

Data exploration for modeling property price

LYT

2019.3.21

Update 25 March

Talked to Chief editor of International Journal of Data Analytics: Web of Science, Scopus, Inspec, PsycINFO, Ei Compendex, and so on

Project objective

price estimation, updating past property worth records in the dataset to the present value

I will try to build a model that is able to predict current prices of houses in record, based on their attribute, spatial information and historical transactions to reflect the market value of them. This helps to capture the economic risk of tidal flooding in a more intuitive way.

Timeline

late Mar: a trial machine learning algorithm (Langyi more on this part)

May: a good visualization (also for class project, Bin and Julia will be more involved here)

Methodology

Can try simple tree, random forest, and boosting) first for now. Random variable selection limitation can help to reduce correlated features issue. Can have variable importance measure. More robust model.

No dimensionality reduction for now, cuz there are not so many variables yet.

Will use package randomForest and caret.

Feed multiple models across parcels. Consider ZIP/county/metrollevel or other indicators of neighbourhood. But whether to use the same group of variables can be an issue since some variables miss by county.

How to deal with variables that change over time? Variable of time can be put into model itself, but introducing others will bring covariates.

If use CV: training vs. testset, how to divide? Across parcels? Look at descriptive statistics across counties to think about heterogeneity (there's difference between small/big/rural/urban counties). What about 10-fold or LOOCA?

literature review:

In terms of property valuation, mainstream method is parametric hedonic regression. Machine learning came into application recently. 2 papers are the most relevant for now.

Barr et al. (2017) used gradient boosting trees (offers some interpretability) to estimate individual home price at each periods to construct a house price index. They suggested that local aggregation (metro, county, state, etc.) is more appropriate than global aggregation, as local trends depart from general trend from time to time. They raised the idea of "submarket" as cohort of houses that competing for the same group of people. Therefore, they run many millions of models across geographic hierarchies (but didn't say more specifically). They didn't mention the variables they are using and whether they perform data reduction though.

Garcia-Magarino et al. (2019) tested several machine learning and dimensionality reduction methods to address the problem of estimating the missing prices of a sample of houses. They tried OLS, KNN, SVR (an adaptation of SVM), and Artificial neural networks. Dimensionality reduction methods included Non-negative Matrix Factorization, Recursive Feature Elimination, and Forward Selection.

Data description

There are three sets of data records utilized in the project:

1. home attribute data (codebook: 5.0 Tax Assessor Layout)
2. sales records/transactions (codebook: 5.0 Recorder Layout)
3. flooding risk / environmental variables parcel level variables (parcel risk and spatial data) and one set of polygon parcel boundaries for most of the parcels in these three datasets (sef_parcel.zip)

The parcel attribute and sales data both have an identifier (attomid) for each property. The parcel risk / spatial data file can be joined to this data as it also contains attomid. The parcel polygons and parcel risk / spatial data both have another id (fsid / firststreetid) that can be used to combine each unique parcel.

The parcel risk / spatial data file contains fields that represent the inundation risk with field lengths of 6 or 8, e.g. ltc118, rdk127, mdc118. The first two characters (lt, rd, md, np) represent whether the statistic is about the proportion of the lot, the proportion of roads nearby, the

max depth of inundation on the lot (ceiled to feet), or the proportion of nearby properties impacted. The next two characters (kt, em, c1, c3, c5) represent the risk type: kt for repeated king tides, em for highest annual tide, and c1, c3, c5 for hurricane types. The next two characters represent the year for the risk, 18 for 2018, 23 for 2023, etc. If you find te or qu as characters 7 and 8, it identifies the spatial radius used for the measure, tenth of a mile or a quarter mile.

Considering the input variables, tax data can be used to adjust for market price in cases of missing information; the home attributes data vary by county so we should consider hierarchical modelling, if the trial model reveals significance of the unique variables; need to find a proper way to aggregate environment data.

The home characteristics data

In local desktop I only imported 10000 obs for trial.

```
##
## 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

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##   between, first, last
```

Brief summary

```
## Data include state: FL ,and county: Broward Miami-Dade Palm Beach Collier Lee Monroe Hendry In each state
the number of samples are

## # A tibble: 7 x 4
##   situscounty `n()` `mean(assessorlastsaleamount~`mean(areabuilding, na.~
##   <chr>      <int>      <dbl>      <dbl>
## 1 Broward    759392      344927.    2207.
## 2 Collier    290821      886415.    1500.
## 3 Hendry      35908      576729.     814.
## 4 Lee        574084      347050.    1390.
## 5 Miami-Dade 931150      441260.    2406.
## 6 Monroe     91360      316645.    1007.
## 7 Palm Beach 645208      492636.    2237.
```

Considerable level of heterogenity by county might exist in data.

