

Report Code

This report contains all the code and visualizations from the original report.

```
library(dplyr)

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

df <- read.csv("C:/Users/jeffb/OneDrive/Desktop/WSU-Spring-2022/Big Data Analytics/data-capstone-final-project.csv")
head(df)

##   Date.Created Borough Map.Atlas Block Lot      Address Street.Number
## 1 10/10/2019        1     NA    2    1 4 SOUTH STREET                 4
## 2 10/10/2019        1     NA    2    2 10 SOUTH STREET                10
## 3 10/10/2019        1     NA    2    3 MARGINAL STREET
## 4 10/10/2019        1     NA    2   23      1 PIER 6                  1
## 5 10/10/2019        1     NA    3    1 10 BATTERY PARK                10
## 6 10/10/2019        1     NA    3    2 PETER MINUIT PLAZA
## 
##   Street.Name          Parcel.Name Agency
## 1 SOUTH STREET          SI FERRY TERMINAL DOT;DSBS
## 2 SOUTH STREET          BATTERY MARITIME BLDG      DSBS
## 3 MARGINAL STREET       MTA SUBSTATION        DSBS
## 4 PIER 6                  PIER 6            DSBS
## 5 BATTERY PARK          BATTERY PARK        PARKS
## 6 PETER MINUIT PLAZA  PETER MINUIT PLAZA/BATTERY PK  PARKS
## 
##   Current.Uses
## 1 FERRY TERMINAL;NO USE;WATERFRONT PROPERTY
## 2 IN USE-TENANTED;LONG-TERM AGREEMENT;WATERFRONT PROPERTY
## 3 NO USE-NON RES STRC;TRANSIT SUBSTATION
## 4 IN USE-TENANTED;FINAL COMMITMNT-DISP;LONG-TERM AGREEMENT;NO USE;FINAL COMMITMNT-DISP
## 5                                         PARK
## 6                                         PARK
## 
##   Total.Area Open.Petroleum.Spill Govt.Clean.Up.Program Structure.Completed
## 1      209215                               1908
## 2      191502                               1900
## 3      38800                                1900
## 4      510025                               1987
```

## 5	945425				1951	
## 6	39900				0	
##	Number.Structures	Total.Gross.Area.Structures	Ratio.Building.to.Floor.Area			
## 1	2	280000			1.34	
## 2	1	158197			0.83	
## 3	1	7500			0.19	
## 4	1	24346			0.05	
## 5	6	945425			1.00	
## 6	3	0			0.00	
##	Allowable.Building.to.Floor.Area	Land.Use.Category	Community.Board			
## 1	6.5	7	101			
## 2	10.0	7	101			
## 3	10.0	7	101			
## 4	10.0	7	101			
## 5	0.0	9	101			
## 6	0.0	9	101			
##	Census.Ttract	Census.Block	School.Dist	Council.District	Postcode	Fire.Comp
## 1	9	1025	2	1	10004	L015
## 2	9	1025	2	1	10004	L015
## 3	9	1	NA	1	NA	
## 4	9	1025	2	1	10004	L015
## 5	13	2005	2	1	10004	E010
## 6	9	1023	2	1	10004	L015
##	Health.Area	Health.Ctr	Police.Prct	Major.Use	Number.of.Easements	
## 1	7700	15	1	Y7	0	
## 2	7700	15	1	Y7	0	
## 3	NA	NA	NA	Y7	0	
## 4	7700	15	1	T2	0	
## 5	7700	15	1	Q1	0	
## 6	7700	15	1	Q1	0	
##	Commercial.Floor.Area	Residential.Floor.Area	Office.Floor.Area			
## 1	280000	0	0			
## 2	158197	0	0			
## 3	7500	0	0			
## 4	24346	0	0			
## 5	945425	0	0			
## 6	0	0	0			
##	Retail.Floor.Area	Garage.Floor.Area	Storage.Floor.Area	Factory.Floor.Area		
## 1	0	0	0	0		
## 2	0	0	0	0		
## 3	0	0	0	0		
## 4	0	0	0	0		
## 5	0	0	0	0		
## 6	0	0	0	0		
##	Other.Floor.Area	Num.Floors	Residential.Units			
## 1	280000	5	0			
## 2	158197	5	0			
## 3	7500	1	0			
## 4	24346	2	0			
## 5	945425	1	0			
## 6	0	0	0			
##	Residential.and.Non.Residential.Units	Lot.Front	Lot.Depth	Bldg.Front		
## 1		1	314.28	564.84	268	
## 2		1	311.61	555.84	205	

```

## 3 0 91.90 564.56 50
## 4 0 793.67 551.00 85
## 5 4 323.00 1260.00 89
## 6 0 496.00 76.00 0
##   Bldg.Depth Proximity.Code Irr.Lot.Code Lot.Type.Code Bsmt.Code Assess.Land
## 1    210          1           Y          2          0 18818100
## 2    270          0           Y          2          0 18787500
## 3    75           0           N          2          5 735750
## 4    551          0           Y          5          0 15255450
## 5    57           1           Y          1          0 127100700
## 6    0            0           Y          1          0 5060250
##   Exempt.Land Exempt.Tot Year.Alter.1 Year.Alter.2 His.Dist      Landmark
## 1        NA 35179200       2006       2006
## 2        NA 26151750       2001          0      INDIVIDUAL LANDMARK
## 3        NA 898200          0          0
## 4        NA 18723600          0          0
## 5        NA 141688800         0          0      INDIVIDUAL LANDMARK
## 6        NA 5060250          0          0
##   Condominium.Number Coordinates E.Designation.Number
## 1              0 980670/194440
## 2              0 981037/194506
## 3              0 981224/194590
## 4              0 981687/194732
## 5              0 9799916/195328
## 6              0 980470/194954
##   Industrial.Business.Zone Zone.Dist.1 Zone.Dist.2 Overlay.1 Overlay.2
## 1                  M1-4
## 2                  C4-6
## 3                  C4-6
## 4                  C4-6
## 5                  PARK
## 6                  PARK      C5-5
##   SP.Dist.1 SP.Dist.2 Potential.Urbn.Ag
## 1        LM
## 2        LM
## 3        LM
## 4        LM
## 5
## 6        LM
##   Contact EDC...Occupied
## 1 Saundra Malanowicz (212)386-0612 smalanowicz@dcas.nyc.gov 100
## 2 Saundra Malanowicz (212)386-0612 smalanowicz@dcas.nyc.gov 100
## 3 Saundra Malanowicz (212)386-0612 smalanowicz@dcas.nyc.gov NA
## 4 Saundra Malanowicz (212)386-0612 smalanowicz@dcas.nyc.gov 100
## 5 Saundra Malanowicz (212)386-0612 smalanowicz@dcas.nyc.gov NA
## 6 Saundra Malanowicz (212)386-0612 smalanowicz@dcas.nyc.gov NA
##   Pluto.Version Latitude Longitude     BIN        BBL
## 1    19v1 40.70037 -74.01291 1085792 1000020001
## 2    19v1 40.70055 -74.01159 1000003 1000020002
## 3    19v1 40.70078 -74.01091 1089745 1000020003
## 4    19v1 40.70117 -74.00924 1075697 1000020023
## 5    19v1 40.70281 -74.01563 1077335 1000030001
## 6    19v1 40.70178 -74.01363 1089623 1000030002
##   NTA

