ONT</b>	N	1,070,994	100	612	228,815	23.0	35.6	0	7,575	
<b>BLDDE PTH</b>	N	1,070.994	100	621	228,853	39.9	42.7	0	9,393	
<b>AVLAN D2</b>	N	282,726	26.4	58,591	0	246,235	6,178,963	3	2,371,05,000	
<b>AVTOT 2</b>	N	282,732	26.4	111,360	0	713,911	11,652,529	3	4,501,180,002	
<b>EXLAN D2</b>	N	87,449	8.2	22,195	0	351,235	10,802,213	1	2,371,05,000	
<b>EXTOT 2</b>	N	130,828	12.2	48,348	0	656,768	16,072,510	7	4,501,180,002	
<b>EXCD2</b>	C	92,948	8.7	60	0					1017
<b>PERIOD</b>	C	1,070,994	100	1	0					FINAL
<b>YEAR</b>	Year	1,070,994	100	1	0					2010/11
<b>VALTYPE</b>	C	1,070,994	100	1	0					AC-TR

Next detailed information about each field would be introduced.

### 1) RECORD

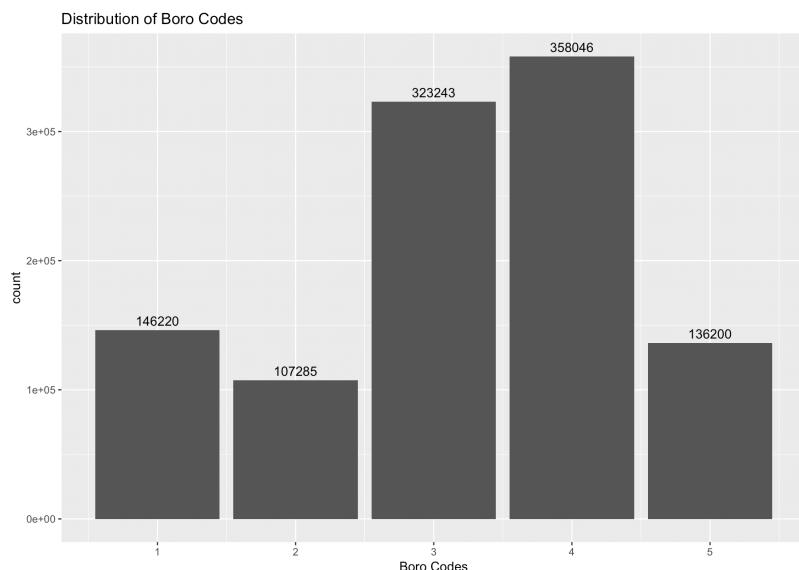
This is the unique identifier for each data record.

### 2) BBLE

BBLE is the concatenation of the property's borough code, block code, and number of lots within the block. Each record of AV has a unique BBLE.

### 3) B

B represents the borough code of New York. Specifically, 1 refers to Manhattan; 2 refers to Bronx; 3 refers to Brooklyn; 4 refers to Queens; 5 refers to Staten Island. According to the following distribution, most of our data records are in Queens borough.



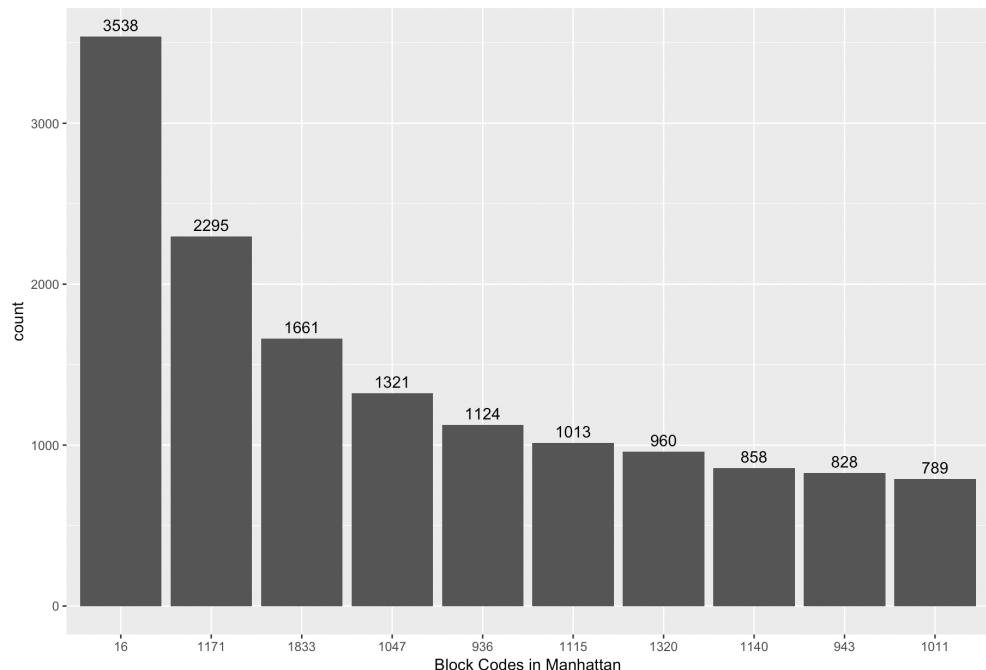
#### 4) BLOCK

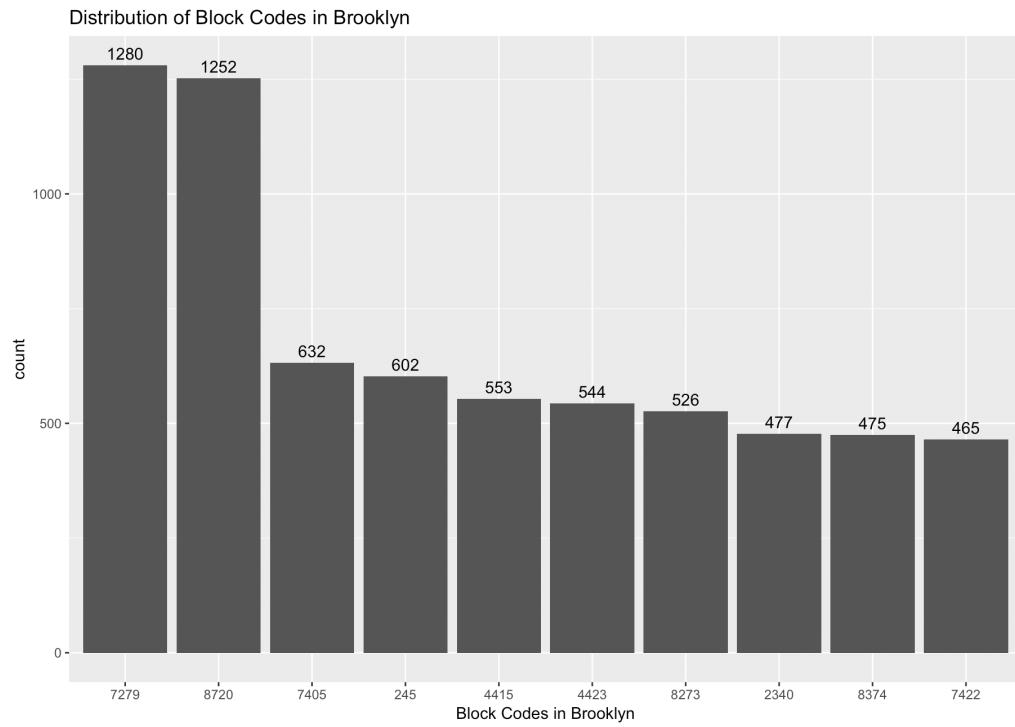
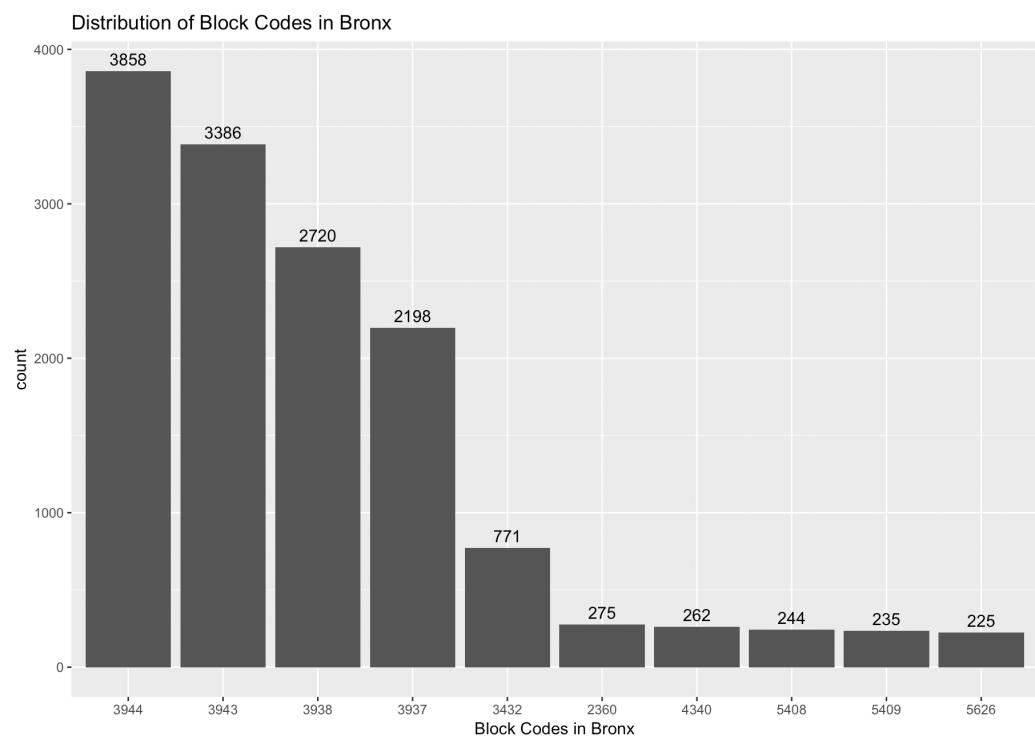
Below is a chart summarizing the block code with in each borough.