Select and recode useful variables

Modeling by county

```
## $Broward
##      attomid situsstatecode situscounty ownertypedescription1
##      1: 141117808           FL      Broward      INDIVIDUAL
##      2: 141189293           FL      Broward      INDIVIDUAL
##      3:   178194           FL      Broward      INDIVIDUAL
##      4:   879797           FL      Broward      INDIVIDUAL
##      5:  2408813           FL      Broward      INDIVIDUAL
##      ---
## 759388: 15106042           FL      Broward      INDIVIDUAL
## 759389: 162575311          FL      Broward      INDIVIDUAL
## 759390:  1997509          FL      Broward      INDIVIDUAL
## 759391:  7165857          FL      Broward      INDIVIDUAL
## 759392:  51459910          FL      Broward      INDIVIDUAL
```

##	ownertypedescription2	yearbuilt	propertyusegroup	deedlastsaledate			
##	1:	UNKNOWN	1955	Residential			
##	2:	INDIVIDUAL	1957	Residential			
##	3:	UNKNOWN	1957	Residential			
##	4:	UNKNOWN	1956	Residential			
##	5:	UNKNOWN	1953	Residential			
##	---						
##	759388:	UNKNOWN	1965	Residential			
##	759389:	UNKNOWN	1968	Residential			
##	759390:	INDIVIDUAL	1963	Residential			
##	759391:	UNKNOWN	1964	Residential			
##	759392:	UNKNOWN	1963	Residential			
##	areabuilding	parkinggarage	parkinggaragearea	hvacheatingdetail			
##	1:	1292	1	0			
##	2:	2178	1	0			
##	3:	1133	1	0			
##	4:	1363	1	0			
##	5:	1192	1	0			
##	---						
##	759388:	1485	1	0			
##	759389:	2639	1	0			
##	759390:	1235	1	0			
##	759391:	1433	1	0			
##	759392:	5407	1	0			
##	hvacheatingfuel	construction	plumbingfixturescount	bathcount			
##	1:	380	23	0			
##	2:	380	23	0			
##	3:	380	23	0			
##	4:	380	23	0			
##	5:	380	23	0			
##	---						
##	759388:	380	23	0			
##	759389:	380	23	0			
##	759390:	380	23	0			
##	759391:	380	23	0			
##	759392:	380	23	0			
##	bathpartialcount	bedroomscount	roomscount	storiescount	unitscount		
##	1:	0	3	0	1		
##	2:	0	4	0	1		
##	3:	0	2	0	1		
##	4:	0	3	0	1		
##	5:	0	3	0	1		
##	---						
##	759388:	0	2	0	1		
##	759389:	0	3	0	1		
##	759390:	0	2	0	1		
##	759391:	0	3	0	1		
##	759392:	0	5	0	2		
##	fireplacecount	exteriorlcode	roofmaterial	viewdescription			
##	1:	0	5G	105			
##	2:	0	5G	105			
##	3:	0	5G	105			
##	4:	0	5G	105			
##	5:	0	5G	109			
##	---						
##	759388:	0	5G	114			
##	759389:	0	5G	105			
##	759390:	0	5G	104			
##	759391:	0	5G	105			
##	759392:	0	5G	125			
##	porchcode	porcharea	patioarea	deckflag	deckarea	drivewayarea	pool
##	1:	0	0	0	0	0	999
##	2:	0	0	0	0	0	999
##	3:	0	120	0	0	0	0
##	4:	0	0	0	0	0	999
##	5:	0	0	0	0	0	0
##	---						
##	759388:	120	0	0	0	0	0
##	759389:	735	0	0	0	0	999
##	759390:	0	0	0	0	0	0
##	759391:	0	0	0	0	0	999
##	759392:	0	450	0	0	0	999

```

##      poolarea fencearea arenaflag buildingscount shedcode
##      1:      448          0          0              1          0
##      2:      265          0          0              1          0
##      3:         0          0          0              1          0
##      4:      450          0          0              1          0
##      5:         0          0          0              1          0
##      ---
## 759388:         0          0          0              1          0
## 759389:      450          0          0              1          0
## 759390:         0          0          0              1          0
## 759391:      288          0          0              1          0
## 759392:      450          0          0              1          0
##      utilitybuildingarea
##      1:              0
##      2:              0
##      3:              0
##      4:              0
##      5:              0
##      ---
## 759388:              0
## 759389:              0
## 759390:              0
## 759391:              0
## 759392:              0
##
## $Collier
##      attomid situsstatecode situscounty ownertypedescription1
##      1: 160335976              FL      Collier              INDIVIDUAL
##      2: 23874236              FL      Collier              INDIVIDUAL
##      3: 20726915              FL      Collier              INDIVIDUAL
##      4: 148407949              FL      Collier              INDIVIDUAL
##      5: 148397678              FL      Collier              INDIVIDUAL
##      ---
## 290817: 253162436              FL      Collier              COMPANY
## 290818: 250970817              FL      Collier              INDIVIDUAL
## 290819: 250970816              FL      Collier              COMPANY
## 290820: 252437433              FL      Collier              COMPANY
## 290821: 250797939              FL      Collier              INDIVIDUAL
##      ownertypedescription2 deedlastsaledate parkinggarage
##      1:              UNKNOWN              1
##      2:              UNKNOWN      2017-07-21              1
##      3:              NP      2018-12-20              1
##      4:              UNKNOWN      2002-06-24              1
##      5:              UNKNOWN      2007-12-12              1
##      ---
## 290817:              UNKNOWN      2018-10-01              1
## 290818:              UNKNOWN      2005-02-08              1
## 290819:              UNKNOWN      2013-08-04              1
## 290820:              UNKNOWN      2016-04-18              1
## 290821:              NP      2018-12-10              1
##      parkinggaragearea hvacheatingdetail hvacheatingfuel
##      1:              0              219              380
##      2:              0              219              380
##      3:              0              219              380
##      4:              0              219              380
##      5:              0              219              380
##      ---
## 290817:              0              219              380
## 290818:              0              219              380
## 290819:              0              219              380
## 290820:              0              219              380
## 290821:              0              219              380
##      plumbingfixturescount bathcount bathpartialcount bedroomscount
##      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
##      ---
## 290817:              0              0              0              0
## 290818:              0              0              0              0
## 290819:              0              0              0              0
## 290820:              0              0              0              0

