

# NY Property Data Quality Report

Cheng Chen

1/31/2017

## 1. File description

**File name:** NY\_property\_data Source

**Url:** <https://data.cityofnewyork.us/Housing-Development/Property-Valuation-and-Assessment-Data/rgy2-tti8>

**Data Provided by:** Department of Finance (DOF)

**Dataset Owner:** NYC OpenData

**Category:** Housing & Development

**Date Created:** September 2, 2011 Metadata Last Updated: September 5, 2014

**Data volume:** 1048575 Fields: 30 fields, 16 continuous, 12 categorical, 2 text

**Details variables:** RECORD, BBLE, BLOCK, LOT, EASEMENT, OWNER, BLDGCL, TAXCLASS, LTFRONT, LTDEPTH, STORIES, FULLVAL, AVLAND, AVTOT, EXLAND, EXTOT, EXCD1, STADDR, ZIP, EXMPTCL, BLDFRONT, BLDDEPTH, AVLAND2, EXLAND2, EXLAND2, EXTOT2, EXCD2, PERIOD, YEAR, VALTYPE.

## 2. List of info for each field

### "RECORD"

**Field name:** RECORD

**Description:** categorical no metric. Track data order

### "BBLE"

**Field name:** BBLE

**Description:** categorical no metric. Concatenation of AV\_BORO, AV\_BLOCK, AV\_LOT, AV\_EASEMENT

No missing data, totally unique number.

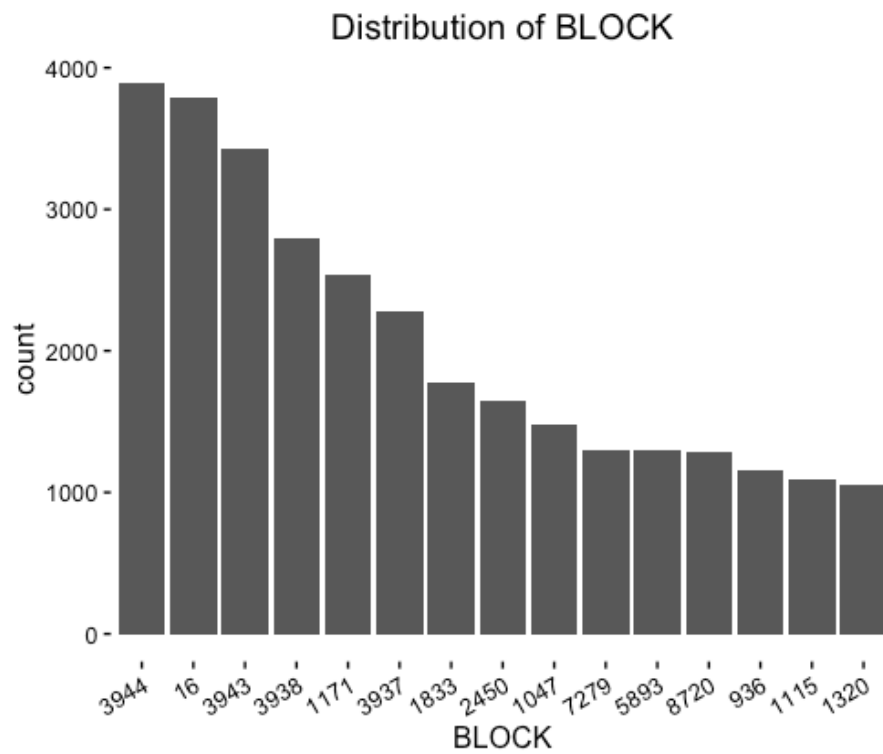
## "BLOCK"

**Field name:** BLOCK

**Description:** categorical no metric. Valid block ranges by boro

No missing data

Unique block number: 13949



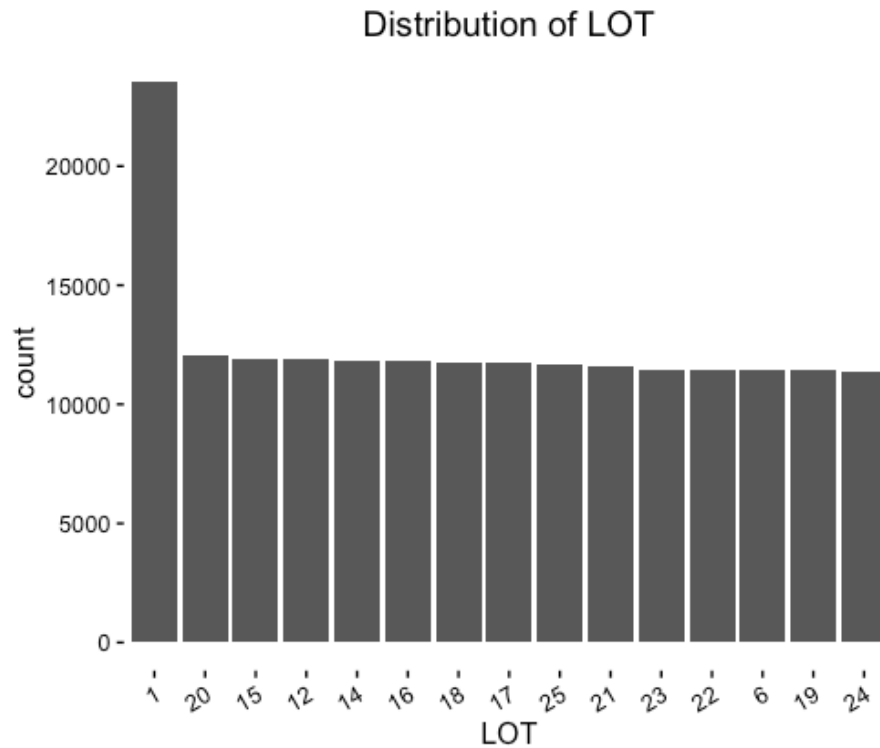
## "LOT"

**Field name:** LOT

**Description:** categorical no metric. Unique number within block or boro

No missing data

Unique LOT number: 6366



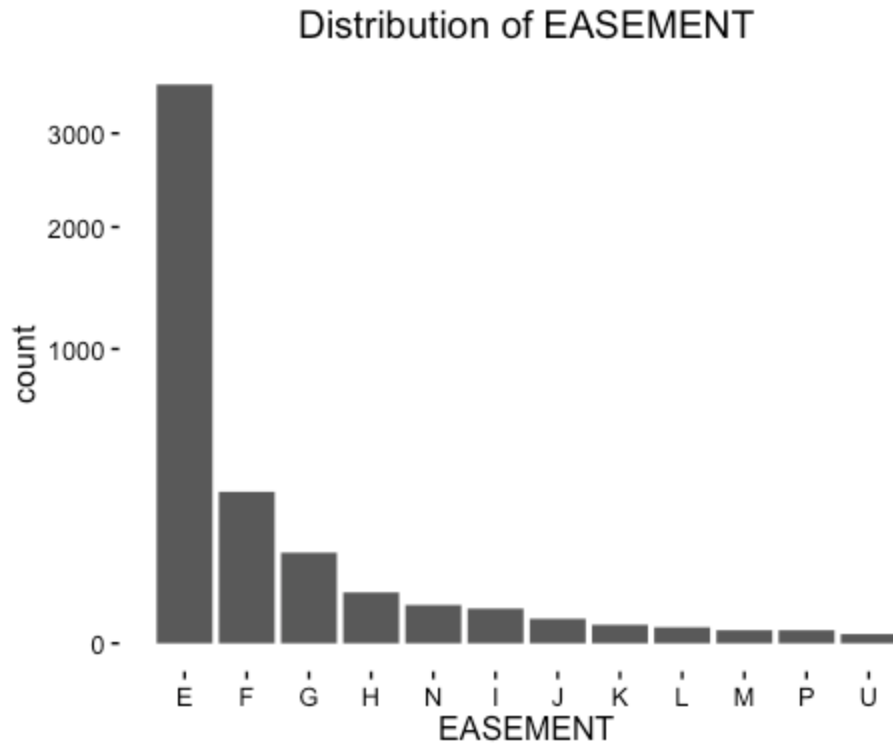
## "EASEMENT"