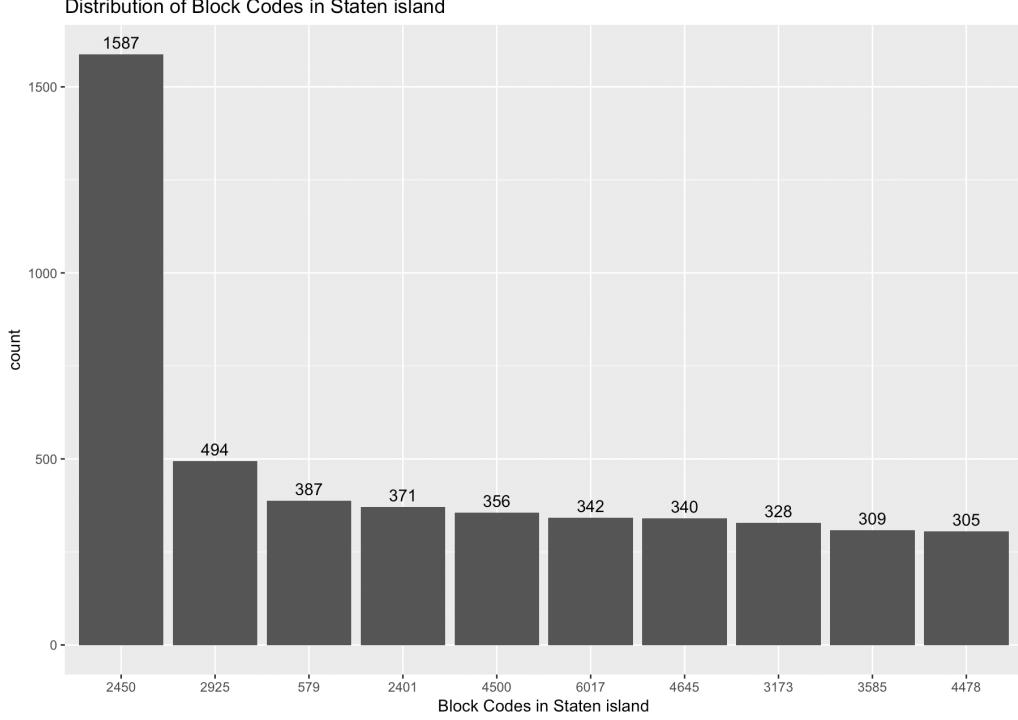
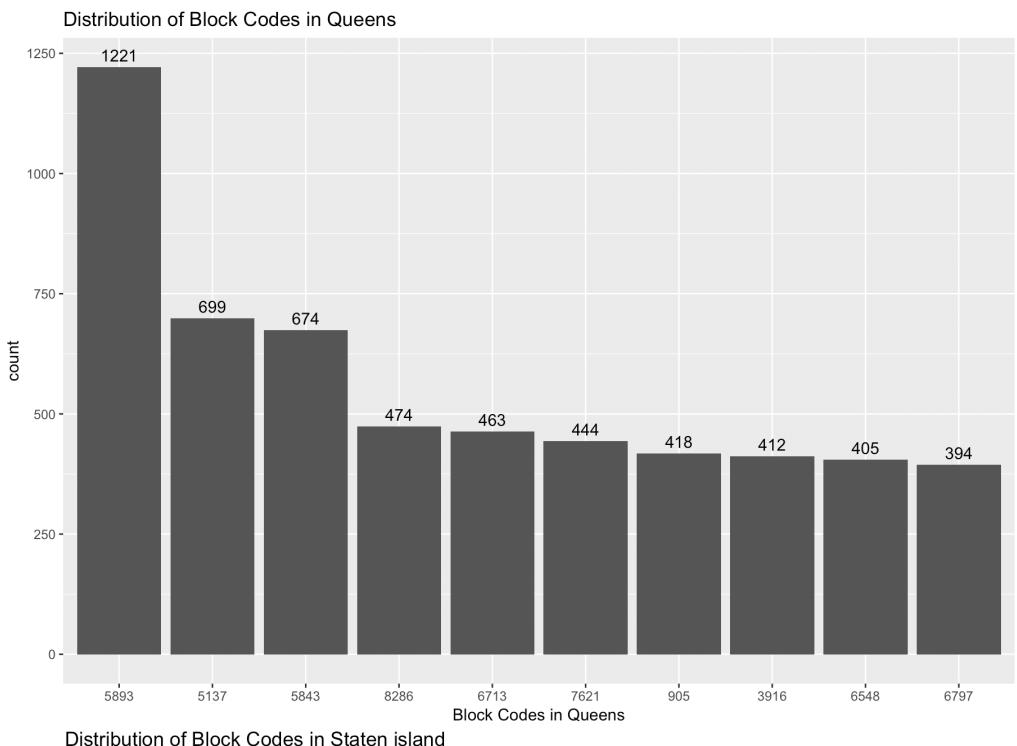
Borough	Block range
Manhattan	1 to 2255
Bronx	2260 to 5958
Brooklyn	1 to 8955
Queens	1 to 16350
Staten island	1 to 8050

For each borough, the top 10 block codes would be listed in the following graphs.

Distribution of Block Codes in Manhattan

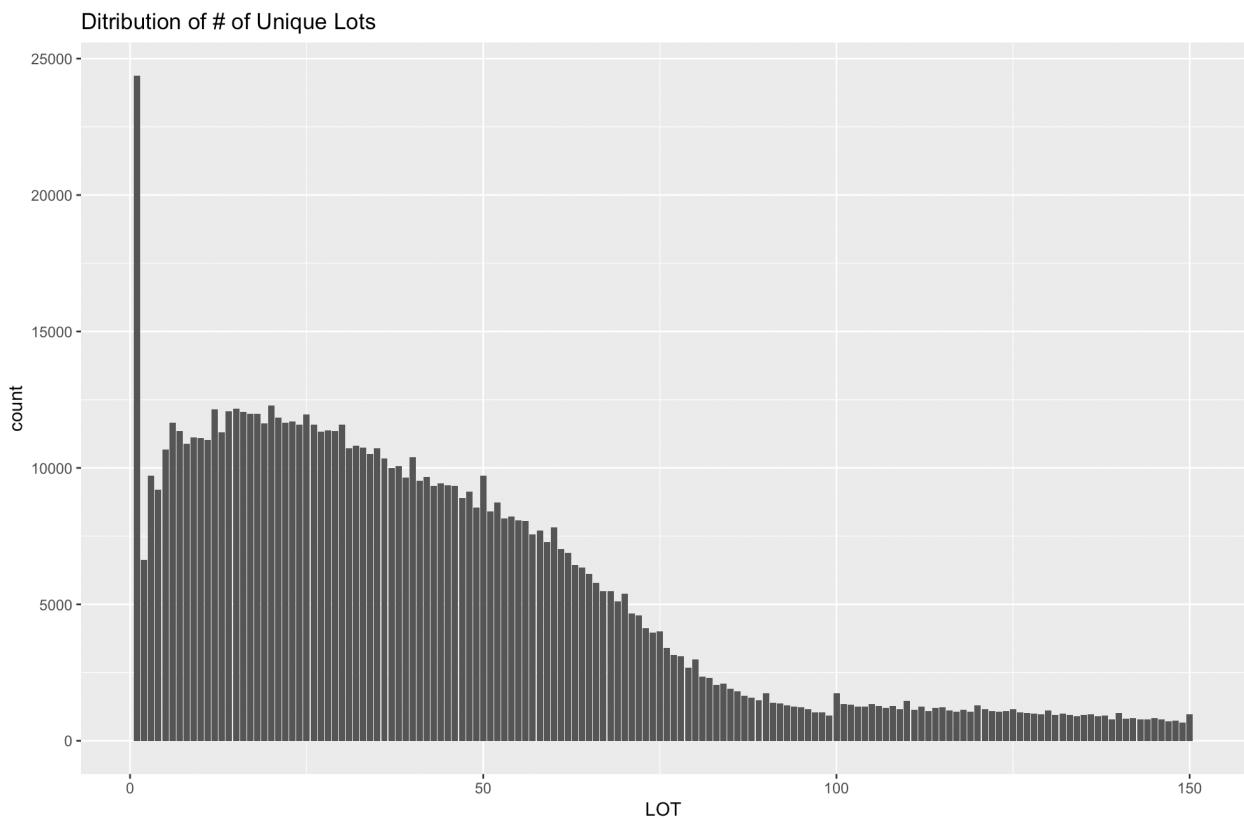




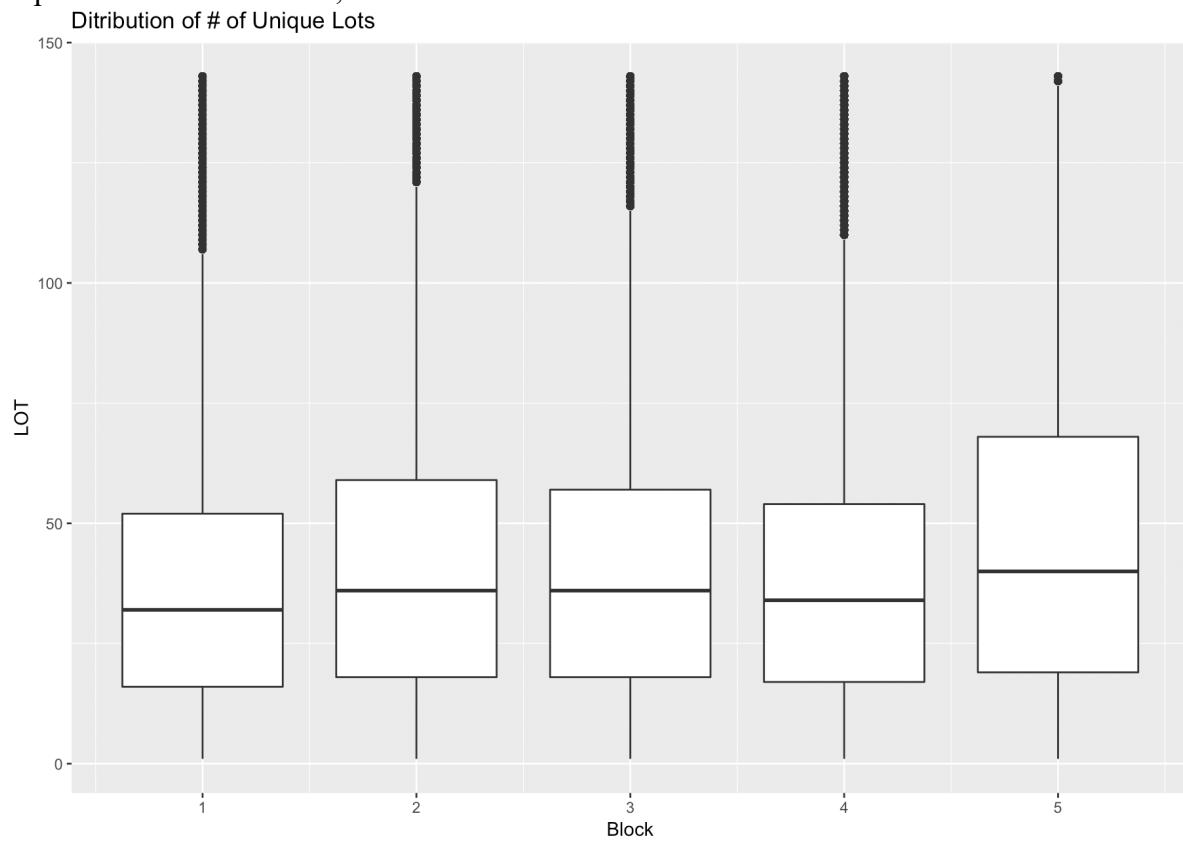


## 5) LOT

LOT refers to the unique number of lots within the specified block. The third quartile of LOT is 143, while the maximum is 9978. In the following graph data records above 150 are treated as outliers and are omitted.



