City of New York Property Valuation and Assessment Data

Data Quality Report

Yu Dong
01/28/2017

Summary

File description:

The City of New York Property Valuation and Assessment Data file is a public available dataset posted by the Department of Finance on the City of New York Open Data website. The dataset contains the records of more than 1 million properties across the city of New York and information on their sizes, values, owners, building classes, tax classes, etc.

File Name:

City of New York Property Valuation and Assessment Data

Data Source:

City of New York Open Data Website (https://data.cityofnewyork.us/Housing-Development/Property-Valuation-and-Assessment-Data/rgy2-tti8)

Number of Records:

1,048,575 records

Number of Fields:

30 variables in total – 13 categorical variables, 14 numeric variables, 2 text variables, 1 date variables

Time of Records:

November 2011

Fields Explanation

Field 1

Field Name: RECORD

Description:
RECORD is a categorical variable. It works as the ordinal reference number for each property record.

Unique Values:
1,048,575 unique values, ranging from 1 to 1,048,575. No repeated values or missing values exist.

Field 2

Field Name: BBLE

Description:
BBLE is a nominal categorical variable with 10 or 11 digits. It is the concatenation of BORO code (1 digit), BLOCK code (5 digit), LOT code (4 digit) and EASEMENT code (1 digit if exists).

Unique Values:
1,048,575 unique values. No repeated values or missing values exist.

Field 3

Field Name: BLOCK

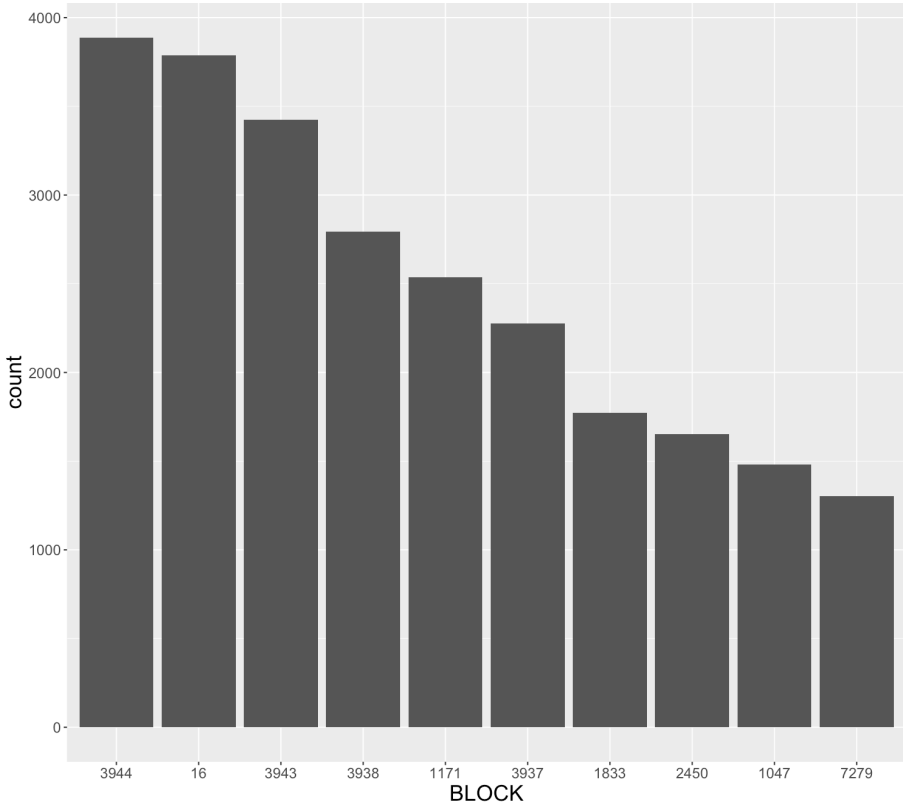
Description:

BLOCK is a categorical variable with 1 to 5 digits. It represents the property's corresponding block code in a certain borough. For each borough, there is a valid block code range:

MANHATTAN	1 TO 2,255
BRONX	2,260 TO 5,958
BROOKLYN	1 TO 8,955
QUEENS	1 TO 16,350
STATEN ISLAND	1 TO 8,050

Unique Values:

BLOCK has 13949 unique values, ranging from 1 to 16350. No missing values. The top 10 most frequently appeared BLOCK code is shown below.



BLOCK	Percentage(%)
3944	0.37
16	0.36
3943	0.32
3938	0.27
1171	0.24
3937	0.21
1833	0.17
2450	0.16
1047	0.14
7279	0.12

Field 4

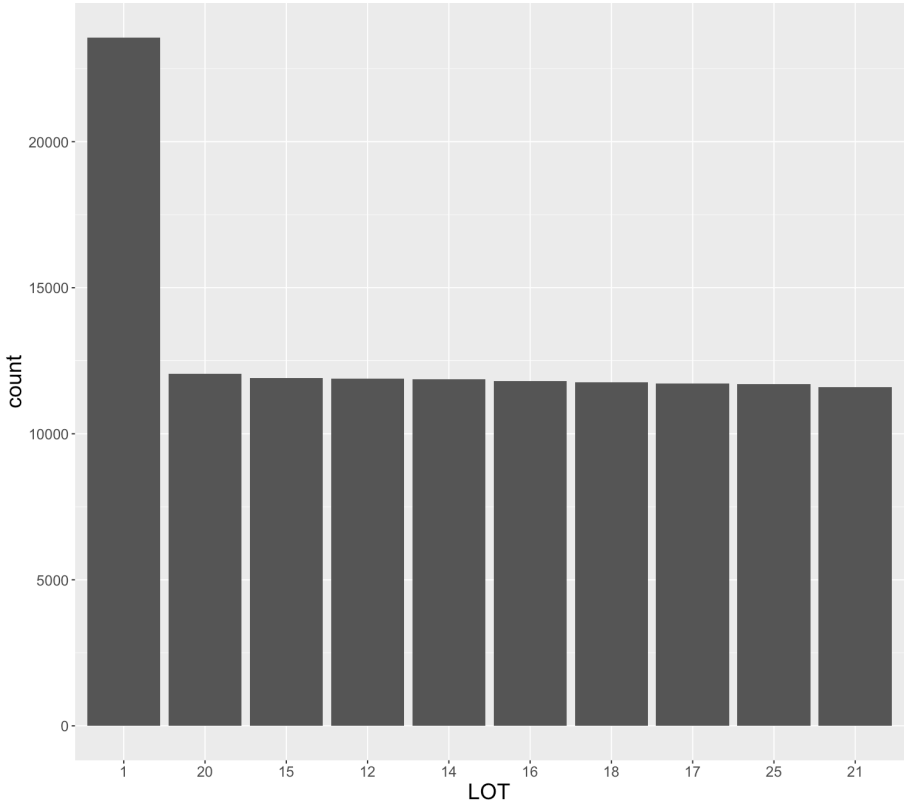
Field Name: LOT

Description:

LOT is a categorical variable with 1 to 4 digits. It represents the property's lot code within its borough and block.

Unique Values:

LOT has 6366 unique values, ranging from 1 to 9978. No missing values. The top 10 most frequently appeared LOT code is shown below.



LOT	Percentage(%)
1	2.25
20	1.15
15	1.14
12	1.13
14	1.13
16	1.13
18	1.12
17	0.12
25	1.12
21	1.11

Field 5

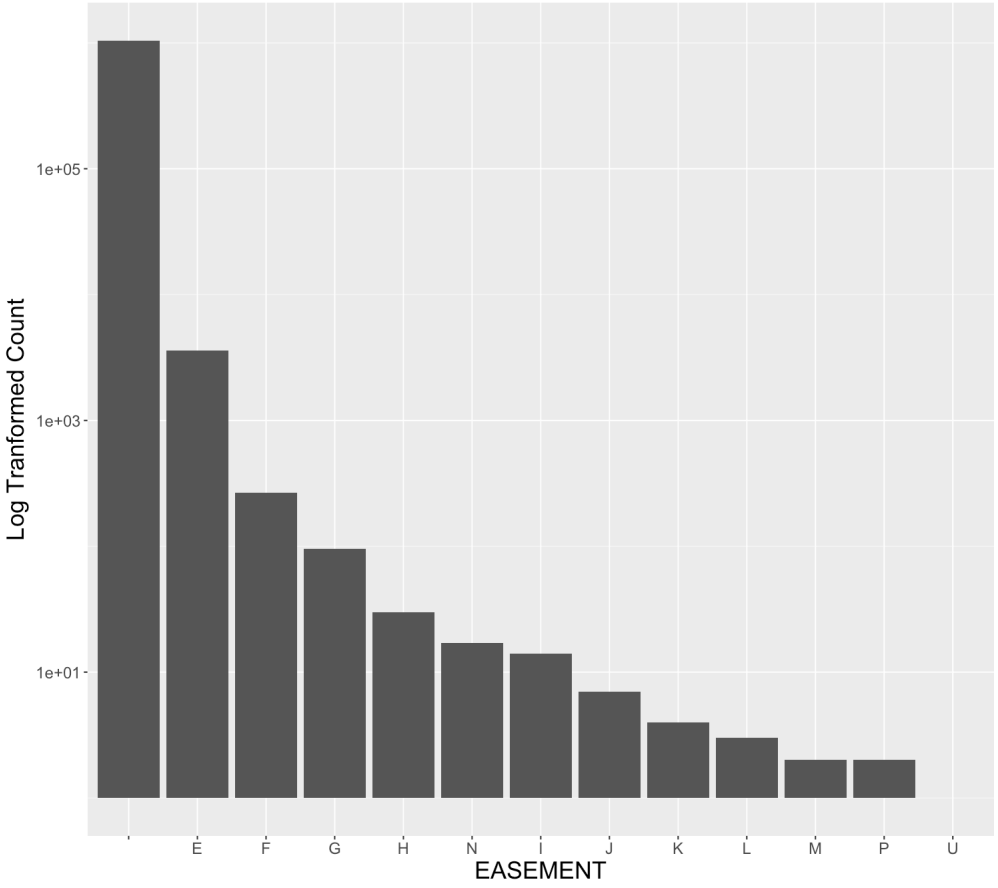
Field Name: EASEMENT

Description:

EASEMENT is a nominal categorical variable representing the property's easement type.