**Field name:** EASEMENT

**Description:** categorical with metric, is a field that is used to describe easement

**Levels table:**

	E	F	G	H	I	J	K	L
	3603	265	95	30	14	7	4	3
M	N	P	U	NA's	Unique #			
2	17	2	1	1044532	13			



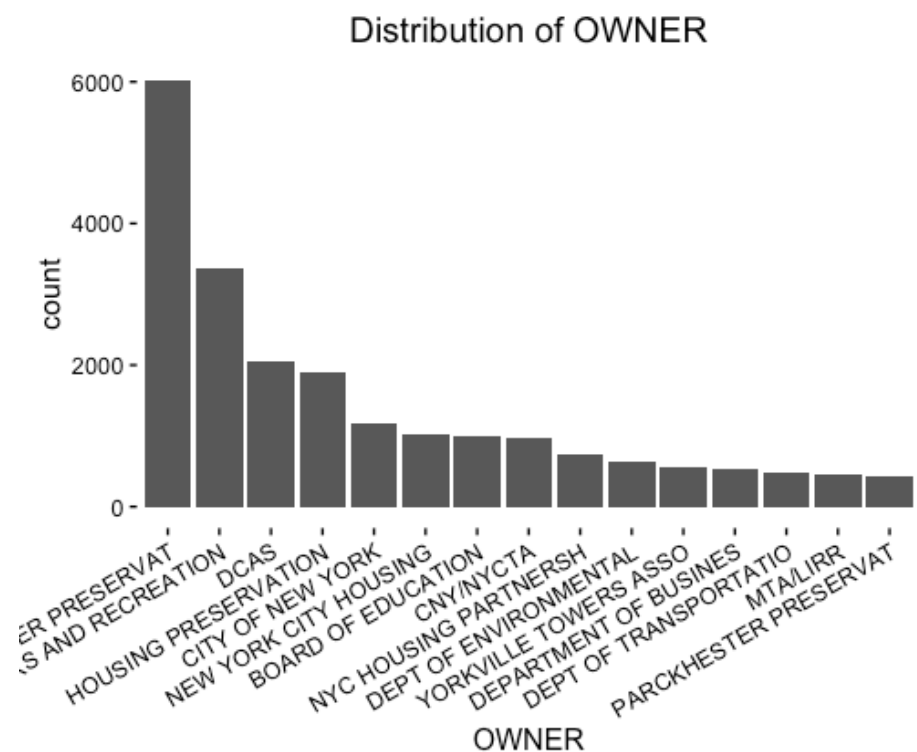
## "OWNER"

**Field name:** OWNER

**Description:** character without metric. The Owner's Name.

No missing data

Blow figure shows top 15 owners; total unique owner number is 847055.



## "BLDGCL"

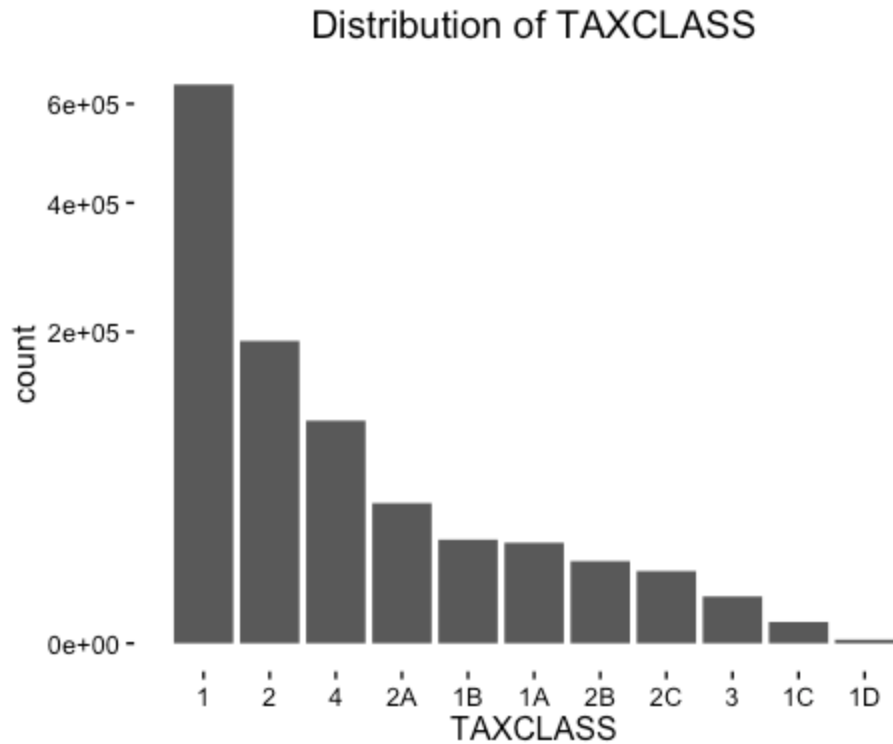
**Field name:** BLDGCL

**Description:** categorical with metric. Building class

**Levels table:**

R4	A1	A5	B1	B2	C0	B3	A2	A9
139879	119340	92896	84054	73156	73077	59091	49085	25931
B9	R5	V0	R3	C3	C1	S2	C2	R2
25235	23950	21520	20899	16332	15070	14480	13632	10558
R1	C7	K1	V1	A0	S1	A3	R0	G7
8015	7544	7529	6836	6815	6133	5742	5681	5562
D1	S9	K9	A4	M1	K2	D4	C4	C6
5101	4120	4051	3874	3841	3607	3531	3299	3212
G9	S3	O9	E9	G0	C5	S4	S5	K4
3083	3011	2991	2989	2906	2867	2556	2529	2515
Z9	F9	G2	D7	U7	Z7	F4	E1	E3
2436	2371	2031	1786	1786	1435	1319	1306	1262
O7	W1	Q1	D9	V9	A6	U9	D3	M9
1150	1150	1147	1094	1050	1020	994	935	923
R6	G6	R8	U6	D6	L8	W2	F1	K5
917	766	749	740	720	676	577	576	565
O8	Q2	M3	O6	G4	O1	G1	U1	C9
563	553	545	543	510	479	467	452	431
O3	L9	I5	Z0	V3	W9	R9	A7	H9





## "LTFRONT"

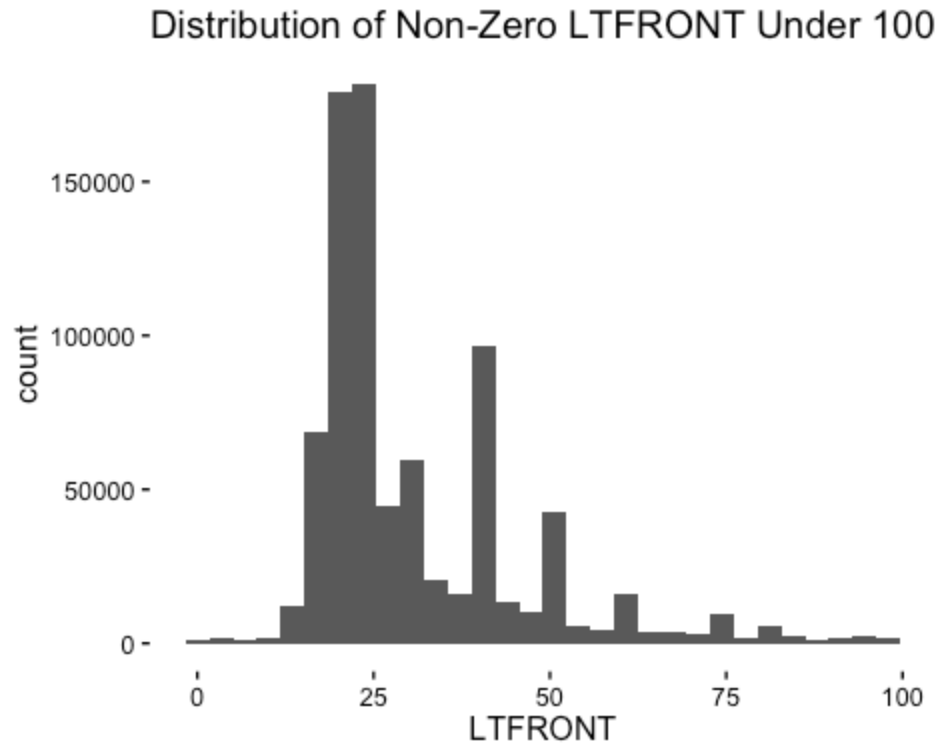
**Field name:** LTFRONT

**Description:** continuous with metric, Lot Frontage in feet.

**Summary statistics:**

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's	mode	SD	Unique #
1.00	20.00	25.00	43.12	40.00	9999.00	168867	20	78.62	1276