This graph shows the comparison of LOT distributions in different blocks. As we can see, compared with other blocks, there are much more lots in Staten islands.

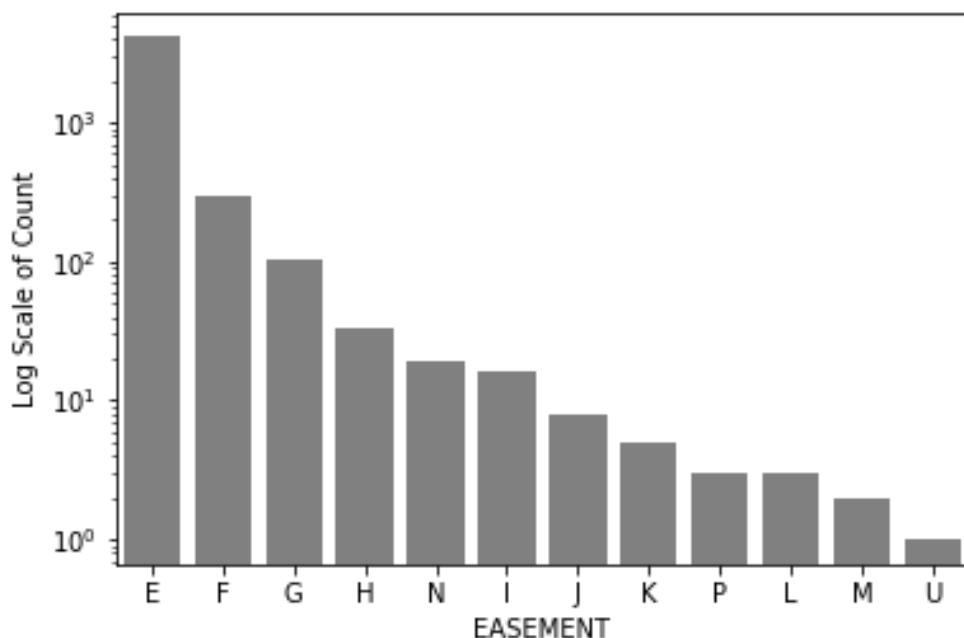


## 6) EASEMENT

EASEMENT describes the conditions of easement, which is specifically explained in the following chart.

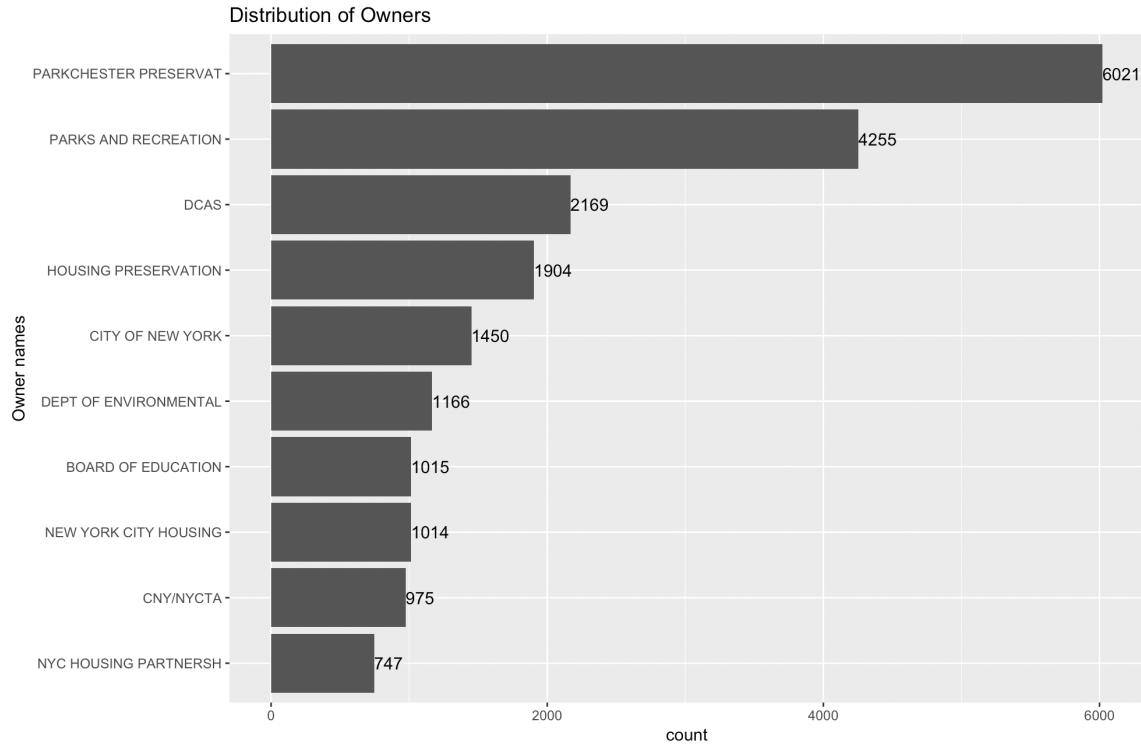
Code	Easement
	No Easement
E through M	The portion of the lot that has a Land Easement
N	Non-Transit Easement
P	Piers
U	U.S. government

1,066,358 data records in the dataset have no easement. The distribution of the remaining 4636 records is shown below.



## 7) OWNER

This field represents owner's name. The following chart gives the distribution of top 10 field values.

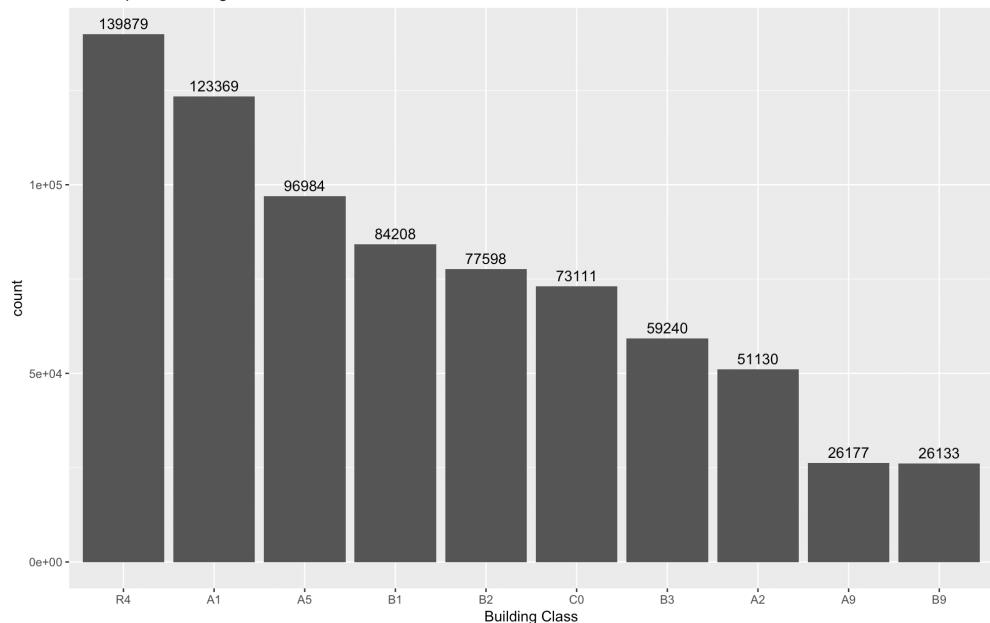


## 8) BLDGCL

BLDGCL is the code of building class. There is a direct correlation between the building class and the tax class.

Tax Class	Building Class
1	A0 - A9, B1 - B9, C0, G0, R3, R6, R7, S0 - S2, V0, V2, V3, Z0
2	C1 - C9, D0 - D9, R0, R1, R2, R4, R8, R9, S3, S4, S5, S9
3	U1 - U2, U4 - U9
4	ALL OTHERS

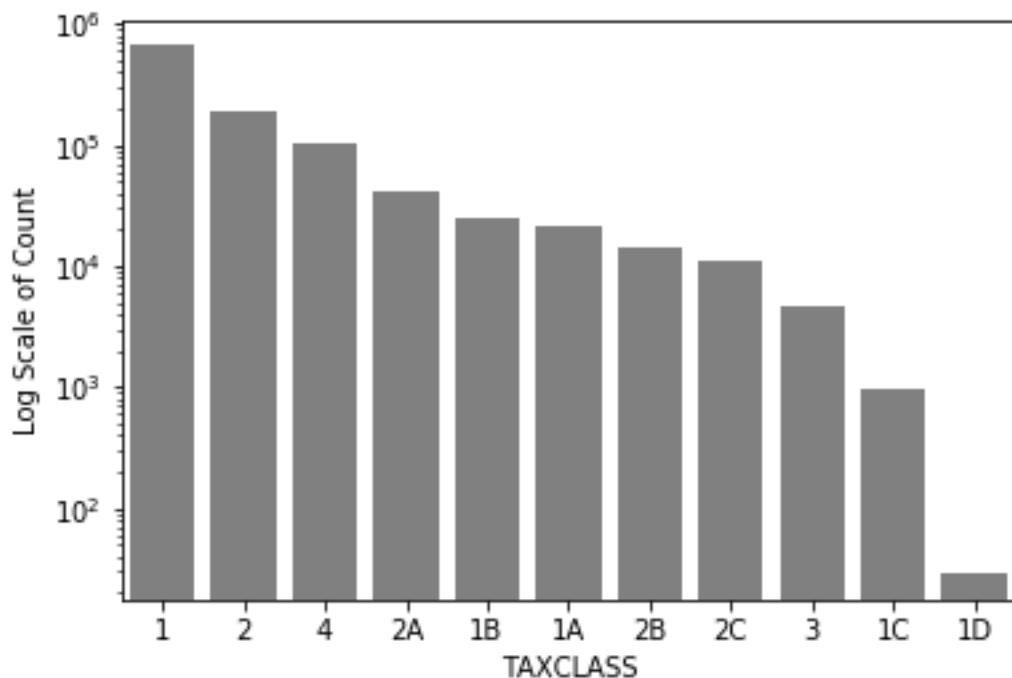
The Top 10 Building Classes



## 9) TAXCLASS

This field is the current property tax class code (NYS Classification).

