



ANALYSIS OF BUILDING PERMIT DELAY

Lu Pang



Nationally, many cities are struggling with a lack of affordable housing. During the last two decades San Francisco has been stingy when it comes to homebuilding and prices have risen. In the recent past, research found that main discrepancy in demand and supply in real estate industry is due to delays in issuing building permits.

A building permit is an official approval document issued by a governmental agency that allows you or your contractor to proceed with a construction or remodeling project on one's property. For the city of San Francisco, permit issuing is taken care by www.sfdbi.org/.

CONTEXT

What is our topic?
And why is it important?

Building Permits Data

Retrieved from

<https://data.sfgov.org/Housing-and-Buildings/Building-Permits-on-or-after-January-1-2013/p4e4-a5a7>

For every month delay in approving new building permits, a housing market's ability to meet housing demand falters. More time-consuming means adding more cost and risk to a project. Facing with the competitive market, real estate companies are looking for a tool to estimate waiting time taken by permit types.

CONTEXT

Problem
Identification

Data Introduction

This data set pertains to all types of structural permits. Data includes details on application/permit numbers, job addresses, supervisorial districts, and the current status of the applications.

Field Name	Data Type	Definition	Notes (optional)	Completed Date	Date & Time	The date on which project was completed, applicable if Current Status = "completed"
Permit Number	Text	Permit Number		First.Construction.Document.Date	Date & Time	Date on which construction was documented
Permit Type	Numeric	Type of permit	1 = NEW CONSTRUCTION 2 = NEW CONSTRUCTION WOOD FRAME 3 = ADDITIONS, ALTERATIONS OR REPAIRS 4 = SIGN - ERECT 5 = GRADE, QUARRY, FILL OR EXCAVATE 6 = DEMOLITIONS 7 = WALL OR PAINTED SIGN 8 = OTC ALTERATIONS PERMI	Structural.Notification Number of Existing Stories Number of Proposed Stories Voluntary Soft-Story Retrofit Fire Only Permit Permit.Expiration.Date Estimated Cost Revised Cost Existing.Use Existing Units Proposed.Use Proposed Units Plansets	Text Numeric Numeric Text Date & Time Numeric Numeric Text Numeric Text Numeric Text	Structural Notification Number of Existing Stories Number of Proposed Stories Voluntary Soft-Story Retrofit Fire Only Permit Permit Expiration Date Estimated Cost Revised Cost Existing Use Existing Units Proposed Use Proposed Units Plansets
Permit.Type.Definition	Text	The type of permit definition	Use the values for APPLICATION_TYPE as a lookup	TIDF.Compliance	Text	TIDF compliant or not, this is a new legal requirement
Permit Creation Date	Date & Time			Existing Construction Type	Numeric	Construction type, existing, as categories represented numerically
Block	Text	Block		Existing.Construction.Type.Description	Text	Description of the above, for example, wood or other construction types
Lot	Text	Lot		Proposed Construction Type	Numeric	Construction type, proposed, as categories represented numerically
Street Number	Numeric	Street Number		Proposed.Construction.Type.Description	Text	Description of the above
Street Number Suffix	Text	Street Number Suffix		Site Permit	Text	Permit for site
Street Name	Text	Street Name		Supervisor District	Numeric	Supervisor district the permit falls within
Street Suffix	Text	Street Suffix		Neighborhoods...Analysis.Boundaries	Text	Analysis neighborhood the permit falls within
Unit	Numeric	Unit		Zipcode	Text	Zip code the permit falls within
Unit Suffix	Text	Unit Suffix		Location	Coordinates	The latitude and longitude associated with the permit
Description	Text	Description		Record ID	Text	Application Record ID
Current Status	Text	Current Status				
Current Status Date	Date & Time	Date at which current status was entered				
Filed Date	Date & Time	Filed date for the permit				
Issued Date	Date & Time	Issued date for the permit				

APPROACH

Research Questions:

How different variables affect the building permit approval time?

How can we predict the approval time by a given permit type?

Approach Strategy:

Building Histogram, Box Plot, Correlation Matrix, Bar chart, Linear Regression Analysis to understand all the variables relation to the building permit approval time. Then select variables that play important role in time cost to build a prediction tool.

Load Data (CONT.)

```
library(readxl)
data<-read.csv("Building_Permits_on_or_after_January_1__2013.csv")
head(data)
dim(data)
```

Permit Number	Permit Type	Permit Type Definition	Permit Creation Date	Block	Lot	Street Number	Street Suffix	Street Name	Street Suffix	Unit	Unit Suffix	Description	Current Status	Current Status Date	Filed Date	Issued Date	Completed Date	First Construction Document Date	Structural Notification	Number Existing Stories	Number Proposed Stories	Number of Voluntary Soft-Story Retrofit
2.013E+11	8	otc alterations permit	1/2/13	697	39	1400		Geary		Bl		add 8-inch double detector check backflow preventer at main riser per plumbing inspection compliant	complete	10/15/13	1/2/13	1/2/13	10/15/13			26	26	
M364227	8	otc alterations permit	1/2/13	1007	26	3225		Clay	St			street space	issued	1/2/13	1/2/13	1/2/13						
2.013E+11	8	otc alterations permit	1/2/13	1102	17	1247		Broderick	St	0		rear of bldg install christy box. outside rear of bldg.	complete	2/7/13	1/2/13	1/2/13	2/7/13			3	3	
M364228	8	otc alterations permit	1/2/13	164	011A	920		Montgomery	St			street space and mta parking	issued	1/2/13	1/2/13	1/2/13						
M364229	8	otc alterations permit	1/2/13	105	19	350		Union	St			street space	issued	1/2/13	1/2/13	1/2/13						

[1] 294498 43

Load Data

```
library(readxl)
data<-read.csv("Building_Permits_on_or_after_January_1__2013.csv")
head(data)
dim(data)
```

Fire Only Permit	Permit Expiration Date	Estimated Cost	Revised Cost	Existing Use	Existing Units	Proposed Use	Proposed Units	Plansets	TIDF Compliance	Existing Construction Type	Existing Construction Description	Proposed Construction Type	Proposed Construction Description	Site Permit	Supervisor District	Neighborhoods - Analysis Boundaries	Zipcode	Location	Record ID
Y	12/28/13	9600	9600	misc group residns.	300	misc group residns.	300	2	1	constr type 1	1	constr type 1	1	5	Japantown	94109	(37.785542 418360315, - 122.427133 99508852)	1.2921E+11	
				1										2	Presidio Heights	94115	(37.788920 41199593,- 122.446173 1214486)	1.2921E+11	
	12/28/13	500	500	2 family dwelling	2	2 family dwelling	2	0	5	wood frame (5)	5	wood frame (5)	5	2	Western Addition	94115	(37.781885 10830172,- 122.441125 31894436)	1.2921E+12	
				1										3	North Beach	94133	(37.797691 71403776,- 122.403399 94530932)	1.2921E+12	
				1										3	North Beach	94133	(37.801213 254049166, - 122.405297 39174583)	1.2921E+12	