Proportion of non-zero value covered by distribution figure: 0.9264529



## "LTDEPTH"

**Field name:** LTDEPTH

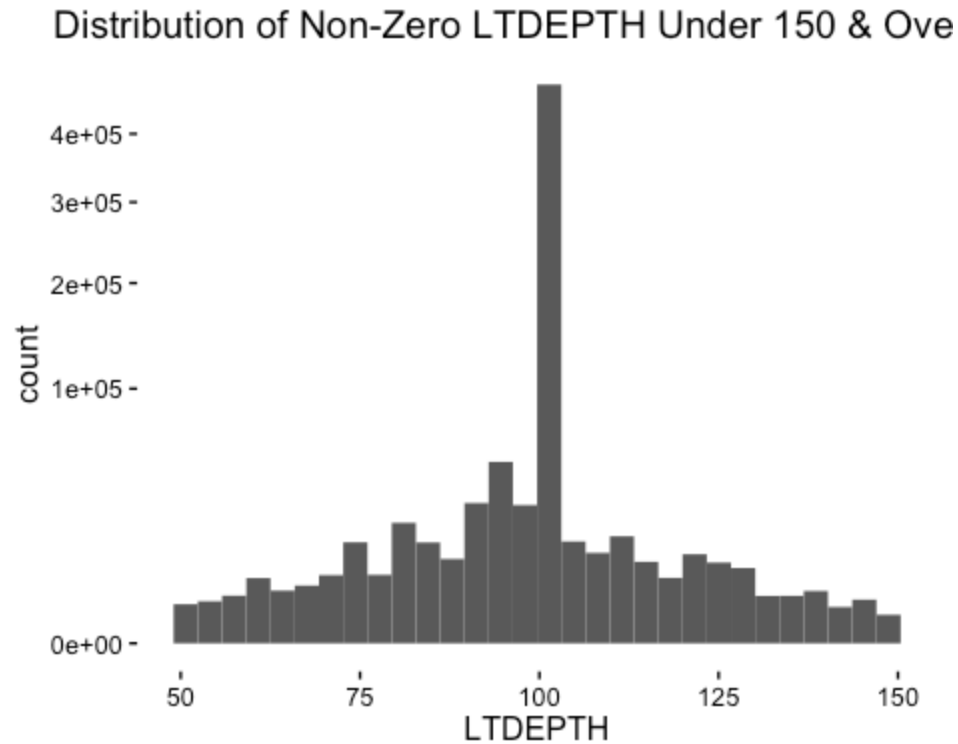
**Description:** continuous with metric, Lot Depth in feet.

**Summary statistics:**

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's	mode	SD	Unique #
1.0	97.0	100.0	105.3	100.0	9999.0	169888	100	70.71	1335

[1] 0.9190918





## "STORIES"

**Field name:** STORIES

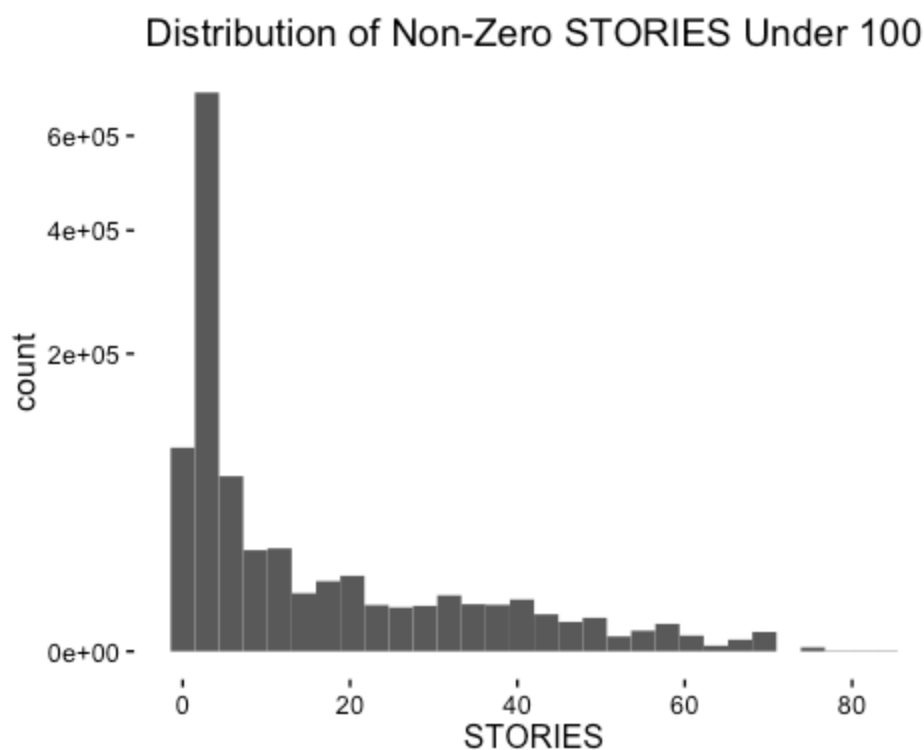
**Description:** continuous with metric. The number of stories for the building

**Summary statistics:**

Proportion of missing value: 0.04972653

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's	mode	SD	Unique #
1.00	2.00	2.00	5.06	3.00	119.00	52142	2	8.43	111

Proportion of non-zero value covered by distribution figure: 0.999993



## "FULLVAL"

**Field name:** FULLVAL

**Description:** continuous with metric, full value of the property

### Summary statistics:

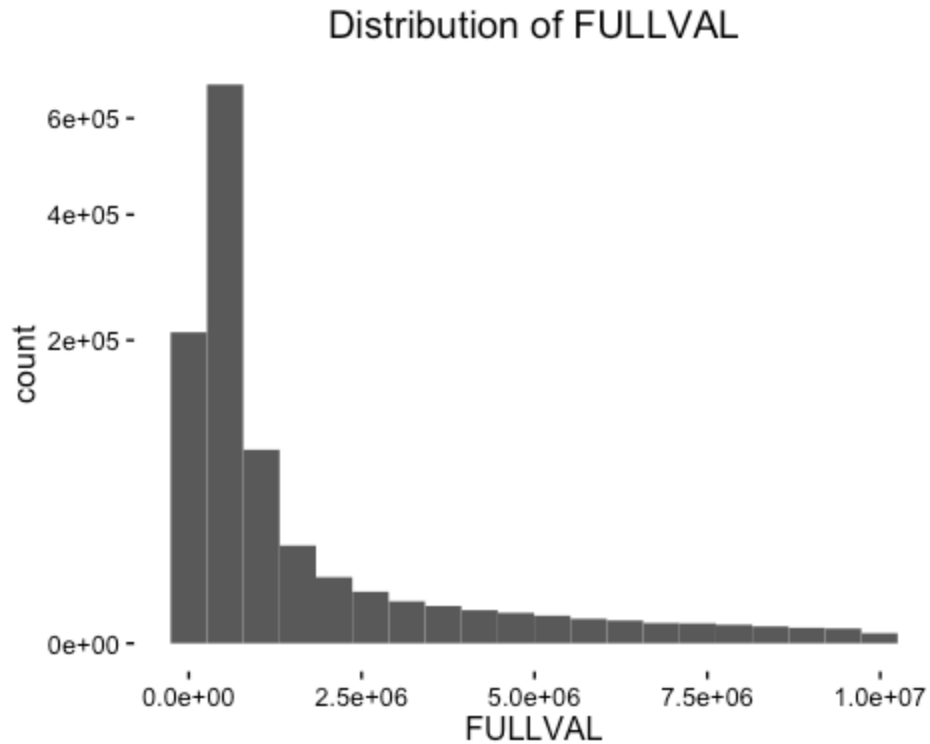
Proportion of missing value plus zero value: 0.0121708