Unique values:

EASEMENT has 13 levels – "", "E", "F", "G", "H", "I", "J", "K", "L", "M", "N", "P", "U". The null value indicates the property does not have any special easement. No missing values exist. The sorted bar chart with log transformed y axis is shown below.



Level	Counts
	1044532 (99.6%)
E	3603
F	265
G	95
H	30
I	14
J	7
K	4
L	3
M	2
N	17
P	2
U	1

Field 6

Field Name: OWNER

Description:

OWNER is a text variable indicating the owner of the property.

Unique Values:

OWNER has 847055 unique values. In the OWNER field, 31081 properties have the value “”, indicating possible missing values. No missing values exist. The top 10 most frequently occurred OWNER names are:

Owner	Count	Percentage(%)
	31081	2.96
PARKCHESTER PRESERVAT	6021	0.57
PARKS AND RECREATION	3358	0.32
DCAS	2053	0.20
HOUSING PRESERVATION	1900	0.18
CITY OF NEW YORK	1189	0.11
NEW YORK CITY HOUSING	1014	0.10
BOARD OF EDUCATION	1003	0.10
CNY/NYCTA	975	0.09
NYC HOUSING PARTNERSH	747	0.07

Field 7

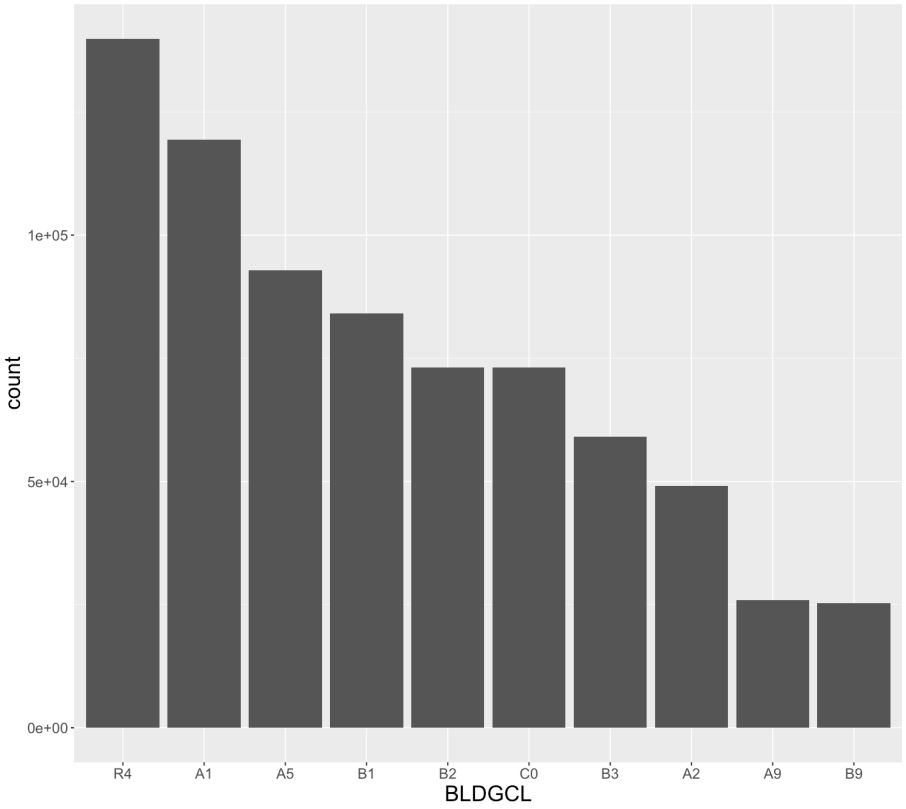
Field Name: BLDGCL

Description:

BLDGCL is a nominal categorical variable indicating the building class.

Unique Values:

BLDGCL has 200 unique levels. Each level has 2 digits – the first digit is a character from A to Z, the second digit is a number from 0 to 9. No missing values exist. The top 10 most frequently occurred BLDGCL is shown in below:



BLDGCL	Percentage(%)
R4	13.3
A1	11.4
A5	8.9
B1	8
B2	7
C0	8
B3	5.6
A2	4.7
A9	2.5
B9	2.4

Field 8

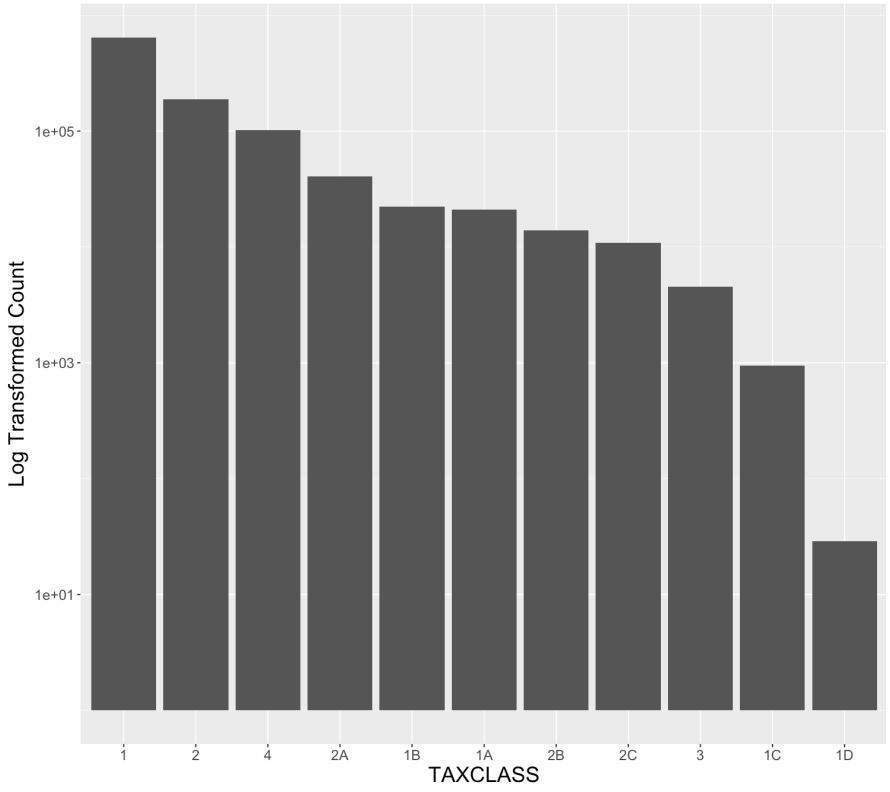
Field Name: TAXCLASS

Description:

TAXCLASS is a categorical variable indicating the tax class of the property.

Unique Values:

TAXCLASS has 11 unique levels – “1”, “1A”, “1B”, “1C”, “1D”, “2”, “2A”, “2B”, “2C”, “3”, and “4”. No missing values exist. Sorted TAXCLASS levels are shown below:



Level	Percentage(%)
1	61.4
2	18.0
4	9.8
2A	3.9
1B	2.1
1A	2.0
2B	1.3
2C	1.0
3	0.4
1C	0.1
1D	0.0

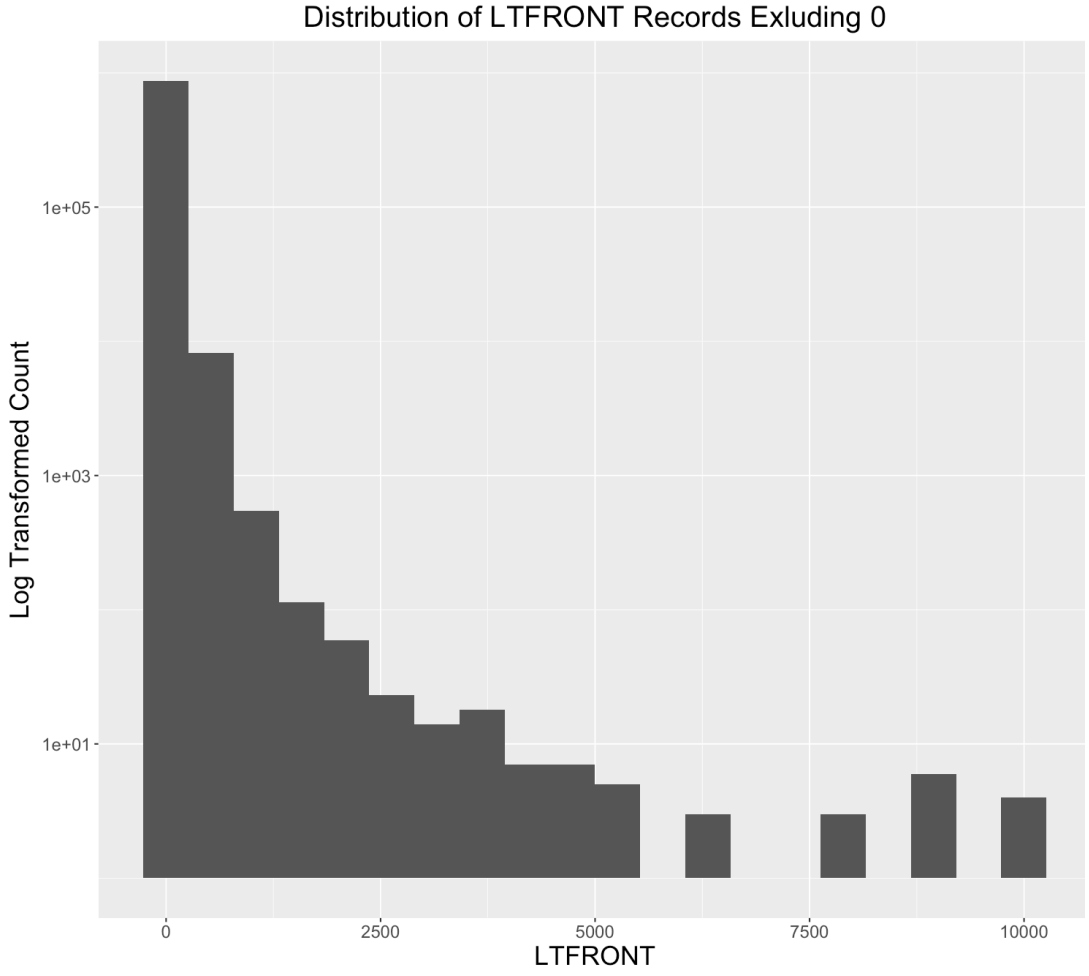
Field 9

Field Name: LTFRONT

Description:
LTFRONT is a numeric variable representing the length of lot frontage in feet.

