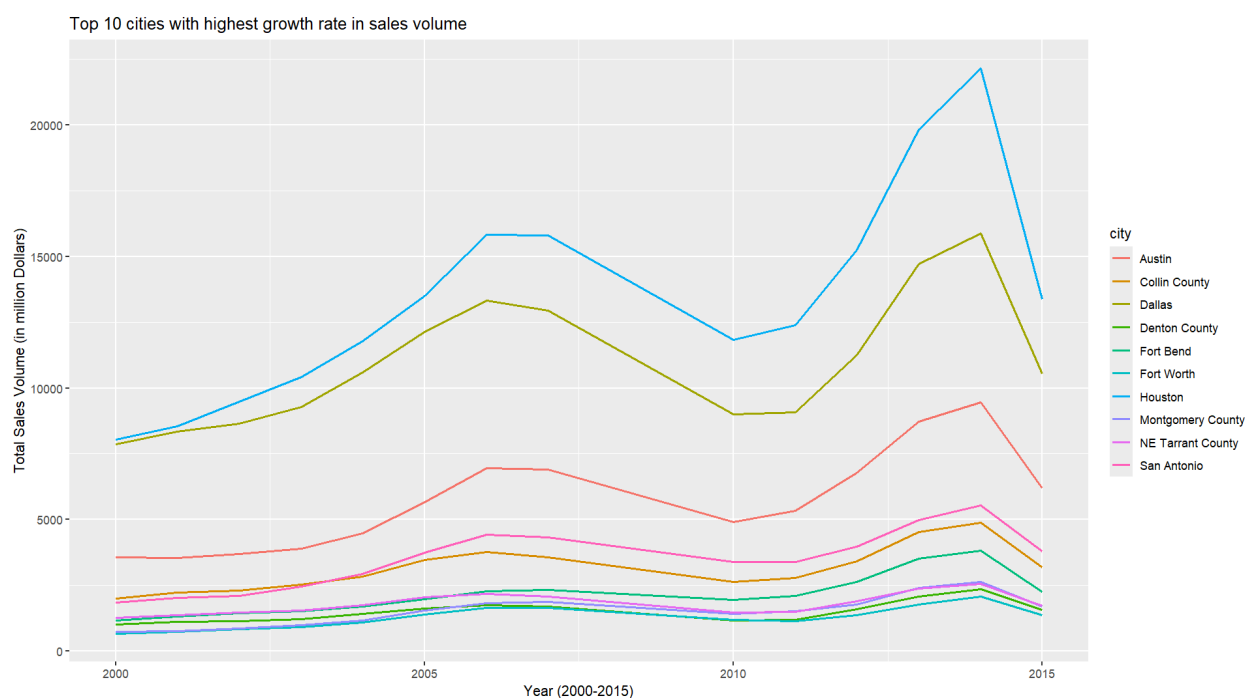


TXHOUSING DATASET RECOMMENDATION

The following recommendation is targeted towards the interests of potential buyers outlined in the OBJECTIVES.md file. The approach to solving this problem and the coding methodology have been documented in the accompanying R script file. For conciseness, I will refer to key findings here without detailing all steps taken to reach these conclusions.