```

```

## 1 Battery Park City-Lower Manhattan
## 2 Battery Park City-Lower Manhattan
## 3 Battery Park City-Lower Manhattan
## 4 Battery Park City-Lower Manhattan
## 5 Battery Park City-Lower Manhattan
## 6 Battery Park City-Lower Manhattan

names(df)

## [1] "Date.Created"
## [2] "Borough"
## [3] "Map.Atlas"
## [4] "Block"
## [5] "Lot"
## [6] "Address"
## [7] "Street.Number"
## [8] "Street.Name"
## [9] "Parcel.Name"
## [10] "Agency"
## [11] "Current.Uses"
## [12] "Total.Area"
## [13] "Open.Petroleum.Spill"
## [14] "Govt.Clean.Up.Program"
## [15] "Structure.Completed"
## [16] "Number.Structures"
## [17] "Total.Gross.Area.Structures"
## [18] "Ratio.Building.to.Floor.Area"
## [19] "Allowable.Building.to.Floor.Area"
## [20] "Land.Use.Category"
## [21] "Community.Board"
## [22] "Census.Ttract"
## [23] "Census.Block"
## [24] "School.Dist"
## [25] "Council.District"
## [26] "Postcode"
## [27] "Fire.Comp"
## [28] "Health.Area"
## [29] "Health.Ctr"
## [30] "Police.Prct"
## [31] "Major.Use"
## [32] "Number.of.Easements"
## [33] "Commercial.Floor.Area"
## [34] "Residential.Floor.Area"
## [35] "Office.Floor.Area"
## [36] "Retail.Floor.Area"
## [37] "Garage.Floor.Area"
## [38] "Storage.Floor.Area"
## [39] "Factory.Floor.Area"
## [40] "Other.Floor.Area"
## [41] "Num.Floors"
## [42] "Residential.Units"
## [43] "Residential.and.Non.Residential.Units"
## [44] "Lot.Front"
## [45] "Lot.Depth"

```

```

## [46] "Bldg.Front"
## [47] "Bldg.Depth"
## [48] "Proximity.Code"
## [49] "Irr.Lot.Code"
## [50] "Lot.Type.Code"
## [51] "Bsmt.Code"
## [52] "Assess.Land"
## [53] "Exempt.Land"
## [54] "Exempt.Tot"
## [55] "Year.Alter.1"
## [56] "Year.Alter.2"
## [57] "His.Dist"
## [58] "Landmark"
## [59] "Condominium.Number"
## [60] "Coordinates"
## [61] "E.Designation.Number"
## [62] "Industrial.Business.Zone"
## [63] "Zone.Dist.1"
## [64] "Zone.Dist.2"
## [65] "Overlay.1"
## [66] "Overlay.2"
## [67] "SP.Dist.1"
## [68] "SP.Dist.2"
## [69] "Potential.Urban.Ag"
## [70] "Contact"
## [71] "EDC...Occupied"
## [72] "Pluto.Version"
## [73] "Latitude"
## [74] "Longitude"
## [75] "BIN"
## [76] "BBL"
## [77] "NTA"

```

```
library("reshape2")
```

```
## Warning: package 'reshape2' was built under R version 4.1.2
```