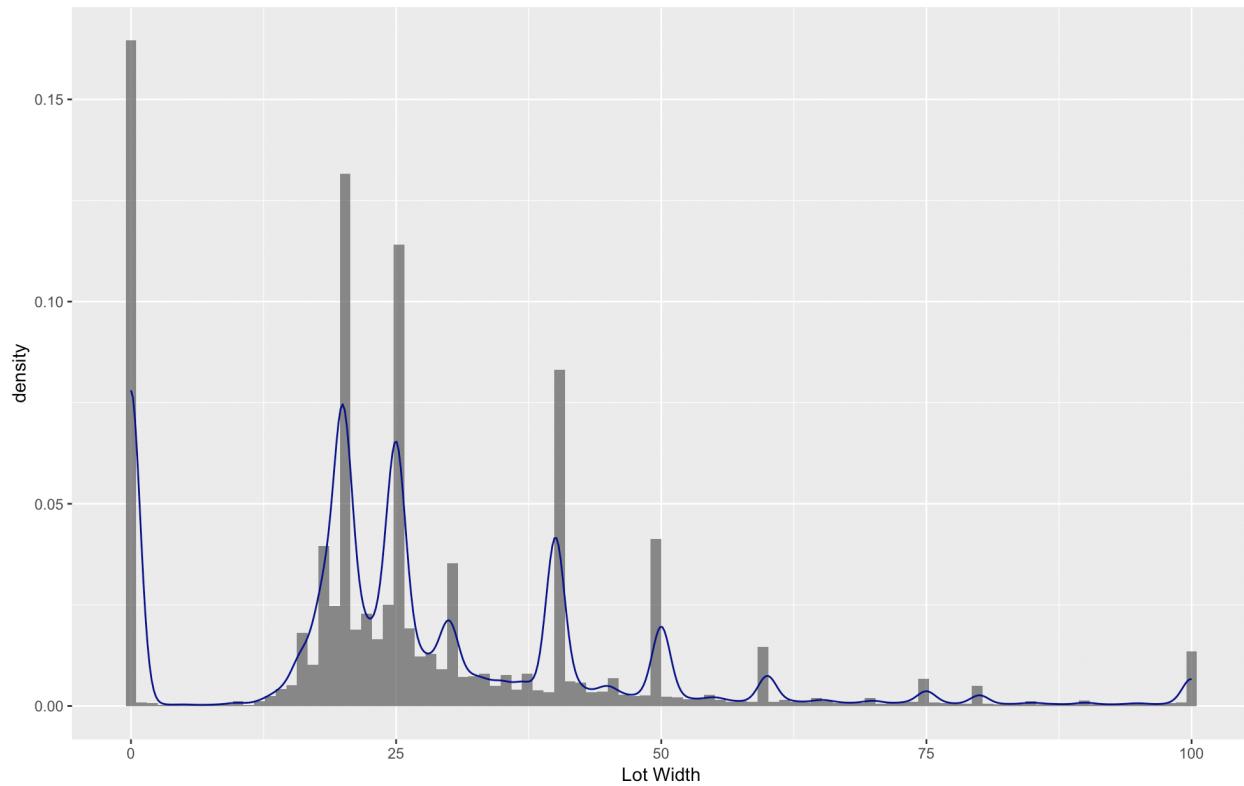
Tax Class	Building type
1	1-3 Unit Residences
1A	1-3 Story Condominiums
1B	Residential Vacant Land
1C	1-3 Unit Condominiums
1D	Select Bungalow Colonies
2	Apartments
2A	Apartments With 4-6 Units
2B	Apartments With 7-10 Units
2C	Coops/Condos With 2-10 Units
3	Utilities (Except Ceiling RR)
4	All Others



## 10) LTFRONT

This field refers to the width of lots. The third quartile of LTFRONT is 40, while the maximum is 9999. In the following graph data records between 100 and 9999 are treated as outliers and are omitted.

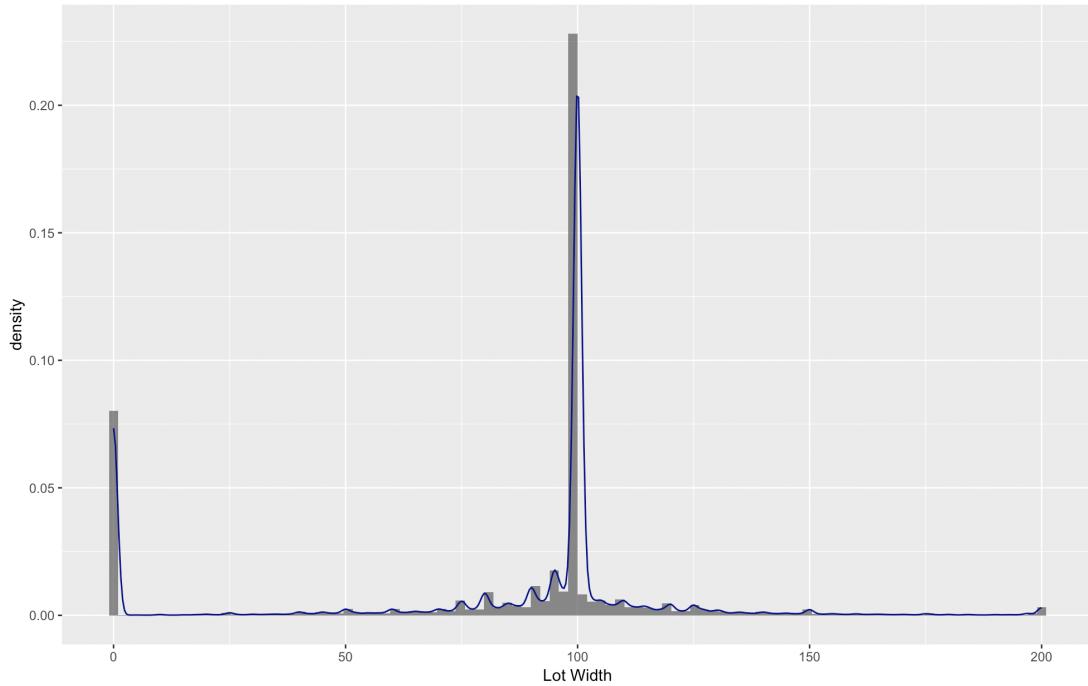
Distribution of Lot Width



## 11) LTDEPTH

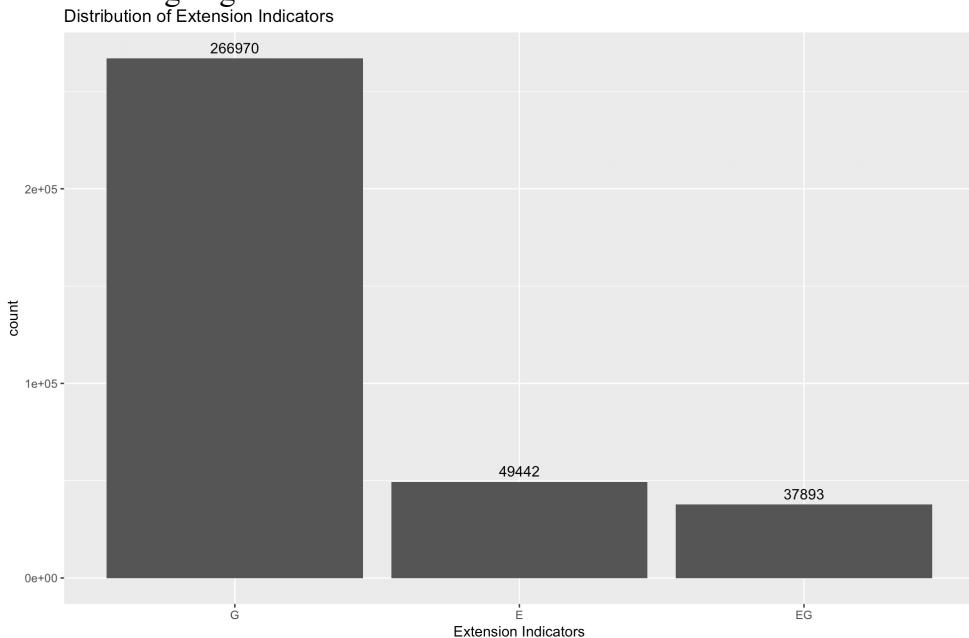
This field refers to the depth of lots. The third quartile of LTDEPTH is 100, while the maximum is 9999. In the following graph data records between 100 and 9999 are treated as outliers and are omitted.

