**Data Quality Report**

**1. Introduction**

The New York City department of Finance values properties in NYC every year to calculate the property tax. This report provides property tax data such as market and assessed values, exemptions, and abatements from the assessment year 2010/11. The information is listed by categories, such as borough, tax class, and type of building. There are 1048575 records, 30 columns, which includes 14 categorical variables and 16 numerical variables. The table below shows some basic information of this dataset, like data type, mean, standard deviation, maximum and minimum.

**2. Dataset Description**

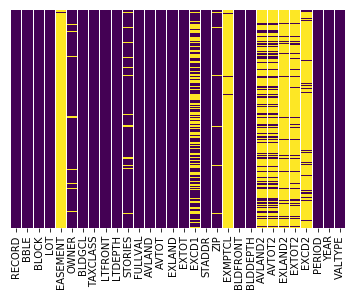
**Acronyms:**

|  |  |
| --- | --- |
| **Abbreviation** | **Description** |
| LTFRONT | Lot frontage in feet |
| LTDEPTH | Lot depth in feet |
| FULLVAL | Total market value of property |
| AVLAND | Market value of the land |
| AVTOT | Total market value |
| EXLAND | Exempt land value |
| EXTOT | Exempt total value |
| EXCD1 | Exempt condo value |
| BLDFRONT | Building frontage in feet |
| BLDDEPTH | Building depth in feet |
| AVLAND2 | 2nd Market value of the land |
| AVTOT2 | 2nd Total market value |
| EXLAND2 | 2nd Exempt land value |
| EXTOT2 | 2nd Exempt total value |
| EXCD2 | 2nd Exempt condo value |
| BLDGCL | Building class |

**Summary table:**



**Heat map of missing values in the dataset:**



**3. Numerical Data Analysis**

**BLOCK – Block #**

As the block id in different cities may be same, it cannot effectively show the real number of properties in a specific block.

Bock # to area mapping:

Manhattan - 1 to 2,255

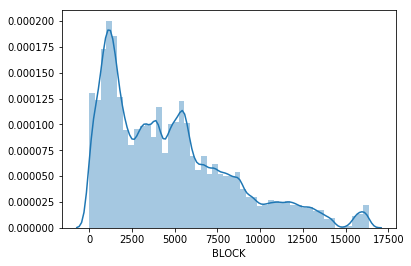
Bronx - 2,260 to 5,958

Brooklyn - 1 to 8,955

Queens - 1 to 16,350

Staten Island - 1 to 8,050

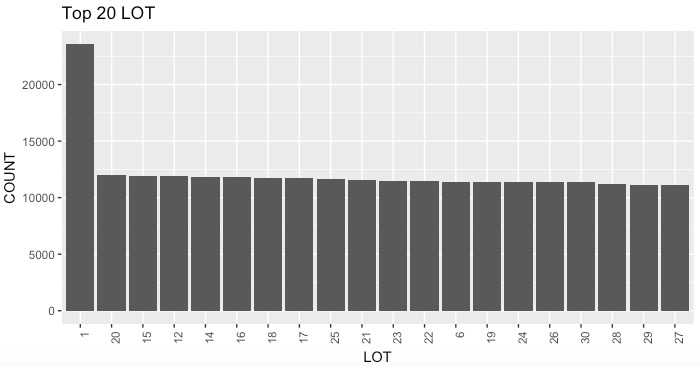


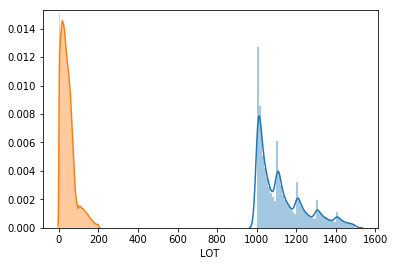
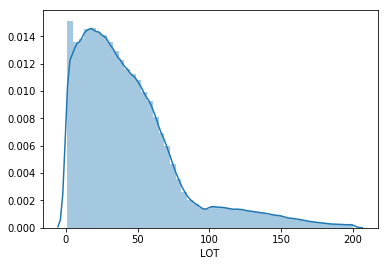


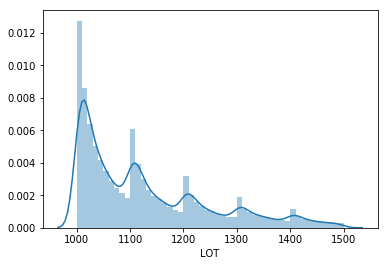
**LOT – Lot # within Block**

Every record has a lot id but some of them share the same lot id. Value 1 has the highest frequency in this dataset, but lot 1 may not be the lot with highest number of properties as lot id is only unique within block.



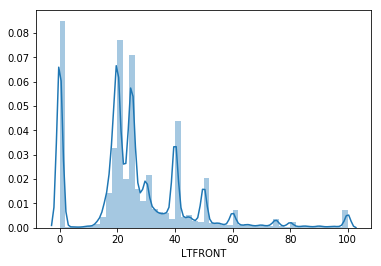
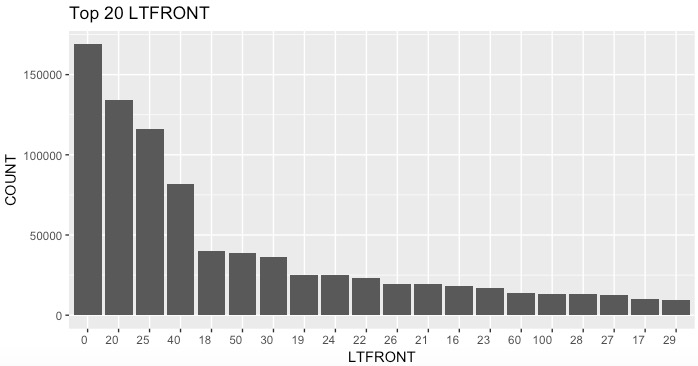