```

```

## 290821:          0          0          0          0
##      roomscount storiescount unitscount fireplacecount exteriorlcode
##      1:          0          0          1          0
##      2:          0          1          1          0
##      3:          0          0          0          0
##      4:          0          1          1          0
##      5:          0          1          1          0
##      ---
## 290817:          0          0          0          0
## 290818:          0          1          0          0
## 290819:          0          1          0          0
## 290820:          0          1          0          0
## 290821:          0          2          0          0
##      roofmaterial viewdescription porchcode porcharea patioarea
##      1:          101          0          0          0
##      2:          101          0          0          0
##      3:          101          0          0          0
##      4:          101          0          0          0
##      5:          101          0          0          0
##      ---
## 290817:          101          0          0          0
## 290818:          101          0          0          0
## 290819:          101          0          0          0
## 290820:          101          0          0          0
## 290821:          101          0          0          0
##      deckflag deckarea drivewayarea pool poolarea fencearea arenaflag
##      1:          0          0          0  0          0          0          0
##      2:          0          0          0  0          0          0          0
##      3:          0          0          0  0          0          0          0
##      4:          0          0          0  0          0          0          0
##      5:          1      1740          0  0          0          0          0
##      ---
## 290817:          0          0          0  0          0          0          0
## 290818:          0          0          0  0          0          0          0
## 290819:          1      425          0  0          0          0          0
## 290820:          1      169          0  0          0          0          0
## 290821:          1      318          0  0          0          0          0
##      buildingscount shedcode utilitybuildingarea
##      1:          0          0          0
##      2:          0          0          0
##      3:          0          0          0
##      4:          0          0          0
##      5:          0          0          0
##      ---
## 290817:          0          0          0
## 290818:          0          0          0
## 290819:          0          0          0
## 290820:          0          0          0
## 290821:          0          0          0
##
## $Hendry
##      attomid situsstatecode situscounty ownertypedescription1
##      1: 141049468          FL      Hendry      INDIVIDUAL
##      2: 163117513          FL      Hendry      INDIVIDUAL
##      3: 163122117          FL      Hendry      INDIVIDUAL
##      4: 148843959          FL      Hendry      INDIVIDUAL
##      5: 28687187          FL      Hendry      INDIVIDUAL
##      ---
## 35904: 159627697          FL      Hendry      INDIVIDUAL
## 35905: 153662938          FL      Hendry      INDIVIDUAL
## 35906: 163107964          FL      Hendry      INDIVIDUAL
## 35907: 163112053          FL      Hendry      INDIVIDUAL
## 35908: 19145191          FL      Hendry      INDIVIDUAL
##      ownertypedescription2 propertyusegroup deedlastsaledate
##      1:      INDIVIDUAL      Residential      2018-03-02
##      2:      UNKNOWN      Residential      2009-07-27
##      3:      INDIVIDUAL      Residential      2018-08-31
##      4:      UNKNOWN      Residential      2016-06-27
##      5:      UNKNOWN      Residential      2017-01-09
##      ---
## 35904:      UNKNOWN      Residential      2017-03-13
## 35905:      UNKNOWN      Residential      2017-02-21
## 35906:      UNKNOWN      Residential      1998-10-01