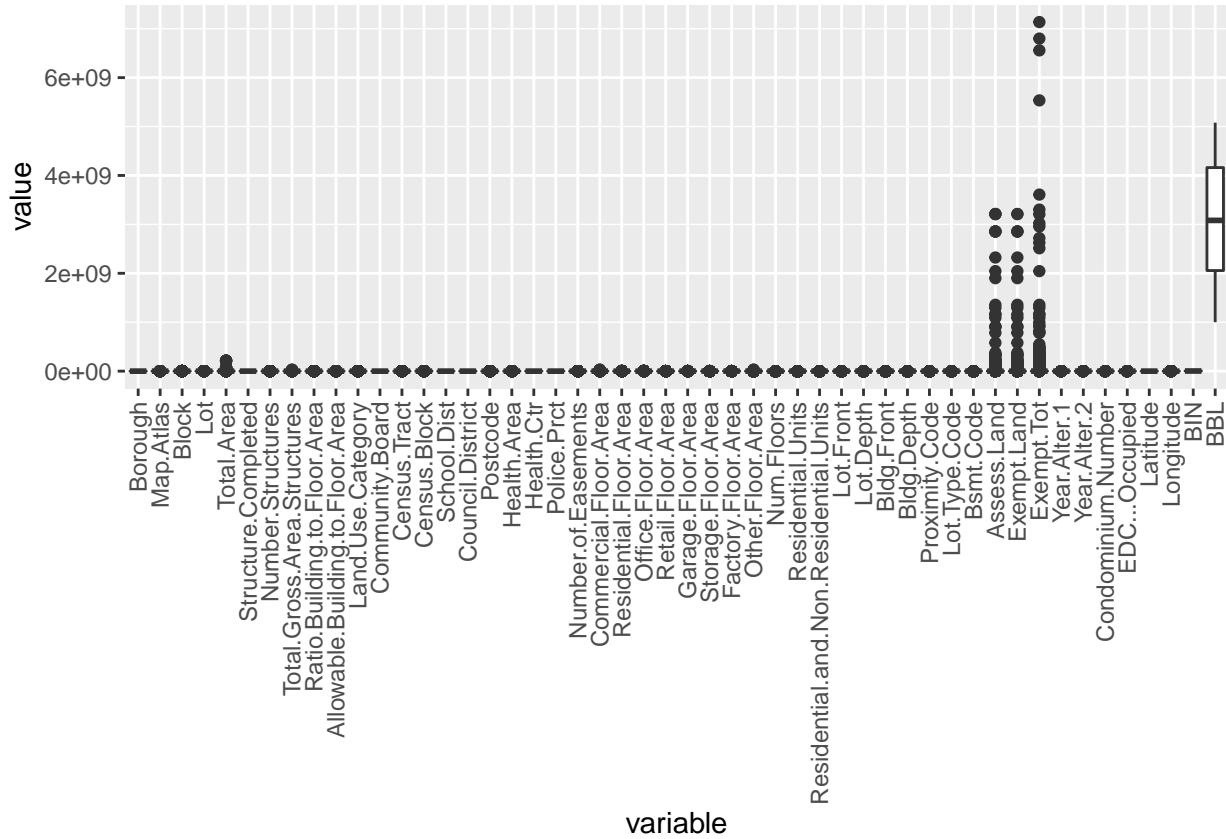
```
library(ggplot2)
```

```
data_long <- melt(df) # Reshaping data frame
```

```
## Using Date.Created, Address, Street.Number, Street.Name, Parcel.Name, Agency, Current.Uses, Open.Pet
```

```
ggplot(data_long, aes(x = variable, y = value)) + # Applying ggplot function
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(angle = 90))
```

```
## Warning: Removed 351291 rows containing non-finite values (stat_boxplot).
```



```

col1 <- c('Assess.Land', 'Exempt.Land', 'Exempt.Tot', 'BBL')
df1 <- df[!(names(df)) %in% col1]

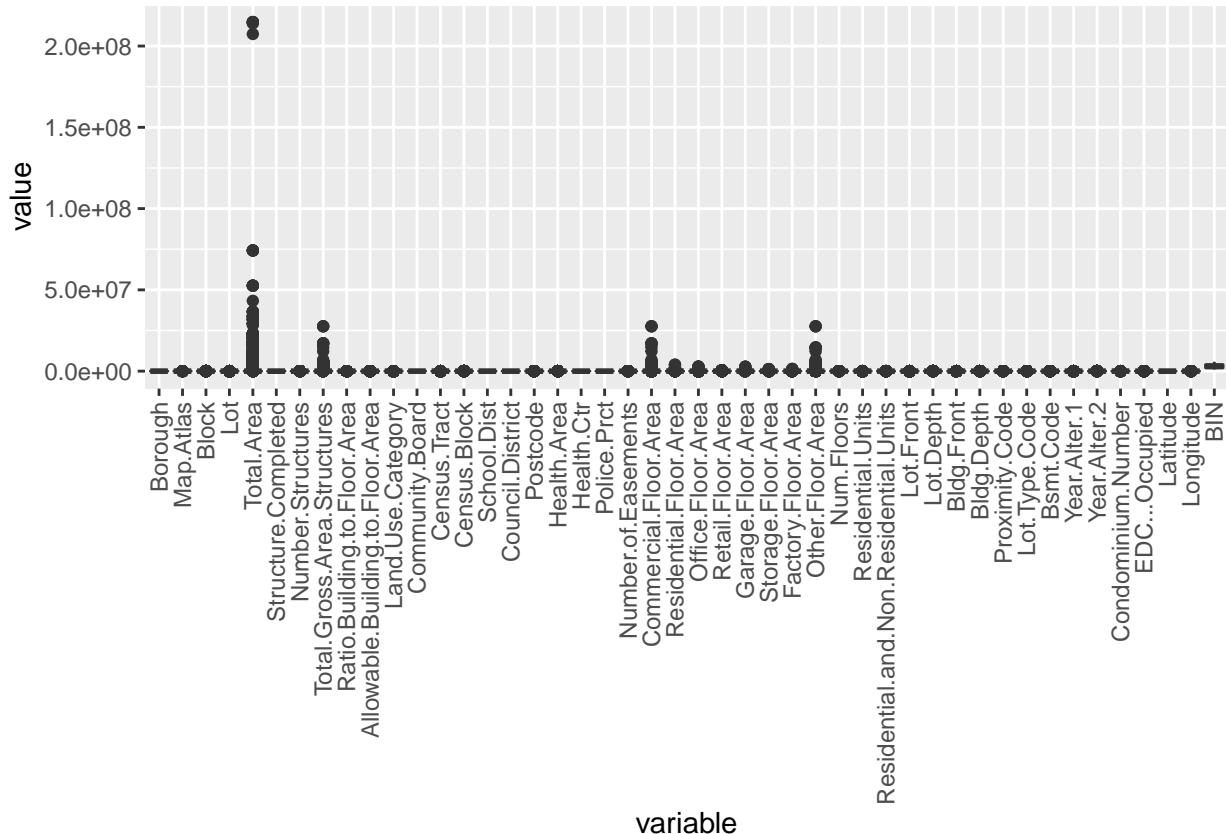
#removing columns with large numbers
data_long <- melt(df1)                                     # Reshaping data frame

## Using Date.Created, Address, Street.Number, Street.Name, Parcel.Name, Agency, Current.Uses, Open.Pet...

#head(data_long)
ggplot(data_long, aes(x = variable, y = value)) +          # Applying ggplot function
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(angle = 90))

## Warning: Removed 337684 rows containing non-finite values (stat_boxplot).