Unique Values:
LTFRONT has 1277 unique values ranging from 0 to 9999. No missing values exist. There are 168,867 records of 0 LTFRONT, and a LTFRONT of 0 may indicate missing value. The statistics and distribution excluding 0 records are shown as below.

Minimum	1
Maximum	9999
Median	25
Mean	43.12
Mode	20
SD	78.62



Field 10

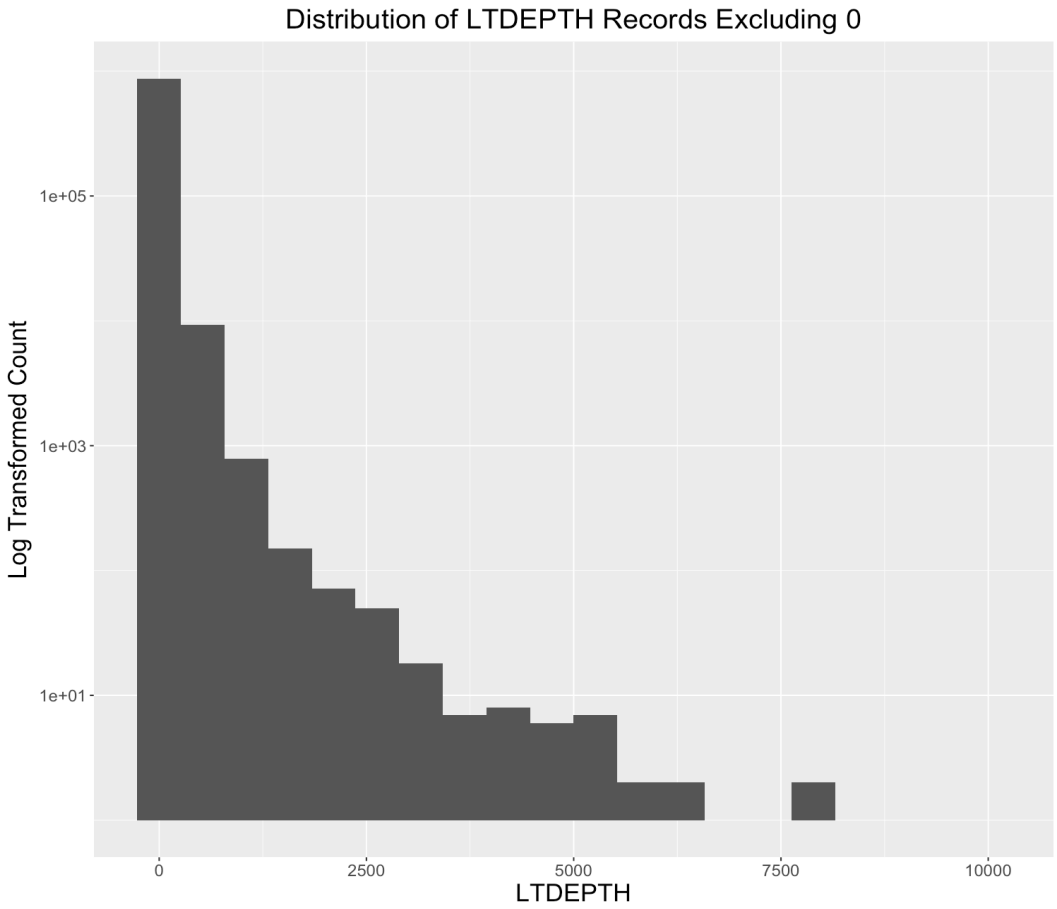
Field Name: LTDEPTH

Description:

LTDEPTH is a numeric variable representing the length of lot depth in feet.

Unique Values:
LTDEPTH has 1336 unique values ranging from 0 to 9999. No missing values exist. There are 169,888 records of 0 LTDEPTH, and a LTDEPTH of 0 may indicate missing value. The statistics and distribution excluding 0 records are shown as below.

Minimum	1
Maximum	9999
Median	100
Mean	105.34
Mode	100
SD	70.71



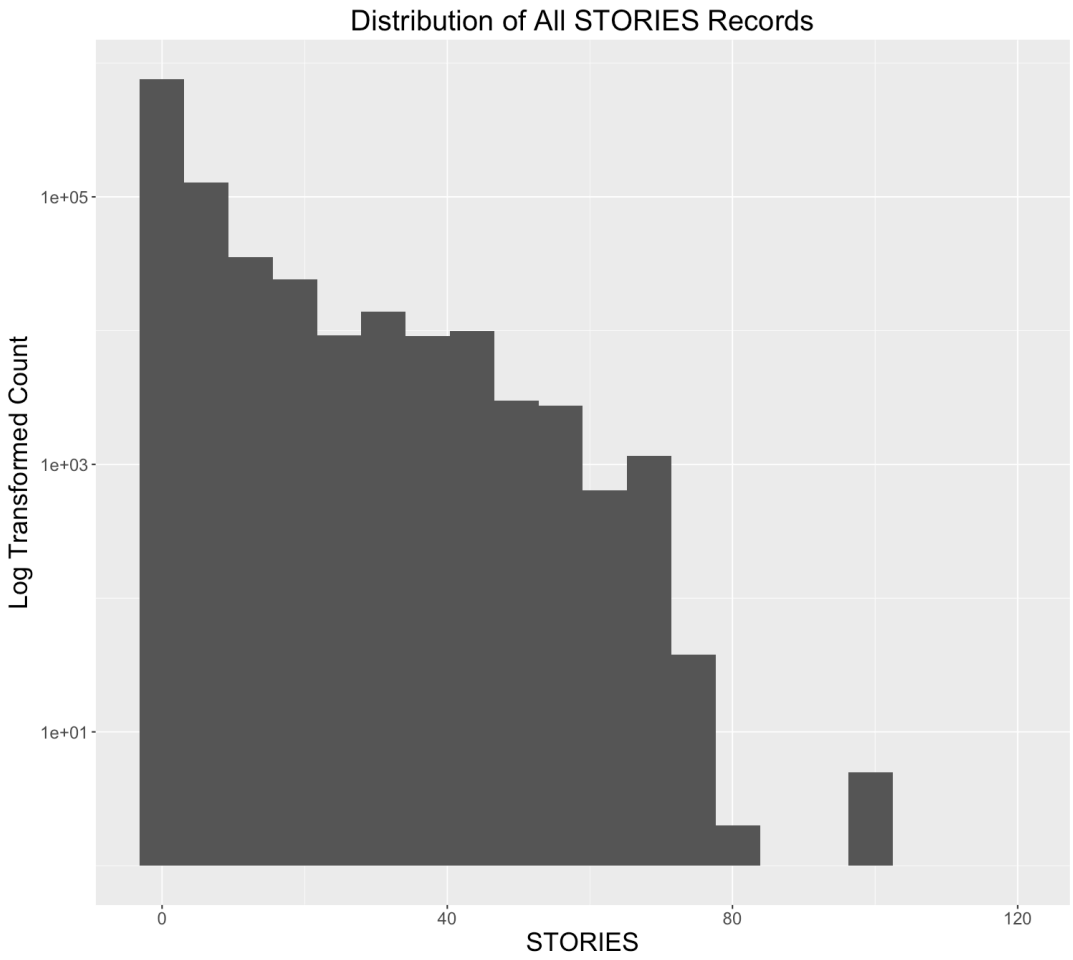
Field 11

Field Name: STORIES

Description:
STORIES is a numeric variable representing the number of stories of the property.

Unique Values:
STORIES has 112 unique values ranging from 1 to 119. There are 52,142 missing values in the STORIES field. The statistics and distribution are shown as below.

Minimum	1
Maximum	119
Median	2
Mean	5.06
Mode	2
SD	8.43



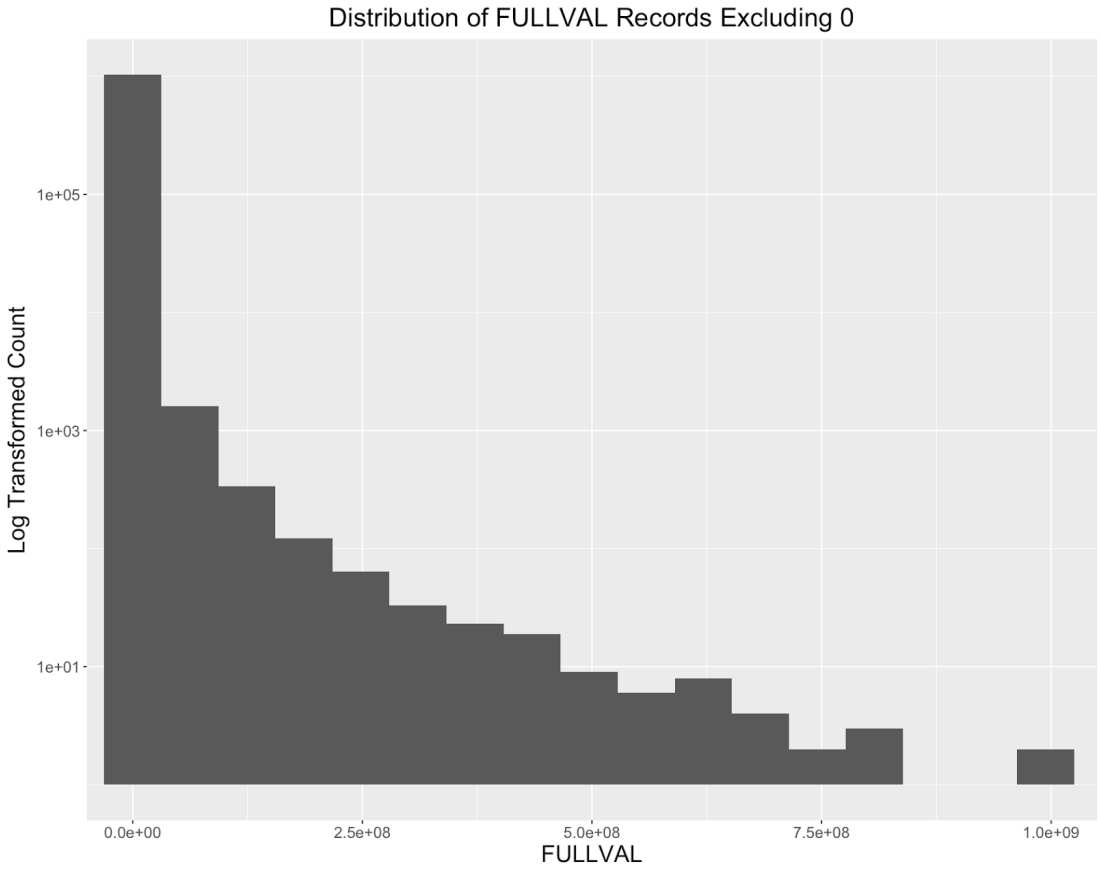
Field 12

Field Name: FULLVAL

Description:
FULLVAL is a numeric variable representing the full value of the property.