Distribution of Lot Width



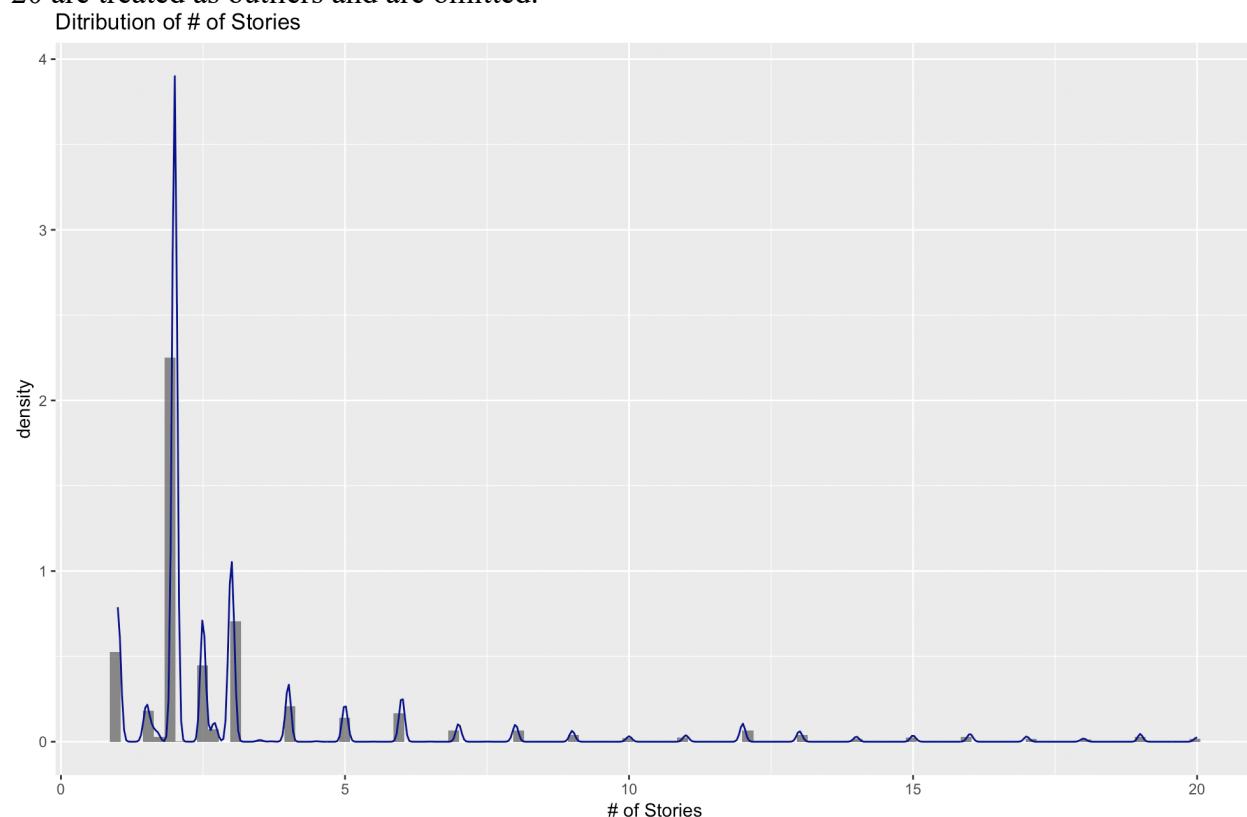
## 12) EXT

This field is the extension indicator, among which E refers to extension; G refers to garage; EG refers to extension and garage.



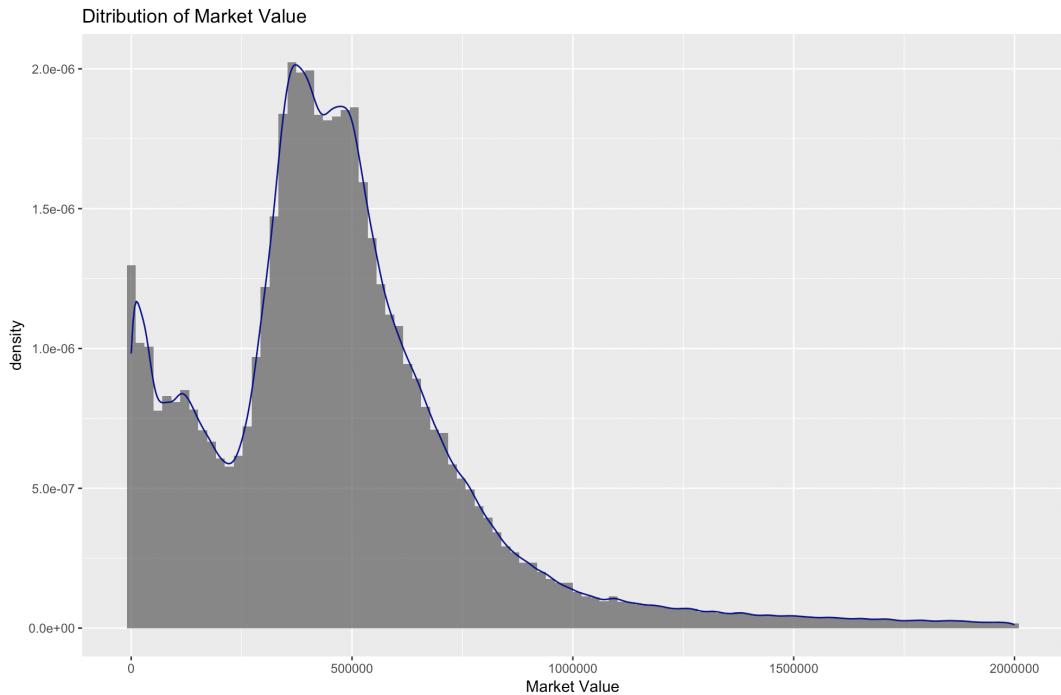
## 13) STORIES

STORIES refers to number of stories in the building. In the following graph data records above 20 are treated as outliers and are omitted.



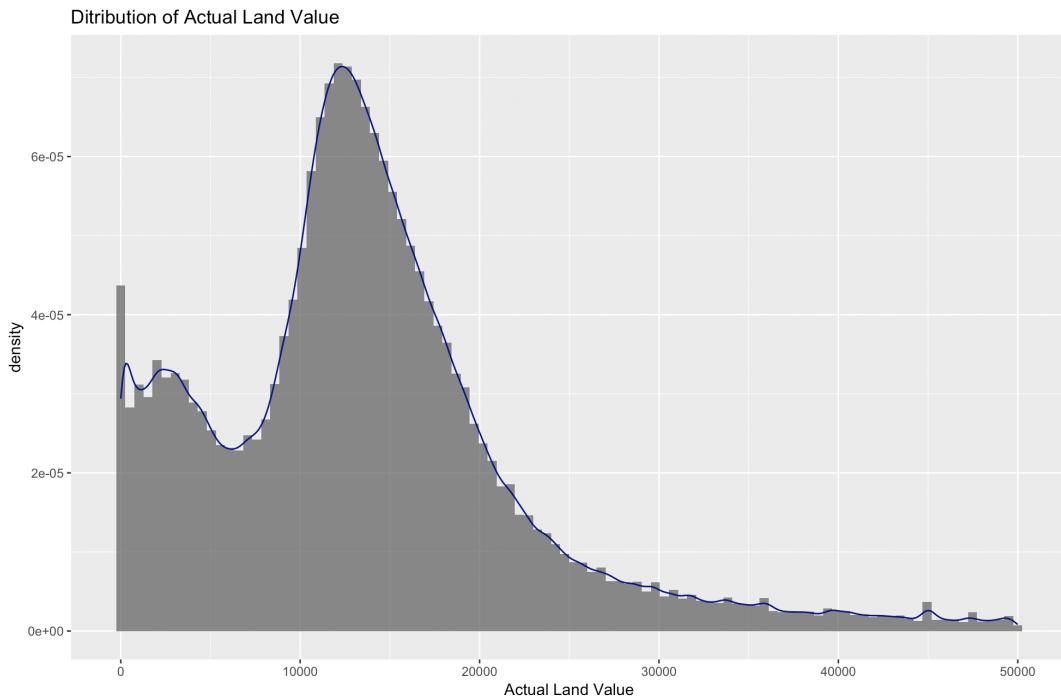
## 14) FULLVAL

FULLVAL refers to the market value. The third quartile of LTDEPTH is 619000, while the maximum is 6150000000. In the following graph data records above 2,000,000 are treated as outliers and are omitted.



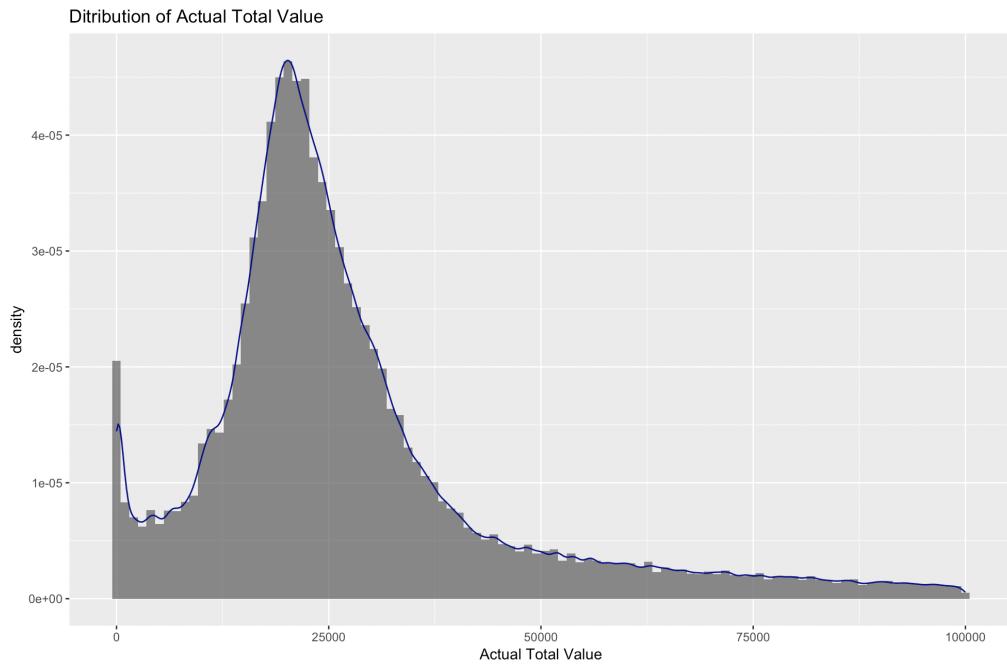
## 15) AVLAND

AVLAND represents the actual land value. In the following graph data records above 50,000 are treated as outliers and are omitted.



