Analysis of Pierce County House Sales

Henry Shaw & Saransh Rakshak

Due: September 3rd, 2024

Below is a breakdown of four variables selected within the Pierce County House Sales dataset provided. We decided to choose House Square Feet, Bedrooms, Bathrooms, and Year Built for our four fields to review.

Initially, we ran a correlation analysis to see which variables have stronger relationships with sales price. This correlation analysis showed that the four fields selected have a positive correlation to sales price. House Square Feet has a strong correlation, Bathrooms are moderate, while Bedrooms and Year Built is somewhat weaker. These can be seen below:

Correlation of Variables with Sale Price - Entire Dataset

	sale_price	house_square_feet	year_built	bedrooms	bathrooms
Correlation	1	0.5263257	0.4094384	0.2288812	0.2054946

- When we break down the correlation analysis to just the houses under $1,000,000 subset, we can see all four variables increase in their correlation to sales price. This can be seen below:

Correlation of Variables with Sale Price - Under $1,000,000

	sale_price	house_square_feet	year_built	bedrooms	bathrooms
Correlation	1	0.6277807	0.505363	0.3415674	0.3223927

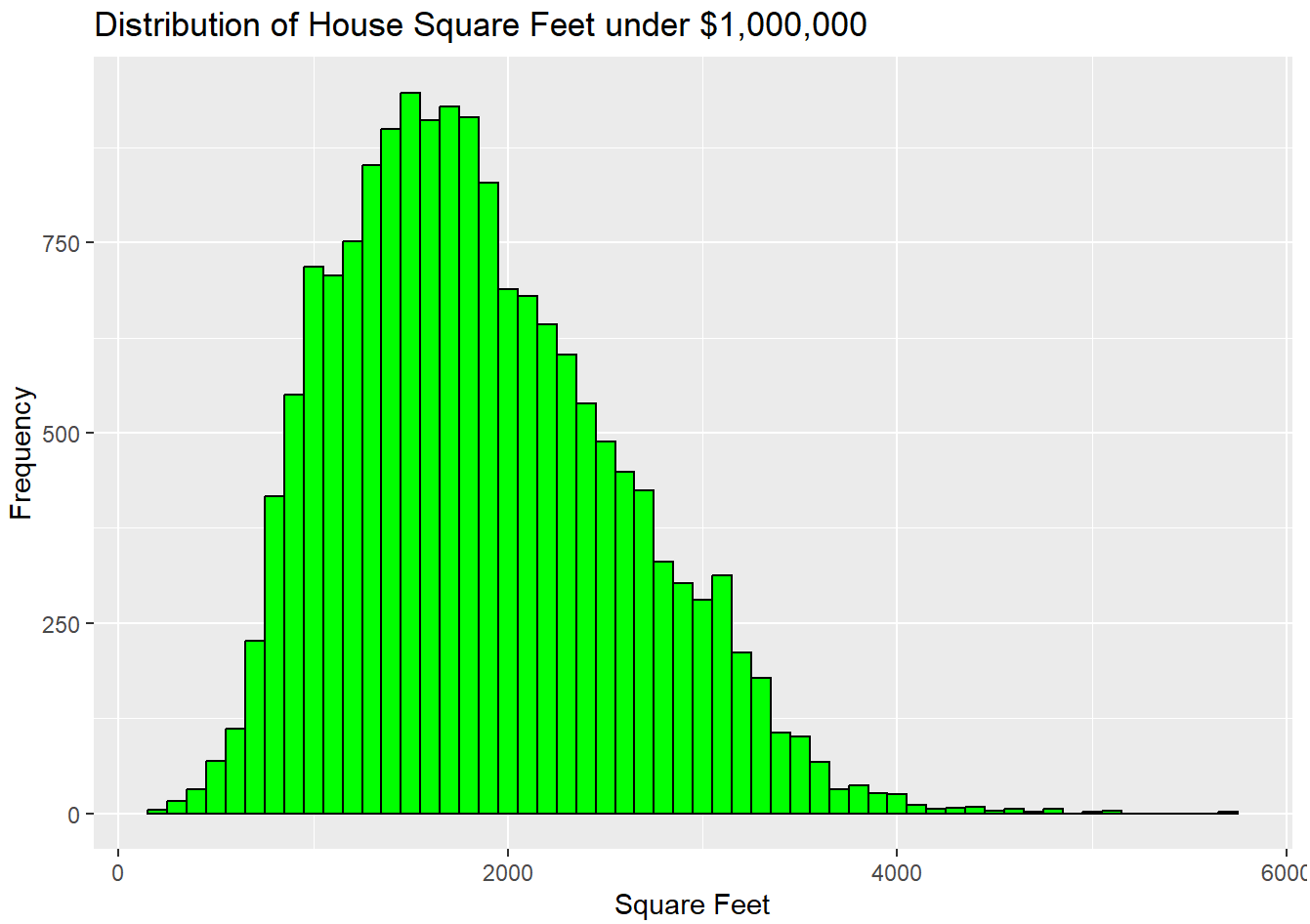
- When we break down the correlation analysis to just the houses over $1,000,000 subset, we can see the chosen variables decrease in their correlation to sales price. House Square Feet and Bathrooms have a negative correlation now. This can be seen below:

Correlation of Variables with Sale Price - Over $1,000,000

	sale_price	house_square_feet	year_built	bedrooms	bathrooms
Correlation	1	-0.1650727	0.0623177	0.0444497	-0.0398135

Histogram for House Square Feet with Sales Price under $1,000,000.

