

DATA SKEPTIC

Skeptical of and with data.

[Home](#) | [Episodes](#) | [Events](#) | [OpenHouse](#) | [Store](#) | [Bios](#) | [Blog](#) | [Resources](#) | [Contact](#)

FEEDBACK #1

Shortly following the announcement of this project on the Data Skeptic mailing list, I recieved this email from listener "CB" who wrote:

I like the podcast a lot. But, I am missing something. Maybe the topic of an episode but personally, the more technical you get on the episodes, the better I like them so this isn't ideal

What do hackers mean when they say data "should" be free? My sensibilities lean toward individual freedoms. Why should anyone be privy to what I sell my house for? It seems similar to a trend I see as overshoot where people think society has precedence over the individual. I am all in favor of gov't protecting us all from polluters, etc. but house data is like the copyright debate. I taught my kids to always respect the artist's right to choose if he or she should sell or give away. It is my house and what I sold it for is no one's business except the buyer's. This "data should be free" movement to me is people getting carried away with the term "rights", and in the process diluting the importance of the real ones.

Set me straight

I appreciated CB's candidate reply. Below is my response

Thanks for taking the time to write. I think there are two concepts you're asking about. First, what does it mean for information to be free? Second, why do I assert that housing sales data should be free?

As for what it means for information to be "free", I mean freely accessible in a format available for analysis. Richard Stallman is famous (in part) for articulating the distinction "Think free as in free speech, not free beer". More formal, he meant the distinction between gratis (for zero price) and libre (with little or no restriction). In this context, I actually mean both.

I definitely mean free as in libre: any analysis I wish to perform using housing sales data should be unrestricted. I don't believe anyone should tell me I'm forbidden from attempting to estimate the value of a home based on comparative recent sales. Further, I don't think I should be required to get an estimate from an appraiser who may not be transparent or effective in their efforts. Naturally, everyone has their own right to interpret my analysis as they see fit, but I should be unrestricted from performing it.

I also mean free as in gratis, but with some caveats. Maintaining and publishing a database has a cost, and I do not think society should necessarily bear the cost burden of providing the bandwidth and hosting for such data. However, if independent parties choose to assemble and curate such a database, and then share it, they should have the right to do so, and in this way, the data can be free as in gratis.

In other words, given that it is agreed (which I will address in a moment) that home sales data should be public, I believe no one should be restricted from gathering, sharing, and analyzing this data. These three acts are the primary goals of my project.

On to your second point - should this information be public or private? Before sharing my opinion, I will point out that this data already is public, at least in the US. I believe this relates to tax law somehow, since property taxes are assessed (in part) based on perceived value. However, I'm not a lawyer and don't presume to know all the in's and out's here, except to say that on an individual basis, one can easily look up the sales history of any property. My complain is my inability to access them in a non case-by-case way.

So should this data be free? Is it only the business of the buyer and seller? In other words, should we change the law? It is my personal opinion that this transactional data should remain public. I'm a big fan of the platitude "your rights end where mine begin". Sale data impacts taxes which impacts one's neighbors. For this reason, it's my opinion that the transactional dataset should be free. To be clear, I'm not interested in who the buyer and seller are. Simply the date and agreed price of any legal parcel along with a description of the features (e.g. bedrooms and bathrooms) of said parcel.

So that's my point of view.

Yet, I would certainly respect (though not necessarily agree with) a well

crafted argument to the contrary. Part of my hope in kicking off this project is to spur conversations about the right to view data and data privacy. So I appreciate you writing, and I hope to consider topics like this open ended conversations. I'm eager to include game theoretic concepts such as one and two sided private information auctions equilibria, beyond my primary goals of using this dataset to teach people about multiple regression.

Thanks for writing,

Kyle

What do you think? Please share in the comments below.



4 Comments

Data Skeptic

1 Login ▾

♥ Recommend

🔗 Share

Sort by Best ▾



Join the discussion...



Dudley • 8 months ago

Hello Data Skeptic,

I truly enjoy your podcast, and love this project....

A couple items to note about- 1) transactional data is not always public in the US. In many states, the buyer will sign an agreement to make their sales price public. Also, there are a number of states which are nondisclosure, so explanatory variables like; gross living area, year built, lot sizes, bath/bed counts, etc. may not be available to the public, or are estimated by municipalities. 2) You mentioned you used Redfin in one of your linear regressions, if you notice, even in Redfin many states will only report the final list price at the time of sale, not the actual sales price (Look at Providence or Texas for an example). Their interface is very user friendly as it allows an easy export of data to excel, but be cautious of the data you are exporting.

There are other pay-per-click sites that can offer you a similar functionality as Redfin, and have actual transaction data. Also, you may be able to export (or scrape) data from a city's assessors website. Many assessor's sites also have GIS mapping capability, etc., so you have a lot of robust data to examine.

Lastly, I would love to help out as best I can as I do have access to MLS data I pay for in my area. Please let me know what I can do to help.

Regard,

Dudley Dawson

^ | v • Reply • Share ▾



Data Skeptic Mod → Dudley • 8 months ago

Hi Dudley,

Thanks for taking the time to share all our insights. Your cautions about using data exported from any service (Redfin or otherwise) are important. I'm unclear about what the myriad of rules across the country might be for what is and is not available. I hope my little project brings together people that can figure that out.

Regarding MLS data, I am extremely eager to work with this. However, I am also a little wary. I don't know their terms of service agreements, and I wouldn't want anyone (yourself or otherwise) to get in trouble for data sharing. I would love to have your help in any capacity. If you're interested in joining the Slack channel (where most activities and discussion are centralized), please send me an email and I'll get you added.

Thanks!

Kyle

^ | v • Reply • Share ›



Dudley → Data Skeptic • 8 months ago

Hi Kyle,

I'd love to participate, please add me.

^ | v • Reply • Share ›



Data Skeptic Mod → Dudley • 8 months ago

I need your email to do so. Would you mind reaching out to me via [kyle@Data Skeptic .com](mailto:kyle@DataSkeptic.com)? Then I can add you. Thanks!

^ | v • Reply • Share ›