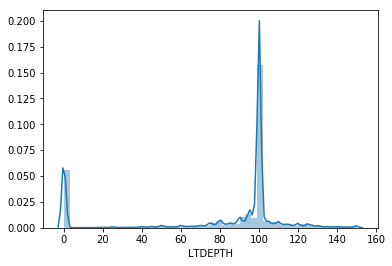
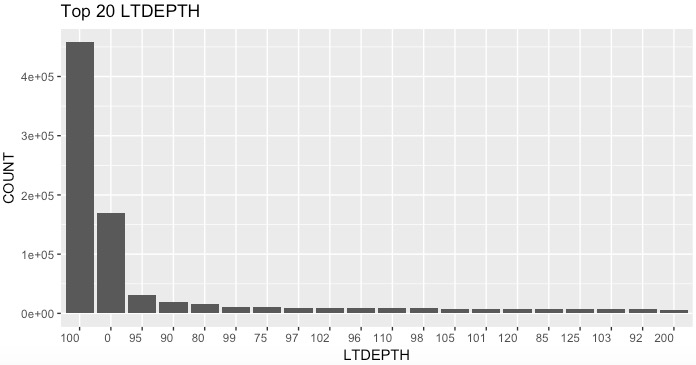
**LTFRONT – Lot Width**





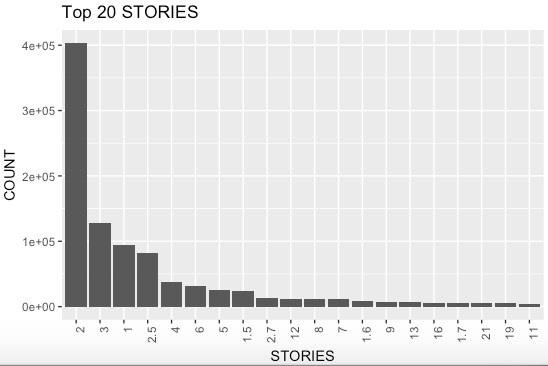
**LTDEPTH – Lot Depth**

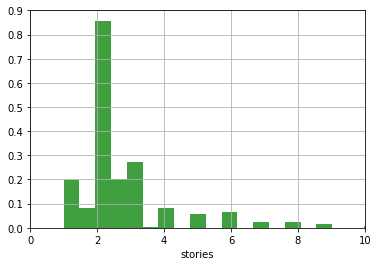
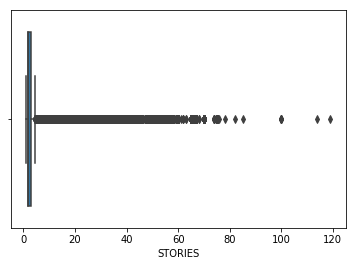


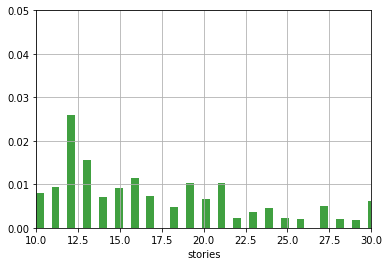
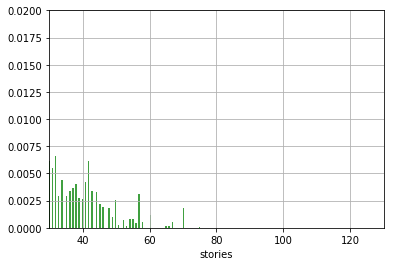