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's	mode	SD
4.00e+0	3.11e+05	4.50e+05	8.91e+05	6.23e+05	6.15e+09	12762	502000	11774390

Unique #

1008276

Proportion of non-zero value covered by distribution figure: 0.9926174



## "AVLAND"

**Field name:** AVLAND

**Description:** continuous with metric, Current year's total market value of the land

### Summary statistics:

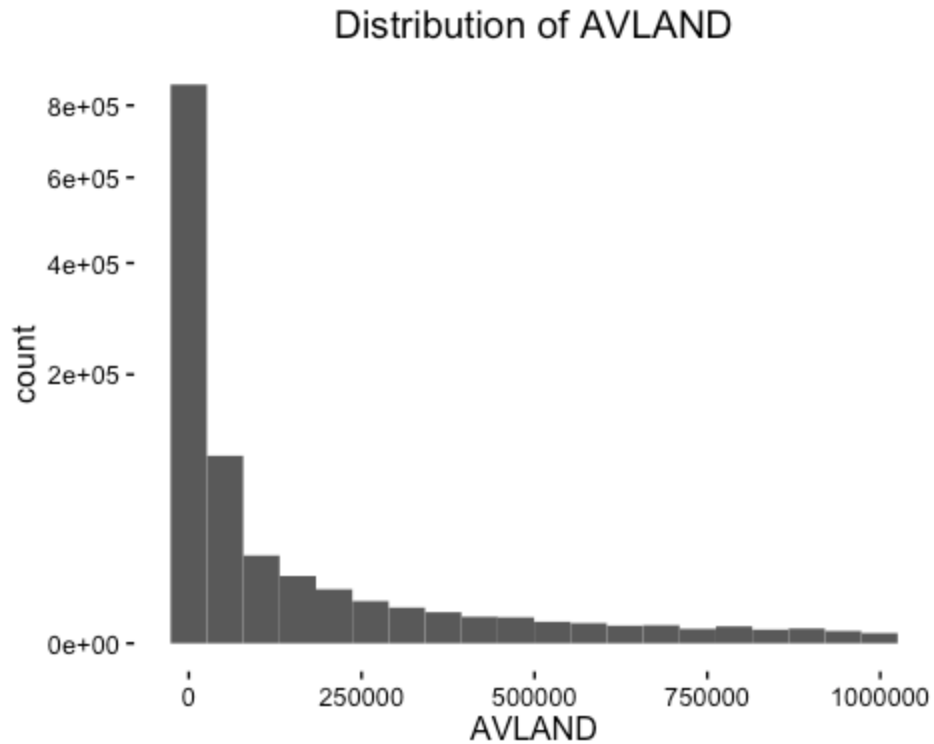
Proportion of missing value plus zero value: 0.01217271

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.000e+00	9.425e+03	1.375e+04	8.705e+04	1.983e+04	2.668e+09	12764

Mode	SD	Unique #
------	----	----------

45000	4125933	70528
-------	---------	-------

Proportion of non-zero value covered by distribution figure: 0.9900619



## "AVTOT"

**Field name:** AVTOT

**Description:** continuous with metric, Current year's total market value

### Summary statistics:

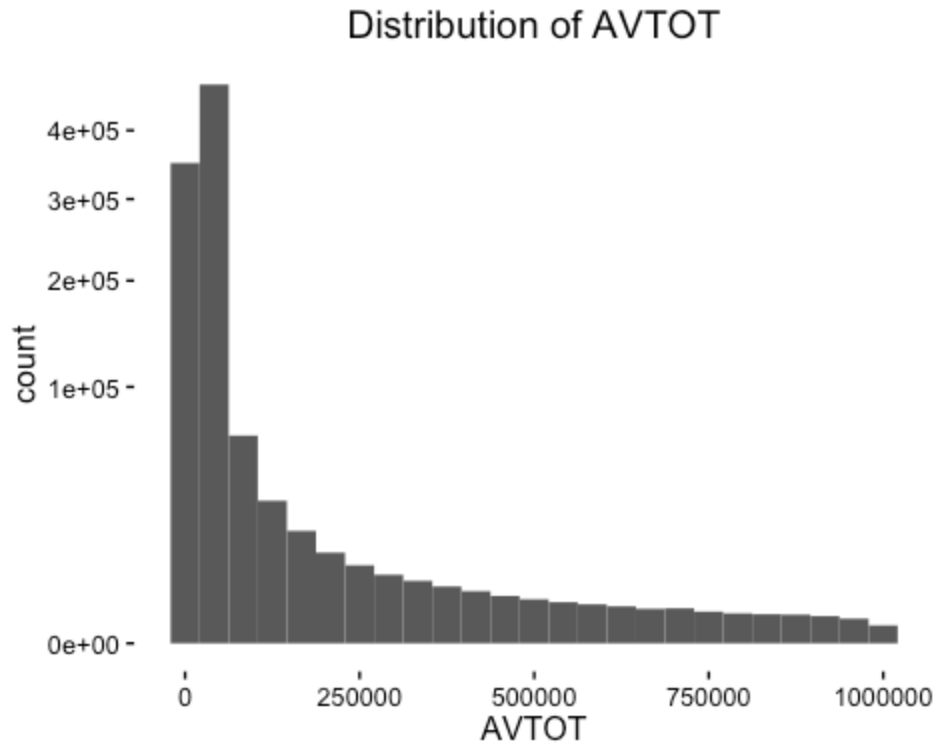
Proportion of missing value plus zero value: 0.0121708

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.000e+00	1.866e+04	2.556e+04	2.336e+05	4.688e+04	4.668e+09	12762

Mode	SD	Unique #
------	----	----------

16588	6993850	112293
-------	---------	--------

Proportion of non-zero value covered by distribution figure: 0.9755419



## "EXLAND"

**Field name:** EXLAND

**Description:** continuous with metric, Current Transitional Exempt Land Value

### Summary statistics:

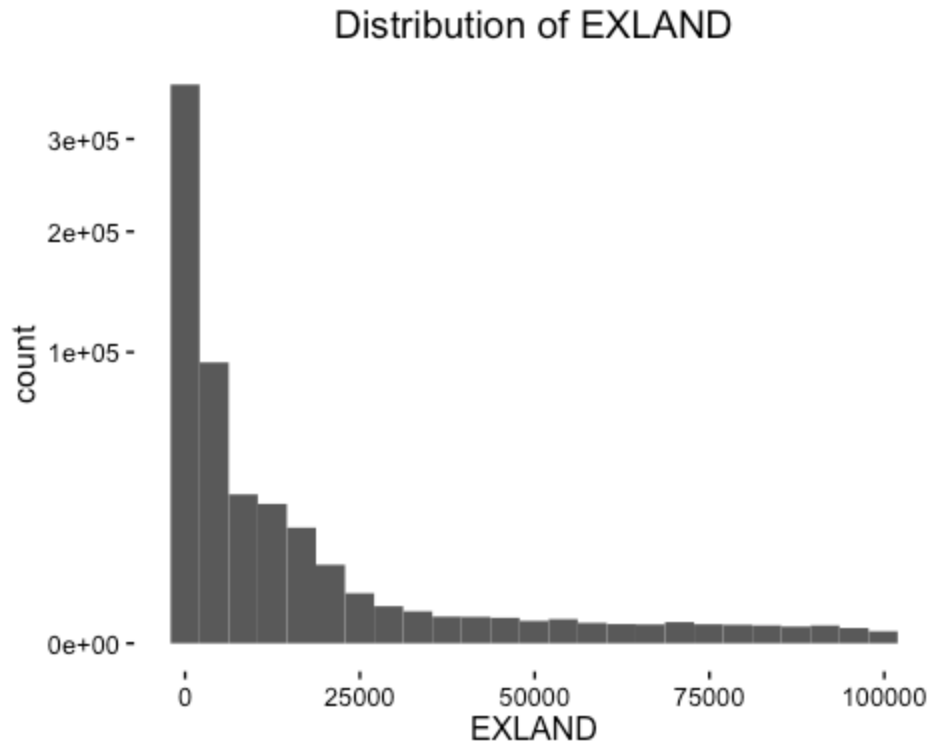
Proportion of missing value plus zero value: 0.4617924