```



```

col2 <- c('Total.Area')
df2 <- df1[!(names(df1)) %in% col2]

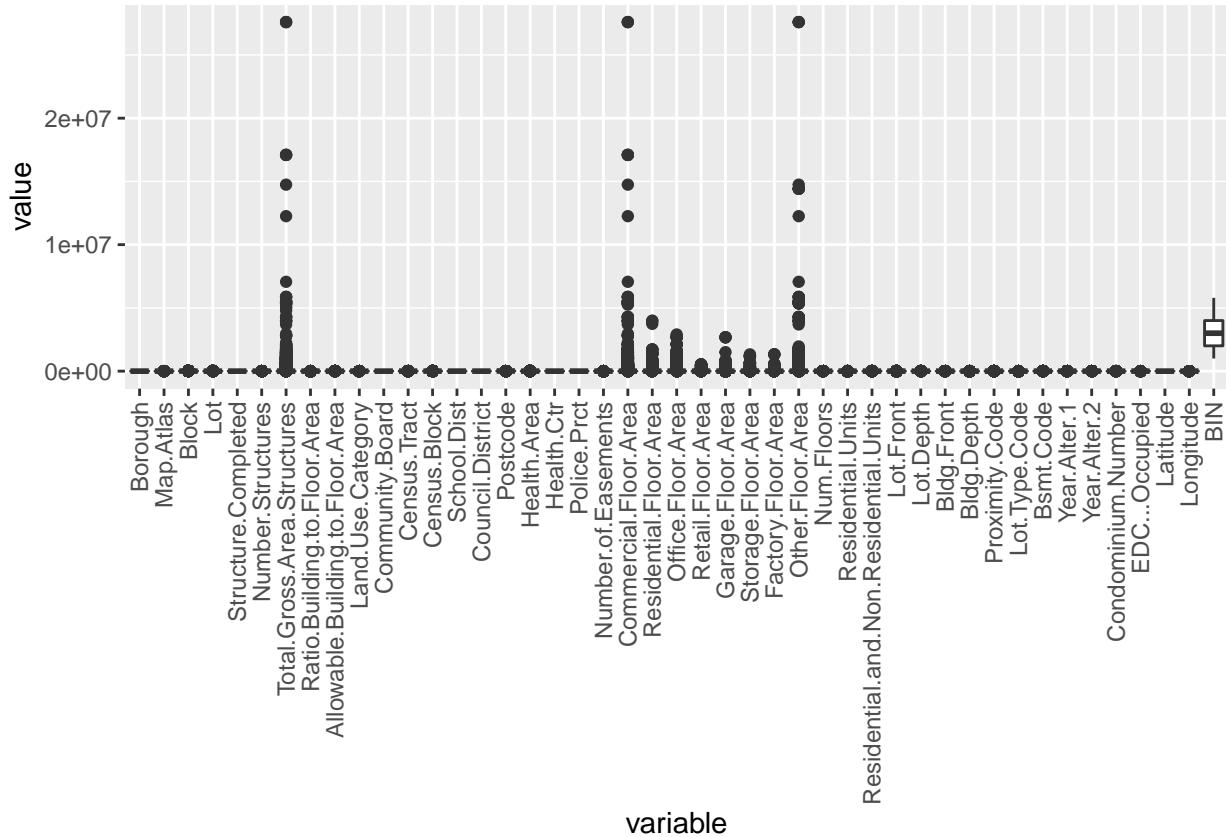
#removing columns with large numbers
data_long <- melt(df2)                                # Reshaping data frame

## Using Date.Created, Address, Street.Number, Street.Name, Parcel.Name, Agency, Current.Uses, Open.Pet

#head(data_long)
ggplot(data_long, aes(x = variable, y = value)) +      # Applying ggplot function
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(angle = 90))

## Warning: Removed 337684 rows containing non-finite values (stat_boxplot).