**STORIES – Number of stories in the building**



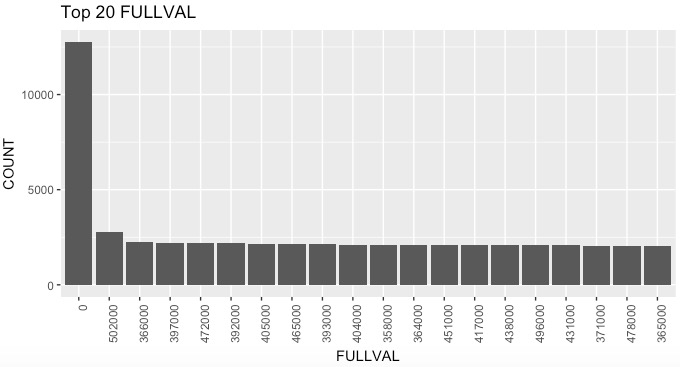


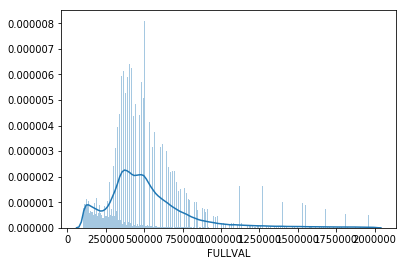




**FULLVAL – Market Value**

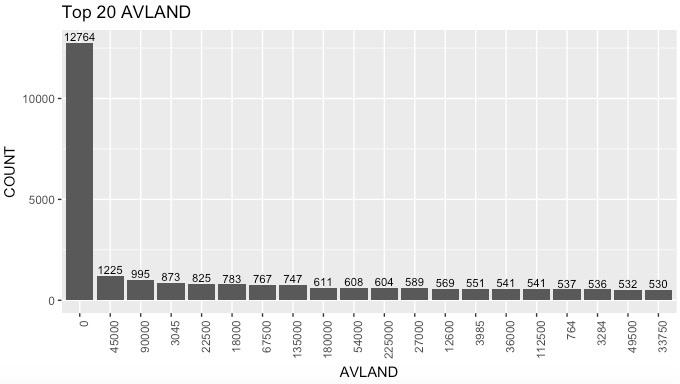


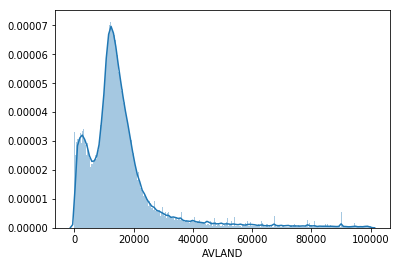




**AVLAND – Actual Land Value**

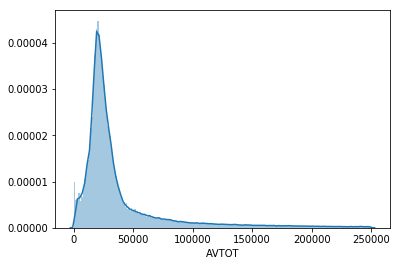
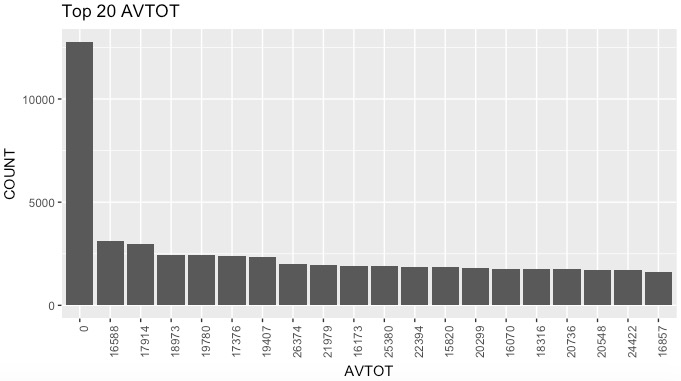






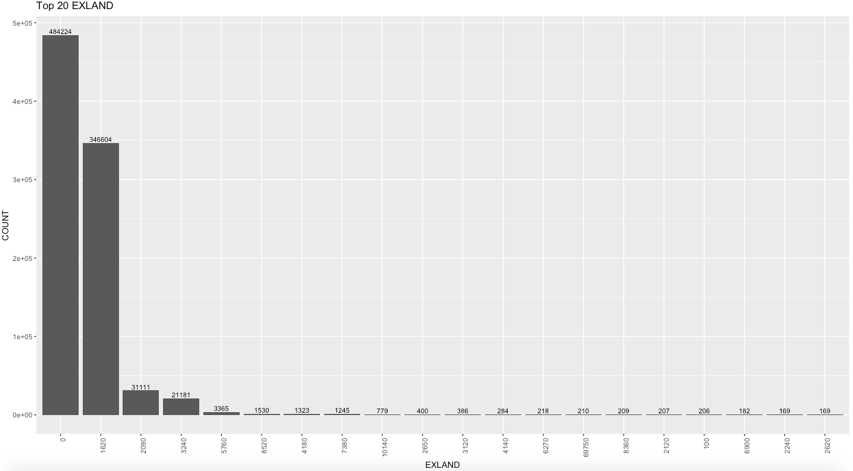
**AVTOT – Actual Total Value**

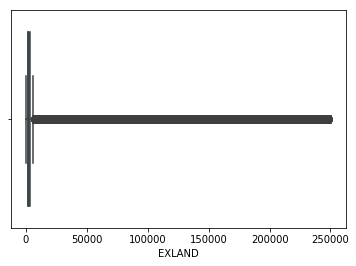


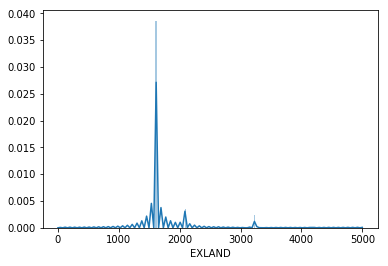


**EXLAND – Actual Exempt Land Value**



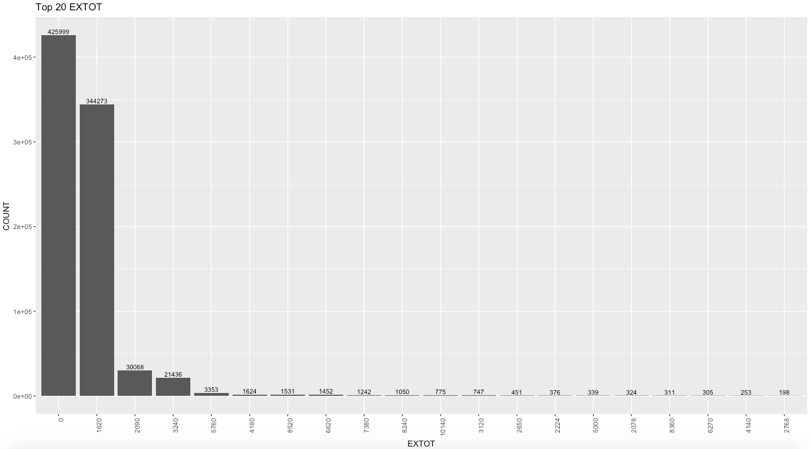


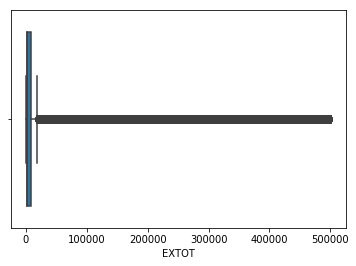
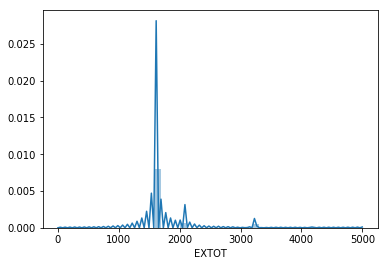




**EXTOT – Actual Exempt Land Total**

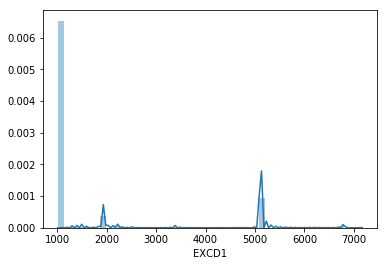
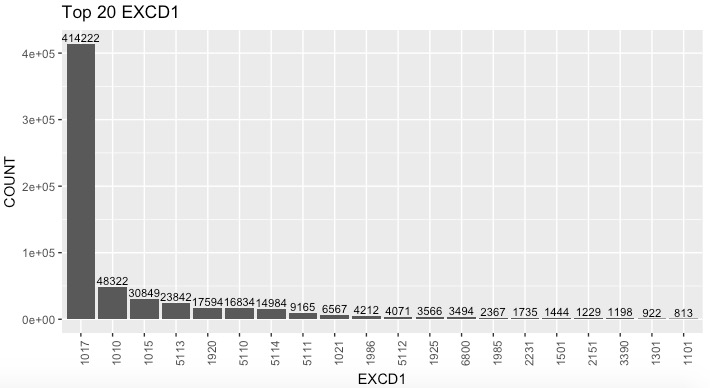