Summarize Data

> summary(data)																	
Permit.Number	Permit.Type	Permit.Type.Definition		Permit.Creation.Date	Block	Lot	Street.Number	Street.Number.Suffix	Street.Name	Street.Suffix							
2.01709E+11: 3606	Min. : 1.000	otc alterations	permit	:265050	9/15/17: 424	3708 : 1794	1 : 15213	Min. : 0	:290791	Market : 7787	St : 204852						
2.01708E+11: 3453	1st Qu.: 8.000	additions	alterations or repairs:	21579	11/3/15: 366	3717 : 994	7 : 7784	1st Qu.: 237	A : 2575	California: 6738	Av : 64309						
2.0191E+11 : 3204	Median : 8.000	sign - erect		: 4228	2/17/16: 364	3735 : 914	3 : 7406	Median : 705	B : 493	Mission : 6304	Bl : 5231						
2.01809E+11: 3133	Mean : 7.526	new construction	wood frame	: 1247	9/14/17: 334	289 : 876	2 : 7390	Mean : 1119	V : 366	Montgomery: 3444	Wy : 4972						
2.01406E+11: 3095	3rd Qu.: 8.000	demolitions		: 966	9/14/18: 322	3721 : 863	8 : 7202	3rd Qu.: 1700	C : 105	Geary : 2879	Dr : 4850						
2.01808E+11: 3088	Max. : 8.000	wall or painted sign		: 690	6/27/14: 306	3707 : 841	9 : 7053	Max. : 8400	E : 33	03rd : 2649	: 4153						
(Other) : 274919	(Other)			: 738	(Other):292382	(Other):288216	(Other):242450	(Other): 135	(Other)	:264697	(Other): 6131						
Unit	Unit.Suffix	Description	Current.Status	Current.Status.Date	Filed.Date	Issued.Date	Completed.Date	First.Construction.Document.Date Structural.Notification									
Min. : 0	:291274	street space	: 39914	complete :149908	8/16/19 : 352	9/15/17: 424	: 21819	:144422 :290463 :284737									
1st Qu.: 0	A : 537	reroofing	: 12917	issued :116999	8/11/20 : 348	11/3/15: 366	6/15/16 : 341	8/16/19 : 237	11/7/16 : 182		Y: 9761						
Median : 0	B : 318	street space permit:	7846	filed : 16784	11/15/19: 347	2/17/16: 364	10/2/18 : 333	11/15/19: 209	11/1/16 : 114								
Mean : 81	C : 304	re-roofing	: 3346	cancelled: 2929	2/7/20 : 339	9/14/17: 332	7/26/16 : 274	10/30/15: 206	5/15/17 : 68								
3rd Qu.: 1	D : 199	reroofing.	: 1550	withdrawn: 2923	10/18/19: 331	9/14/18: 320	12/18/15: 272	2/7/20 : 184	11/9/16 : 34								
Max. : 9100	PARCEL A: 193	streetspace	: 1356	expired : 2726	3/10/20 : 317	6/27/14: 306	10/15/19: 249	8/11/20 : 184	12/31/19: 23								
NA's : 251132	(Other) : 1673	(Other)	:227569	(Other) : 2229	(Other) :292464	(Other):292386	(Other) :271210	(Other) :149056	(Other) : 3614								
Number.of.Existing.Stories	Number.of.Proposed.Stories	Voluntary.Soft.Story.Retrofit	Fire.Only.Permit	Permit.Expiration.Date	Estimated.Cost	Revised.Cost		Existing.Use	Existing.Units								
Min. : 0.00	Min. : 0.00	:294458	:265421	: 79734	Min. : 0	Min. : 0	1 family dwelling:68334	Min. : 0.00									
1st Qu.: 2.00	1st Qu.: 2.00	Y: 40	Y: 29077	9/27/19: 315	1st Qu.: 3500	1st Qu.: 1		:64522	1st Qu.: 1.00								
Median : 3.00	Median : 3.00			2/29/20: 303	Median : 12000	Median : 7500	apartments	:59331	Median : 1.00								
Mean : 5.92	Mean : 5.96			5/31/19: 290	Mean : 185786	Mean : 134945	office	:36463	Mean : 16.39								
3rd Qu.: 4.00	3rd Qu.: 4.00			12/2/18: 258	3rd Qu.: 38000	3rd Qu.: 30000	2 family dwelling:29933	3rd Qu.: 4.00									
Max. : 88.00	Max. : 479.00			4/30/20: 256	Max. : 537958646	Max. : 780500000	retail sales	:10060	Max. : 1907.00								
NA's : 67103	NA's :67095			(Other):213342	NA's :60199	NA's :7617	(Other)	:25855	NA's :85647								
Proposed.Use	Proposed.Units	Plansets	TIDF.Compliance	Existing.Construction.Type	Existing.Construction.Type.Description	Proposed.Construction.Type	Proposed.Construction.Type.Description										
1 family dwelling:66838	Min. : 0.00	Min. : 0.00	:294496	Min. :1.00		: 67827		: 67523			: 67524						
	:66579	1st Qu.: 1.00	1st Qu.: 0.00	P: 1	1st Qu.:3.00	constr type 1 : 41982		1 : 41942			constr type 1 : 41942						
apartments :63215	Median : 2.00	Median : 2.00	Y: 1	Median :5.00	constr type 2 : 5509		2 : 5163			constr type 2 : 5163							
office :35587	Mean : 17.33	Mean : 1.23		Mean :4.06	constr type 3 : 13774		3 : 13411			constr type 3 : 13411							
2 family dwelling:31751	3rd Qu.: 5.00	3rd Qu.: 2.00		3rd Qu.:5.00	constr type 4 : 606		4 : 566			constr type 4 : 566							
retail sales : 7514	Max. : 1911.00	Max. :20.00		Max. :5.00	wood frame (5):164800		5 : 165892			wood frame (5):165892							
(Other) :23014	NA's :84357	NA's :58924		NA's :67827			III: 1										
Site.Permit	Supervisor.District	Neighborhoods...	Analysis.Boundaries	Zipcode		Location	Record.ID										
:286961	Min. : 1.000	Financial District/South Beach: 32830		Min. : 94102		: 2780	Min. :1.294e+10										
Y: 7537	1st Qu.: 3.000	Mission	: 21539	1st Qu.:94109	(37.79226164880319, -122.40348595414692)	: 751	1st Qu.:1.317e+12										
	Median : 6.000	Sunset/Parkside	: 15334	Median :94114	(37.78977769387532, -122.40173378341645)	: 447	Median :1.411e+12										
	Mean : 5.532	West of Twin Peaks	: 12845	Mean :94116	(37.78985529361942, -122.39680054556624)	: 447	Mean :1.205e+12										
	3rd Qu.: 8.000	Pacific Heights	: 12404	3rd Qu.:94122	(37.795606967902906, -122.39922677429354)	: 424	3rd Qu.:1.504e+12										
	Max. :11.000	Marina	: 12272	Max. :94158	(37.79294896729476, -122.39809861205246)	: 397	Max. :1.597e+12										
	NA's : 2808	(Other)	:187274	NA's :2805	(Other)	:289252											

DATA CLEANING

Step 1

- ❖ original dataset is overwhelmed
- ❖ select all fields relevant to topic
- ❖ create a new dataset
- ❖ name it as "delay"

```
delay<-subset(data,select =  
c(Permit.Type,  
  Current.Status,  
  Filed.Date,  
  Issued.Date,  
  Completed.Date,  
  First.Construction.Document.Date,  
  Structural.Notification,  
  Number.of.Existing.Stories,  
  Number.of.Proposed.Stories,  
  Voluntary.Soft.Story.Retrofit,  
  Fire.Only.Permit,  
  Permit.Expiration.Date,  
  Estimated.Cost,  
  Revised.Cost,  
  Existing.Use,  
  Existing.Units,  
  Proposed.Use,  
  Proposed.Units,  
  TIDF.Compliance,  
  Existing.Construction.Type,  
  Proposed.Construction.Type,  
  Supervisor.District,  
  Neighborhoods...Analysis.Boundaries,  
  Zipcode,  
  Location,  
  Record.ID))
```