Growth Rate of Cities

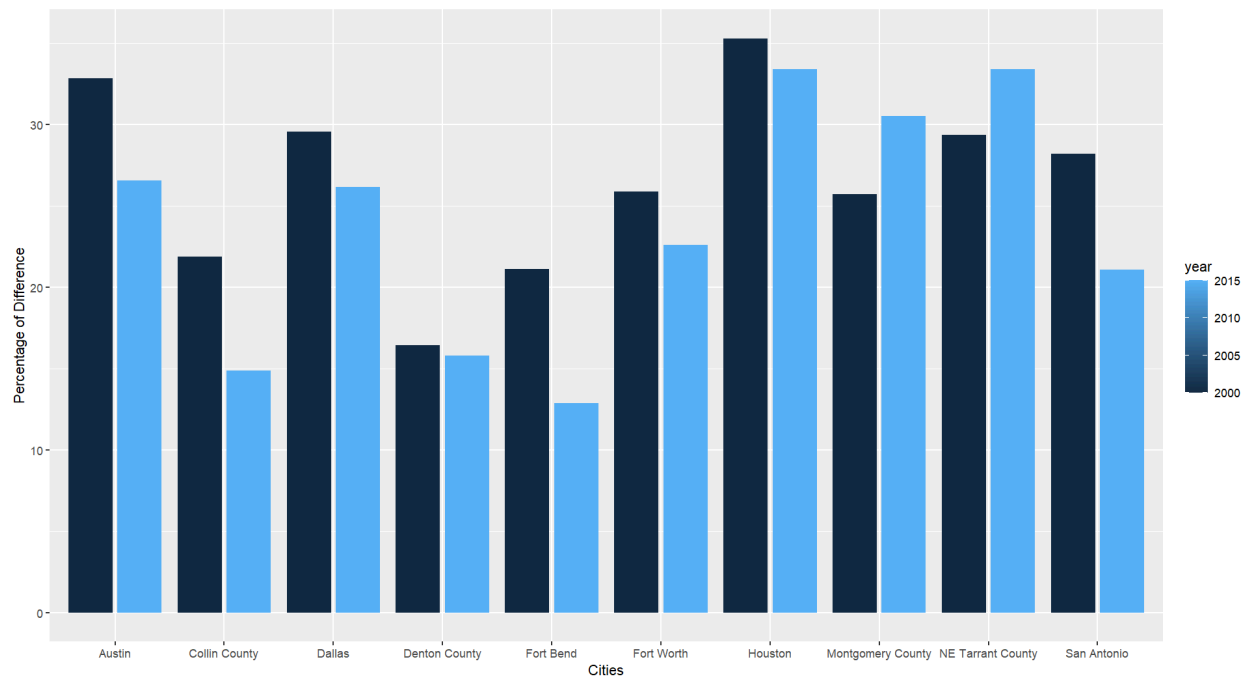
Our clients expressed interest in understanding the growth rate of various cities in Texas. The visualization below respectively highlights the top 10 cities with the highest growth rates: Houston, Dallas, Austin, San Antonio, Collin County, Fort Bend, Montgomery County, Fort Worth, Denton County, and NE Tarrant County.



To investigate whether these growth rates are influenced by outliers, I analyzed the difference between the average house price and the median house price in 2000 and 2015 for each city. The table below indicates the difference between the two prices each year for each city. The gaps between the average and median house prices range from around \$20,000 to over \$70,000. However, when these differences are compared to the cities' median house prices, they appear relatively small. For instance, in San Antonio, the difference increased from about \$27,000 in 2000 to around \$40,000 in 2015, yet remains modest relative to the median prices of \$93,000 and \$190,000, respectively. Similarly, NE Tarrant County exhibits a gap of about \$70,000 against a median house price of \$220,000.

##	city	year	mean_price	avg_median_price	diff
##	<chr>	<int>	<dbl>	<dbl>	<dbl>
## 1	Austin	2000	191238.	143925	47313.
## 2	Austin	2015	327829.	259000	68829.
## 3	Collin County	2000	199498.	163692.	35806.
## 4	Collin County	2015	332479.	289400	43079.
## 5	Dallas	2000	173338.	133758.	39579.
## 6	Dallas	2015	286742.	227271.	59470.
## 7	Denton County	2000	167655.	143958.	23696.
## 8	Denton County	2015	264182.	228143.	36039.
## 9	Fort Bend	2000	161490.	133308.	28181.
## 10	Fort Bend	2015	307532.	272400	35132.
## 11	Fort Worth	2000	105405.	83733.	21671.
## 12	Fort Worth	2015	189657.	154671.	34985.
## 13	Houston	2000	153285.	113283.	40001.
## 14	Houston	2015	278123.	208471.	69651.
## 15	Montgomery County	2000	164757.	131025	33732.
## 16	Montgomery County	2015	311985.	239029.	72956.
## 17	NE Tarrant County	2000	178596.	138058.	40538.
## 18	NE Tarrant County	2015	298328.	223614.	74714.
## 19	San Antonio	2000	119498.	93208.	26289.
## 20	San Antonio	2015	230703.	190543.	40160.