Unique Values:
FULLVAL has 108277 unique values ranging from 0 to about 6,000,000,000. There are 12,762 properties with the FULLVAL of 0 in the dataset. No missing values exist. The statistics and distribution excluding 0 records are shown as below.

Minimum	4
Maximum	6.15E+09
Median	45000
Mean	8.91E+05
Mode	502000
SD	1.17E+07

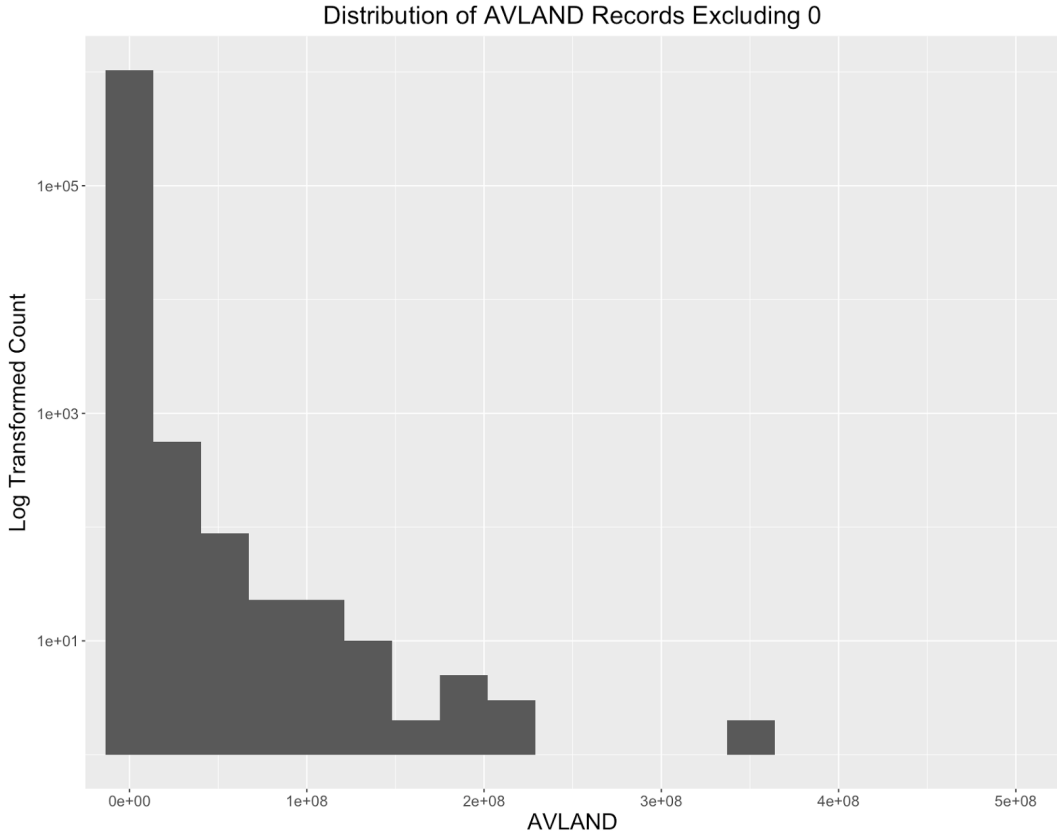


Field 13
Field Name: AVLAND

Description:
AVLAND is a numeric variable representing the assessed value of the land.

Unique Values:
AVLAND has 70,529 unique values ranging from 0 to about 2,700,000,000. There are 12,764 properties with the AVLAND of 0 in the dataset. No missing values exist. The statistics and distribution excluding 0 records are shown as below.

Minimum	1
Maximum	2.67E+09
Median	13751
Mean	86054.72
Mode	45000
SD	4.10E+06



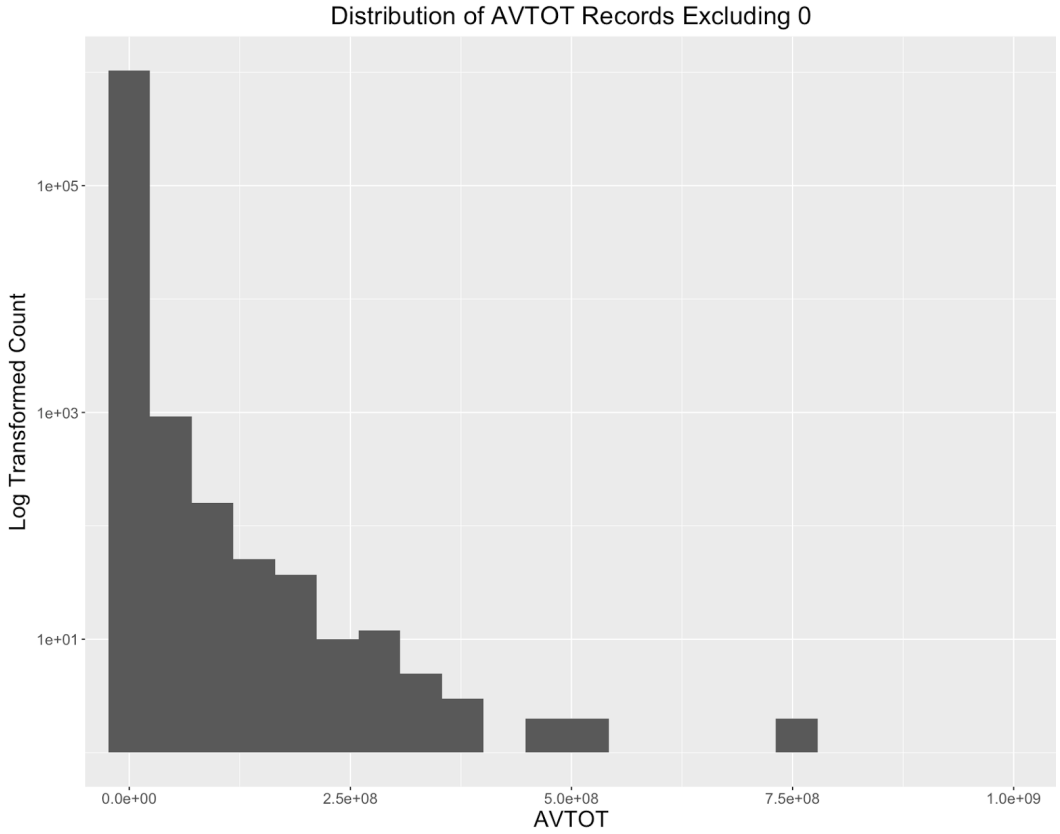
Field 14

Field Name: AVTOT

Description:
AVTOT is a numeric variable representing the assessed total value of the property.

Unique Values:
AVTOT has 112294 unique values ranging from 0 to about 4,700,000,000. There are 12,762 properties with the AVTOT of 0 in the dataset. No missing values exist. The statistics and distribution excluding 0 records are shown as below.

Minimum	1
Maximum	4.67E+09
Median	25560
Mean	233601.3
Mode	16588
SD	6.99E+06



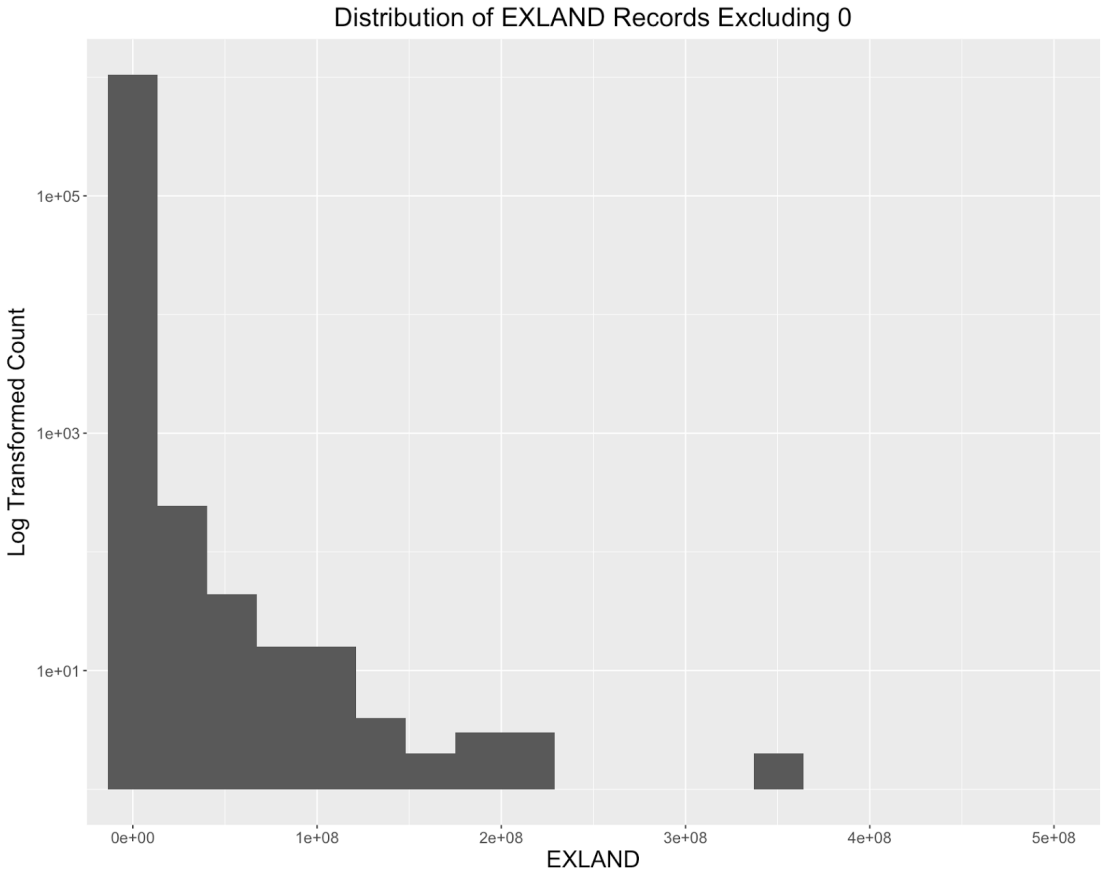
Field 15

Field Name: EXLAND

Description:
EXLAND is a numeric variable representing the value of the exempt land. The value of EXLAND is always smaller or equal to AVLAND.