```



```

col3 <- c('Total.Gross.Area.Structures', 'Commercial.Floor.Area', 'Other.Floor.Area', 'BIN')
df3 <- df2[!(names(df2)) %in% col3]

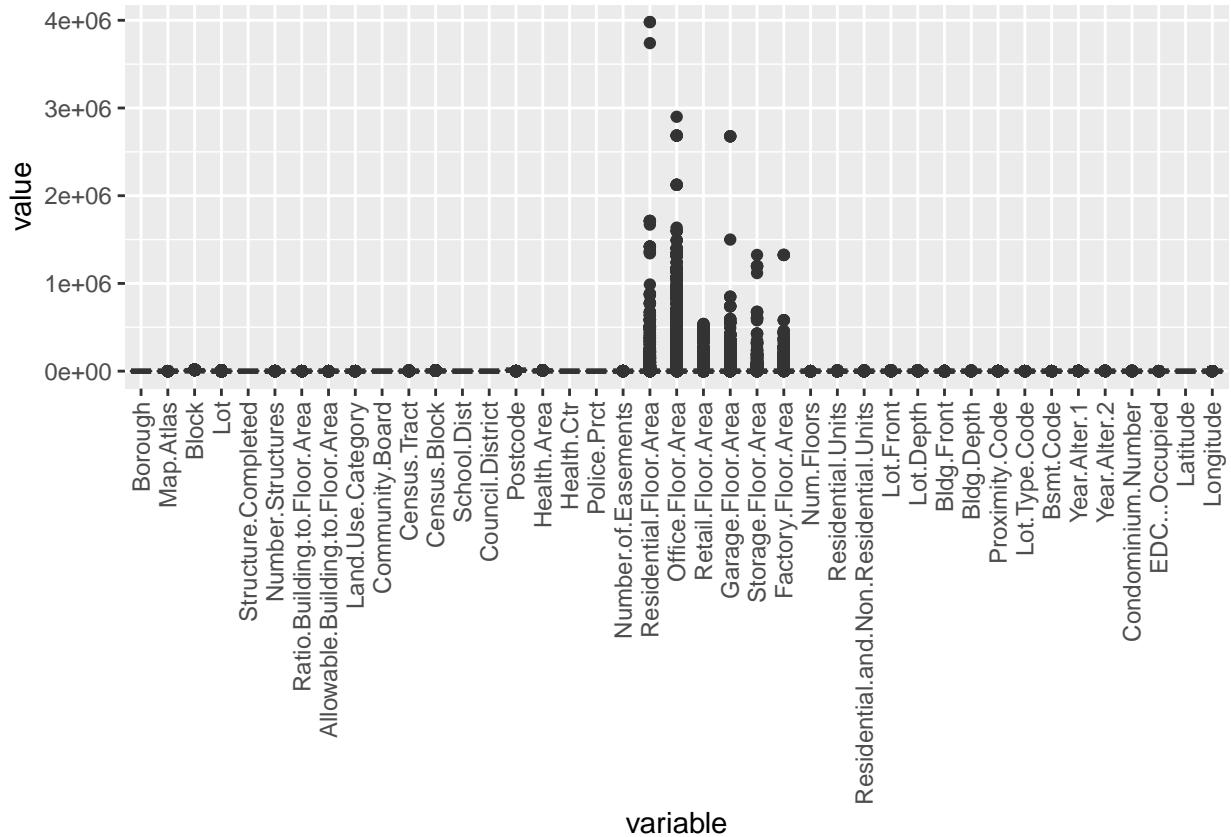
#removing columns with large numbers
data_long <- melt(df3)                                     # Reshaping data frame

## Using Date.Created, Address, Street.Number, Street.Name, Parcel.Name, Agency, Current.Uses, Open.Pet...

ggplot(data_long, aes(x = variable, y = value)) +           # Applying ggplot function
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(angle = 90))

## Warning: Removed 303486 rows containing non-finite values (stat_boxplot).

```



```

col4 <- c('Residential.Floor.Area', 'Office.Floor.Area', 'Retail.Floor.Area', 'Garage.Floor.Area', 'Storage.Floor.Area', 'Factory.Floor.Area')
df4 <- df3[!(names(df3)) %in% col4]

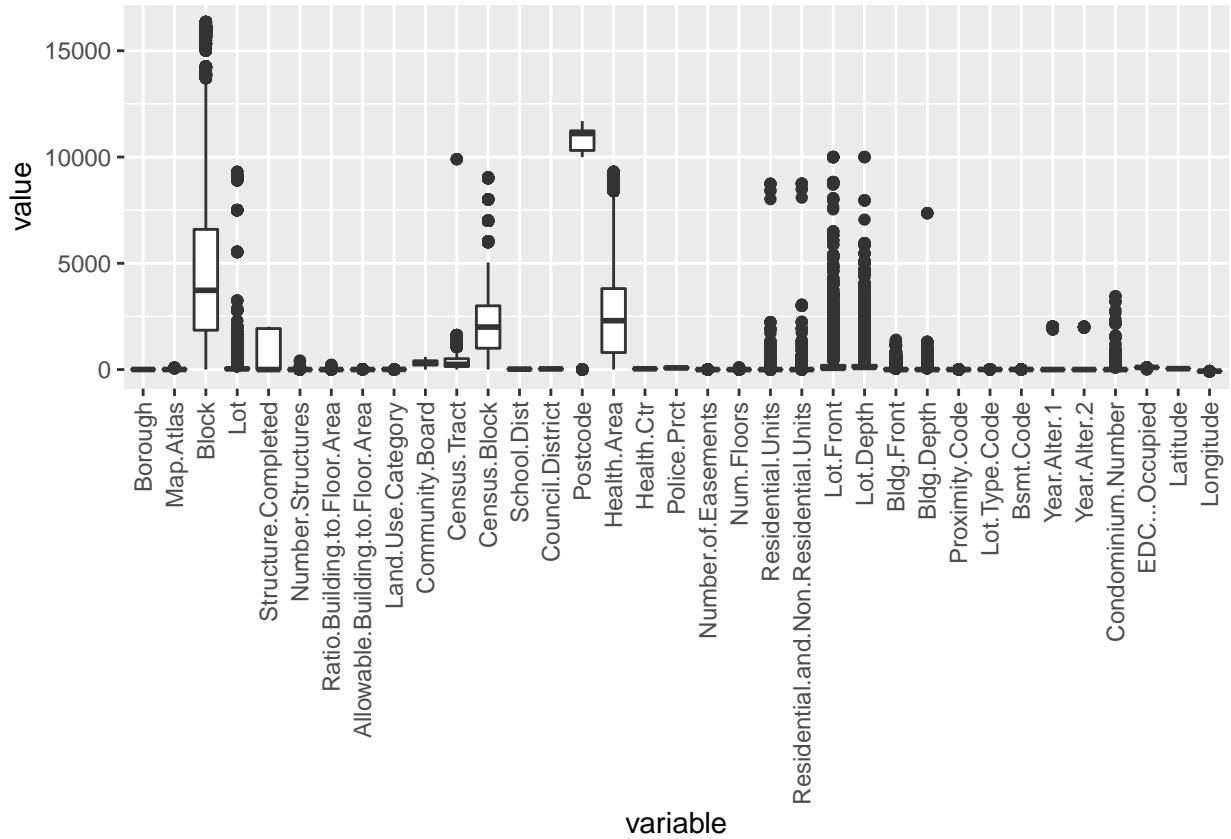
#removing columns with large numbers
data_long <- melt(df4)                                     # Reshaping data frame

## Using Date.Created, Address, Street.Number, Street.Name, Parcel.Name, Agency, Current.Uses, Open.Pet...

#head(data_long)
ggplot(data_long, aes(x = variable, y = value)) +          # Applying ggplot function
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(angle = 90))

## Warning: Removed 303486 rows containing non-finite values (stat_boxplot).