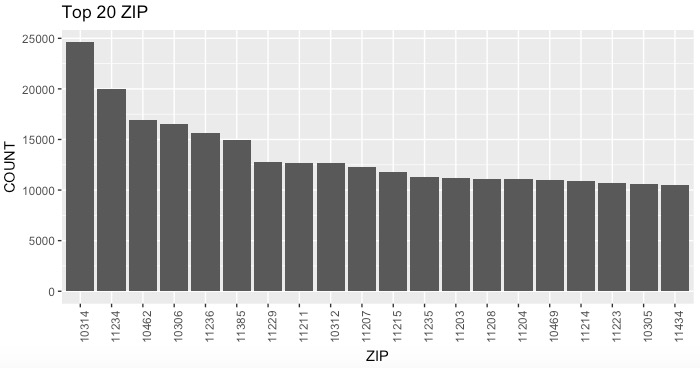
**EXCD1 – Exemption Code 1**

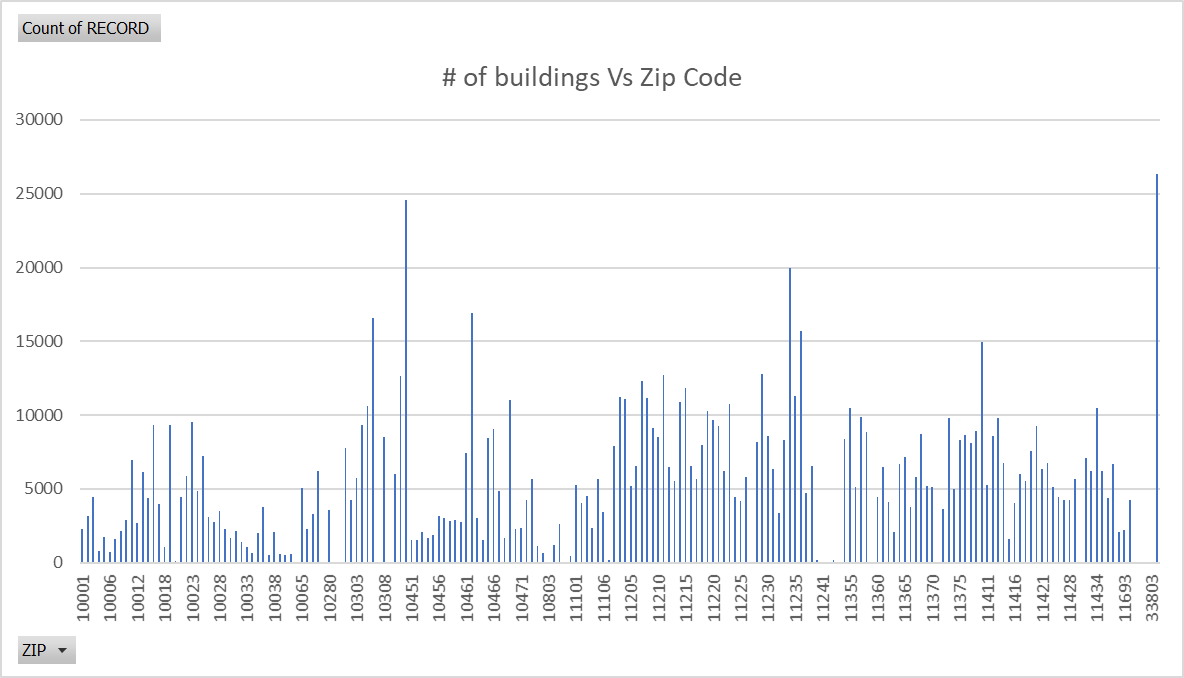




**ZIP – Zip Code**

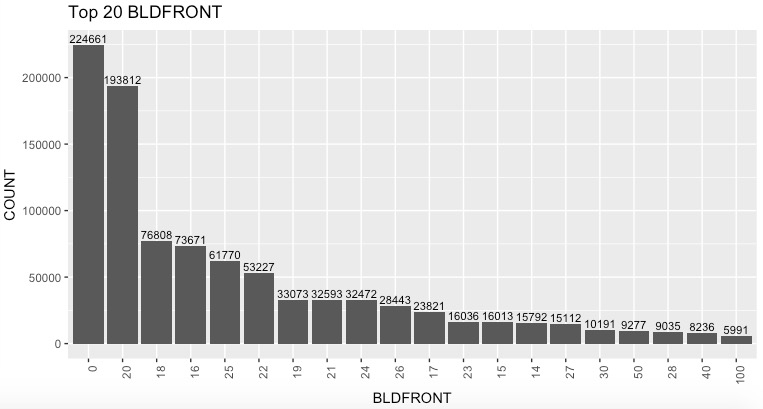


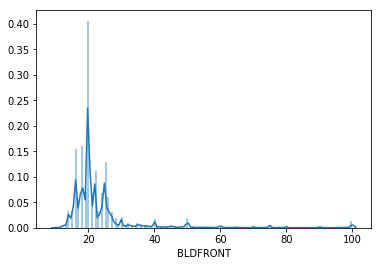




**BLDFRONT – Building Width**

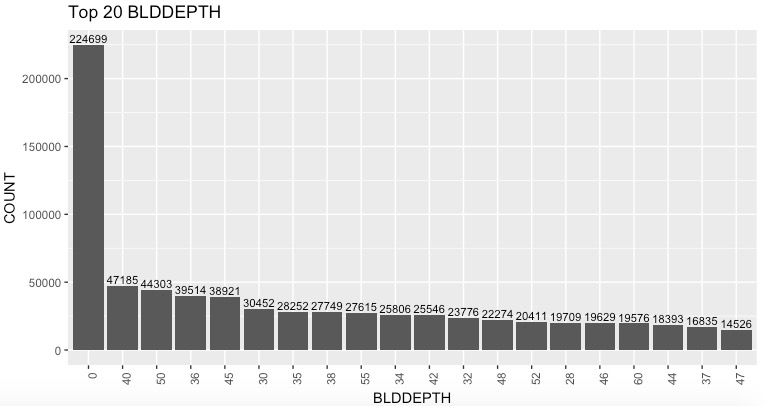


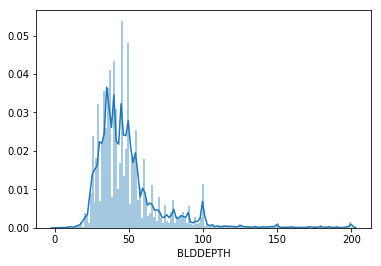




**BLDDEPTH – Building Depth**

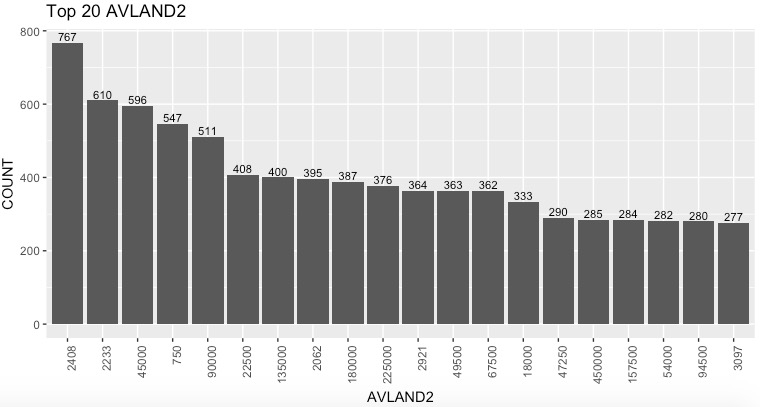


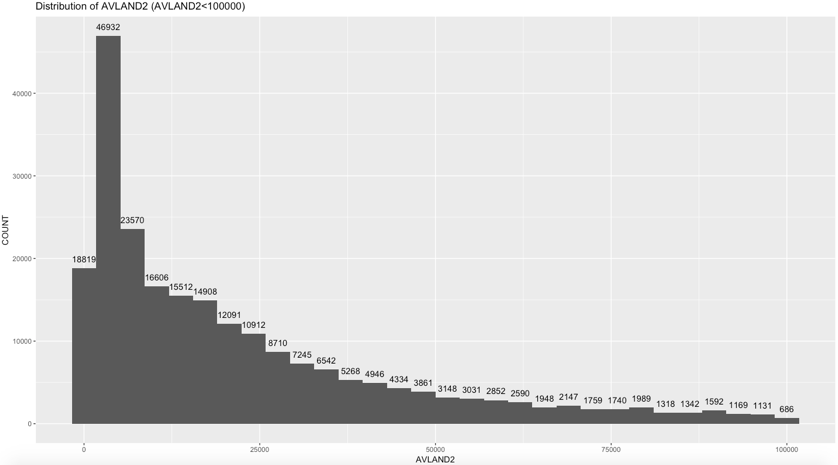