DATA CLEANING

Step 2

- ❖ Manipulate data
- ❖ Add columns, “wtime”, “ctime”,
“diff_stories”, “diff_cost”,
“diff_units”

```
library(dplyr)
library(lubridate)

```{r}

delay$wtime<-as.numeric(
 mdy(delay$Issued.Date)-
 mdy(delay$Filed.Date))

delay$ctime<-as.numeric(
 mdy(delay$Completed.Date)-
 mdy(delay$First.Construction.Document.Date))

delay$diff_stories<-delay$Number.of.Proposed.Stories-
 delay$Number.of.Existing.Stories

delay$diff_cost<-delay$Revised.Cost-
 delay$Estimated.Cost

delay$diff_units<-delay$Proposed.Units-
 delay$Existing.Units

```
```

DATA CLEANING

Step 3

- ❖ get the number of missing data points per column
- ❖ Get the percentage of the values in this dataset were missing
- ❖ Handle missing data
- ❖ Remove columns with extreme high percentage of missing data points, eg, "ctime"

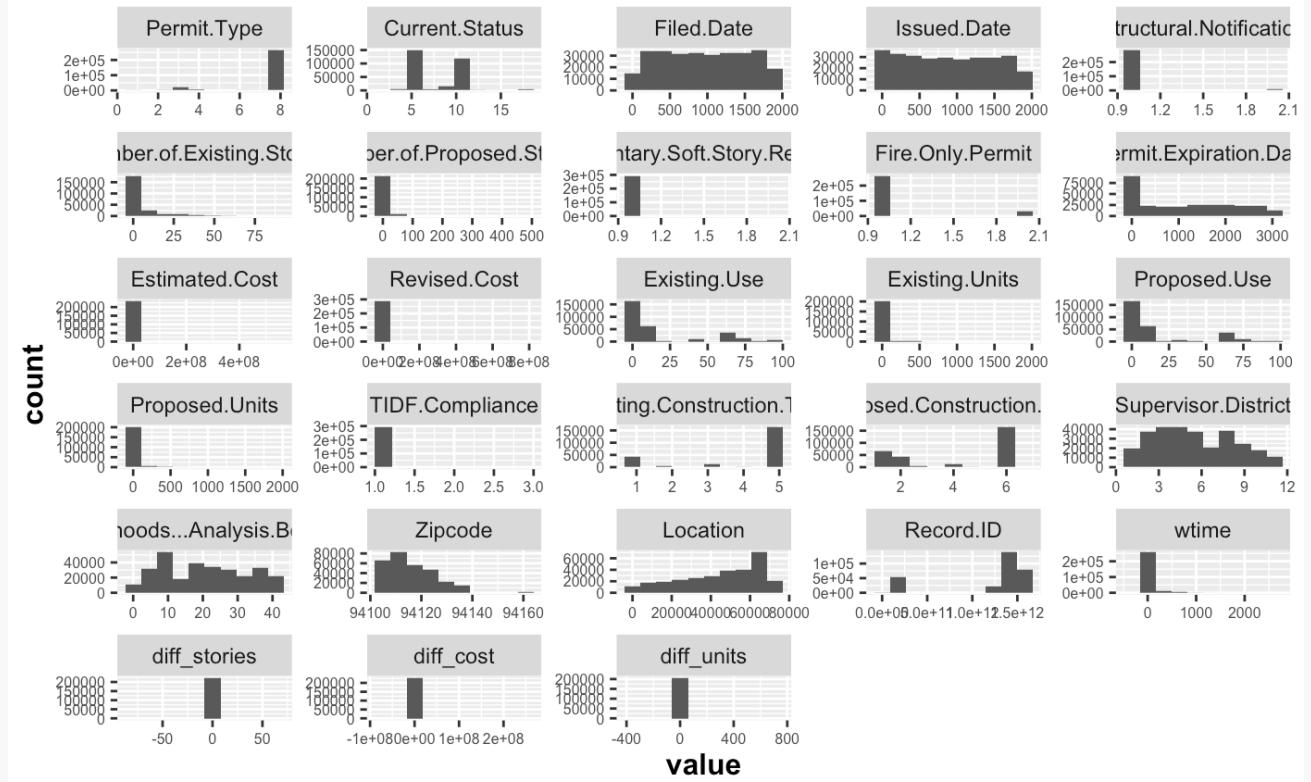
| | |
|-------------------------------------|----------------------------------|
| Permit.Type | Current.Status |
| 0 | 0 |
| Filed.Date | Issued.Date |
| 0 | 0 |
| Completed.Date | First.Construction.Document.Date |
| 0 | 0 |
| Structural.Notification | Number.of.Existing.Stories |
| 0 | 67103 |
| Number.of.Proposed.Stories | Voluntary.Soft.Story.Retrofit |
| 67095 | 0 |
| Fire.Only.Permit | Permit.Expiration.Date |
| 0 | 0 |
| Estimated.Cost | Revised.Cost |
| 60199 | 7617 |
| Existing.Use | Existing.Units |
| 0 | 85647 |
| Proposed.Use | Proposed.Units |
| 0 | 84357 |
| TIDF.Compliance | Existing.Construction.Type |
| 0 | 67827 |
| Proposed.Construction.Type | Supervisor.District |
| 0 | 2808 |
| Neighborhoods...Analysis.Boundaries | Zipcode |
| 0 | 2805 |
| Location | Record.ID |
| 0 | 0 |
| wtime | ctime |
| 21819 | 291999 |
| diff_stories | diff_cost |
| 72309 | 66808 |
| diff_units | |
| 88914 | |

[1] 10.81454

VISUALIZE DATA

Overview-Histogram

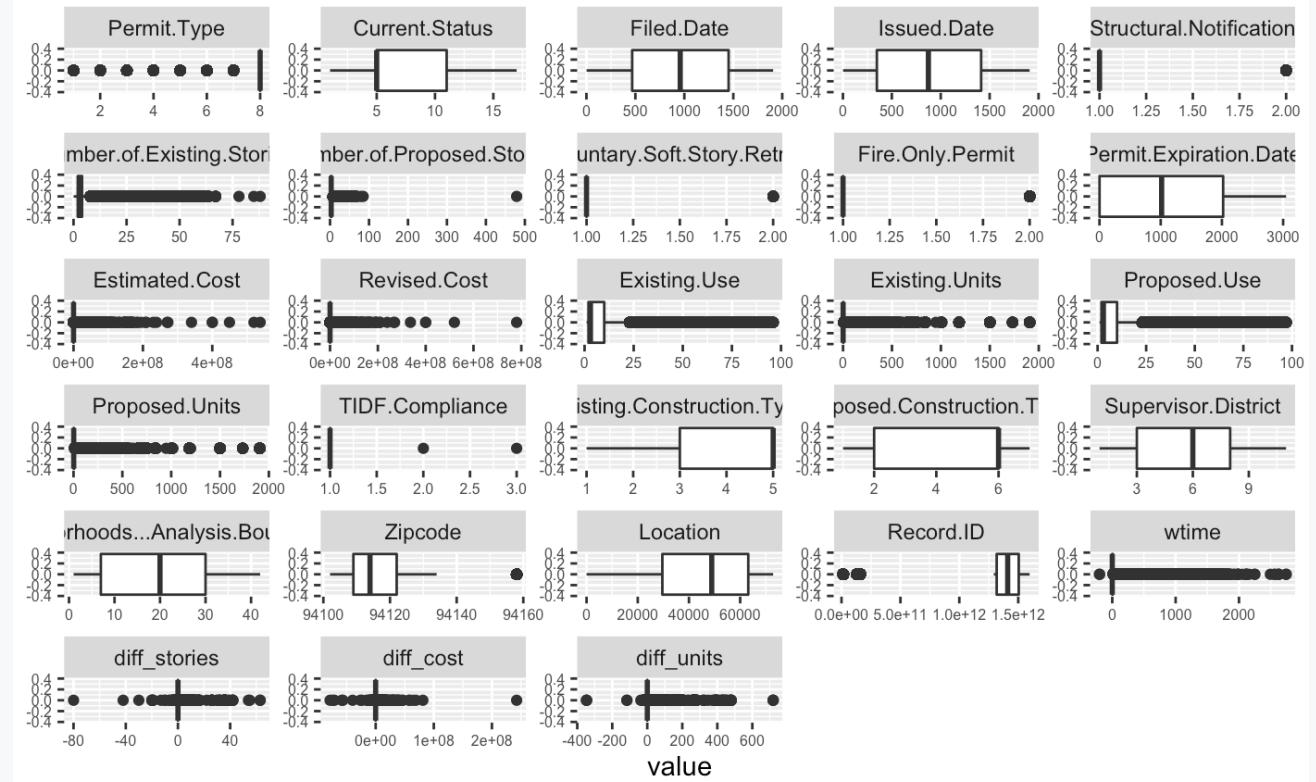
- ❖ `library(ggplot2)`
- ❖ `library(dplyr)`
- ❖ `library(tidyr)`
- ❖ `gather() function`
- ❖ `ggplot() function`
- ❖ `geom_histogram() function`