To visually see the gap between the two measurements in a year for each city, we can refer to the following bar chart that results directly from the table above.



From 2000 to 2015, the gap between the two tends to decrease, with two exceptions of Montgomery County and NE Tarrant County. For most of our highest-growing cities, the difference between the mean and median house prices ranges from 13% to approximately 35%.

These numbers are not relatively high, which means the impact of outliers (if exist) is not big enough to create a great difference between the mean house prices and median house prices for these Texas cities. However, cities such as Austin, Houston, Dallas, and San Antonio have bigger gaps than others, indicating the existence of mild outliers.

Optimal Timing for Investment

To determine the ideal timing for purchasing and selling properties, I examined trends in median house prices across the study period for each city. Median prices were chosen due to their resistance against outliers, offering a representative view of the market. The results below indicate optimal times and cities for investment.

	city	time	median
	<chr>	<chr>	<dbl>
1	Harlingen	2001-3	50000
2	Nacogdoches	2005-2	54000
3	Nacogdoches	2005-3	55000
4	Paris	2002-12	55000
5	Paris	2010-11	55000
6	Port Arthur	2002-3	55000
7	Brownsville	2000-4	55800
8	Nacogdoches	2005-4	55800
9	Paris	2001-1	56000
10	Port Arthur	2000-7	56200
11	Paris	2000-11	57000
12	Paris	2012-10	57500
13	Abilene	2000-3	58100
14	Paris	2002-11	58300
15	Paris	2000-5	58600
16	Abilene	2000-2	58700
17	Harlingen	2001-8	58800
18	Port Arthur	2000-10	58800
19	Abilene	2000-10	59300
20	Abilene	2002-11	59400

The notable thing about this finding is that the best buying opportunities occurred in years that are not recent and with cities that are not necessarily on the high-growth list. The reason may be due to inflation and changes in income or purchasing power of people throughout the years.

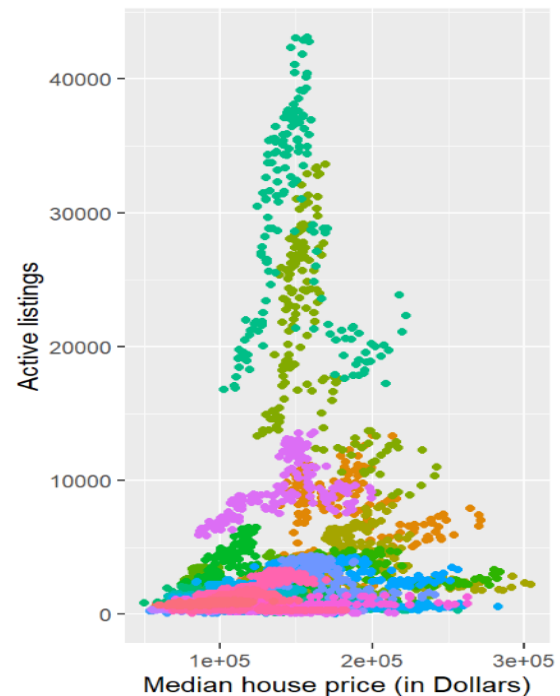
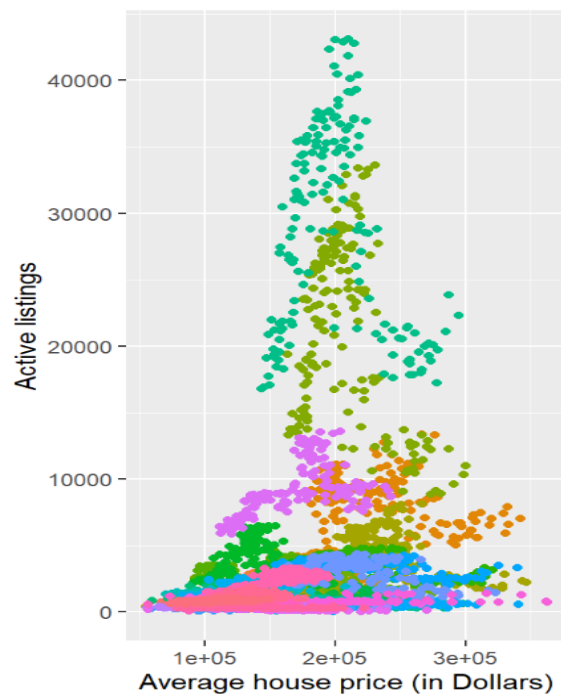
An interesting exception is Paris, Texas, where median house prices in 2010 and 2012 were comparable to prices in other cities nearly a decade earlier. If our clients are interested, further research could explore recent prices and property quality in Paris to identify similarly affordable options.