**AVLAND2 – Transitional Land Value**

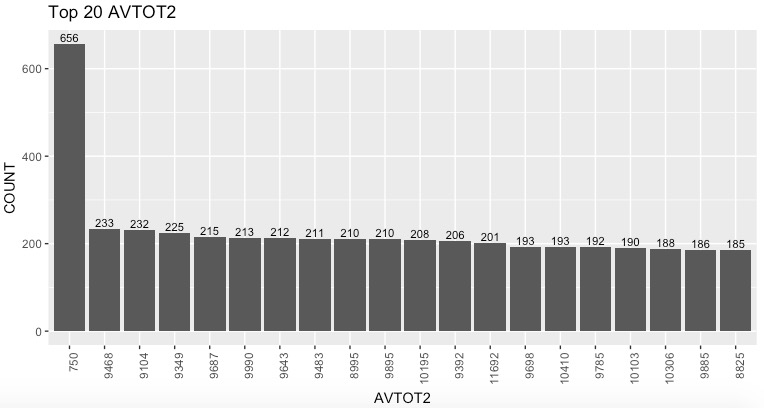


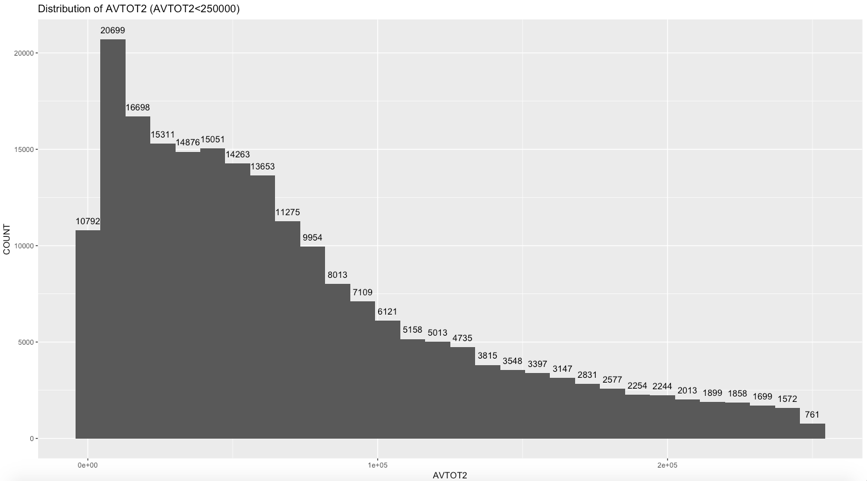




**AVTOT2 – Transitional Total Value**

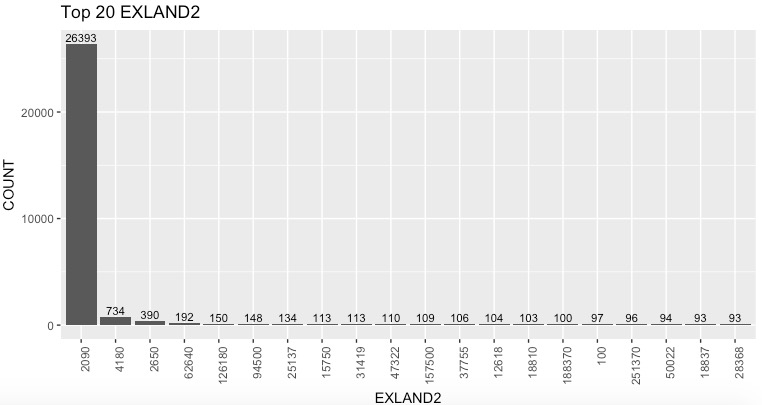


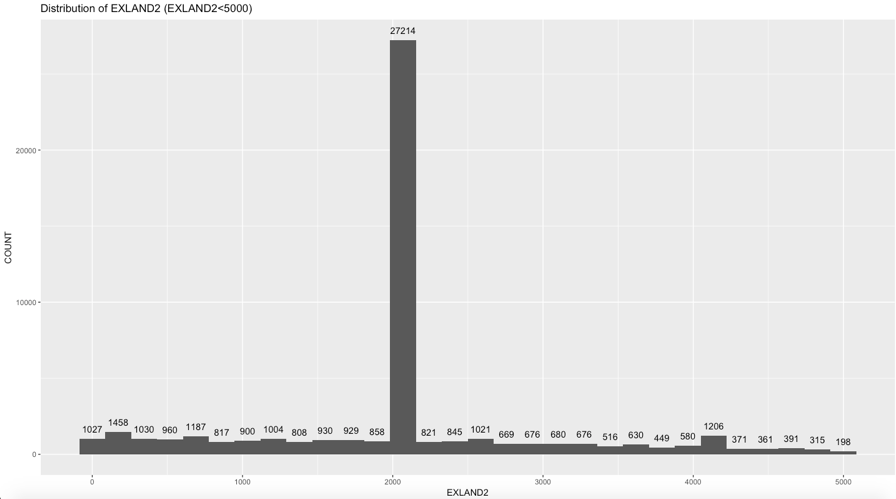




**EXLAND2 – Transitional Exempt Land Value**

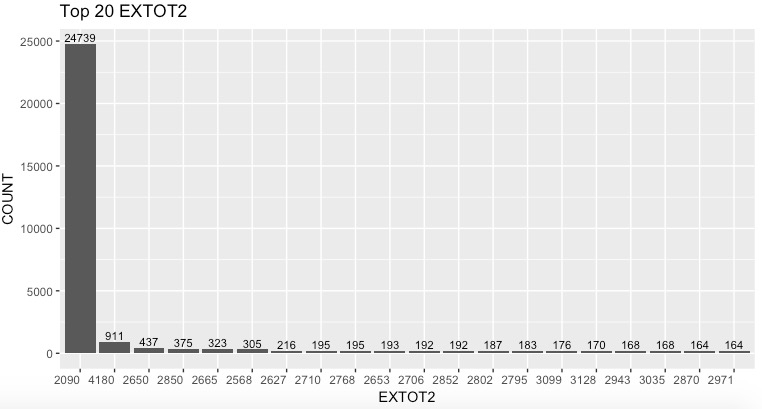


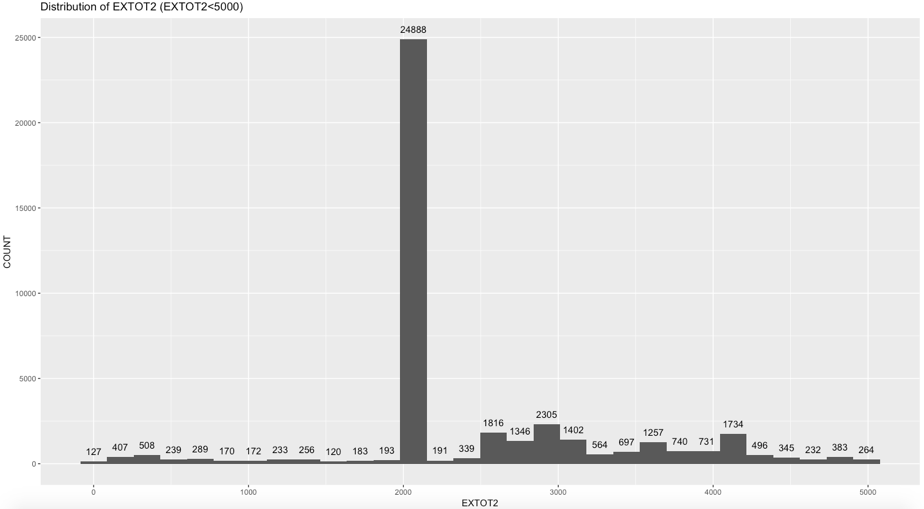




**EXTOT2 – Transitional Exempt Land Total**

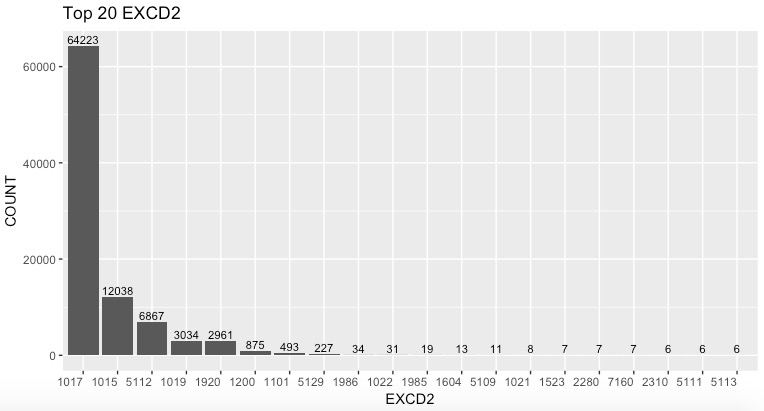


