FINDING A BETTER ESTIMATE

Case Study of Zestimate

by Kylie Zhang

DATA (#1)

Information Provided

Key variable

“PropertyID”: ID (identifier of a home)

Descriptive variable

“Street”: address (location of a home)

Actual transactions

“SaleDate”: Sale date
“SaleAmount”: Sale price

Zestimate

“ZestimateOnSaleDate”: Zestimate of homes sold
“ZestimateAmount”: Zestimate on July 1 in a year

Preliminary Findings

Almost every home has a Zestimate

Among 14,190 homes on file, 13,998 (98.6%) have Zestimate, including Zestimate when a home is sold and Zestimate routinely performed to all homes between 1997 and 2007.

Most homes were NOT sold

Among 14,190 homes on file, 8,075 (56.9%) were never sold between 1996 and 2007.

* Sources: Homes.csv, Transaction.csv, and ZestimateHistory.csv.

TOOLS AND ANALYSES (#1 CONT.)

SAS/SQL

Powerful tool to process large datasets

Data manipulation

Standardize values in "Street"

Data consolidation

Merge data from multiple sources

Statistical test

Compare two populations with paired T-test

Summary statistics

Calculate max, min, mean, median, by groups

Tableau/Excel

Intuitive tools to explore/analyze data

Initial exploration

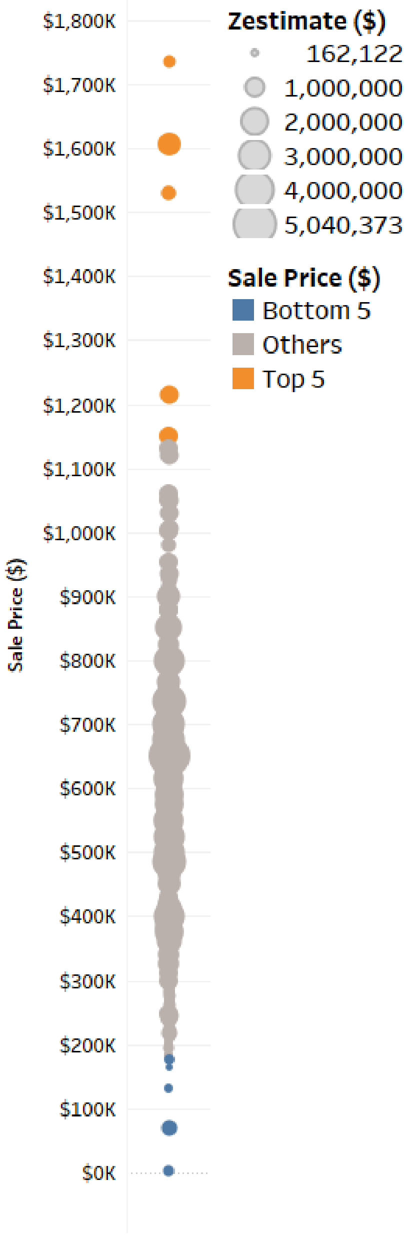
Familiarize with the data quickly

Data visualization

Visualize findings in an intuitive way

TOP 5 & BOTTOM 5 SALES IN 2007 (#2)

ID	Street	Sale Date	Sale Price	Zestimate	Notes
49140066	PHINNEY AVE N	1/23/2007	$1,735,000	$449,225	Top 5
48833184	CORLISS AVE N	6/18/2007	$1,607,775	$1,529,901	Top 5
48920199	N 34TH ST	1/30/2007	$1,530,000	$641,184	Top 5
48692876	N 48TH ST	8/24/2007	$1,215,000	$1,001,712	Top 5
49144220	WOODLAWN AVE N	7/25/2007	$1,150,000	$993,723	Top 5
48981414	N 73RD ST	3/26/2007	$175,000	$266,157	Bottom 5
60971762	AURORA AVE N	3/6/2007	$165,000	$162,122	Bottom 5
48920442	N 35TH ST	5/11/2007	$131,000	$225,567	Bottom 5
49140284	LINDEN AVE N	8/31/2007	$70,000	$712,132	Bottom 5
48847142	N 91ST ST	7/23/2007	$1,000	$420,084	Bottom 5



UNREALISTIC EXTREME VALUES (#2 CONT.)

Step 1. Compare to Zestimate

The sale prices of 4 homes (i.e. 49140066, 48920199, 49140284, and 48847142) are substantially different from Zestimate.

Step 2. Compare to sale prices of the same home in other years

No other transactions available for those 4 homes.