Unique Values:
EXLAND has 33186 unique values ranging from 0 to about 2,700,000,000. There are 484,224 properties with the EXLAND of 0 in the dataset. No missing values exist. The statistics and distribution excluding 0 records are shown as below.

Minimum	1
Maximum	2.67E+09
Median	1620
Mean	68397.01
Mode	1620
SD	5485336



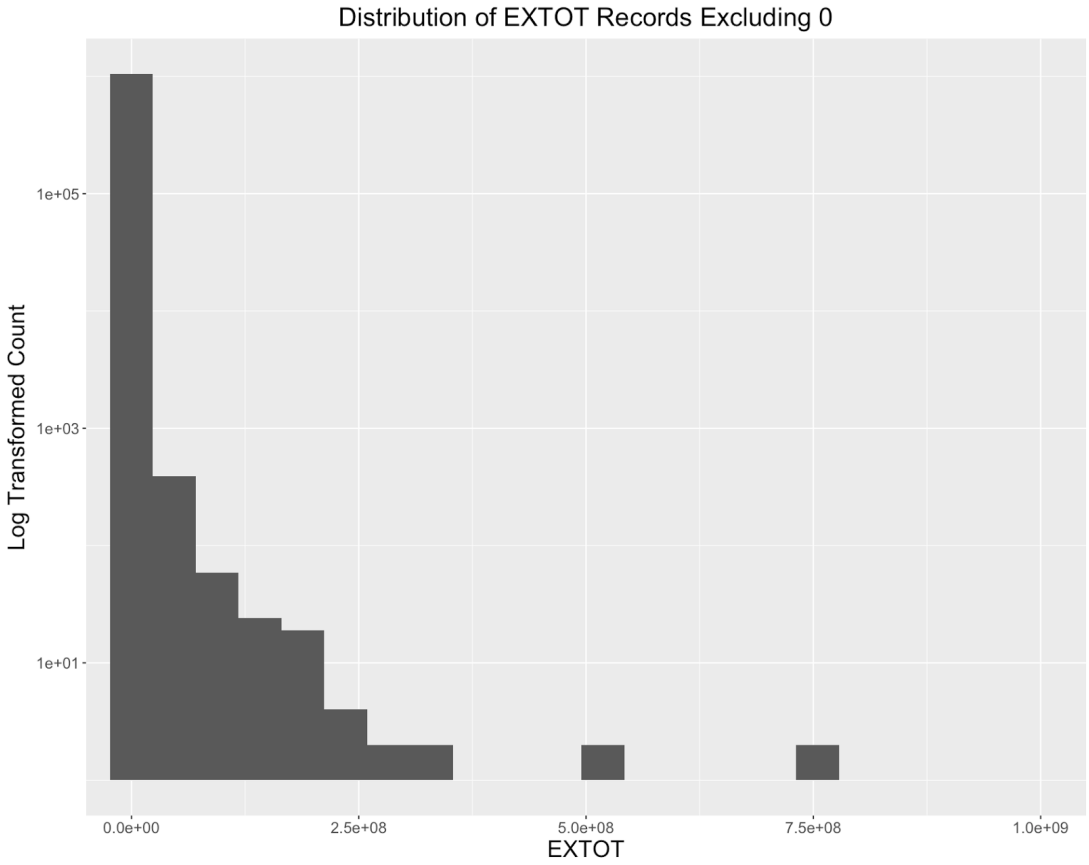
Field 16

Field Name: EXTOT

Description:
EXTOT is a numeric variable representing the total value of the exempt property. The value of EXTOT is always smaller or equal to AVTOT.

Unique Values:
EXTOT has 63805 unique values ranging from 0 to about 4,700,000,000. There are 425,999 properties with the EXTOT of 0 in the dataset. No missing values exist. . The statistics and distribution excluding 0 records are shown as below.

Minimum	1
Maximum	4.67E+09
Median	1620
Mean	155867.1
Mode	1620
SD	8536636

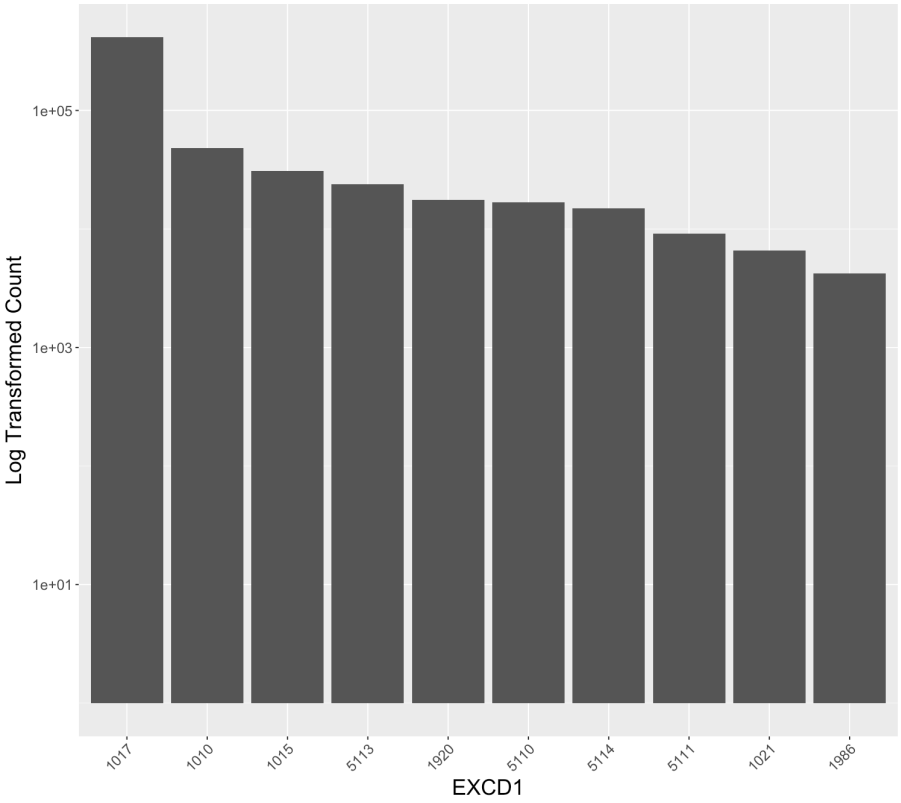


Field 17

Field Name: EXCD1

Description:
EXCD1 is a categorical variable, possibly representing the code for the exempt reasons.

Unique Values:
EXTOT has 130 levels, taking 4-digit numbers from 1010 to 7170. There are 425,933 missing values exist. For all properties that hold missing value in EXCD1, their values in both EXLAND and EXTOT fields are 0. The top 20 most frequently occurred EXCD1 values are shown below.



EXCD1	Percentage(%)
1017	39.5
1010	4.6
1015	2.9
5113	2.3
1920	1.7
5110	1.6
5114	1.4
5111	8.7
1021	6.3
1986	4.0

Field 18

Field Name: STADDR

Description:

STADDR is a text variable, representing the street address of the property.

Unique values:

STADDR has 820,638 unique values. No missing values exist in this field. However, there are 641 records with the value of "" (null) in the STADDR field, indicating missing values.
The top 10 most frequently occurred STADDR values are:

Address	Counts
501 SURF AVENUE	902
330 EAST 38 STREET	817
322 WEST 57 STREET	720
155 WEST 68 STREET	671
20 WEST 64 STREET	657
1 IRVING PLACE	650
	641
220 RIVERSIDE BOULEVARD	628
360 FURMAN STREET	599
200 EAST 66 STREET	585

Field 19

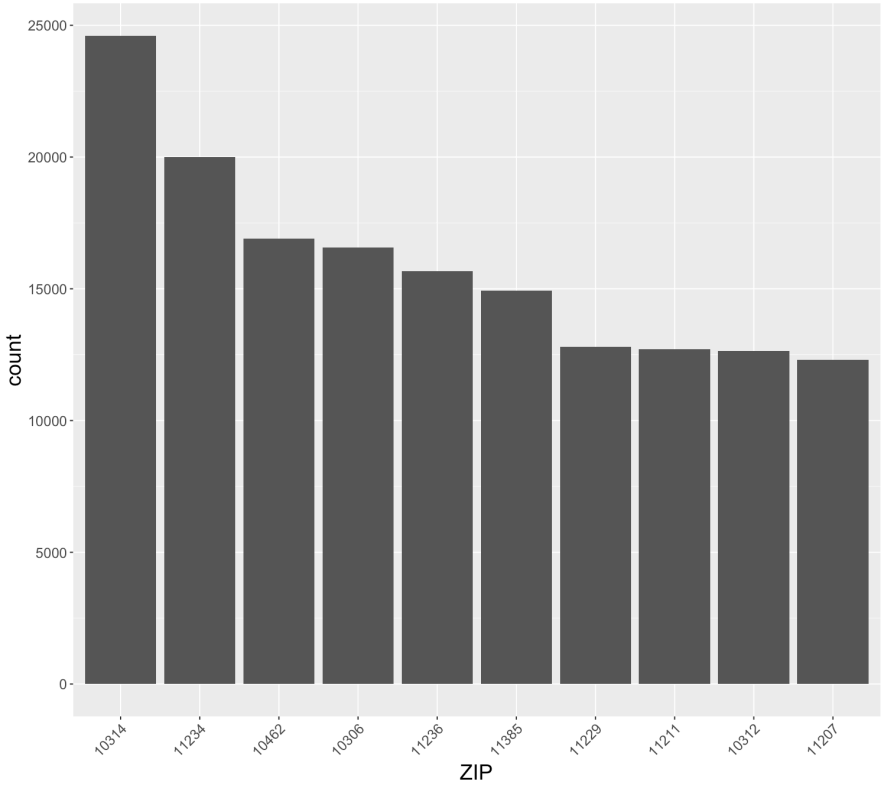
Field Name: ZIP

Description:

ZIP is a categorical variable, recording the zipcode of the property.