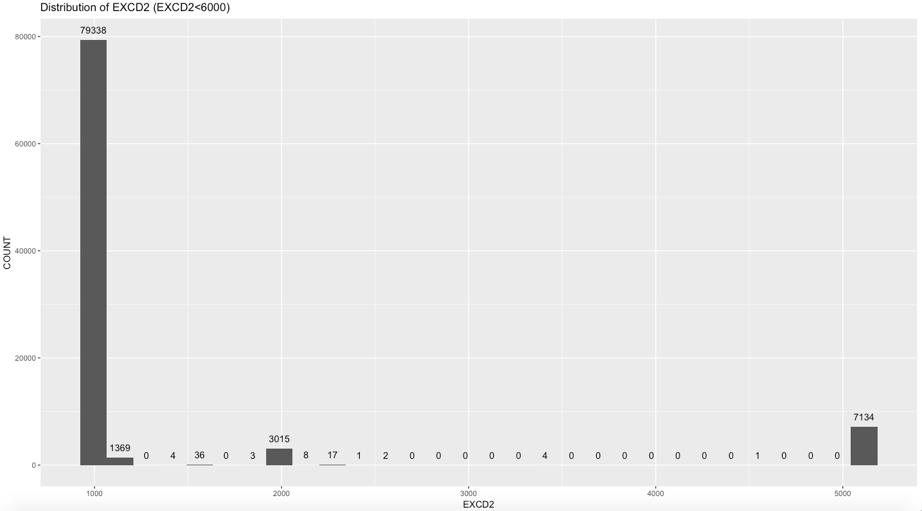




**EXCD2 – Exemption Code 2**



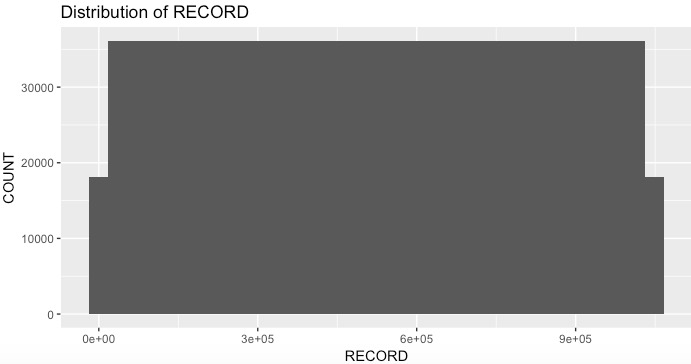




**4. Categorical Data Analysis**

**RECORD – Record ID**

There are 1048575 records in this dataset, so the record id varies from 1 to 1048575.



**BBLE - Concatenation of AV\_BORO, AV\_BLOCK, AV\_LOT, AV\_EASEMENT**

There are 1048575 different records in this dataset, which means every record have a unique BBLE.