Step 3. Compare to sale prices of homes in the same street around similar time

The sale prices of 49140066 and 48847142 are unrealistic compared to their Zestimate values and the median sale prices in the same street round similar time.

ID	Street	Sale Date	Sale Price	Zestimate		Comparison Year	# of Transactions	Median Sale Price	Comments
49140066	PHINNEY AVE N	1/23/2007	$1,735,000	$449,225		2006-2007	60	$405,000	Too high
48920199	N 34TH ST	1/30/2007	$1,530,000	$641,184		2006-2007	1	$1,530,000	
49140284	LINDEN AVE N	8/31/2007	$70,000	$712,132		2007	36	$400,000	
48847142	N 91ST ST	7/23/2007	$1,000	$420,084		2007	6	$382,750	Too low

ZESTIMATE IN 2007 (#3)

Street	# of Homes	Min Zestimate	Median Zestimate	Mean Zestimate	Max Zestimate
GREENWOOD AVE N	706	$158,457	$332,862	$411,811	$1,370,230
PHINNEY AVE N	427	$191,007	$437,023	$486,250	$5,847,887
WALLINGFORD AVE N	392	$193,543	$525,045	$530,900	$1,144,839
DAYTON AVE N	383	$199,132	$477,849	$502,300	$950,336
MERIDIAN AVE N	376	$207,991	$592,791	$610,722	$4,222,919
FREMONT AVE N	335	$176,610	$437,024	$451,532	$1,242,065
LINDEN AVE N	334	$255,152	$445,084	$471,956	$1,263,391
PALATINE AVE N	331	$256,408	$566,341	$577,511	$1,103,613
ASHWORTH AVE N	326	$277,655	$575,752	$573,192	$1,485,387
CORLISS AVE N	320	$321,426	$588,101	$619,624	$1,627,064
DENSMORE AVE N	312	$275,633	$555,005	$560,277	$1,086,792

Here is a summary of Zestimate in 2007 for selected streets on the following statistics:

- Min
- Median
- Mean
- Max

The full list is attached as **Appendix 1**.

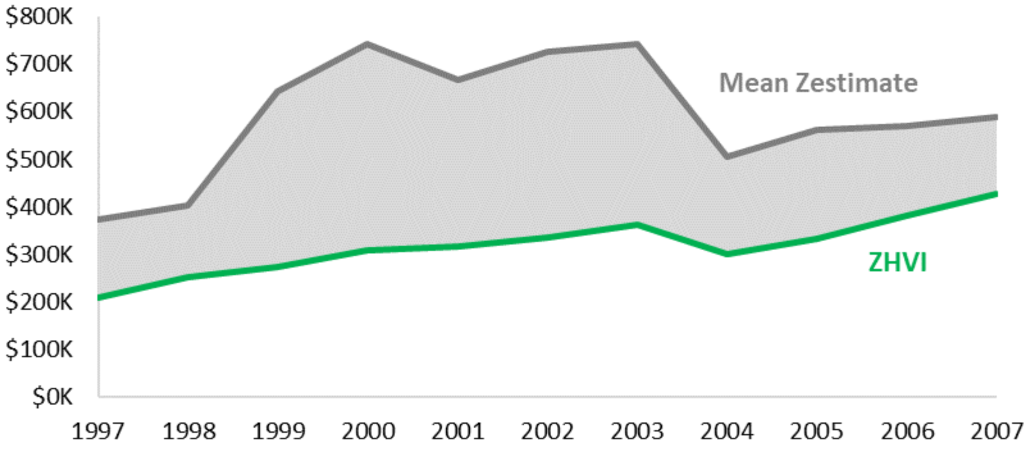
* There is a ±5% difference to the median Zestimate when including Zestimate of homes sold in 2007 in “Transaction.csv.”

ZHVI: MEDIAN VS MEAN ZESTIMATE? (#4)

Year	# of Homes	ZHVI (Median Zestimate)
1997	12,355	$191,443
1998	12,485	$230,973
1999	12,630	$253,508
2000	12,821	$286,300
2001	12,934	$295,501
2002	13,092	$314,400
2003	13,228	$326,639
2004	13,403	$367,900
2005	13,452	$424,186
2006	13,705	$491,862
2007	13,997	$523,024

Median Zestimate is preferred than mean Zestimate because median eliminates extreme values that could drive an index up and down.