```



```

col5 <- c('Block', 'Lot', 'Structure.Completed', 'Census.Block', 'Census.Ttract', 'Postcode', 'Health.Area')
df5 <- df4[!(names(df4)) %in% col5]

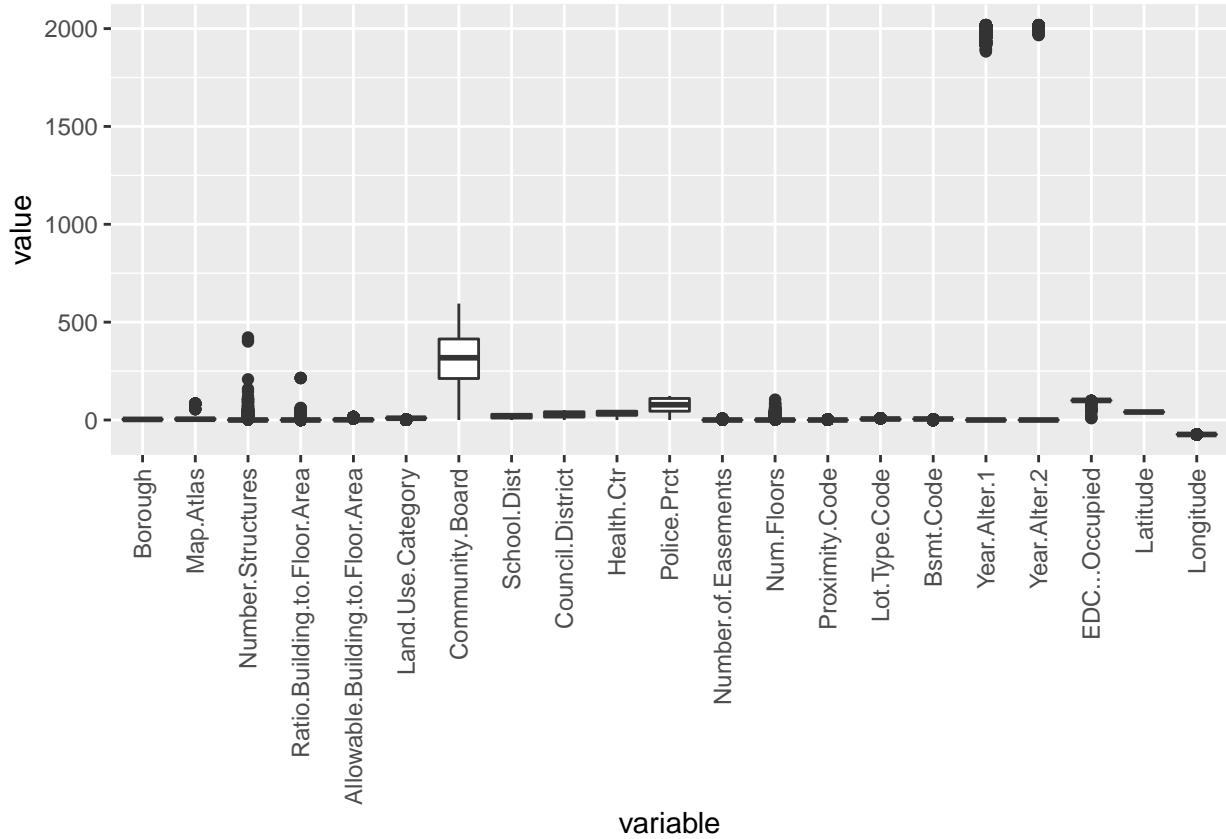
#removing columns with large numbers
data_long <- melt(df5)                                     # Reshaping data frame

## Using Date.Created, Address, Street.Number, Street.Name, Parcel.Name, Agency, Current.Uses, Open.Pet...

#head(data_long)
ggplot(data_long, aes(x = variable, y = value)) +          # Applying ggplot function
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(angle = 90))