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.000e+00	1.620e+03	1.620e+03	6.840e+04	3.240e+03	2.668e+09	484224

Mode	SD	Unique #
------	----	----------

1620	5485336	33185
------	---------	-------

Proportion of non-zero value covered by distribution figure: 0.9702738



## "EXTOT"

**Field name:** EXTOT

**Description:** continuous with metric, Current Transitional Exempt Total Value

**Summary statistics:**

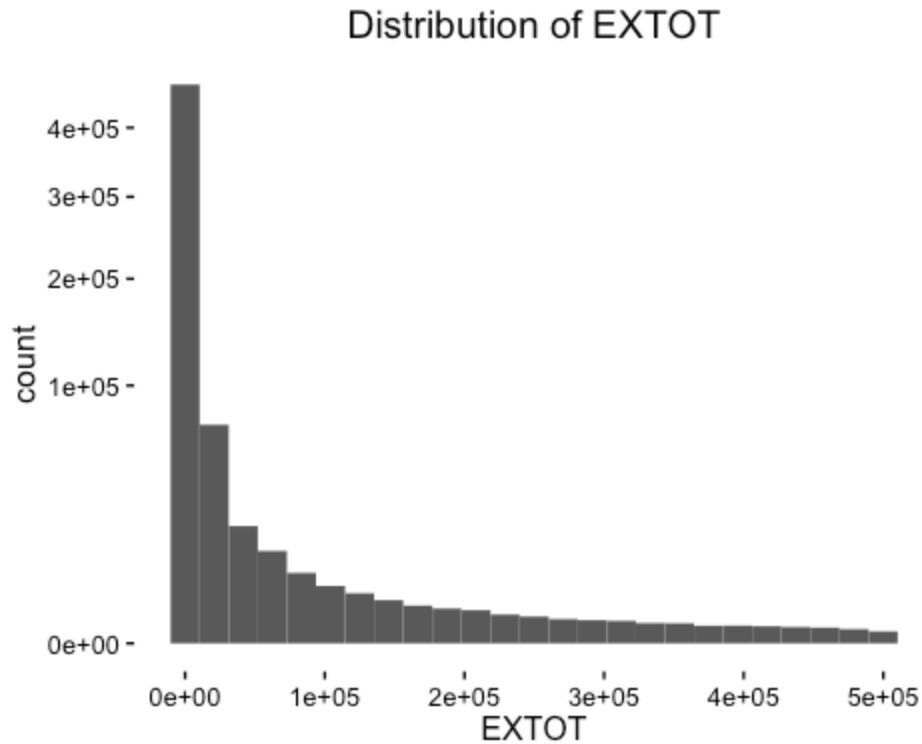
Proportion of missing value plus zero value: 0.4062647

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.000e+00	1.620e+03	1.620e+03	1.559e+05	9.786e+03	4.668e+09	425999

Mode	sd	Unique #
------	----	----------

1620	8536636	63804
------	---------	-------

Proportion of non-zero value covered by distribution figure: 0.9779192



## "EXCD1"

**Field name:** EXCD1

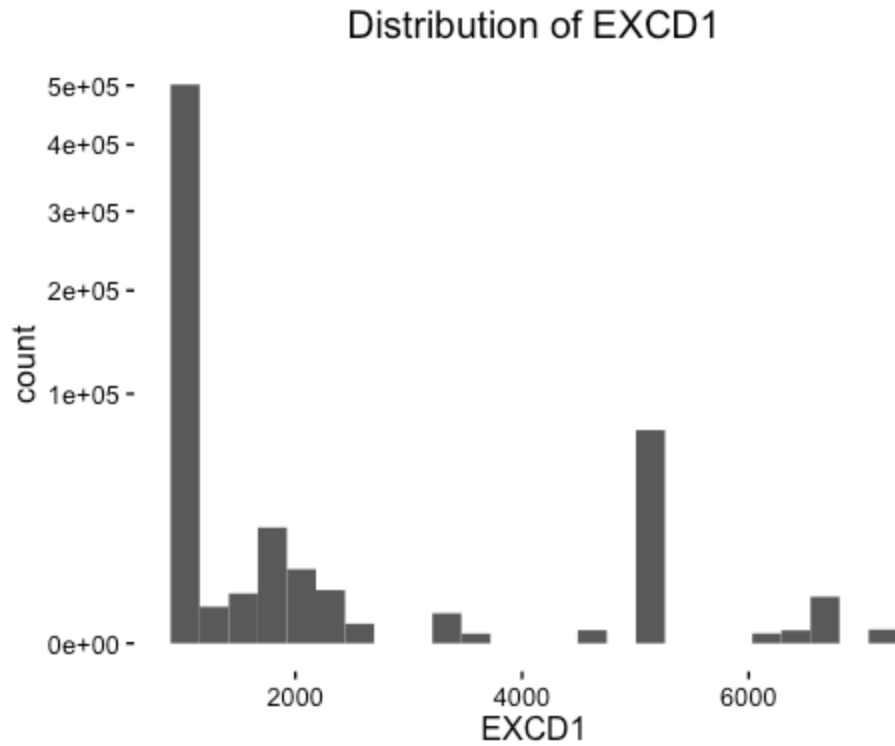
**Description:** continuous with metric

### Summary statistics:

Proportion of missing value plus zero value: 0.4062017

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's	Mode	SD	Unique #
1010	1017	1017	1605	1017	7170	425933	1017	1388.13	129

Proportion of non-zero value covered by distribution figure: 1



## "STADDR"

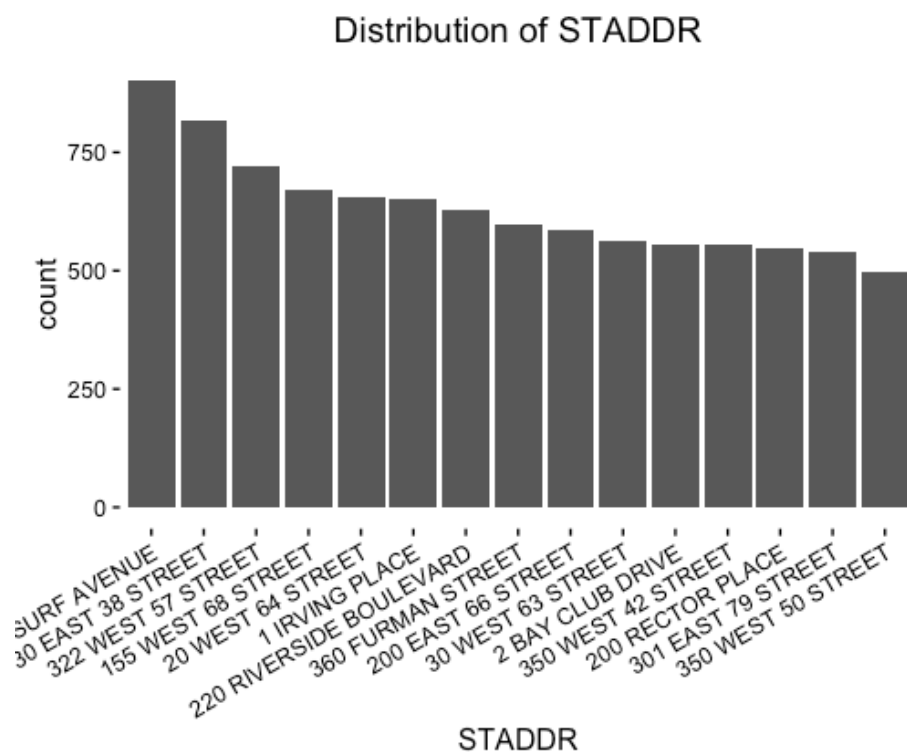
**Field name:** STADDR

**Description:** character without metric, Street name for the property.

No missing data

Below shows top 15 STADDR.