VISUALIZE DATA

Overview-Boxplot

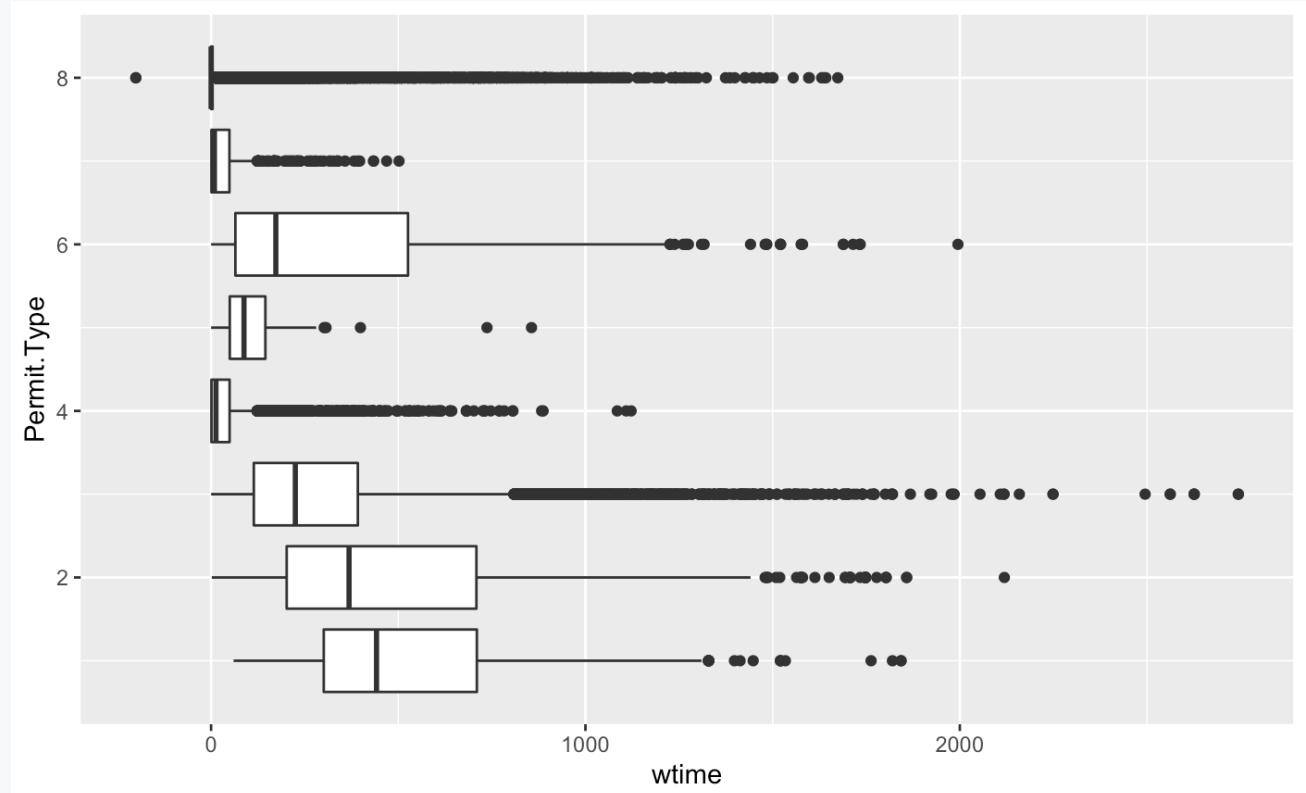
- ❖ `library(ggplot2)`
- ❖ `library(dplyr)`
- ❖ `library(tidyr)`
- ❖ `gather() function`
- ❖ `ggplot() function`
- ❖ `geom_boxplot() function`



VISUALIZE DATA

Relationship

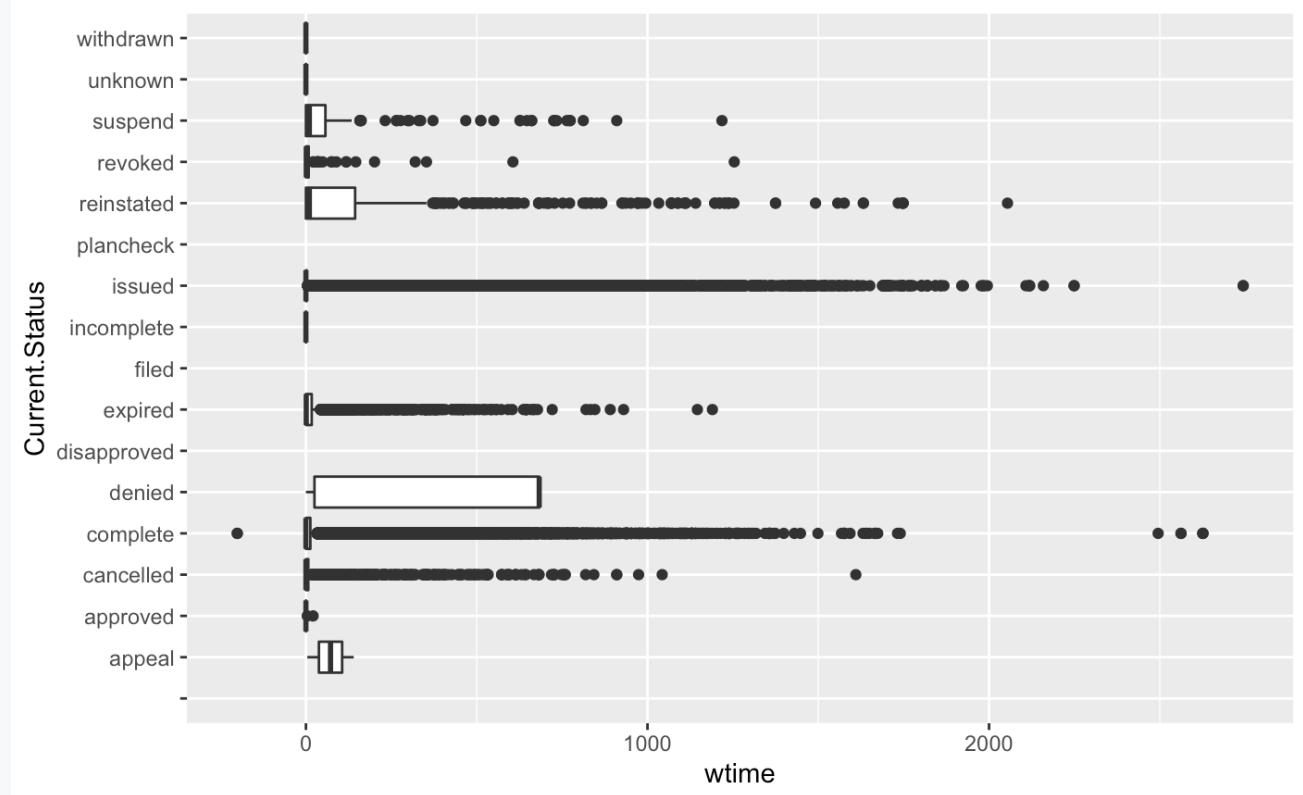
- ❖ Wait time vs. Permit Type
- ❖ Boxplot