Unique Values:

ZIP has 197 unique values and 26,356 missing values. There are three obvious anomaly records with ZIP of 33803, which should be in Florida. The top 20 most frequently occurred ZIP values are:



ZIP	Percentage(%)
10314	2.3
11234	1.9
10462	1.6
10306	1.6
11236	1.5
11385	1.4
11229	1.2
11211	1.2
10312	1.2
11207	1.2

Field 20

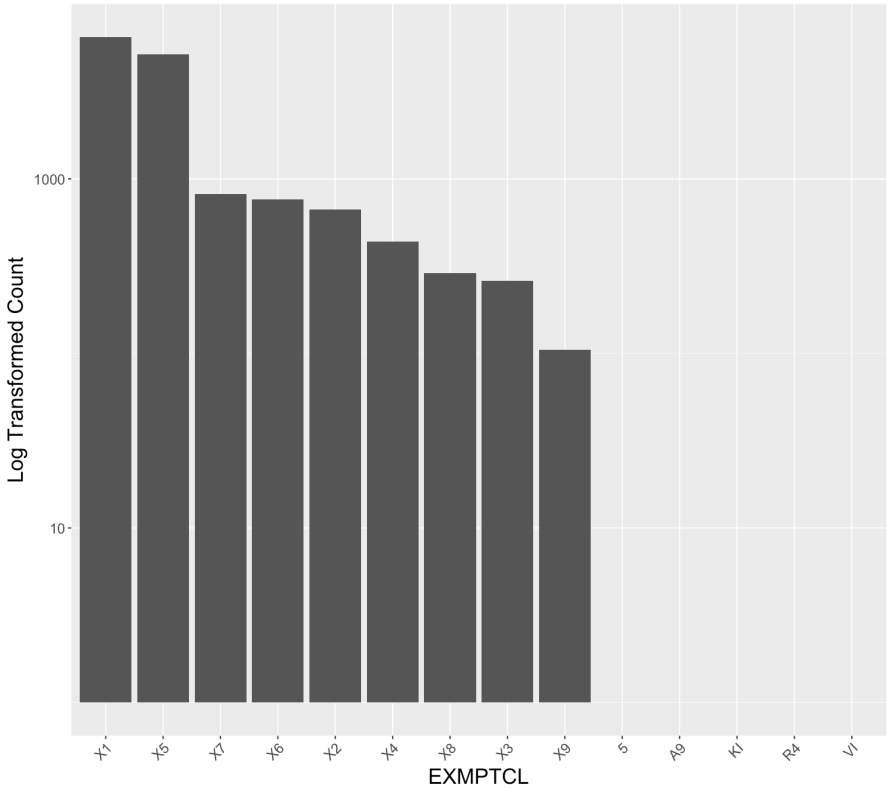
Field Name: EXMPTCL

Description:

EXMPTCL is a nominal categorical variable, representing the exempt class, which is used for exempt properties only.

Unique Values:

EXMPTCL has 15 levels- “”, “5”, “A9”, “KI”, “R4”, “VI”, “X1”, “X2”, “X3”, “X4”, “X5”, “X6”, “X7”, “X8”, and “X9”. 1,033,583 properties take “” value in EXMPTCL. No missing values exist. The sorted bar chart is shown below.



EXMPTCL	Percentage(%)
X1	43.41
X5	34.41
X7	5.46
X6	5.07
X2	4.44
X5	2.92
X8	1.93
X3	1.73
X9	0.70
5	0.01
A9	0.01
KI	0.01
R4	0.01
VI	0.01

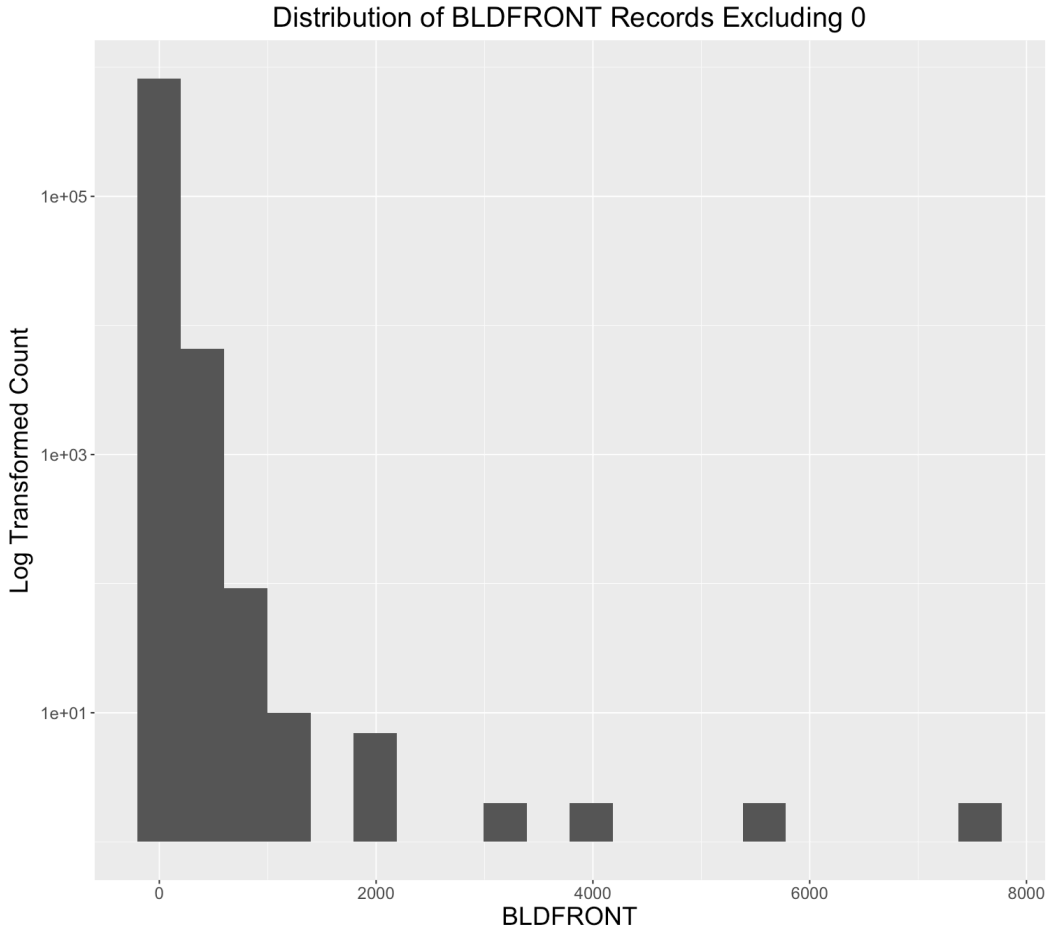
Field 21

Field Name: BLDFRONT

Description:
BLDFRONT is a numeric variable representing the length of building frontage in feet.

Unique Values:
BLDFRONT has 610 unique values ranging from 0 to 7575. No missing values exist. However, there are 224,661 records with value 0, which could be in fact missing values. The statistics and distribution excluding all records with 0 BLDFRONT are shown as below.

Minimum	1
Maximum	7575
Median	20
Mean	29.29
Mode	20
SD	38.03



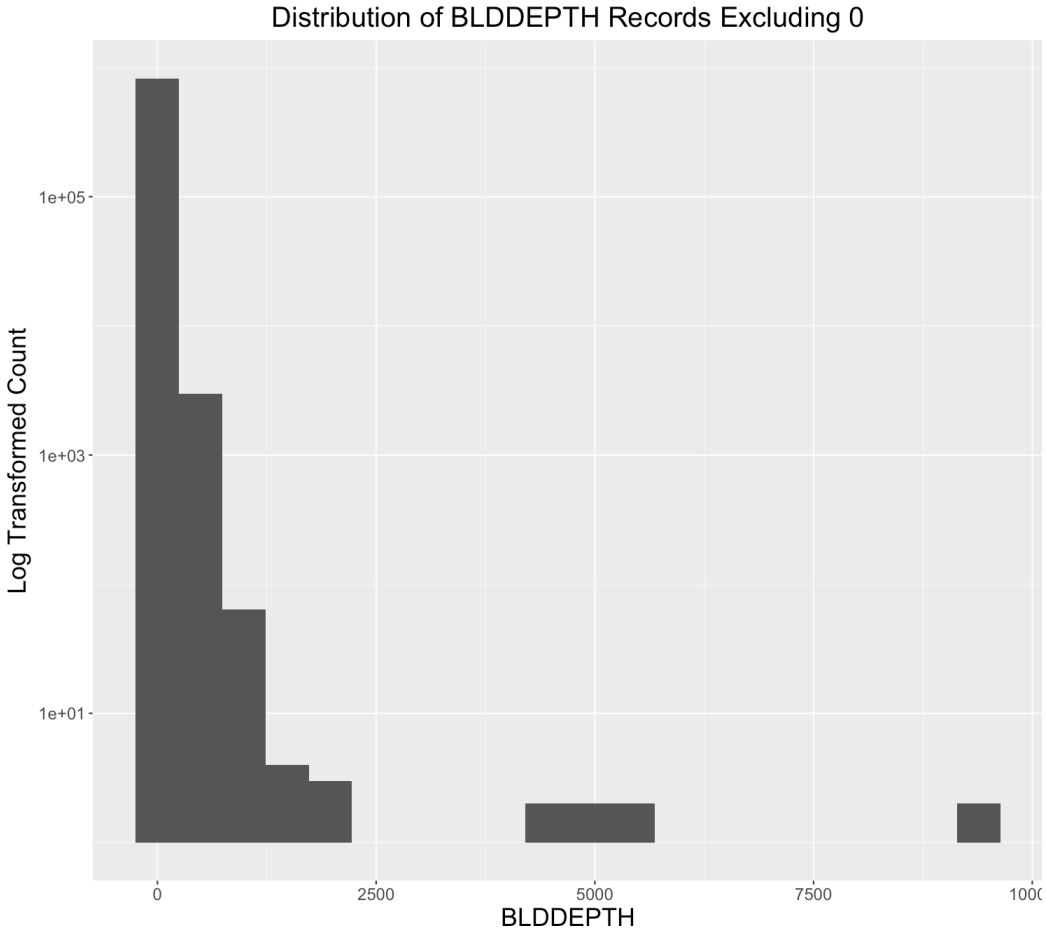
Field 22

Field Name: BLDDEPTH

Description:
BLDDEPTH is a numeric variable representing the length of building depth in feet.