```

##	35907:	UNKNOWN	Residential	1997-09-04				
##	35908:	INDIVIDUAL	Residential	2010-04-23				
##		parkinggarage	parkinggaragearea	hvacheatingdetail	hvacheatingfuel			
##	1:	1	0	102	80			
##	2:	NA	576	102	80			
##	3:	1	0	102	80			
##	4:	1	0	102	80			
##	5:	1	0	102	80			
##	---							
##	35904:	1	0	102	80			
##	35905:	1	0	102	80			
##	35906:	NA	960	102	80			
##	35907:	1	0	102	80			
##	35908:	1	0	102	80			
##		construction	plumbingfixturescount	bathcount	bathpartialcount			
##	1:	15	0	2	0			
##	2:	15	0	2	0			
##	3:	15	0	2	0			
##	4:	15	0	2	0			
##	5:	15	0	2	0			
##	---							
##	35904:	20	0	2	0			
##	35905:	15	0	2	0			
##	35906:	15	0	2	0			
##	35907:	15	0	2	0			
##	35908:	25	0	3	0			
##		bedroomscount	roomscount	storiescount	unitscount	fireplacecount		
##	1:	3	0	1	1	0		
##	2:	3	0	1	1	1		
##	3:	3	0	1	1	0		
##	4:	3	0	1	1	0		
##	5:	3	0	1	1	0		
##	---							
##	35904:	3	0	1	1	0		
##	35905:	3	0	1	1	0		
##	35906:	3	0	1	1	0		
##	35907:	3	0	1	1	0		
##	35908:	4	0	2	1	0		
##		exteriorlcode	roofmaterial	viewdescription	porchcode	porcharea		
##	1:	17A	114	0	P00	207		
##	2:	17A	109	0	P00	1080		
##	3:	17A	109	0		0		
##	4:	17A	109	0		0		
##	5:	17A	109	0		0		
##	---							
##	35904:	9J	112	0		186		
##	35905:	17A	109	0		0		
##	35906:	17A	109	0	P00	230		
##	35907:	17A	109	0	P00	333		
##	35908:	27A	114	0	P00	1493		
##		patioarea	deckflag	deckarea	drivewayarea	pool	poolarea	fencearea
##	1:	207	0	0	0	0	0	0
##	2:	312	0	0	0	0	0	0
##	3:	192	0	0	0	0	0	0
##	4:	48	0	0	0	0	0	0
##	5:	0	0	0	0	0	0	0
##	---							
##	35904:	56	0	0	0	0	0	0
##	35905:	518	0	0	0	0	0	0
##	35906:	0	0	0	0	0	0	0
##	35907:	0	0	0	0	0	0	0
##	35908:	0	0	0	0	999	2736	0
##		arenaflag	buildingscount	shedcode	utilitybuildingarea			
##	1:	0	0	0				
##	2:	0	0	0				
##	3:	0	0	0				
##	4:	0	0	0				
##	5:	0	0	0				
##	---							
##	35904:	0	0	0				
##	35905:	0	0	0				
##	35906:	0	0	0				

```

## 35907:      0      0      0      0
## 35908:      0      0      0      0
##
## $Lee
##      attomid situsstatecode situscounty ownertypedescription1
##      1: 26605522      FL      Lee      COMPANY
##      2: 20123873      FL      Lee      COMPANY
##      3: 20182236      FL      Lee      COMPANY
##      4: 20320009      FL      Lee      INDIVIDUAL
##      5: 22127199      FL      Lee      INDIVIDUAL
##      ---
## 574080: 49262118      FL      Lee      COMPANY
## 574081: 147605396      FL      Lee      INDIVIDUAL
## 574082: 24587995      FL      Lee      INDIVIDUAL
## 574083: 149078865      FL      Lee      INDIVIDUAL
## 574084: 147603518      FL      Lee      COMPANY
##      ownertypedescription2 propertyusegroup deedlastsaledate
##      1:      UNKNOWN      Residential      2016-10-19
##      2:      NP      Residential      2017-10-11
##      3:      UNKNOWN      Residential      2013-12-03
##      4:      NP      Residential      2004-07-19
##      5:      UNKNOWN      Residential      2014-12-03
##      ---
## 574080:      UNKNOWN      Residential      2018-09-27
## 574081:      UNKNOWN      Residential      2011-12-05
## 574082:      COMPANY      Residential      1981-07-01
## 574083:      UNKNOWN      Residential      2012-09-12
## 574084:      NP      Residential      2015-03-11
##      parkinggaragearea hvacheatingdetail hvacheatingfuel
##      1:      0      219      380
##      2:      441      102      80
##      3:      0      219      380
##      4:      0      219      380
##      5:      847      102      80
##      ---
## 574080:      0      219      380
## 574081:      0      102      80
## 574082:      0      219      380
## 574083:      0      219      380
## 574084:      448      102      80
##      plumbingfixturescount bathcount bathpartialcount bedroomscount
##      1:      0      0      0      0
##      2:      0      2      0      3
##      3:      0      2      0      2
##      4:      0      0      0      0
##      5:      0      3      0      3
##      ---
## 574080:      0      0      0      0
## 574081:      0      2      0      3
## 574082:      0      0      0      0
## 574083:      0      0      0      0
## 574084:      0      2      0      3
##      roomscount storiescount unitscount fireplacecount exteriorlcode
##      1:      0      0      0      0
##      2:      0      1      0      0      80A
##      3:      0      2      1      0
##      4:      0      0      0      0
##      5:      0      1      1      0      80A
##      ---
## 574080:      0      0      1      0
## 574081:      0      1      1      0      80A
## 574082:      0      0      0      0
## 574083:      0      0      0      0
## 574084:      0      1      0      0      80A
##      roofmaterial viewdescription porchcode porcharea patioarea
##      1:      101      0      0      0
##      2:      109      0      P00      256      0
##      3:      101      V01      0      0
##      4:      101      0      0      0
##      5:      109      0      P00      631      1914
##      ---
## 574080:      101      0      0      400
## 574081:      109      0      160      0