**EASEMENT – Easement Description**

SPACE Indicates the lot has no Easement;

'A' Indicates the portion of the Lot that has an Air Easement;

'B' Indicates Non-Air Rights;

'E' Indicates the portion of the lot that has a Land Easement;

'F' THRU 'M' are duplicates of 'E';

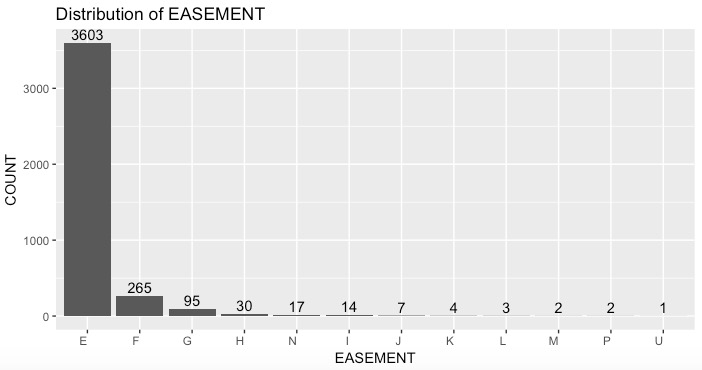
'N' Indicates Non-Transit Easement;

'P' Indicates Piers;

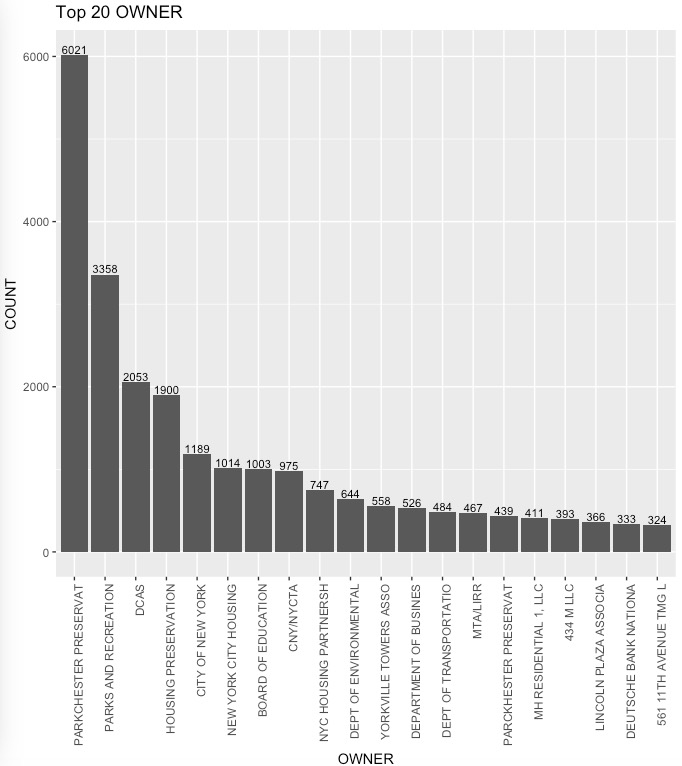
'R' Indicates Railroads;

'S' Indicates Street;

'U' Indicates U.S. Government;