VISUALIZE DATA

Relationship

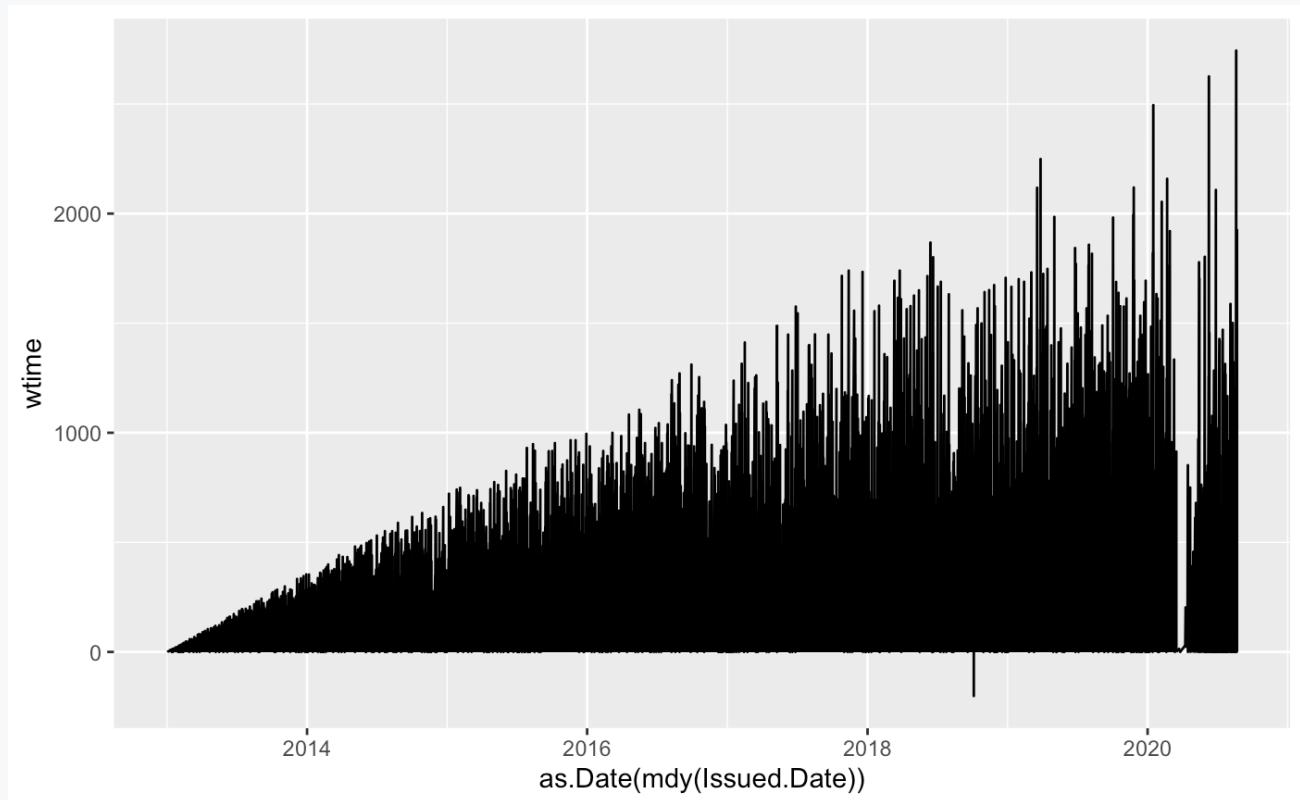
- ❖ Wait time vs. Current Status
- ❖ Boxplot



VISUALIZE DATA

Relationship

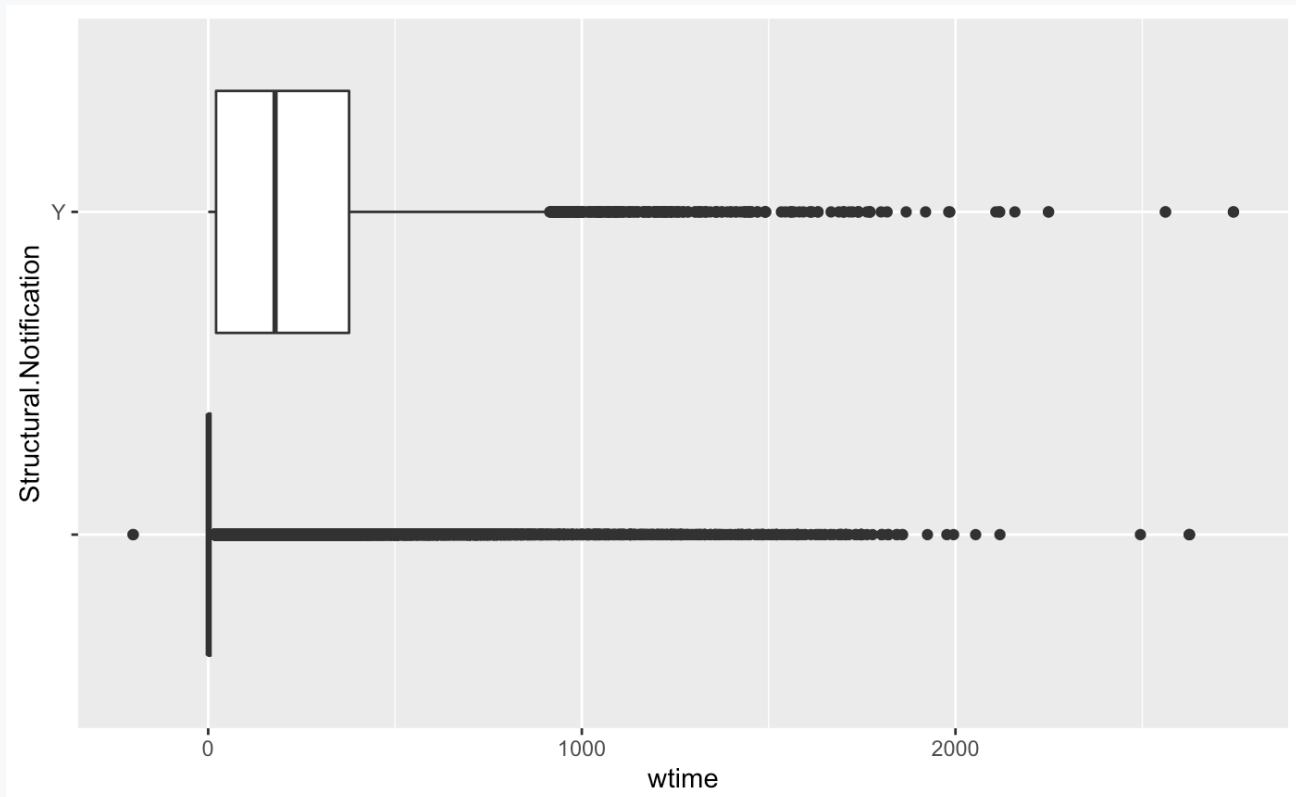
- ❖ Wait time vs. Issued Date
- ❖ Timeseries ggplot