```

```

## 574082:      101      0      0      0
## 574083:      101      0      0      0
## 574084:      109      0      701      0
##      deckflag deckarea drivewayarea pool poolarea fencearea arenaflag
##      1:      0      0      0      0      0      0      0
##      2:      0      0      0      0      0      0      0
##      3:      0      0      0      999      0      0      0
##      4:      0      0      0      0      0      0      0
##      5:      0      0      0      999      392      0      0
##      ---
## 574080:      0      0      0      0      0      0      0
## 574081:      0      0      0      0      0      0      0
## 574082:      0      0      0      0      0      0      0
## 574083:      0      0      0      0      0      0      0
## 574084:      0      0      0      0      0      0      0
##      buildingscount shedcode utilitybuildingarea
##      1:      0      0      0
##      2:      0      0      0
##      3:      0      0      0
##      4:      0      0      0
##      5:      0      0      0
##      ---
## 574080:      0      0      0
## 574081:      0      0      0
## 574082:      0      0      0
## 574083:      0      0      0
## 574084:      0      0      0
##
## $`Miami-Dade`
##      attomid situsstatecode situscounty ownertypedescription1
##      1:  14950802      FL  Miami-Dade      COMPANY
##      2:  15037579      FL  Miami-Dade      INDIVIDUAL
##      3:  15137977      FL  Miami-Dade      INDIVIDUAL
##      4:  15990253      FL  Miami-Dade      INDIVIDUAL
##      5:  16109222      FL  Miami-Dade      INDIVIDUAL
##      ---
## 931146: 162811015      FL  Miami-Dade      COMPANY
## 931147:  22090001      FL  Miami-Dade      INDIVIDUAL
## 931148: 149116489      FL  Miami-Dade      INDIVIDUAL
## 931149: 147616397      FL  Miami-Dade      COMPANY
## 931150: 140765760      FL  Miami-Dade      COMPANY
##      ownertypedescription2 yearbuilt propertyusegroup deedlastsaledate
##      1:      UNKNOWN      1980      Residential      2016-12-27
##      2:      UNKNOWN      1980      Residential      2007-10-15
##      3:      UNKNOWN      1980      Residential      2016-09-21
##      4:      UNKNOWN      1980      Residential      2013-07-27
##      5:      UNKNOWN      1980      Residential      2009-07-09
##      ---
## 931146:      UNKNOWN      1958      Residential      2016-01-12
## 931147:      UNKNOWN      1980      Residential      2005-01-28
## 931148:      COMPANY      1946      Residential
## 931149:      UNKNOWN      2007      Residential      2009-12-30
## 931150:      UNKNOWN      1972      Residential      2015-03-23
##      areabuilding parkinggaragearea hvacheatingdetail hvacheatingfuel
##      1:      1240      0      219      380
##      2:      896      0      219      380
##      3:      1240      0      219      380
##      4:      896      0      219      380
##      5:      1008      0      219      380
##      ---
## 931146:      720      0      219      380
## 931147:      1165      0      219      380
## 931148:      798      0      219      380
## 931149:      832      0      219      380
## 931150:      3630      0      219      380
##      plumbingfixturescount bathcount bathpartialcount bedroomscount
##      1:      0      2      0      2
##      2:      0      1      0      1
##      3:      0      2      0      2
##      4:      0      1      0      1
##      5:      0      2      0      2
##      ---
## 931146:      0      1      0      1