**OWNER – Owner of property**



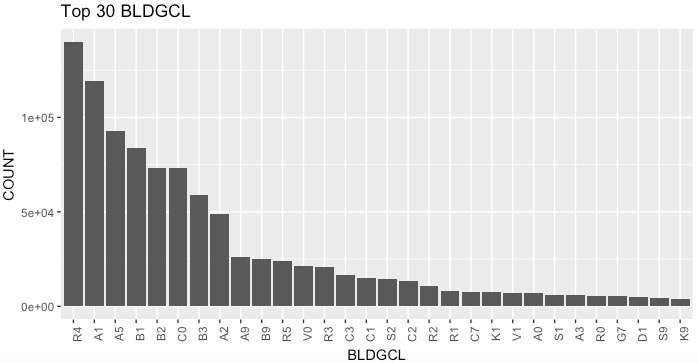
**Class A** represents the highest quality buildings in their market;

Buildings of **Class B** are generally a little older, but still have good quality management and tenants;

Buildings of **Class C** are older buildings (usually more than 20 years), and are located in less desirable areas and are in need of extensive renovation.

**BLDGCL – Building Class**





TAX CLASS 1 = 1-3 UNIT RESIDENCES;

TAX CLASS 1A = 1-3 STORY CONDOMINIUMS

ORIGINALLY A CONDO;

TAX CLASS 1B = RESIDENTIAL VACANT LAND;

TAX CLASS 1C = 1-3 UNIT CONDOMINUMS

ORIGINALLY TAX CLASS 1;

TAX CLASS 1D = SELECT BUNGALOW COLONIES;

TAX CLASS 2 = APARTMENTS;

TAX CLASS 2A = APARTMENTS WITH 4-6 UNITS;

TAX CLASS 2B = APARTMENTS WITH 7-10 UNITS;

TAX CLASS 2C = COOPS/CONDOS WITH 2-10 UNITS;

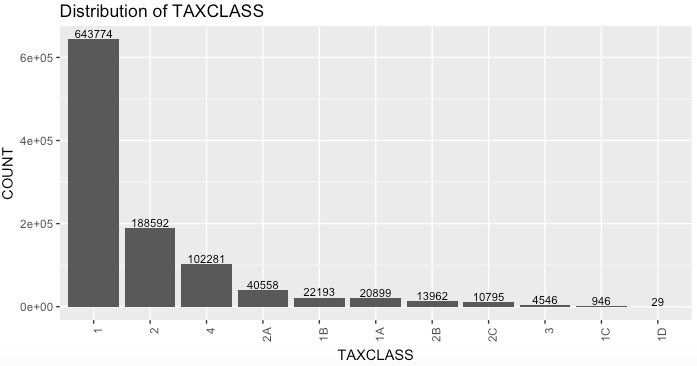
TAX CLASS 3 = UTILITIES (EXCEPT CEILING RR);

TAX CLASS 4A = UTILITIES - CEILING RAILROADS;

TAX CLASS 4 = ALL OTHERS

**TAXCLASS – Tax Class**

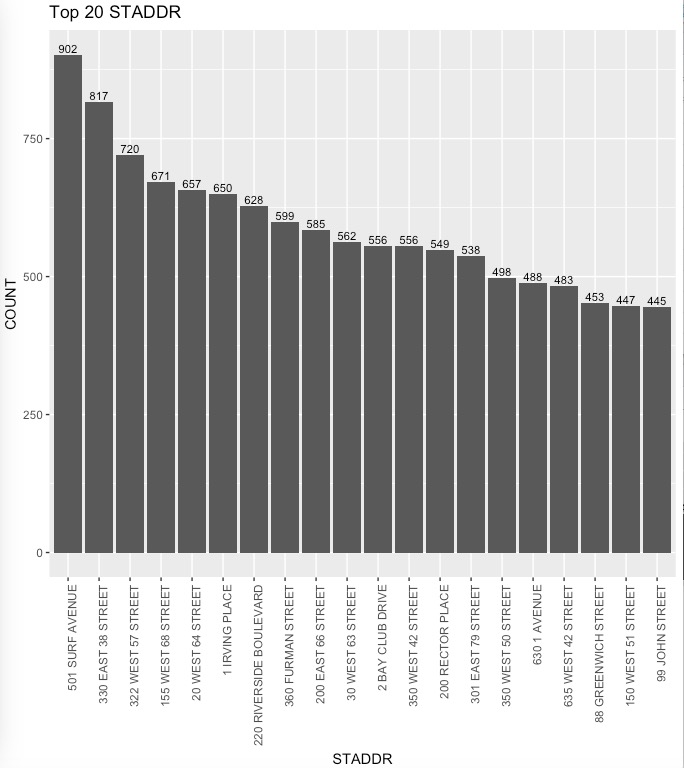




**STADDR – Street Address**

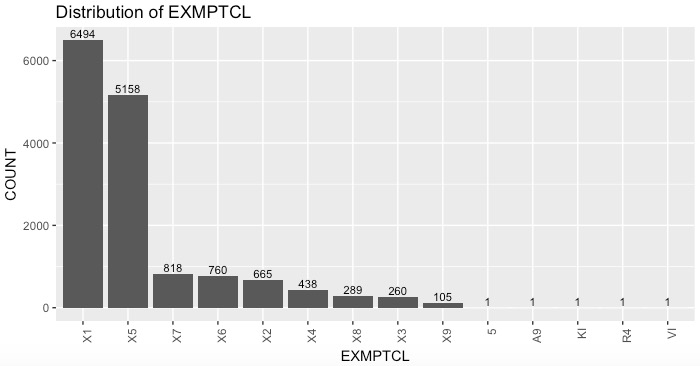
Too many unique value therefore not visualizing this categorical variable. A better way to visualize will be to put it on google map and create heat map visualization to see which areas are more poplar.





**EXMPTCL – Exempt Class**





**PERIOD – Assessment Period**



Single variable – Final.

**YEAR – Assessment Year**



Single Variable – 2010/11

**VALTYPE**



Single Variable – AC-TR