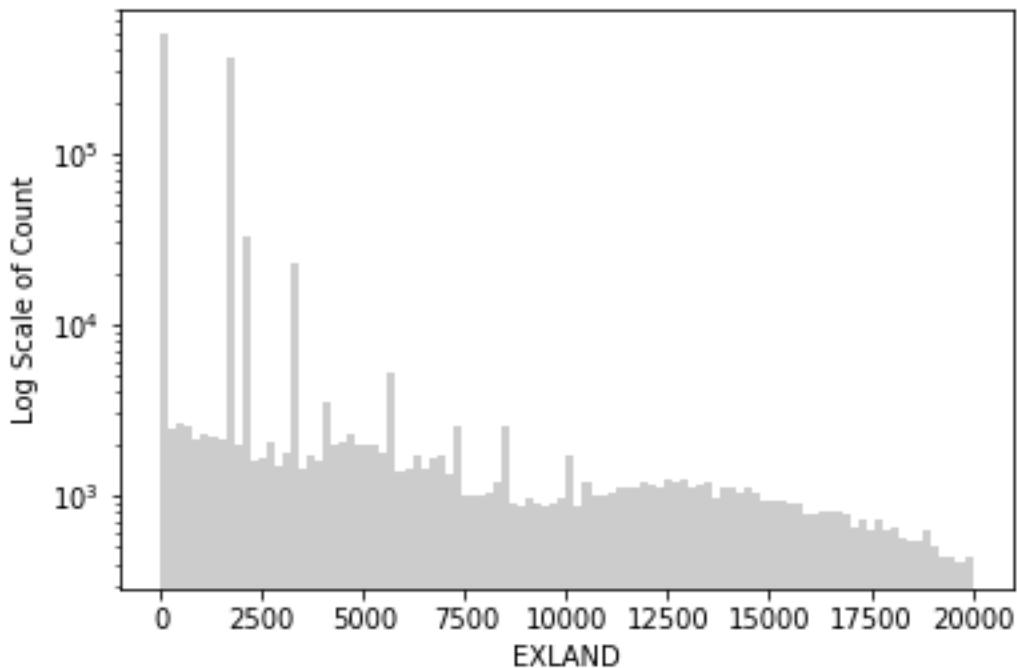
## 16) AVTOT

AVTOT represents the actual total value. In the following density plot data records above 100,000 are treated as outliers and are omitted.



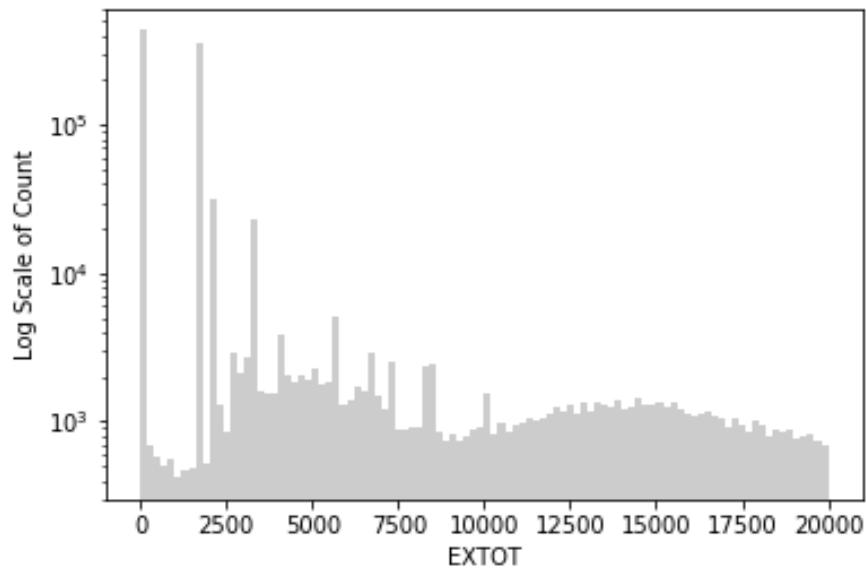
## 17) EXLAND

EXLAND represents the actual exempt land value. The third quartile of EXLAND is 1620, and the maximum value of EXLAND reaches 2668500000. Only data records below 20,000 are included in the following graph.



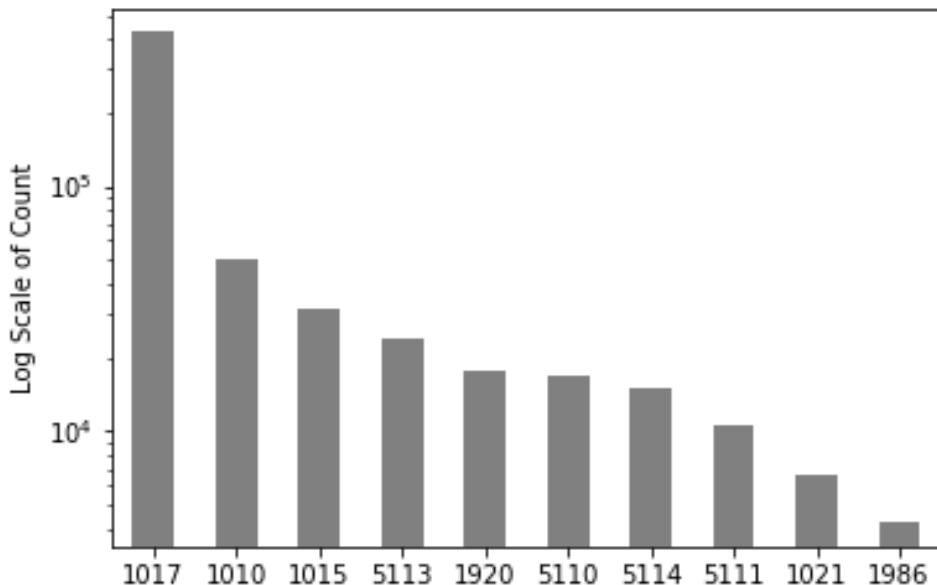
### **18) EXTOT**

EXTOT represents the actual exempt land total. The third quartile of EXLAND is 2090, and the maximum value of EXLAND reaches 4668308947. Only data records below 5000 are included in the following graph.



### **19) EXCD1**

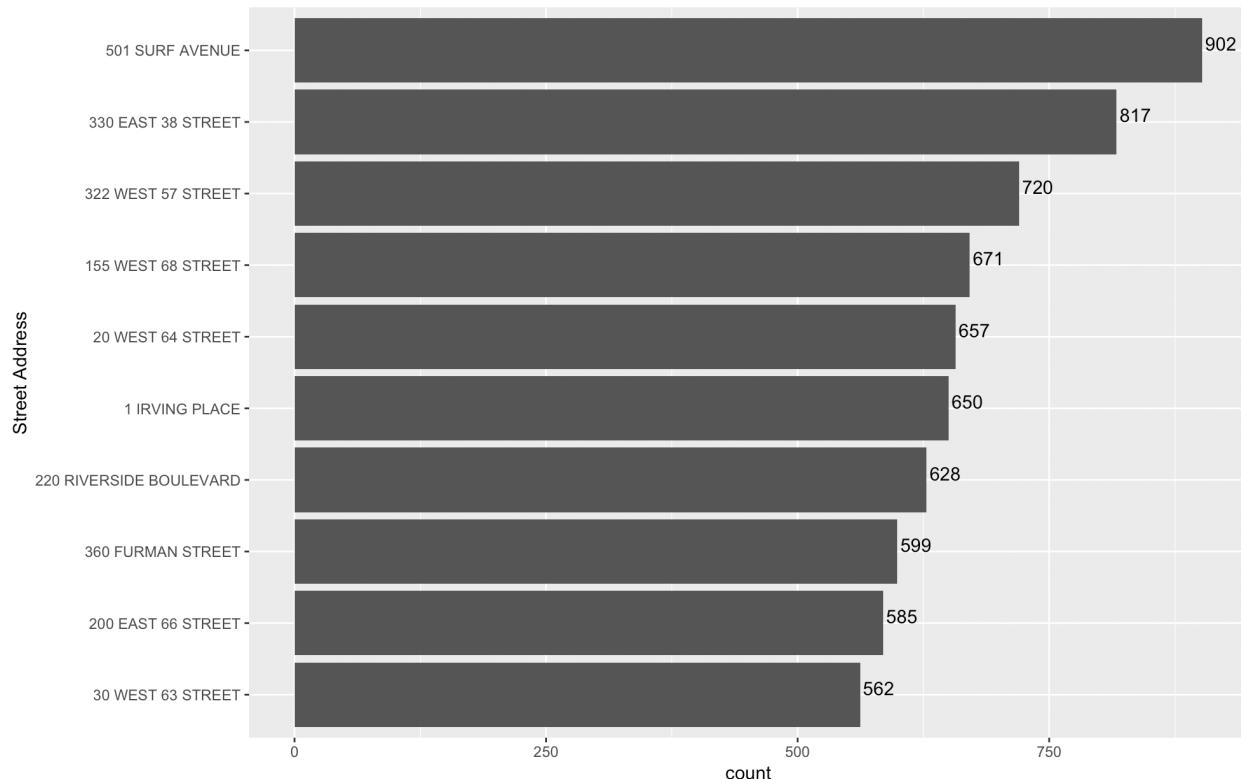
This field is the exemption code. The following graphs shows the top ten exemption codes, among which 1017 is the top one field value.



## 20) STADDR

This refers to the street address. The following graph shows the top 10 street address.