Unique Values:
BLDDEPTH has 620 unique values ranging from 0 to 9393. No missing values exist. However, there are 224,699 records with value 0, which could be in fact missing values. The statistics and distribution excluding all records with 0 BLDFRONT are shown as below.

Minimum	1
Maximum	9393
Median	44
Mean	51.00
Mode	40
SD	42.42



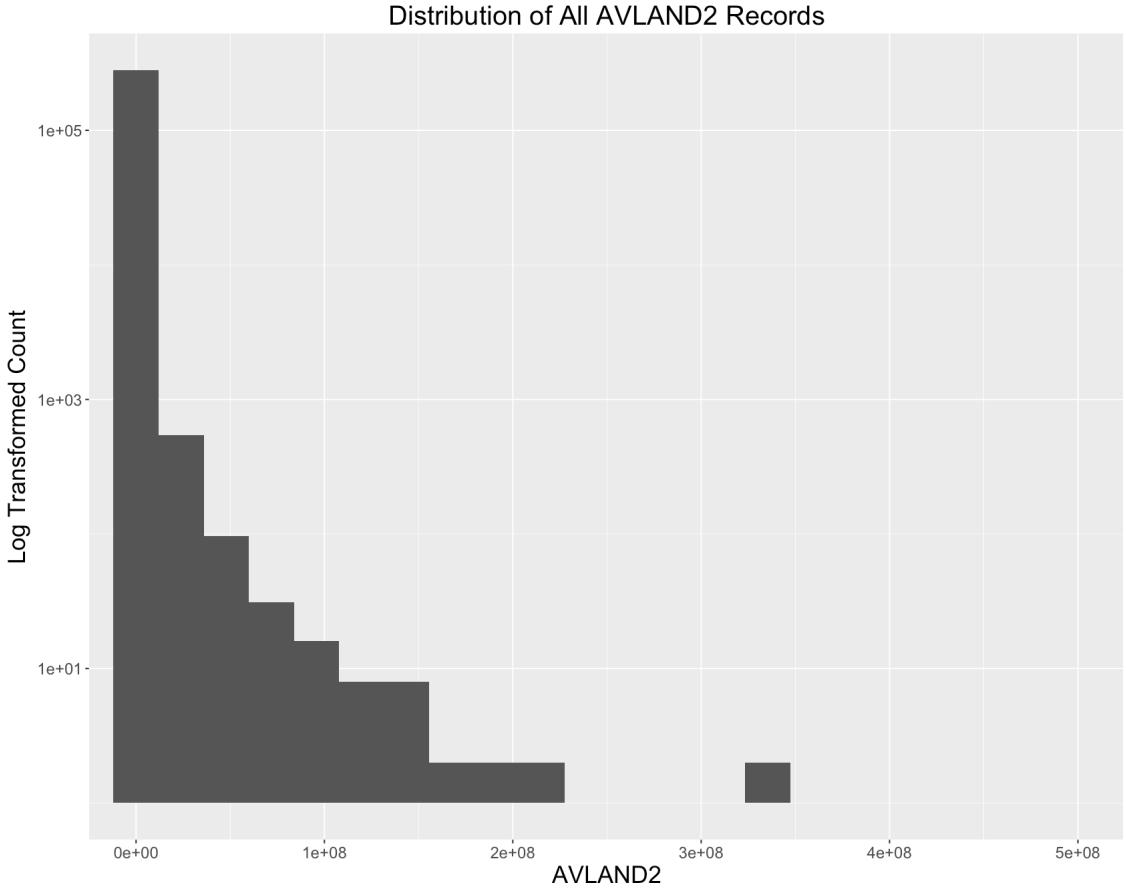
Field 23

Field Name: AVLAND2

Description:
AVLAND2 is a numeric variable representing the assessed value of the land. It could be the updated assessed value compared to AVLAND. Most values of AVLAND2 are lower than their corresponding values of AVLAND.

Unique Values:
AVLAND has 58,170 unique values ranging from 3 to about 2,300,000,000. There are 767,609 records of missing values in the AVLAND2 field. The statistics is shown as below.

Minimum	3
Maximum	2.37E+09
Median	20059
Mean	2.46E+05
Mode	2408
SD	6.20E+06



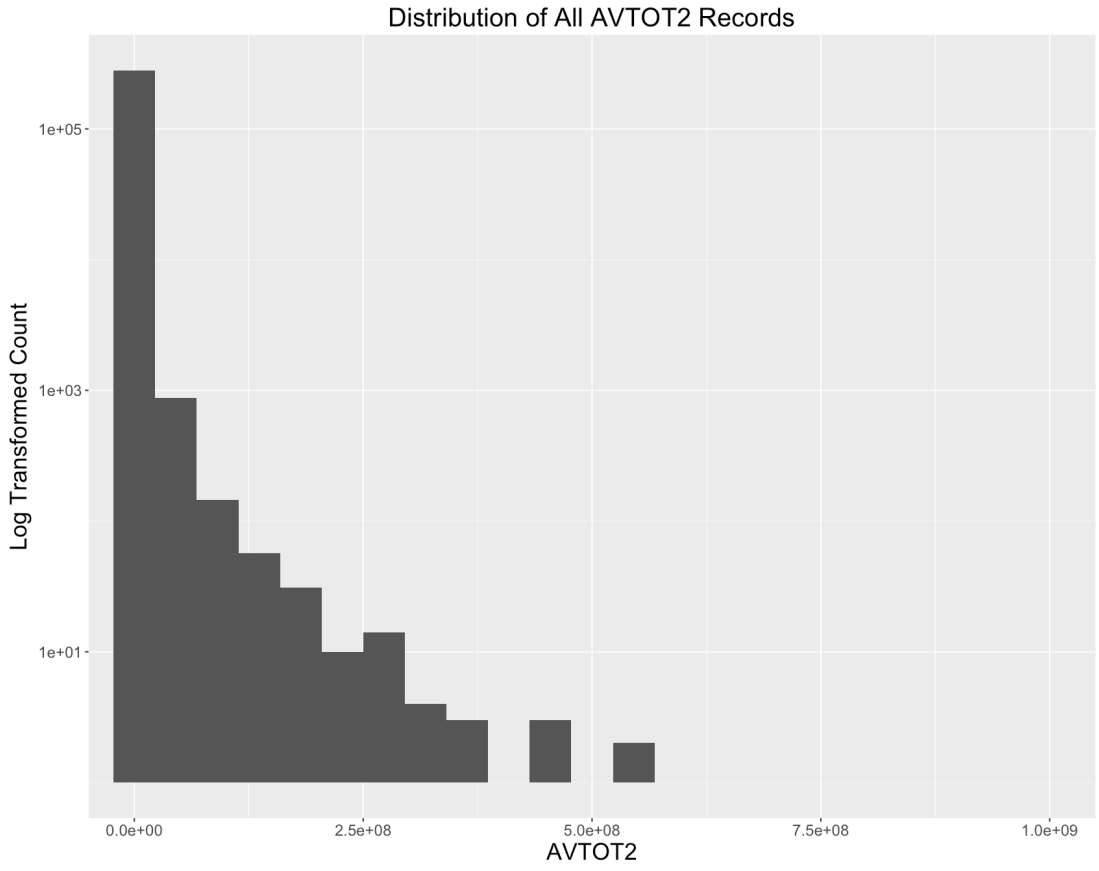
Field 24

Field Name: AVTOT2

Description:
AVTOT2 is a numeric variable representing the assessed total value of the property. It could be the updated assessed value compared to AVTOT. Most AVTOT2 values are smaller than or equal to their corresponding AVTOT value.

Unique Values:
AVTOT2 has 110,891 unique values ranging from 3 to about 4,500,000,000. There are 767,603 missing values in the AVTOT2 field. The statistics is shown as below.