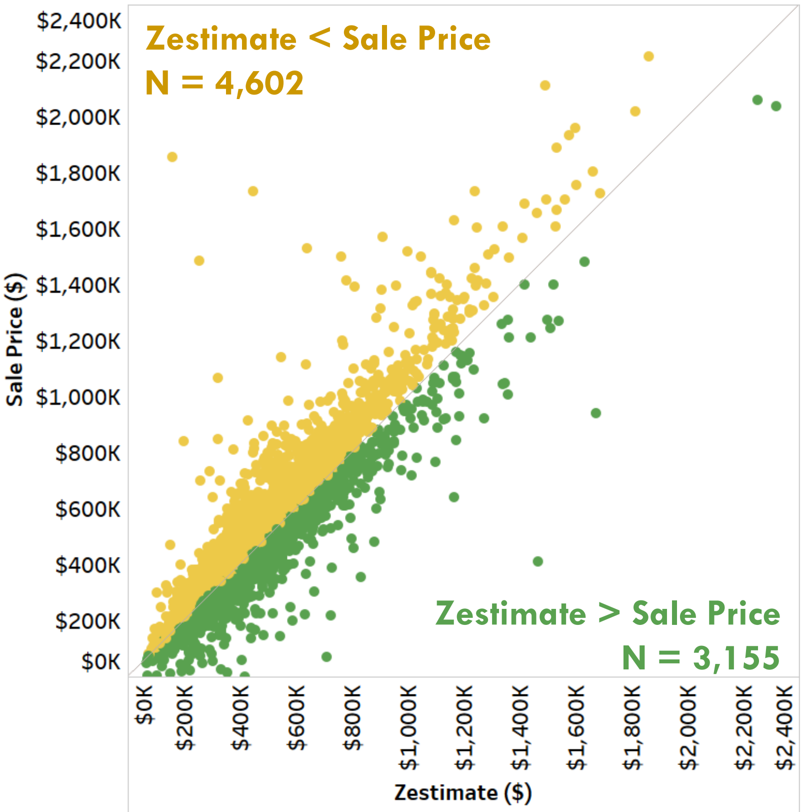
For example, for street "STONE WAY N":



* There is a ±1% difference to ZHVI when including Zestimate of homes sold in "Transaction.csv."

ZESTIMATE: BIASED? (#5)

Year	# of Transactions	Median % Error (Zestimate - Sale Price)/ Sale Price
1996	398	-0.8%
1997	470	-3.0%
1998	547	-2.6%
1999	636	-3.0%
2000	604	-3.4%
2001	541	-2.3%
2002	699	-3.0%
2003	788	-3.7%
2004	877	-4.8%
2005	880	-3.1%
2006	789	-1.5%
2007	528	0.6%



Both the annual Zestimate percentage error and the scatter plot of actual transactions suggest that Zestimate underestimates actual sale price.

* Two data points where sale price exceeds $2M are removed from the scatter plot.

ADJUSTED ZHVI (#6)

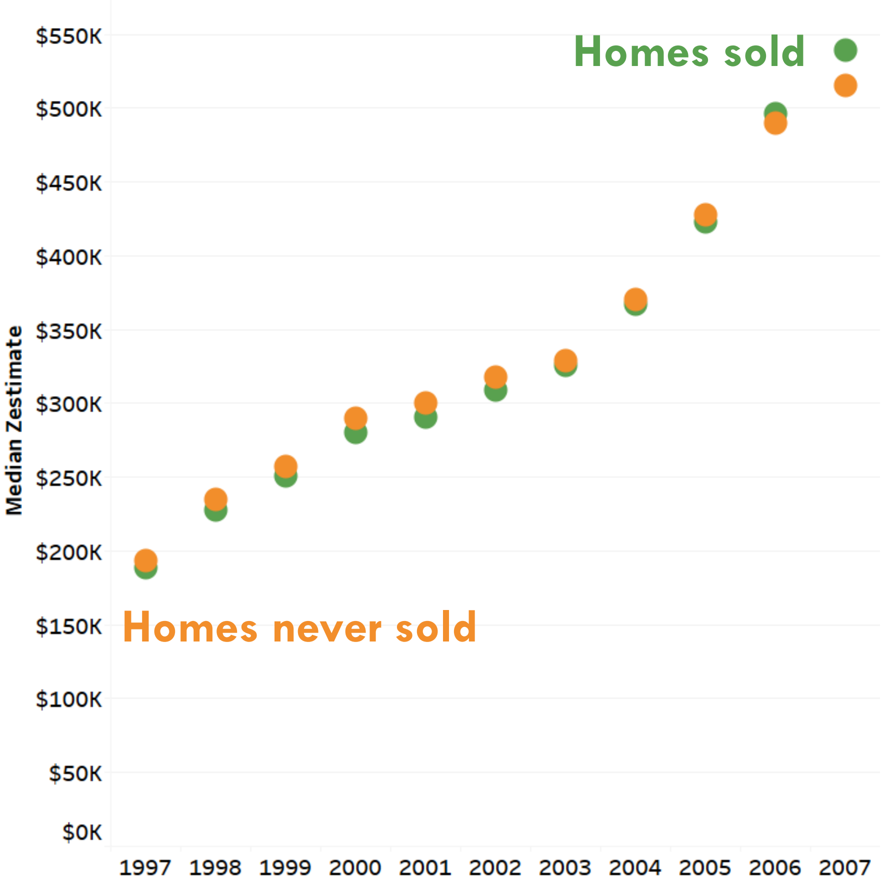
		Step 1		Step 2
Year	**ZHVI**	**ZHVI _1**	**Median % Error**	**Adjusted ZHVI = ZHVI _1/ (1 + Median % Error)**
1996	n/a	n/a	-0.8%	n/a
1997	$191,443	$193,635	-3.0%	$199,542
1998	$230,973	$237,414	-2.6%	$243,679
1999	$253,508	$258,288	-3.0%	$266,349
2000	$286,300	$292,320	-3.4%	$302,506
2001	$295,501	$301,018	-2.3%	$308,087
2002	$314,400	$319,878	-3.0%	$329,842
2003	$326,639	$329,581	-3.7%	$342,420
2004	$367,900	$374,476	-4.8%	$393,371
2005	$424,186	$431,687	-3.1%	$445,589
2006	$491,862	$498,447	-1.5%	$505,910
2007	$523,024	$529,198	0.6%	$526,014