```



```

## 931147:      0      2      0      2
## 931148:      0      1      0      2
## 931149:      0      1      0      1
## 931150:      0      2      0      4
##      roomscount storiescount unitscount fireplacecount exteriorlcode
##      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
##      ---
## 931146:      0      0      0      0      0A
## 931147:      0      0      0      0
## 931148:      0      1      1      0      5G
## 931149:      0      0      0      0
## 931150:      0      2      1      0      5G
##      roofmaterial viewdescription porchcode porcharea patioarea
##      1:      101      0      0      0
##      2:      101      0      0      0
##      3:      101      0      0      0
##      4:      101      0      0      0
##      5:      101      0      0      0
##      ---
## 931146:      101      0      0      0
## 931147:      101      0      0      0
## 931148:      114      0      P00      90      0
## 931149:      101      0      0      0
## 931150:      114      0      0      718
##      deckflag deckarea drivewayarea pool poolarea fencearea arenaflag
##      1:      0      0      0      0      0      0      0
##      2:      0      0      0      0      0      0      0
##      3:      0      0      0      0      0      0      0
##      4:      0      0      0      0      0      0      0
##      5:      0      0      0      0      0      0      0
##      ---
## 931146:      0      0      0      0      0      0      0
## 931147:      0      0      0      0      0      0      0
## 931148:      0      0      0      0      0      0      0
## 931149:      0      0      0      0      0      0      0
## 931150:      0      0      0      0      0      0      0
##      buildingscount shedcode utilitybuildingarea
##      1:      1      0      0
##      2:      1      0      0
##      3:      1      0      0
##      4:      1      0      0
##      5:      1      0      0
##      ---
## 931146:      1      0      0
## 931147:      1      0      0
## 931148:      1      0      0
## 931149:      1      0      0
## 931150:      1      0      0
##
## $Monroe
##      attomid situsstatecode situscounty ownertypedescription1
##      1: 148408544      FL      Monroe      INDIVIDUAL
##      2: 46411568      FL      Monroe      INDIVIDUAL
##      3: 144511137      FL      Monroe      INDIVIDUAL
##      4: 144499266      FL      Monroe      COMPANY
##      5: 144500214      FL      Monroe      INDIVIDUAL
##      ---
## 91356: 13544033      FL      Monroe      COMPANY
## 91357: 149616518      FL      Monroe      INDIVIDUAL
## 91358: 320463113      FL      Monroe      COMPANY
## 91359: 225465923      FL      Monroe      COMPANY
## 91360: 243710828      FL      Monroe      INDIVIDUAL
##      ownertypedescription2 deedlastsaledate parkinggaragearea
##      1:      UNKNOWN      2009-10-15      0
##      2:      UNKNOWN      2014-05-14      0
##      3:      NP      2018-02-12      0
##      4:      UNKNOWN      0
##      5:      UNKNOWN      2015-02-24      0