```

Warning: Removed 229579 rows containing non-finite values (stat_boxplot).



```

col6 <- c('Year.Alter.1', 'Year.Alter.2')
df6 <- df5[!(names(df5)) %in% col6]

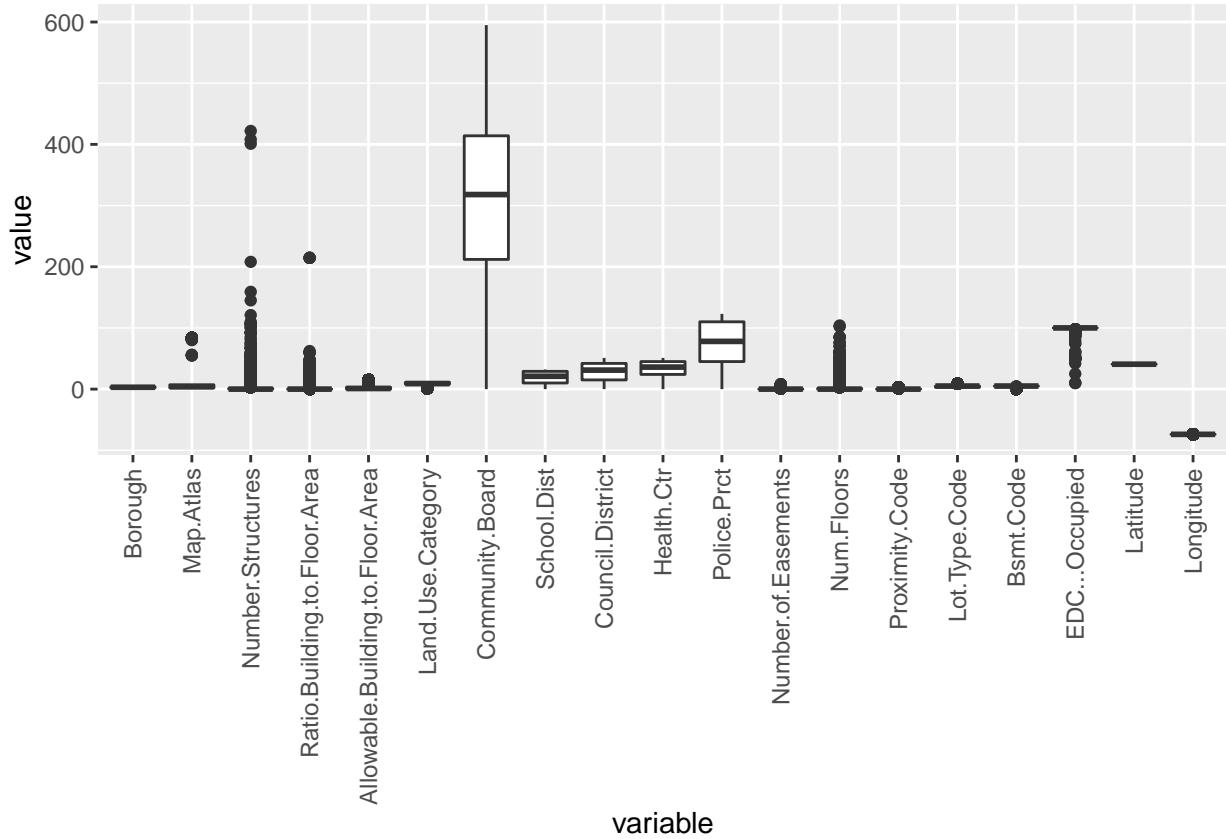
#removing columns with large numbers
data_long <- melt(df6)                                     # Reshaping data frame

## Using Date.Created, Address, Street.Number, Street.Name, Parcel.Name, Agency, Current.Uses, Open.Pet...

#head(data_long)
ggplot(data_long, aes(x = variable, y = value)) +          # Applying ggplot function
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(angle = 90))

## Warning: Removed 229539 rows containing non-finite values (stat_boxplot).

```



```

col7 <- c('Number.Structures', 'Ratio.Building.to.Floor.Area', 'Community.Board')
df7 <- df6[!(names(df6)) %in% col7]

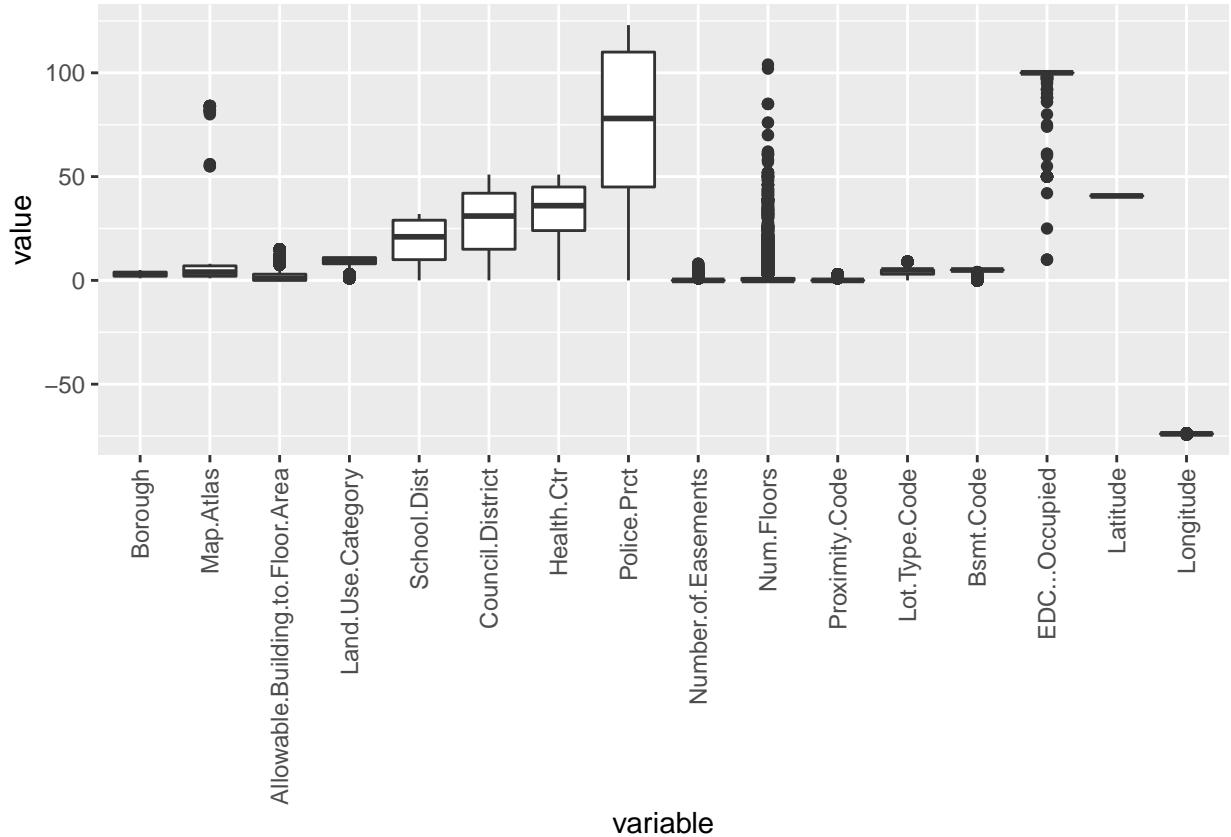
#removing columns with large numbers
data_long <- melt(df7)                                     # Reshaping data frame

## Using Date.Created, Address, Street.Number, Street.Name, Parcel.Name, Agency, Current.Uses, Open.Pet...

#head(data_long)
ggplot(data_long, aes(x = variable, y = value)) +          # Applying ggplot function
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(angle = 90))

## Warning: Removed 229539 rows containing non-finite values (stat_boxplot).