VISUALIZE DATA

Relationship

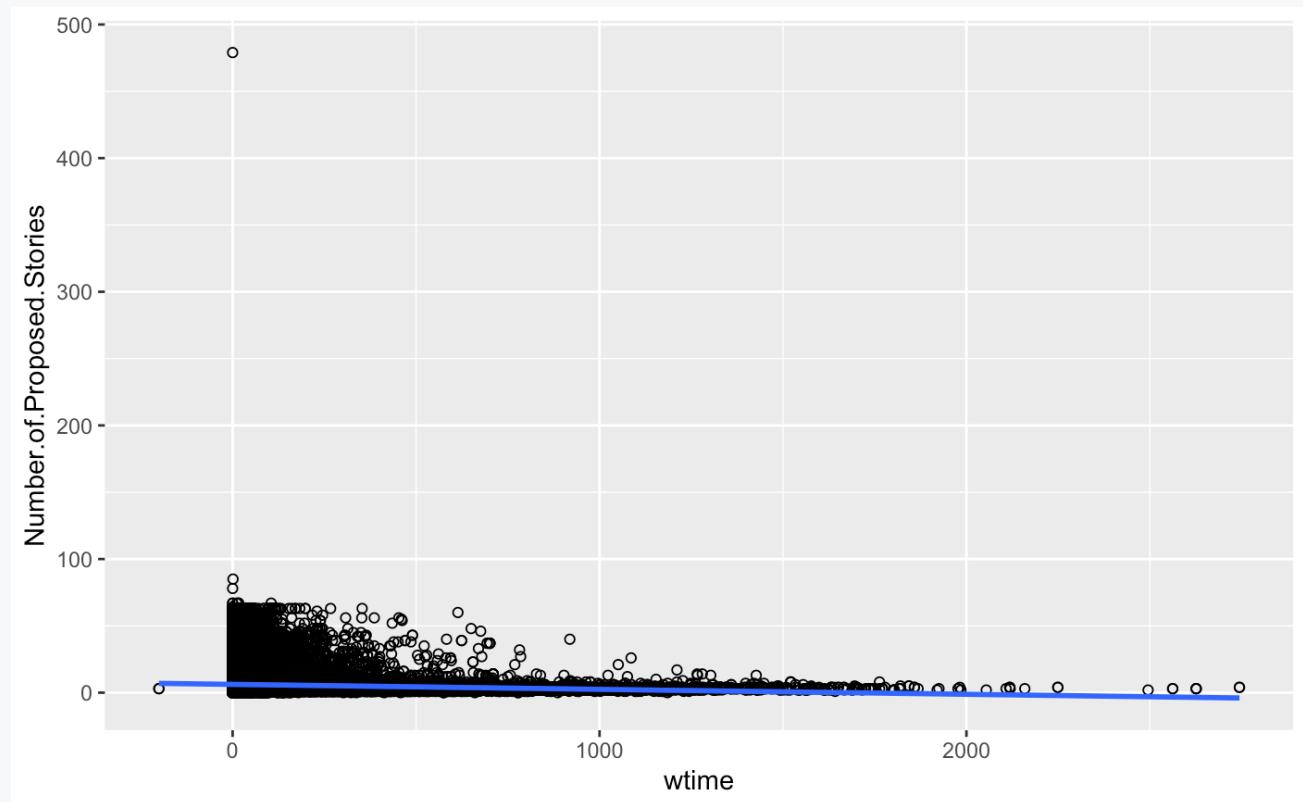
- ❖ Wait time vs.
Structural.Notification
- ❖ Boxplot



VISUALIZE DATA

Relationship

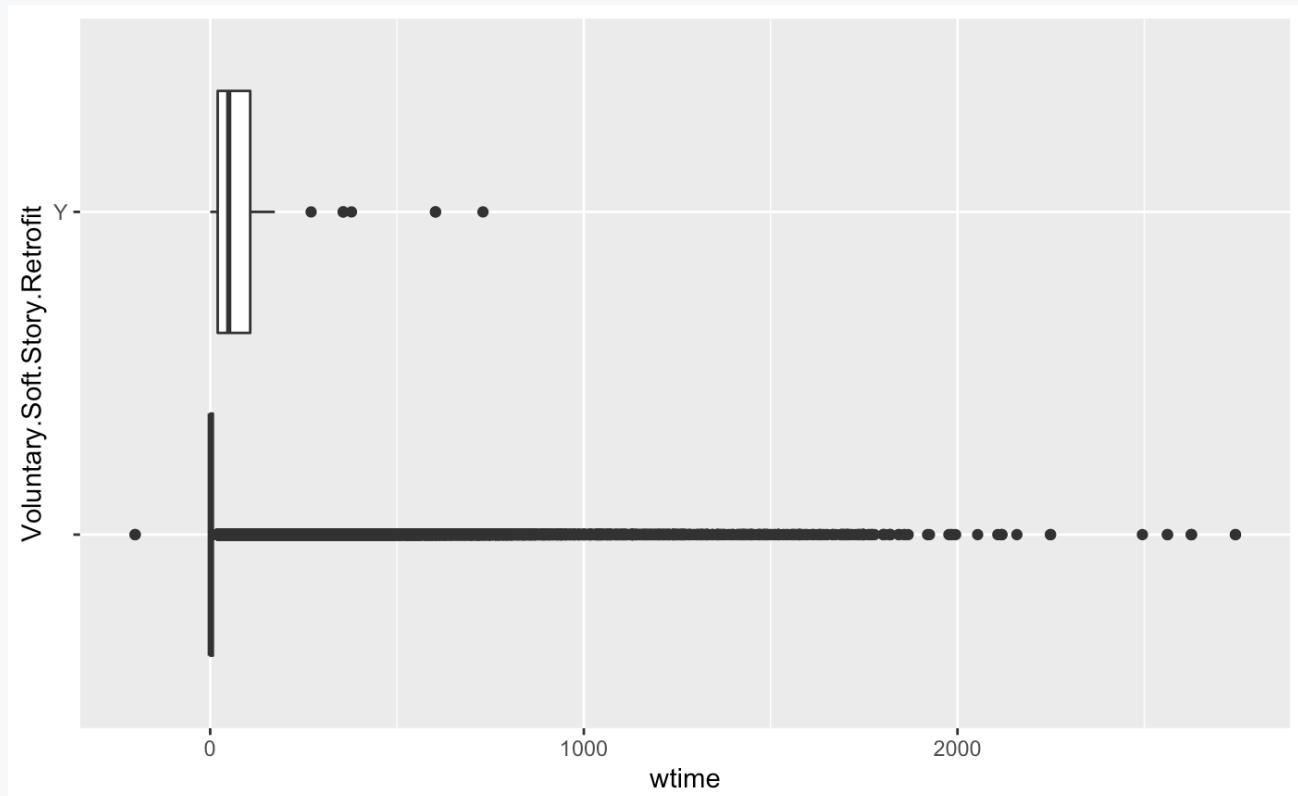
- ❖ Wait time vs.
Number.of.Proposed.Stories
- ❖ Boxplot