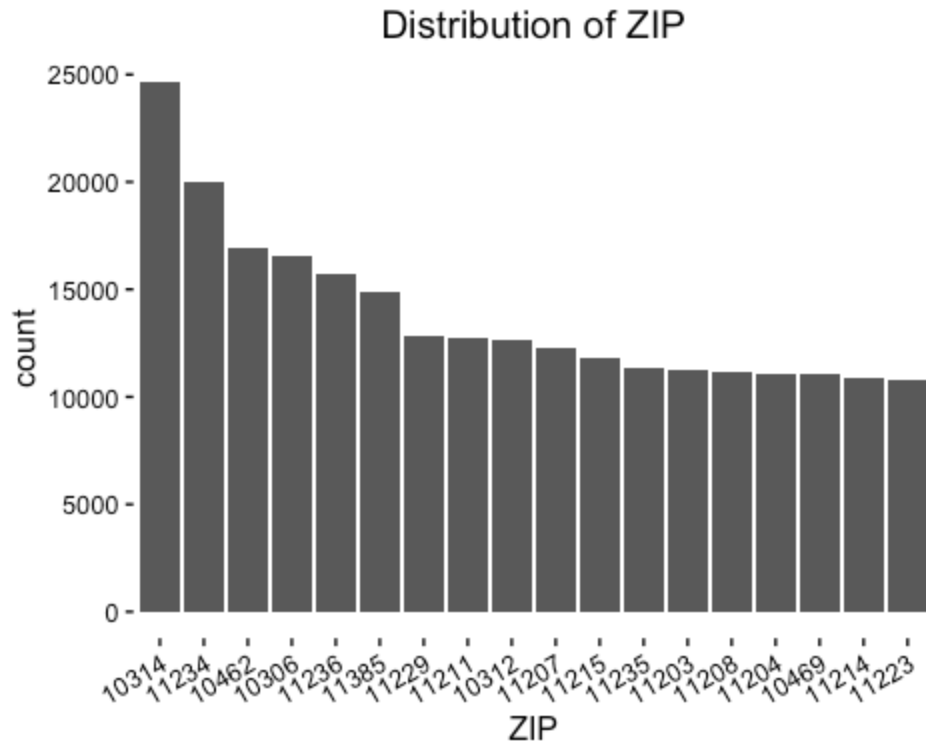


## "ZIP"

**Field name:** ZIP

**Description:** categorical with metric, Postal Zip code of the property

Unique ZIP number: 196

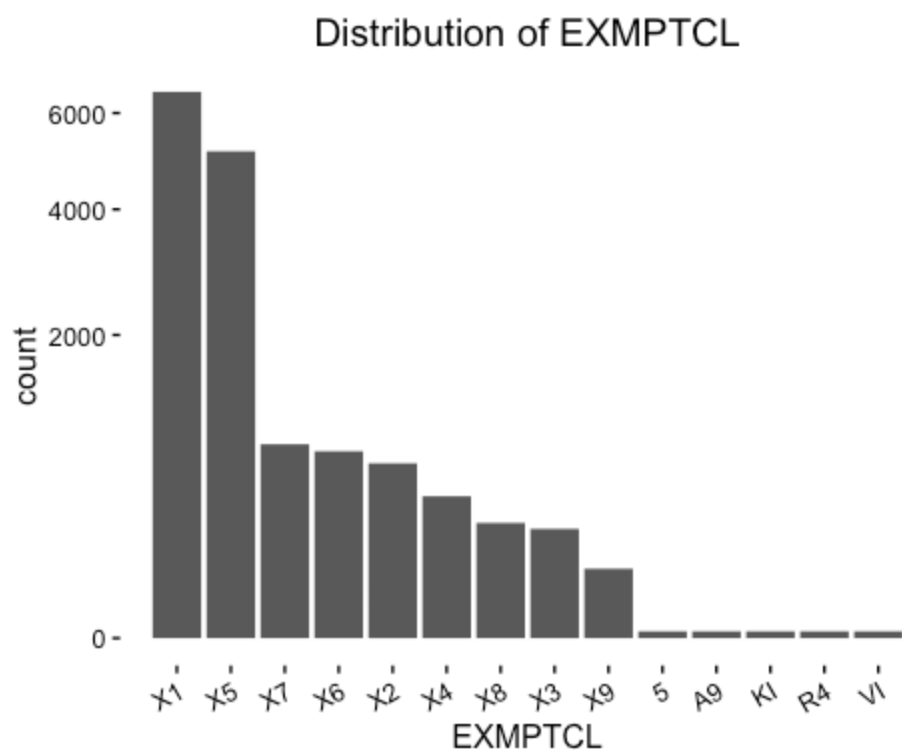


## "EXMPTCL"

**Field name:** EXMPTCL

**Description:** categorical with metric, Exempt Class used for fully exempt properties only.

Unique EXMPTCL number: 14992



## "BLDFRONT"

**Field name:** BLDFRONT

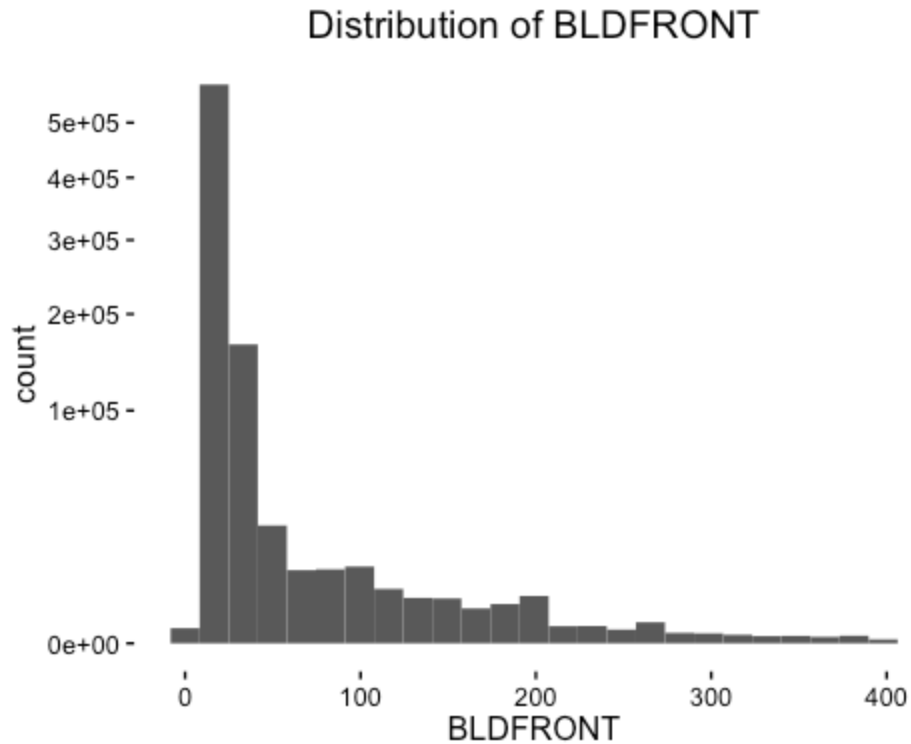
**Description:** continuous with metric

### Summary statistics:

Proportion of missing value plus zero value: 0.2142536

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's	Mode	SD	Unique #
1.0	18.0	20.0	29.3	25.0	7575.0	224661	20	38.029	609