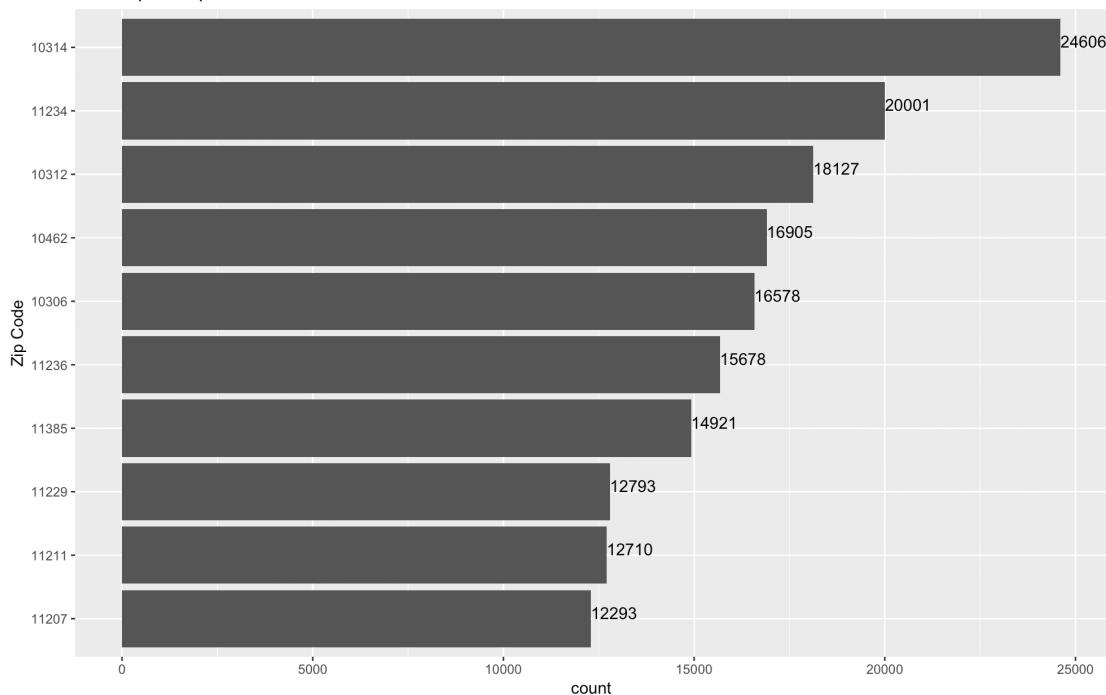
The Top 10 Street Address



## 21) ZIP

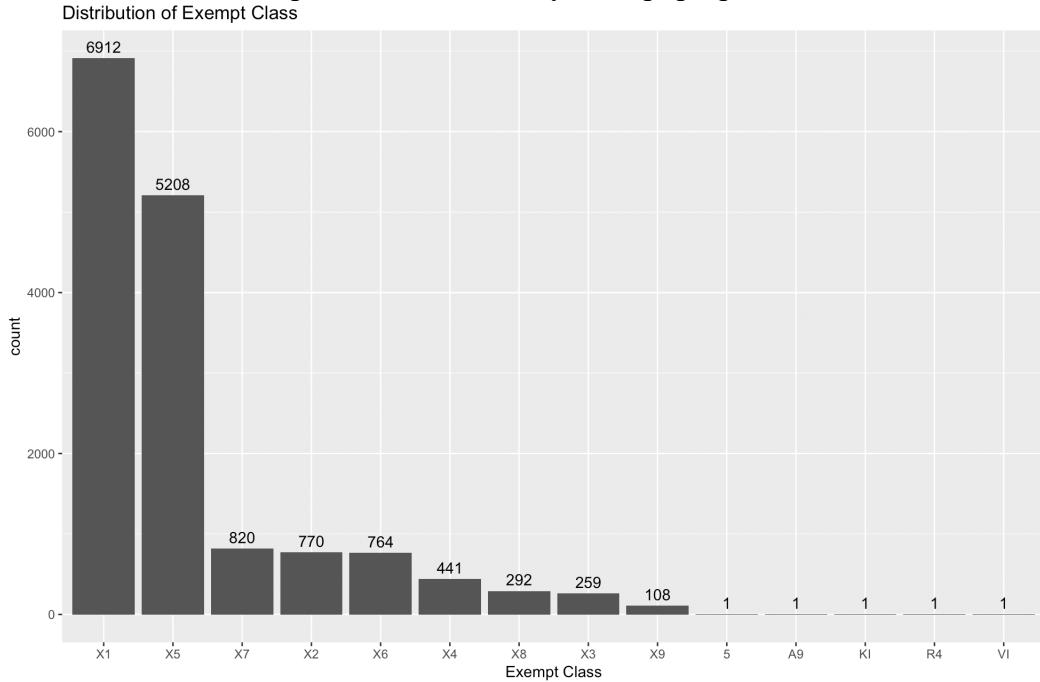
The top 10 zip codes are shown below.

The Top 10 Zip Codes



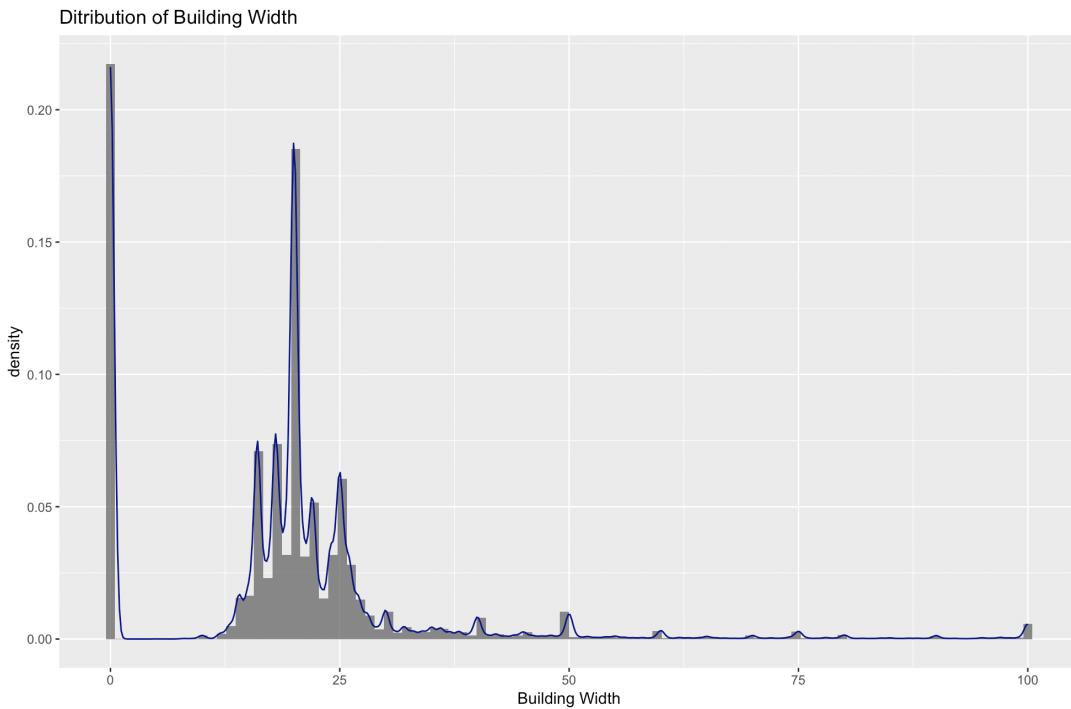
## 22) EXMPTCL

EXMPTCL refers to the exempt class used for fully exempt properties.



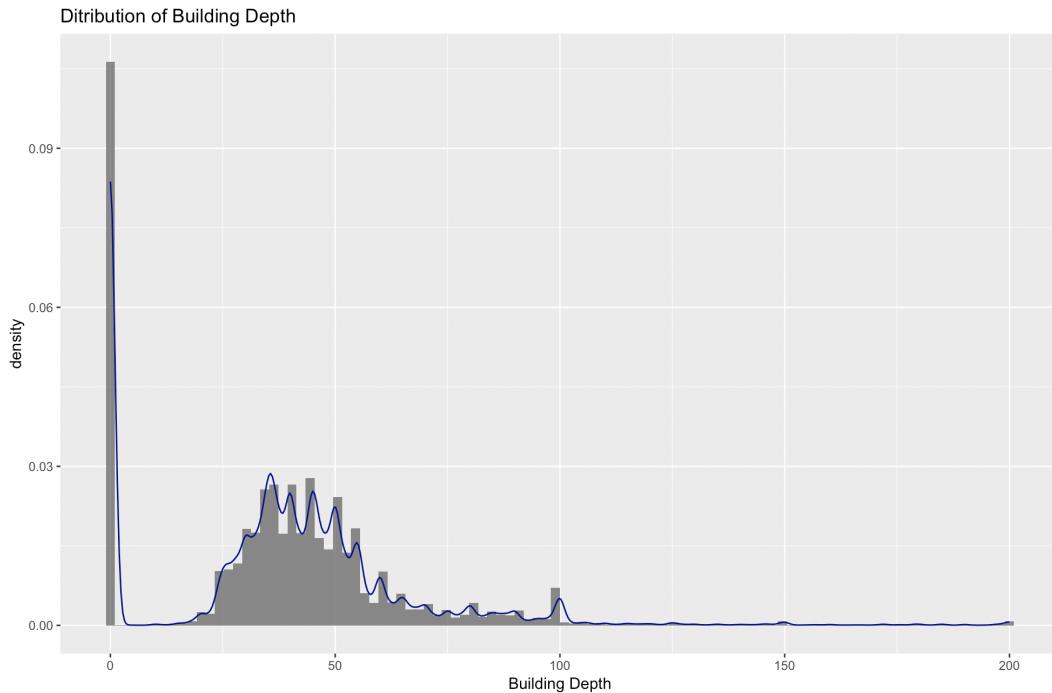
## 23) BLDFRONT

This field refers to the width of building. The third quartile of BLDFRONT is 24, while the maximum is 7575. In the following graph data records above 100 are treated as outliers and are omitted.



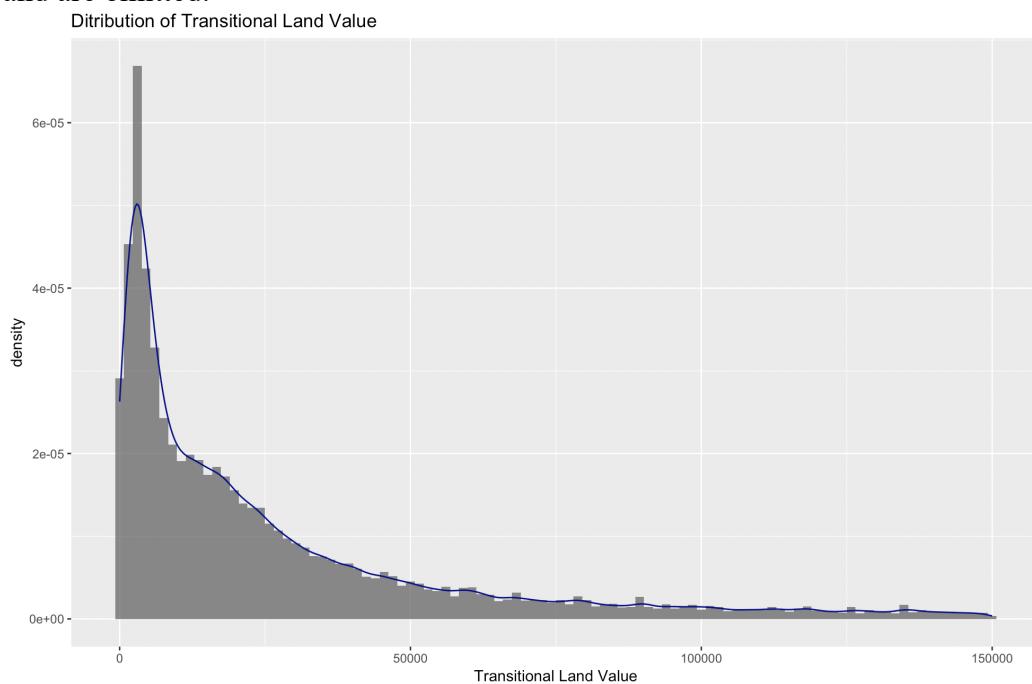
## 24) BLDDEPTH

This field refers to the depth of building. The third quartile of BLDDEPTH is 50, while the maximum is 9393. In the following graph data records above 100 are treated as outliers and are omitted.