On the other hand, unsurprisingly, the optimal times and cities for selling generally involve more recent prices from high-growth cities, which aligns with their current upward trend.

	city	date	median
	<chr>	<dbl>	<dbl>
1	Collin County	2015.	304200
2	Collin County	2015.	300400
3	Collin County	2016.	292600
4	Collin County	2015.	291400
5	Collin County	2015.	285800
6	Fort Bend	2015.	284200
7	Collin County	2015.	283400
8	Midland	2014.	283100
9	Fort Bend	2014.	282300
10	Fort Bend	2016.	280400
11	Fort Bend	2015.	279700
12	Fort Bend	2015.	277300
13	Fort Bend	2015.	276100
14	Fort Bend	2015.	273500
15	Collin County	2015.	272200
16	Collin County	2014.	271600
17	Collin County	2014.	271500
18	Austin	2015.	271200
19	Fort Bend	2015.	270800
20	Austin	2015.	270300

Relationship Between Listings and House Prices

Lastly, I analyzed the relationship between the number of active listings and both the median and mean house prices. The two graphs below respectively show the relationship between active listings and mean house prices and between active listings and median house prices.





There seems to be a nearly identical relationship between both mean and median house prices and active listings with some minor differences along the x-axis. In both figures, there is an unusual spike at around the \$150,000 to \$200,000 range. The cities that are responsible for this spike are mainly Dallas and Houston.

One hypothesis for this spike is demographic. As the middle class—including both lower and upper middle-class segments—represents a significant portion of the population, there is a good chance that the middle class is also the most dominant in Dallas and Houston, considering these cities are also big cities in Texas. Thus, the housing market may have taken notice and likely aims to sell for customers in this class. Consequently, more listings are available within this price range, which is generally affordable for middle-class buyers.

However, for other cities, some have listings of house values clustering around \$100,000 to \$150,000 while some cities offer house values as high as \$250,000 and above. Still, they do not offer as many listings as the previous two cities do. My hypothesis for such clustering for other cities is the target of the housing market. For instance, the housing market in cities with listings whose mean and median house prices are higher may aim to attract people of the upper-class.

Recommendation

For clients with substantial financial resources, we recommend focusing on high-growth cities like Houston and Dallas, where neighborhoods are rapidly developing and property values continue to appreciate. Both cities offer flexibility in listings; however, recent declines in sales volume suggest caution with larger investments.

For clients open to outliers with higher values, besides Houston and Dallas, Austin and San Antonio are also attractive options, though with limited availability of active listings. With high

growth rates and modest price gaps, these cities offer such outliers and potential long-term value in premium properties.

For budget-conscious clients, we suggest exploring lower-priced options in cities where properties are available in the \$100,000 to \$150,000 range. Although individual listings may be fewer in each city, multiple cities offer affordable options that could provide good value.