Proportion of non-zero value covered by distribution figure: 0.9993191



## "BLDDEPTH"

**Field name:** BLDDEPTH

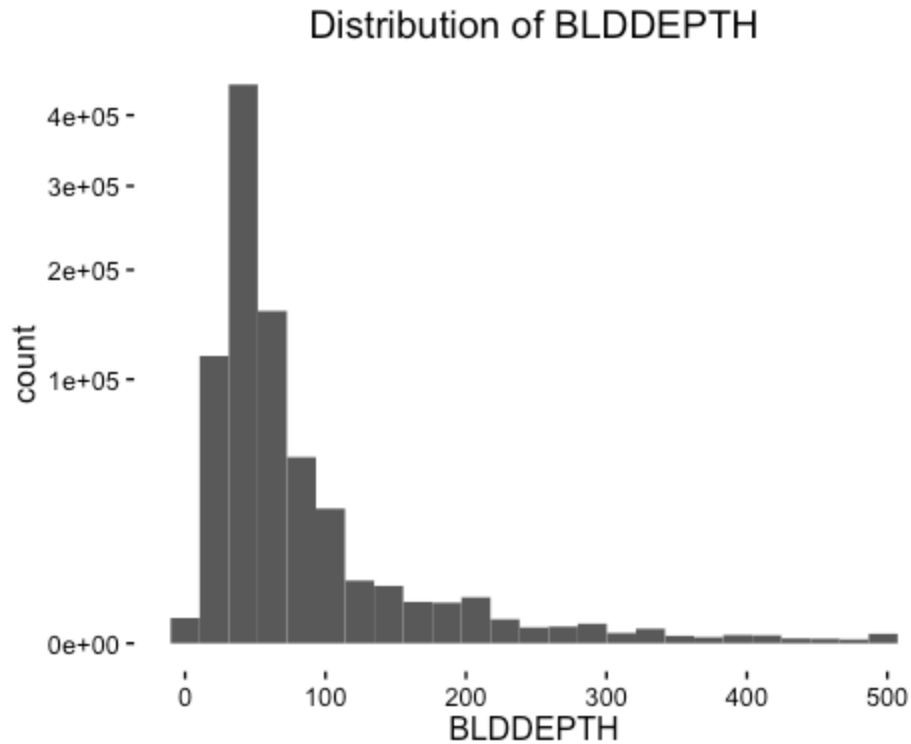
**Description:** continuous with metric

### Summary statistics:

Proportion of missing value plus zero value: 0.2142899

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's	Mode	SD	Unique #
1	35	44	51	55	9393	224699	40	42.424	619

Proportion of non-zero value covered by distribution figure: 0.9989731



## "AVLAND2"

**Field name:** AVLAND2

**Description:** continuous with metric

### Summary statistics:

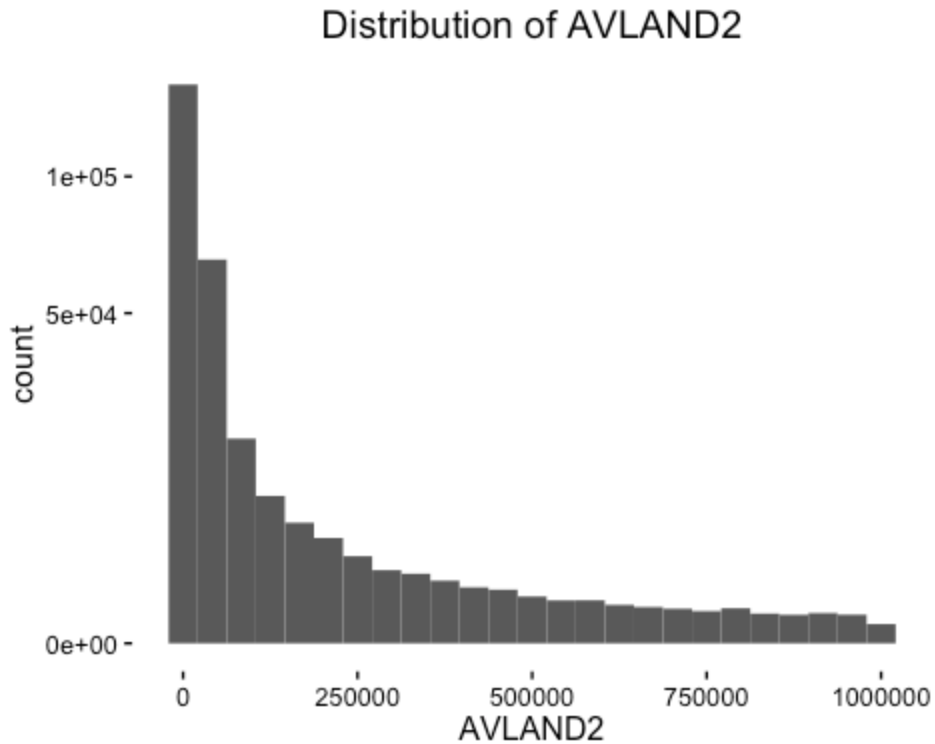
Proportion of missing value plus zero value: 0.7320497

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
3.000e+00	5.705e+03	2.006e+04	2.464e+05	6.234e+04	2.371e+09	767609

Mode	SD	Unique #
------	----	----------

2408	6199390	58169
------	---------	-------

Proportion of non-zero value covered by distribution figure: 0.9677328



## "AVTOT2"

**Field name:** AVTOT2

**Description:** continuous with metric

### Summary statistics:

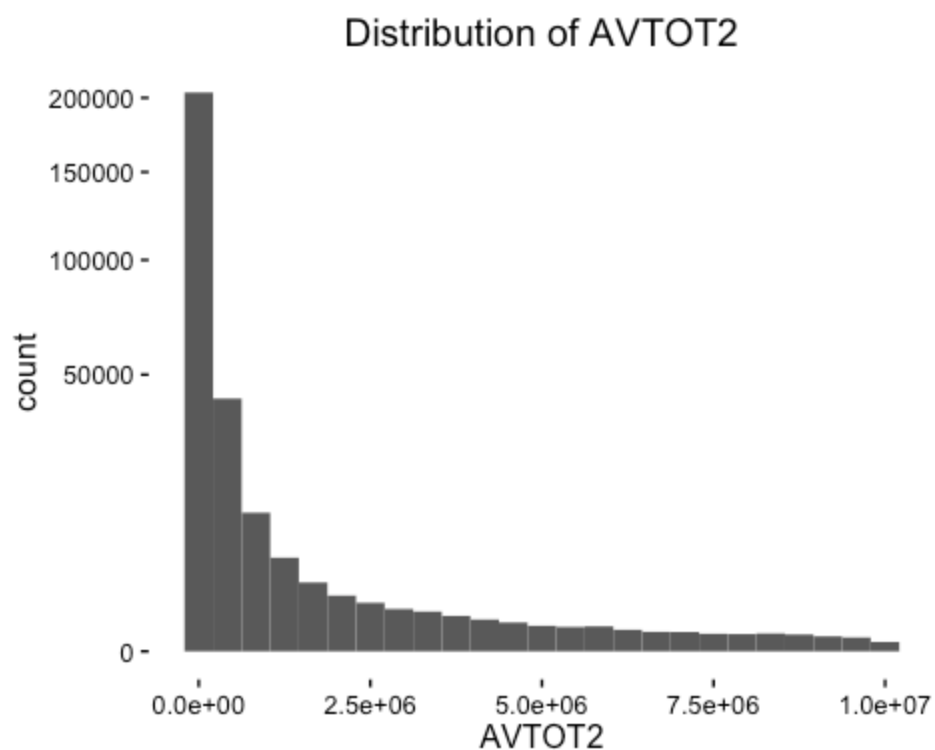
Proportion of missing value plus zero value: 0.732044

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
3.000e+00	3.401e+04	8.001e+04	7.161e+05	2.408e+05	4.501e+09	767603