The adjusted ZHVI is calculated in two steps:

Step 1, replace each home Zestimate with street median in the same year, then calculate the annual median Zestimate as ZHVI_1.
This step eliminates the street-wide extreme values.

Step 2, adjust ZHVI_1 by the annual median Zestimate % error from question #5.
This step reduces Zestimate's systematic error in the index.

ADJUSTED ZHVI: ASSUMPTION (#6 CONT.)



ZHVI consists of two sub populations: Zestimate of homes sold (45.2%) and Zestimate of homes never sold (54.8%).

The underlying assumption in the adjusted ZHVI is that the homes sold are similar to the homes never sold.

Based on T-test (P Value = 28%), the annual median Zestimate of the two sub populations are not statistically different.

Thus, the street median is representative for both sub populations in that street, and the systematic error in Zestimate seen with the homes sold are likely true for homes never sold too.

A BETTER ZESTIMATE (#7)

Year	Median % Error (Zestimate - Sale Price)/ Sale Price
1996	-0.8%
1997	-3.0%
1998	-2.6%
1999	-3.0%
2000	-3.4%
2001	-2.3%
2002	-3.0%
2003	-3.7%
2004	-4.8%
2005	-3.1%
2006	-1.5%
2007	0.6%

Every day, Zestimate is referenced by billions of people who are making housing-related decisions. Zestimate is also the major input to ZHVI, a widely used home price index. Therefore, to provide better information to our customers, Zestimate should be as accurate as possible.

The current systematic error in Zestimate can be mitigated in the following ways:

- **Consider macroeconomic conditions and incorporate forward-looking parameters in Zestimate.**

In question #5, the median Zestimate % error is positive (i.e. Zestimate is greater than the sale price) only in 2007, when housing market was declining. This change of direction in the % error indicates a delay in Zestimate, which can be reduced by controlling for economic leading indicators such as the Consumer Confidence Index and Purchasing Managers Index.

A BETTER ZESTIMATE (#7 CONT.)

- **Consider socioeconomic characteristics (ex. income, education, ethnicity) of a household to better control for idiosyncratic factors of a home (ex. luxury interior, expensive home appliance).**

High income people are more likely to have an expensive taste and invest more in home appliance that cannot be captured by common housing features (ex: number of rooms). Thus, Zestimate will be more accurate by controlling for socioeconomic factors. Information can be collected from US Census, American Community Survey data, or from private sources.

- **Keep Zestimate model inputs up-to-date to reflect home values in real time**

Housing-related information (ex. appraisal, property tax, sale price) is not updated everyday. However, forecasting those sparse data points into the future can help bridge the gap. For example, by extrapolating model inputs (ex: past sale price) to the estimation date rather than using the stale information, Zestimate will be able to better reflect the current home values.

THANK YOU

Kylie Zhang