```

```

##      ---
## 91356:      UNKNOWN      2012-09-18      0
## 91357:      UNKNOWN      2016-06-16      0
## 91358:      UNKNOWN      0
## 91359:      UNKNOWN      0
## 91360:      COMPANY      3920
## hvacheatingdetail hvacheatingfuel construction
##      1:      219      380      0
##      2:      102      380      0
##      3:      219      380      0
##      4:      219      380      0
##      5:      219      380      0
##      ---
## 91356:      219      380      0
## 91357:      219      380      0
## 91358:      219      380      0
## 91359:      219      380      0
## 91360:      219      380      0
## plumbingfixturescount bathcount bathpartialcount bedroomscount
##      1:      0      1      0      2
##      2:      0      2      0      3
##      3:      0      0      0      0
##      4:      0      0      0      0
##      5:      0      2      0      3
##      ---
## 91356:      0      0      0      0
## 91357:      0      1      0      2
## 91358:      NA      0      0      0
## 91359:      0      0      0      0
## 91360:      0      0      0      0
## roomscount storiescount unitscount fireplacecount exteriorlcode
##      1:      0      1      1      0      86A
##      2:      0      1      1      0      5G
##      3:      0      0      0      0      0A
##      4:      0      0      0      0      0A
##      5:      0      1      1      0      11A
##      ---
## 91356:      0      0      0      0      0A
## 91357:      0      1      1      0      86A
## 91358:      0      0      0      NA      0A
## 91359:      0      0      0      0      0A
## 91360:      0      1      0      0      7Q
## roofmaterial viewdescription porchcode porcharea patioarea deckflag
##      1:      107      0      812      50      1
##      2:      101      0      P00      784      1910      1
##      3:      101      0      0      0      0
##      4:      101      0      0      0      0
##      5:      101      0      P00      98      937      0
##      ---
## 91356:      101      0      0      900      1
## 91357:      101      0      180      48      0
## 91358:      101      0      NA      NA      0
## 91359:      101      0      0      0      0
## 91360:      109      0      P00      5356      276      1
## deckarea drivewayarea pool poolarea fencearea arenaflag
##      1:      50      0      0      0      0
##      2:      658      0      230      468      0
##      3:      0      0      0      0      0
##      4:      0      0      0      0      0
##      5:      0      0      230      261      0
##      ---
## 91356:      900      0      0      0      0
## 91357:      0      0      0      0      0
## 91358:      NA      NA      0      NA      0
## 91359:      0      0      0      0      0
## 91360:      24      0      0      0      0
## buildingscount shedcode utilitybuildingarea
##      1:      0      0      0
##      2:      0      0      0
##      3:      0      0      0
##      4:      0      0      0
##      5:      0      0      0
##      ---

```

```

## 91356:          0          0          0
## 91357:          0          0          0
## 91358:         NA          0          0
## 91359:          0          0          0
## 91360:          0          0          0
##
## $`Palm Beach`
##      attomid situsstatecode situscounty ownertypedescription1
##      1: 161476388          FL  Palm Beach          INDIVIDUAL
##      2: 20181118          FL  Palm Beach          INDIVIDUAL
##      3: 53818122          FL  Palm Beach          INDIVIDUAL
##      4: 53818127          FL  Palm Beach          INDIVIDUAL
##      5: 53818130          FL  Palm Beach          INDIVIDUAL
##      ---
## 645204: 140761707          FL  Palm Beach          INDIVIDUAL
## 645205: 140759288          FL  Palm Beach          INDIVIDUAL
## 645206: 161512073          FL  Palm Beach          INDIVIDUAL
## 645207: 149079116          FL  Palm Beach          INDIVIDUAL
## 645208: 43406506          FL  Palm Beach          INDIVIDUAL
##      ownertypedescription2 yearbuilt propertyusegroup deedlastsaledate
##      1:          UNKNOWN      1981      Residential      2004-11-23
##      2:          UNKNOWN      2004      Residential      2014-07-03
##      3:          UNKNOWN      1970      Residential      2009-03-02
##      4:          UNKNOWN      1961      Residential      2010-06-03
##      5:          UNKNOWN      1961      Residential      2010-03-29
##      ---
## 645204:          UNKNOWN      2000      Residential      2012-10-17
## 645205:          UNKNOWN      1993      Residential      2012-12-03
## 645206:          UNKNOWN      1988      Residential      1996-12-03
## 645207:          UNKNOWN      1973      Residential      2014-04-10
## 645208:      INDIVIDUAL      2003      Residential      2005-08-24
##      areabuilding parkinggaragearea hvacheatingdetail hvacheatingfuel
##      1:          1020          297          102          80
##      2:          2689          462          219          380
##      3:          1200           0          219          380
##      4:           854           0          219          380
##      5:           854           0          219          380
##      ---
## 645204:          2763          386          219          380
## 645205:          1829          420          102           80
## 645206:          3039          483          219          380
## 645207:           883           0          219          380
## 645208:          2263          464          102           80
##      construction plumbingfixturescount bathcount bathpartialcount
##      1:           0           0           2           0
##      2:           0           0           3           0
##      3:           0           0           2           0
##      4:           0           0           1           0
##      5:           0           0           1           0
##      ---
## 645204:           0           0           3           0
## 645205:           0           0           2           0
## 645206:           0           0           4           0
## 645207:           0           0           2           0
## 645208:           0           0           2           0
##      bedroomscount roomscount storiescount unitscount fireplacecount
##      1:           2           0           1           1           0
##      2:           4           0           0           0           0
##      3:           2           0           0           1           0
##      4:           2           0           0           0           0
##      5:           2           0           0           0           0
##      ---
## 645204:           4           0           0           1           0
## 645205:           3           0           1           1           0
## 645206:           3           0           1           1           0
## 645207:           2           0           0           1           0
## 645208:           2           0           1           0           0
##      exteriorlcode roofmaterial viewdescription porchcode porcharea
##      1:          80A          109           0           0           0
##      2:           0A          101           0          P00          78
##      3:           0A          101           0           0           0
##      4:           0A          101           0           0           0
##      5:           0A          101           0           0           0