VISUALIZE DATA

Relationship

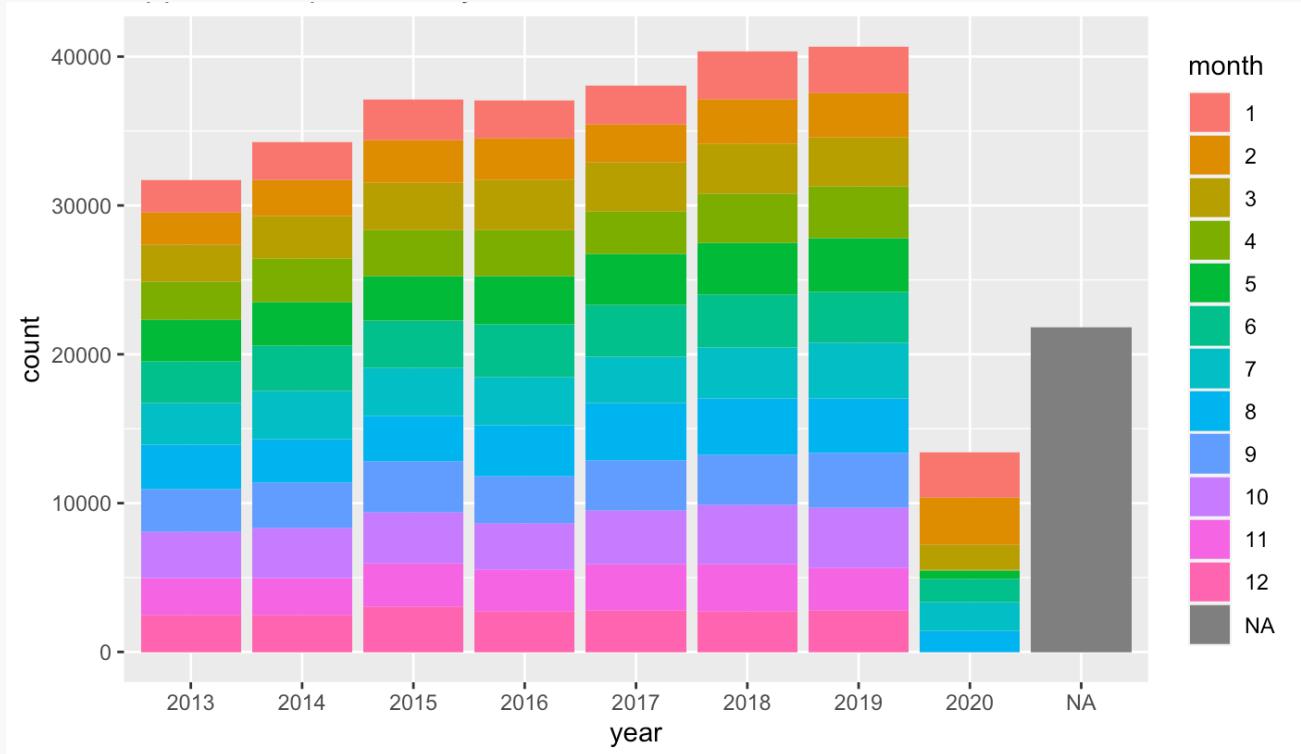
- ❖ Wait time vs.
Voluntary.Soft.Story.Retrofit
- ❖ Boxplot