Minimum	3
Maximum	4.50E+09
Median	80010
Mean	7.16E+05
Mode	750
SD	1.17E+07



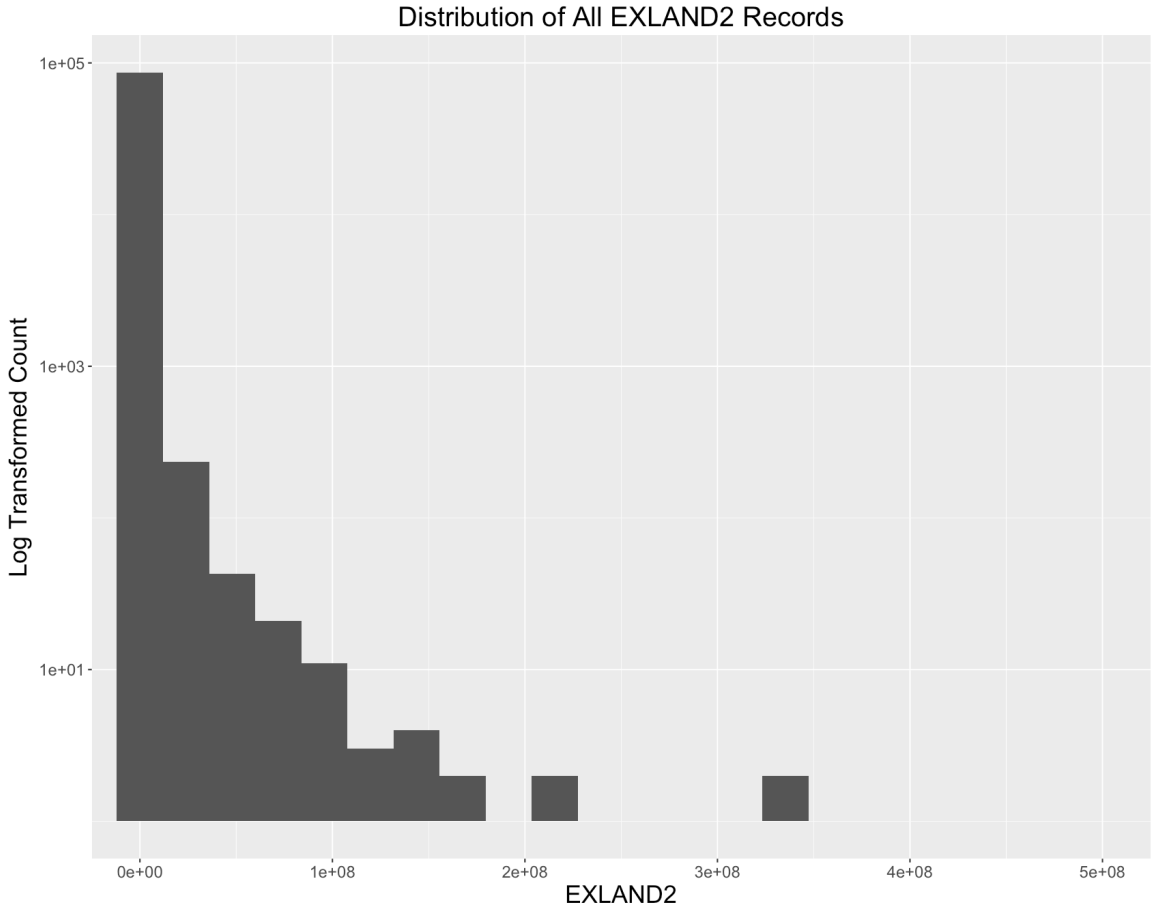
Field 25

Field Name: EXLAND2

Description:
EXLAND2 is a numeric variable representing the value of the exempt land. It could be the updated assessed value compared to EXLAND. Most EXLAND2 values are lower than their corresponding EXLAND values.

Unique Values:
EXLAND2 has 21,997 unique values ranging from 7 to about 2,400,000,000. There are 961,900 missing values in the EXLAND2 field. The statistics is shown as below.

Minimum	7
Maximum	4.50E+09
Median	37116
Mean	6.58E+05
Mode	2090
SD	1.61E+07



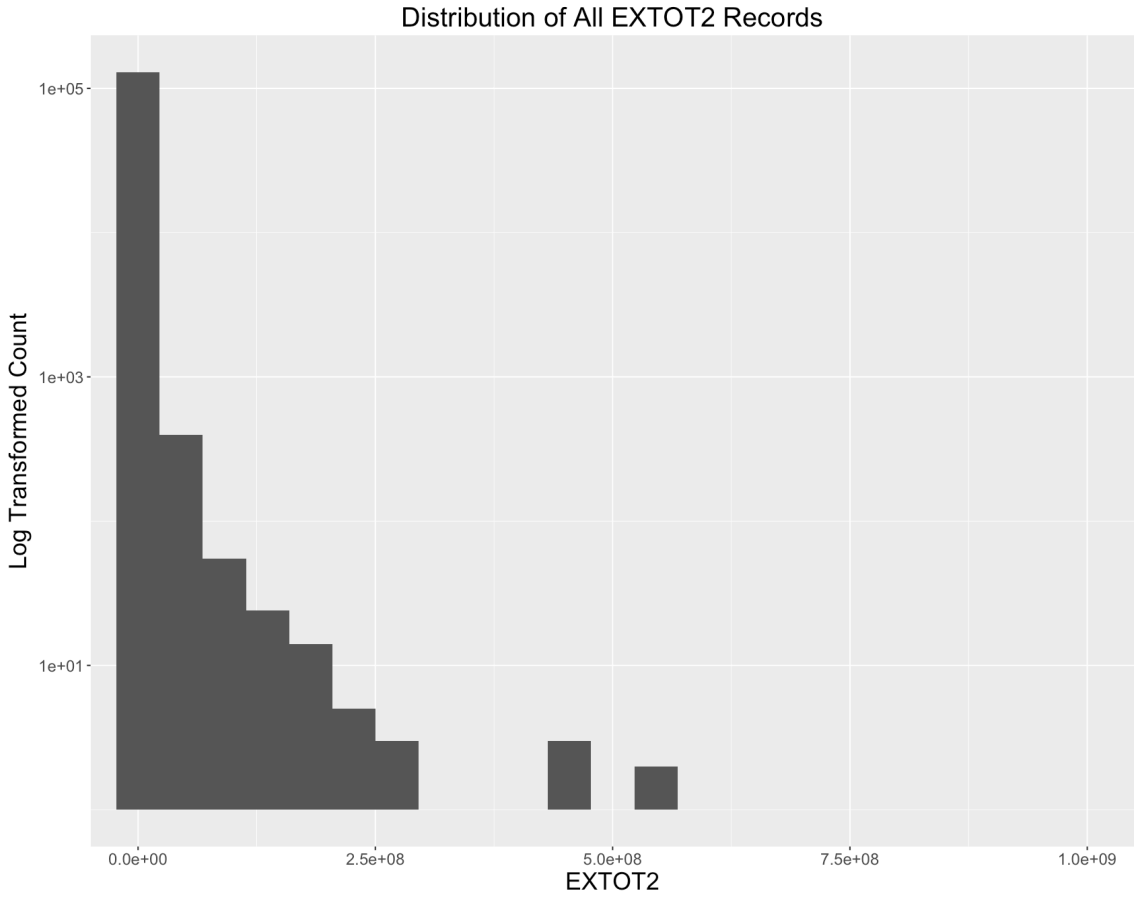
Field 26

Field Name: EXTOT2

Description:
EXTOT2 is a numeric variable representing the total value of the exempt property. It could be the updated assessed value compared to EXTOT. Most EXTOT2 values are lower than their corresponding EXTOT values.

Unique values:
EXTOT2 has 48107 unique values ranging from 7 to about 4,500,000,000. There are 918,642 missing values in the EXTOT2 field. The statistics is shown as below.

Minimum	7
Maximum	4.50E+09
Median	37116
Mean	6.58E+05
Mode	2090
SD	1.61E+07



Field 27

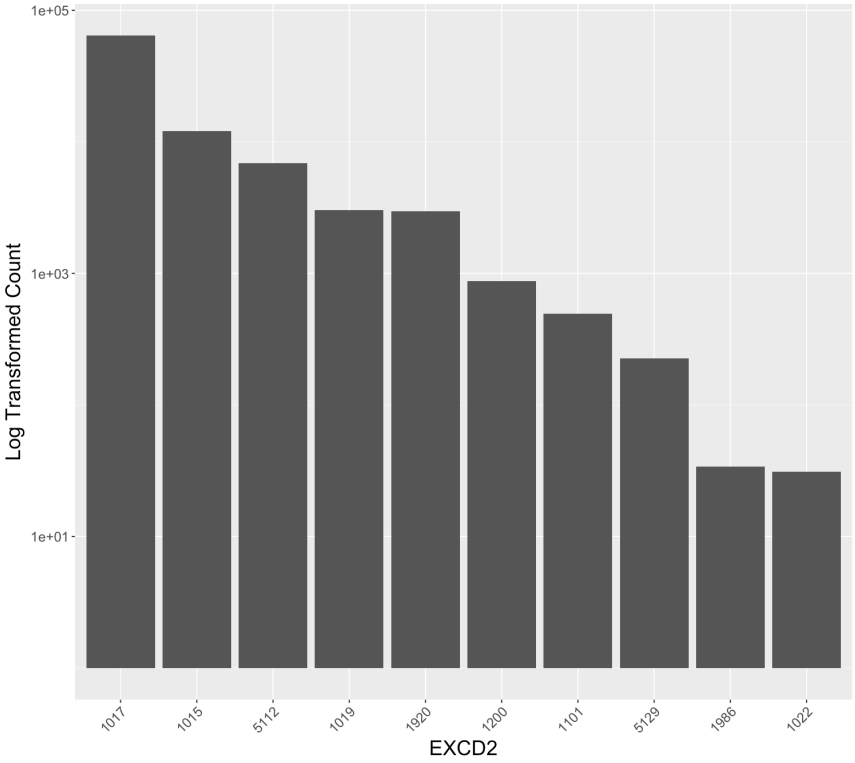
Field Name: EXCD2

Description:

EXCD2 is a categorical variable, possibly representing the code for the exempt reasons for EXLAND2 AND EXTOT2 records.

Unique Values:

EXTOT has 61 levels, taking 4-digit numbers from 1011 to 7160. There are 957,634 missing values exist. The top 10 most frequently occurred EXCD2 values are:



EXCD2	Percentage(%)
1017	70.6
1015	13.2
5112	7.6
1019	3.3
1920	3.3
1200	1.0
1101	0.5
5129	0.2
1986	0.0
1022	0.0

Field 28
Field Name: PERIOD

Description:
PERIOD is a categorical variable, indicating the change period of the record.

Unique Values:
All the records in this dataset take the value of "FINAL" in the PERIOD field.

Field 29
Field Name: YEAR

Description:
YEAR is a date variable, indicating the time that the record is made.

Unique Values:
All the records in this dataset take the value of "2010/11" in the YEAR field.

Field 30
Field Name: VALTYPE

Description:
VALTYPE is a categorical variable, indicating the valid type of the record.

Unique Values:
All the records in this dataset take the value of "AC-TR" in the VALTYPE field.