Mode	SD	Unique #
------	----	----------

750	11690165	110890
-----	----------	--------

Proportion of non-zero value covered by distribution figure: 0.9894901



## "EXLAND2"

**Field name:** EXLAND2

**Description:** continuous with metric

### Summary statistics:

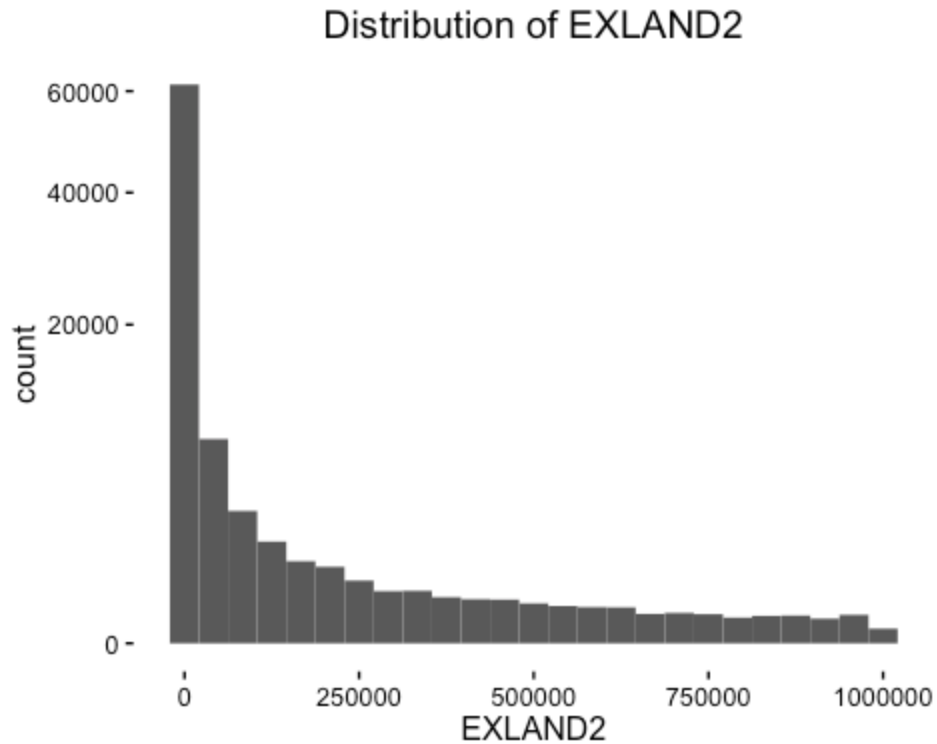
Proportion of missing value plus zero value: 0.9173402

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.000e+00	2.090e+03	3.053e+03	3.518e+05	3.142e+04	2.371e+09	961900

Mode	SD	Unique #
------	----	----------

2090	10852484	21996
------	----------	-------

Proportion of non-zero value covered by distribution figure: 0.9587078



## "EXTOT2"

**Field name:** EXTOT2

**Description:** continuous with metric

### Summary statistics:

Proportion of missing value plus zero value: 0.8760861

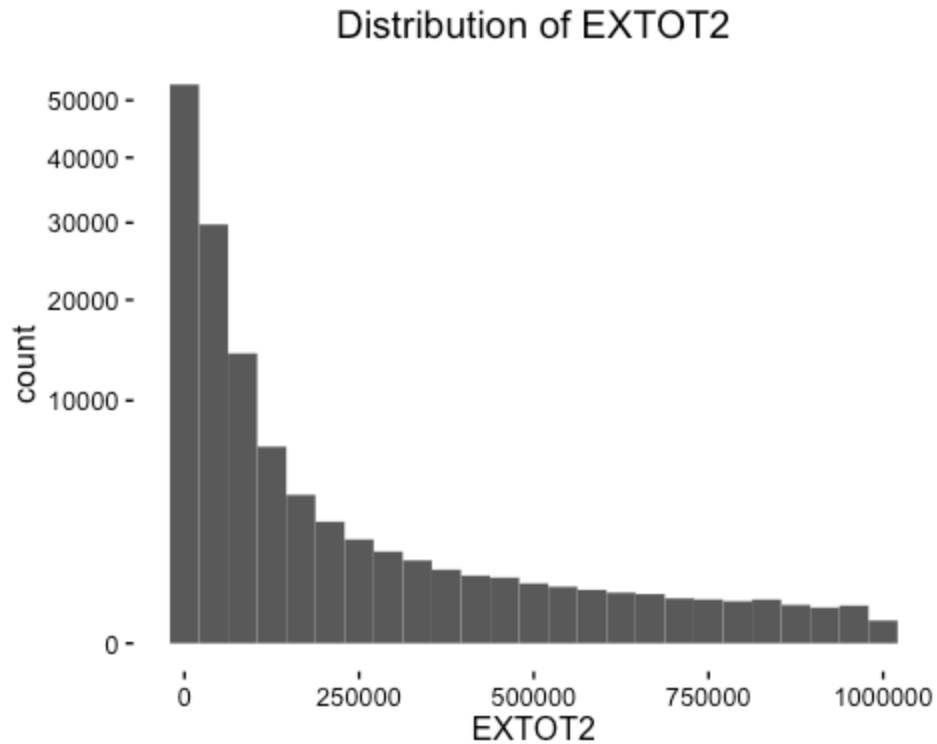
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
7.000e+00	2.889e+03	3.712e+04	6.581e+05	1.066e+05	4.501e+09	918642

Mode	Sd	Unique #
------	----	----------

2090	16129808	48106
------	----------	-------

Proportion of non-zero value covered by distribution figure: 0.933104





## "EXCD2"

**Field name:** EXCD2

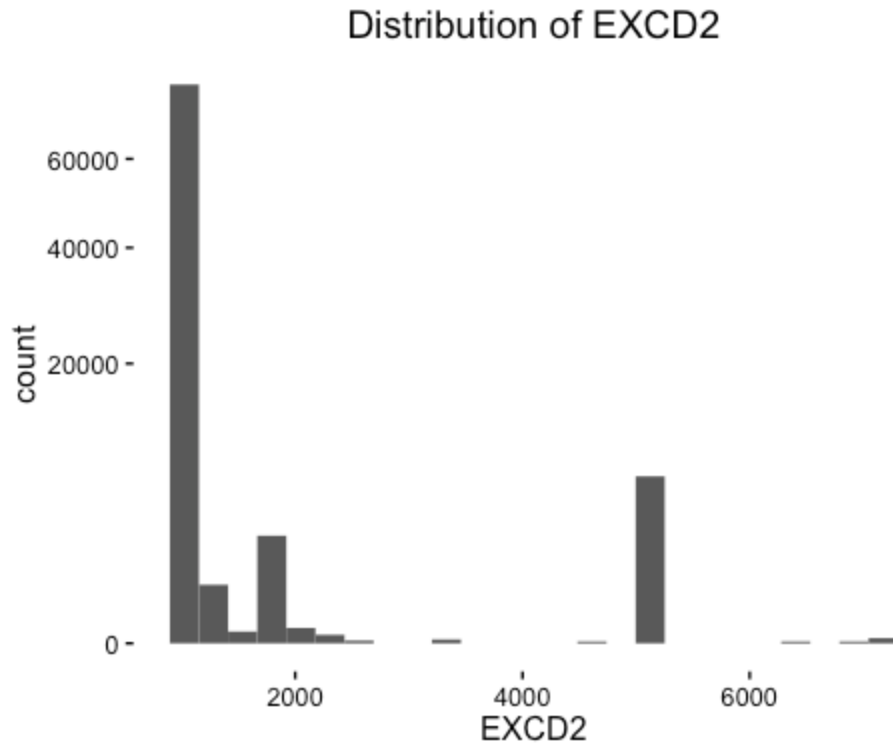
**Description:** continuous with metric

### Summary statistics:

Proportion of missing value plus zero value: 0.9132718

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's	Mode	SD	Unique #
1011	1017	1017	1372	1017	7160	957634	1017	1105.5	60

Proportion of non-zero value covered by distribution figure: 1



## "PERIOD"

**Field name:** PERIOD

**Description:** categorical no metric, Indicator for the Change Period of the File

**Value** = FINAL, no missing data

## "YEAR"

**Field name:** YEAR

**Description:** categorical no metric

**Value** = 2010/11, no missing data

## "VALTYPE"

**Field name:** VALTYPE

**Description:** categorical no metric

**Value** = AC-TR, no missing data