```



```

col8 <- c('Longitude', 'Latitude')
df8 <- df7[!(names(df7)) %in% col8]

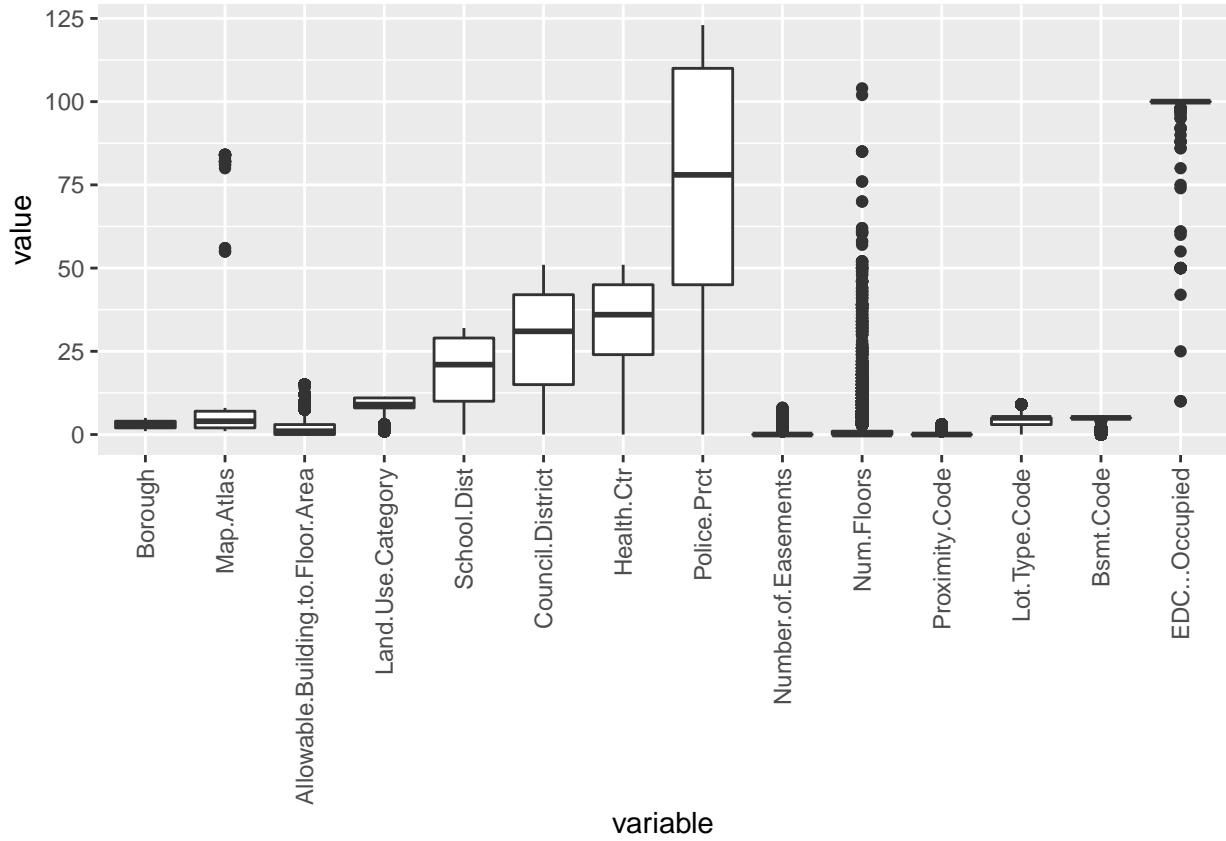
#removing columns with large numbers
data_long <- melt(df8)                                # Reshaping data frame

## Using Date.Created, Address, Street.Number, Street.Name, Parcel.Name, Agency, Current.Uses, Open.Pet

ggplot(data_long, aes(x = variable, y = value)) +          # Applying ggplot function
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(angle = 90))

## Warning: Removed 164621 rows containing non-finite values (stat_boxplot).

```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.