VISUALIZE DATA

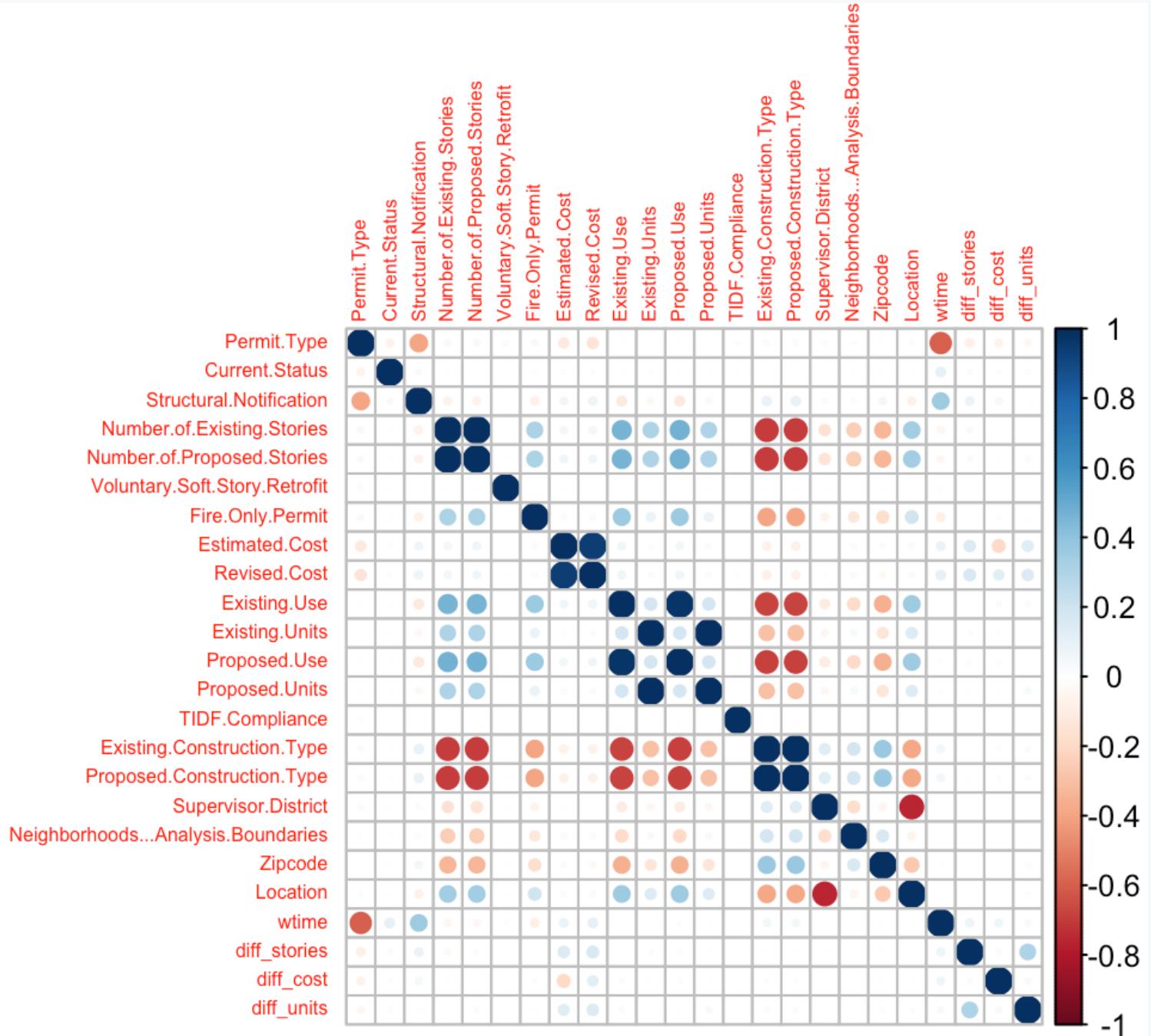
Relationship

- ❖ #applications per year by month
 - ❖ Stacked bar charts



VISUALIZE DATA Relationship

❖ Correlation Matrix



Build Model

Model 1

- ❖ split the dataset into training and testing datasets. 80% training and 20% testing
- ❖ `model1 <- lm(wtime~., data = train)`

```
Call:  
lm(formula = wtime ~ ., data = train)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-679.73 -21.78 -12.93 -1.04 2353.65  
  
Coefficients: (3 not defined because of singularities)  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) 7.983e+03 3.084e+03  2.589  0.00964 **  
Permit.Type -5.323e+01 2.250e-01 -236.537 < 2e-16 ***  
Current.Status 2.934e+00 9.376e-02 31.297 < 2e-16 ***  
Structural.Notification 8.195e+01 1.414e+00 57.938 < 2e-16 ***  
Number.of.Existing.Stories -5.427e+00 5.977e-01 -9.079 < 2e-16 ***  
Number.of.Proposed.Stories 5.462e+00 5.978e-01  9.137 < 2e-16 ***  
Voluntary.Soft.Story.Retrofit -2.135e+01 1.944e+01 -1.098  0.27213  
Fire.Only.Permit -2.326e+01 8.935e-01 -26.033 < 2e-16 ***  
Estimated.Cost -4.188e-05 2.404e-06 -17.424 < 2e-16 ***  
Revised.Cost 4.143e-05 2.294e-06 18.058 < 2e-16 ***  
Existing.Use 2.312e-01 4.913e-02  4.707 2.52e-06 ***  
Existing.Units 2.251e-01 9.650e-02  2.333 0.01967 *  
Proposed.Use -7.982e-03 4.974e-02 -0.160  0.87251  
Proposed.Units -2.224e-01 9.645e-02 -2.306 0.02114 *  
TIDF.Compliance -2.209e+01 4.346e+01 -0.508 0.61130  
Existing.Construction.Type 1.666e+00 1.989e+00  0.837 0.40234  
Proposed.Construction.Type 3.430e+00 1.974e+00  1.738 0.08220 .  
Supervisor.District 7.512e-01 1.548e-01  4.852 1.23e-06 ***  
Neighborhoods...Analysis.Boundaries -6.754e-02 2.237e-02 -3.020 0.00253 **  
Zipcode -8.095e-02 3.276e-02 -2.471 0.01347 *  
Location 3.137e-04 2.345e-05 13.378 < 2e-16 ***  
diff_stories NA NA NA NA  
diff_cost NA NA NA NA  
diff_units NA NA NA NA  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 97.15 on 149740 degrees of freedom  
Multiple R-squared:  0.3851,   Adjusted R-squared:  0.385  
F-statistic: 4689 on 20 and 149740 DF,  p-value: < 2.2e-16
```

Build Model

Model 2

- ❖ fetch the significant variables in the previous result
- ❖ Build the second model

```
Call:  
lm(formula = wtime ~ Permit.Type + Current.Status + Structural.Notification +  
    Number.of.Existing.Stories + Number.of.Proposed.Stories +  
    Fire.Only.Permit + Estimated.Cost + Revised.Cost + Existing.Use +  
    Existing.Units + Proposed.Units + Supervisor.District + Neighborhoods...Analysis.Boundaries +  
    Zipcode + Location, data = train)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-662.05 -20.89 -13.17 -1.82 2354.19  
  
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) 5.093e+03 3.082e+03 1.653 0.098433 .  
Permit.Type -5.332e+01 2.250e-01 -236.971 < 2e-16 ***  
Current.Status 2.951e+00 9.384e-02 31.450 < 2e-16 ***  
Structural.Notification 8.225e+01 1.416e+00 58.104 < 2e-16 ***  
Number.of.Existing.Stories -5.637e+00 5.967e-01 -9.446 < 2e-16 ***  
Number.of.Proposed.Stories 5.285e+00 5.970e-01 8.852 < 2e-16 ***  
Fire.Only.Permit -2.546e+01 8.851e-01 -28.768 < 2e-16 ***  
Estimated.Cost -4.176e-05 2.405e-06 -17.367 < 2e-16 ***  
Revised.Cost 4.083e-05 2.295e-06 17.791 < 2e-16 ***  
Existing.Use 1.075e-01 1.276e-02 8.424 < 2e-16 ***  
Existing.Units 2.348e-01 9.645e-02 2.434 0.014930 *  
Proposed.Units -2.391e-01 9.640e-02 -2.480 0.013136 *  
Supervisor.District 5.844e-01 1.547e-01 3.778 0.000158 ***  
Neighborhoods...Analysis.Boundaries -8.085e-02 2.237e-02 -3.613 0.000302 ***  
Zipcode -5.033e-02 3.274e-02 -1.537 0.124256  
Location 2.714e-04 2.334e-05 11.627 < 2e-16 ***  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 97.25 on 149745 degrees of freedom  
Multiple R-squared: 0.3839, Adjusted R-squared: 0.3838  
F-statistic: 6220 on 15 and 149745 DF, p-value: < 2.2e-16
```

Build Model

Model 3

- ❖ fetch the significant variables in the second result
- ❖ Build the third model

```
Call:  
lm(formula = wtime ~ Permit.Type + Current.Status + Structural.Notification +  
    Number.of.Existing.Stories + Number.of.Proposed.Stories +  
    Fire.Only.Permit + Estimated.Cost + Revised.Cost + Existing.Use +  
    Existing.Units + Proposed.Units + Supervisor.District + Neighborhoods...Analysis.Boundaries +  
    Location, data = train)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-661.67 -21.08 -13.09   -1.82 2354.20  
  
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) 3.555e+02 3.456e+00 102.867 < 2e-16 ***  
Permit.Type -5.331e+01 2.250e-01 -236.965 < 2e-16 ***  
Current.Status 2.951e+00 9.384e-02 31.450 < 2e-16 ***  
Structural.Notification 8.225e+01 1.416e+00 58.107 < 2e-16 ***  
Number.of.Existing.Stories -5.631e+00 5.967e-01 -9.437 < 2e-16 ***  
Number.of.Proposed.Stories 5.288e+00 5.970e-01 8.857 < 2e-16 ***  
Fire.Only.Permit -2.546e+01 8.851e-01 -28.767 < 2e-16 ***  
Estimated.Cost -4.178e-05 2.405e-06 -17.372 < 2e-16 ***  
Revised.Cost 4.083e-05 2.295e-06 17.794 < 2e-16 ***  
Existing.Use 1.099e-01 1.265e-02 8.690 < 2e-16 ***  
Existing.Units 2.332e-01 9.644e-02 2.418 0.015617 *  
Proposed.Units -2.374e-01 9.639e-02 -2.463 0.013772 *  
Supervisor.District 6.675e-01 1.449e-01 4.607 4.1e-06 ***  
Neighborhoods...Analysis.Boundaries -8.005e-02 2.237e-02 -3.578 0.000346 ***  
Location 2.840e-04 2.185e-05 12.999 < 2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 97.25 on 149746 degrees of freedom  
Multiple R-squared:  0.3839,    Adjusted R-squared:  0.3838  
F-statistic: 6664 on 14 and 149746 DF,  p-value: < 2.2e-16
```

Check Accuracy

Model 1&2&3

- ❖ Predict test data
- ❖ Calculate RMSE value
- ❖ The first RMSE is slightly lower

- ❖ Choose Model 1

prediction from a rank-deficient fit may be misleading
the RMSE value of the first model is
100.6784
the RMSE value of the second model is
100.7974
the RMSE value of the third model is
100.7949

Library

| FUNCTION | LIBRARY |
|----------|---|
| read.csv | <pre>library(readxl) *
* `` `r setup, include=FALSE, echo=FALSE}
#set working directory
require("knitr")
opts_knit\$set(root.dir = "~/Downloads/R_datafiles")
` ` `</pre> |
| mdy | <pre>library(lubridate)*
* Sys.setenv(TZ='America/Los_Angeles')</pre> |
| ggplot | library(ggplot2) |
| mutate | library(dplyr) |
| gather | library(tidyr) |
| corrplot | library(corrplot) |