```

```
##      ---
## 645204:      0A      101      0      P00      530
## 645205:      80A      128      0      238
## 645206:      0A      101      0      290
## 645207:      0A      101      0      108
## 645208:      80A      128      0      223
##      patioarea deckflag deckarea drivewayarea pool poolarea fencearea
##      1:      947      0      0      0 999      75      0
##      2:      0      0      0      0 999      0      0
##      3:      0      0      0      0 0      0      0
##      4:      0      0      0      0 0      0      0
##      5:      0      0      0      0 0      0      0
##      ---
## 645204:      304      0      0      0 0      0      0
## 645205:      0      0      0      0 0      0      0
## 645206:      1810      0      0      0 999      83      0
## 645207:      0      0      0      0 0      0      0
## 645208:      132      0      0      0 0      0      0
##      arenaflag buildingscount shedcode utilitybuildingarea
##      1:      0      0      0      0
##      2:      0      1      0      0
##      3:      0      1      0      0
##      4:      0      1      0      0
##      5:      0      1      0      0
##      ---
## 645204:      0      1      0      0
## 645205:      0      0      0      0
## 645206:      0      1      0      0
## 645207:      0      1      0      0
## 645208:      0      0      0      0
```

Transaction data

```
trx_dta<-fread("D:/raw_data/SF_Sales_Transactions_Data.csv")
```

The variable useful here is pretty straightforward: transaciton price 23 TransferAmount. After recoding it with “kick out <\$1000” methods (this already kicks out nearly half in the sample) the result is still not so satisfying. Maybe think about adjust according to price per sq feet?

```
trx_dta[transferamount[trx_dta]transferamount<1000]<-NA trx_dta\
(transferamount%>%sort(decreasing=FALSE)%>%unique())%>%head(10) trx_dta)transferamount%>%summary()
```

Regarding sales time, 2 variables look like compensating each other, but actually not sure: 13 InstrumentDate and 14 RecordingDate.

Parcel/risk data

```
risk_dta<-fread("D:/raw_data/SF_Parcel_Risk_and_Spatial_Data.csv")
```

One thing I’m still not clear is the geographical mapping that has been made. I understand one parcel division in the paper—properties with different levels of flooding risk. But you also mention Inverse Distance Weighted (IDW) to build value surfaces—is this automatic or manually tuned? Can we use it to construct neibourhood parcels that divides between models?

Besides, from 132 Totpopbg to 145 Hisptr seems to record ethnicity background information. How is that recorded and what’s the unit?

A series of interesting variables

```
for (i in 132:180){ a<-risk_dta[,i]%>%length() b<-risk_dta[,i]%>%is.na()%>%sum() if (b/a>0.1){ print(colnames(risk_dta[i])) } }
```

Apart from MedIncbg are all missing, other variables are pretty complete for analysis.