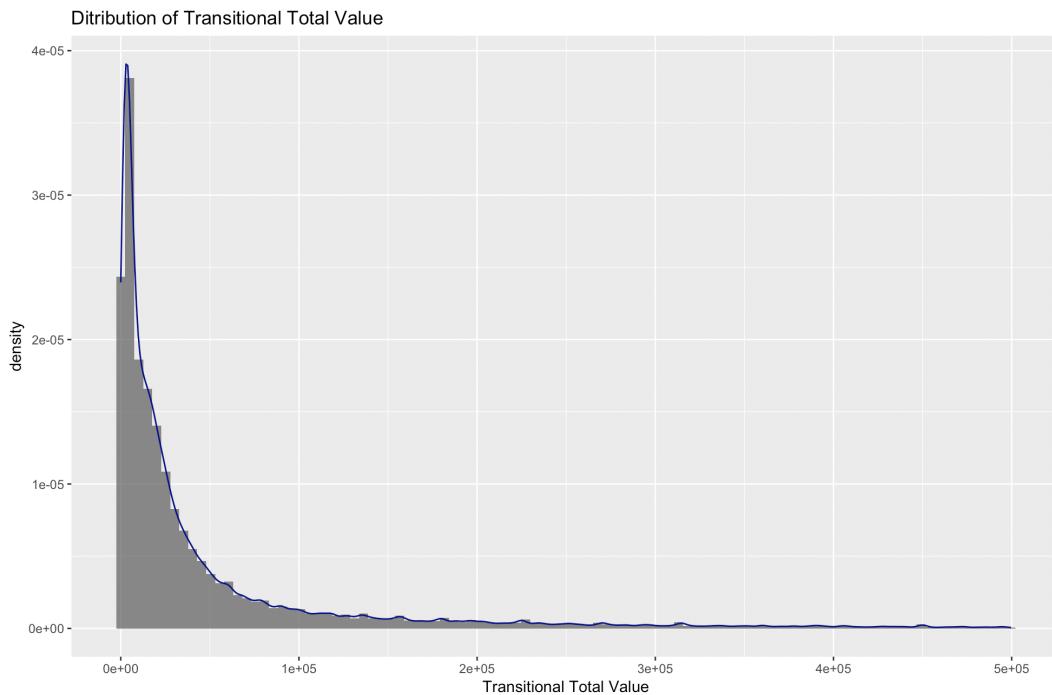
## 25) AVLAND2

AVLAND2 is the transitional land value. The third quartile of AVLAND2 is 62640, while the maximum is 2371005000. In the following graph data records above 150,000 are treated as outliers and are omitted.



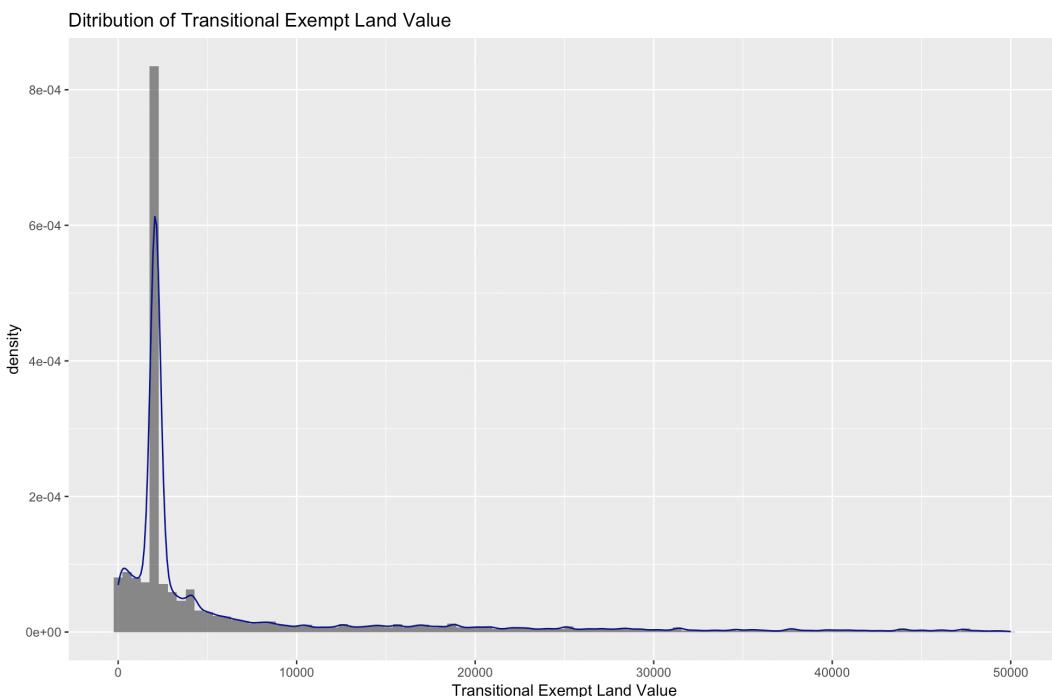
## 26) AVTOT2

AVTOT2 is the transitional total value. The third quartile of AVTOT2 is 240551, while the maximum is 4501180002. In the following graph data records above 500,000 are treated as outliers and are omitted.



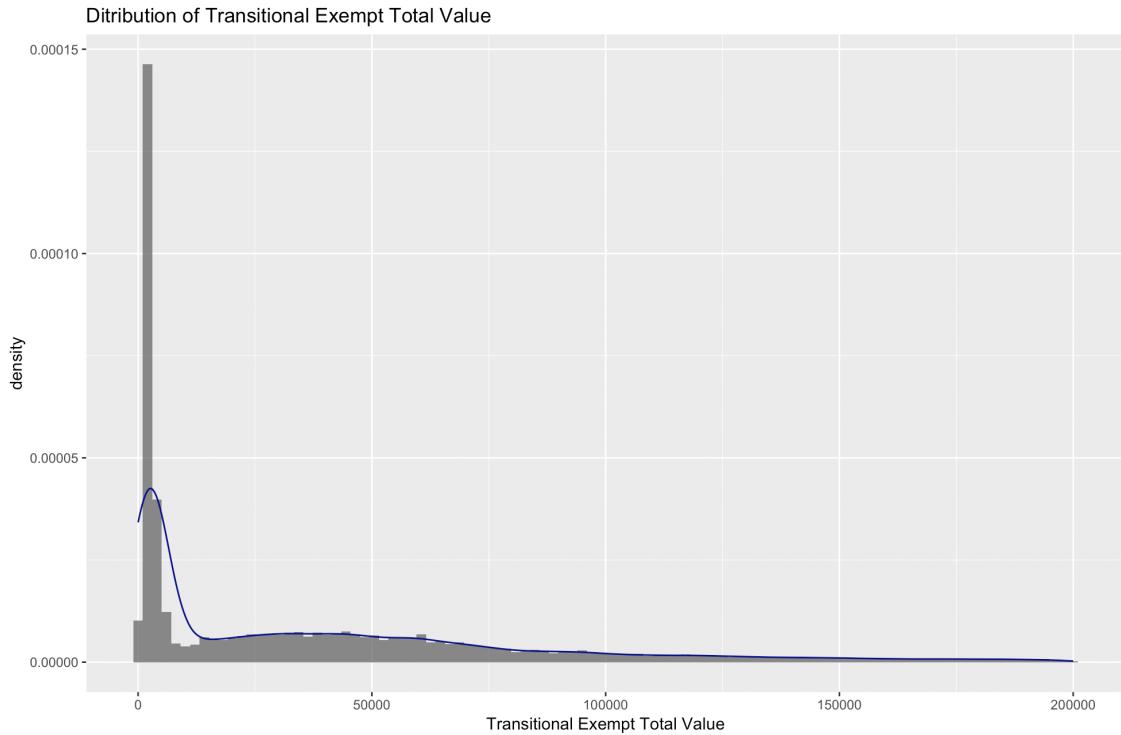
## 27) EXLAND2

EXLAND2 is the transitional exempt land value. The third quartile of EXLAND2 is 31779, while the maximum is 2371005000. In the following graph data records above 50,000 are treated as outliers and are omitted.



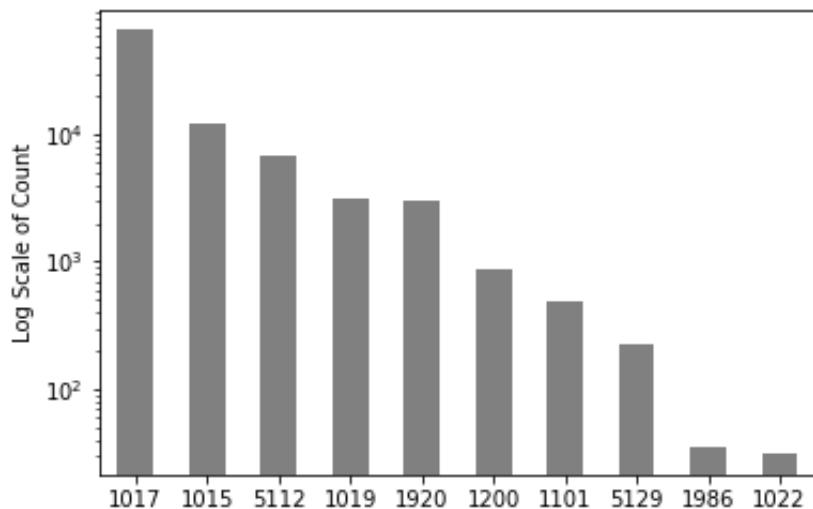
## 28) EXTOT2

EXTOT2 is the transitional exempt land total. The third quartile of EXTOT2 is 106840.8, while the maximum is 4501180002. In the following graph data records above 200,000 are treated as outliers and are omitted.



## 29) EXCD2

This field is another exemption code. The following graphs shows the top ten exemption codes, among which 1017 is the top one field value.



**30) PERIOD**

The assessment periods of all data records are final periods.

**31) YEAR**

All data records are in 2010/11 assessment year.

**32) VALTYPE**

All data records are AC-TR valtype.