For our variable that shows the highest correlation to sales price, we decided to plot a histogram for House Square Feet.



Summary Statistics of House Square Feet under $1,000,000

Mean Square Feet	Median Square Feet	SD of Square Feet	Min Square Feet	Max Square Feet
1,855	1,761	722	200	5,700

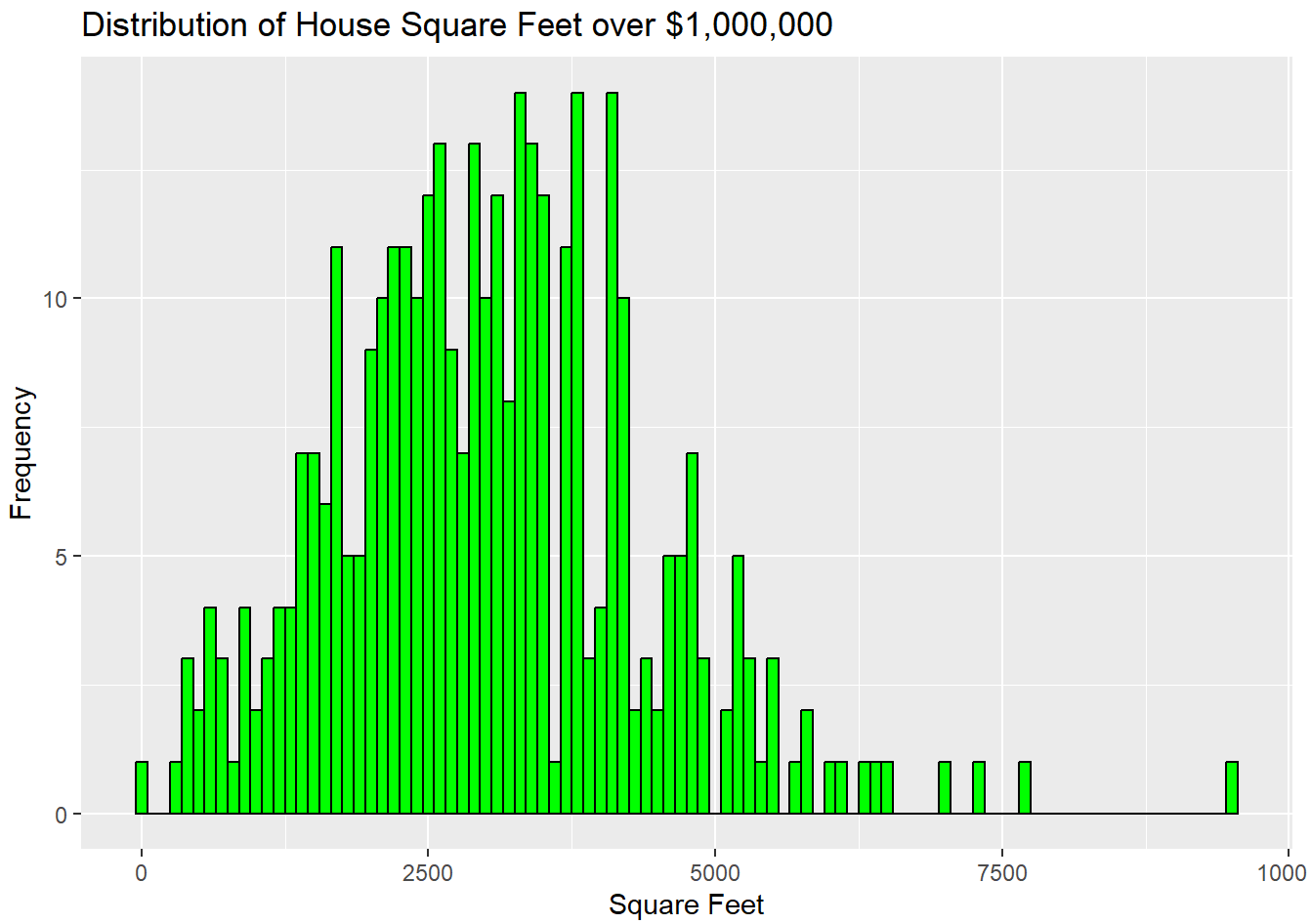
Additional Summary Statistics of House Square Feet under $1,000,000

SD of Square Feet	Range	IQR	Skewness	Kurtosis
722.46	5500.00	1009.00	0.57	0.08

- The histogram for House Square Feet under $1,000,000 is positively or rightly skewed. This can be seen off in the visual and the summary statistics as well. With the median at 1,761, this square feet amount is closer to the min (200) than the max (5,700) which alights with the right skewness of the histogram. This is supported by the skewness value greater than 0.
- Majority of the homes fall closer to the mean, which can be seen with more bars towards the left. The longer tail to the right shows multiple outliers of houses under $1,000,000 that have a higher total of square feet.

Histogram for House Square Feet with Sales Price over

$1,000,000.



Summary Statistics of House Square Feet over $1,000,000

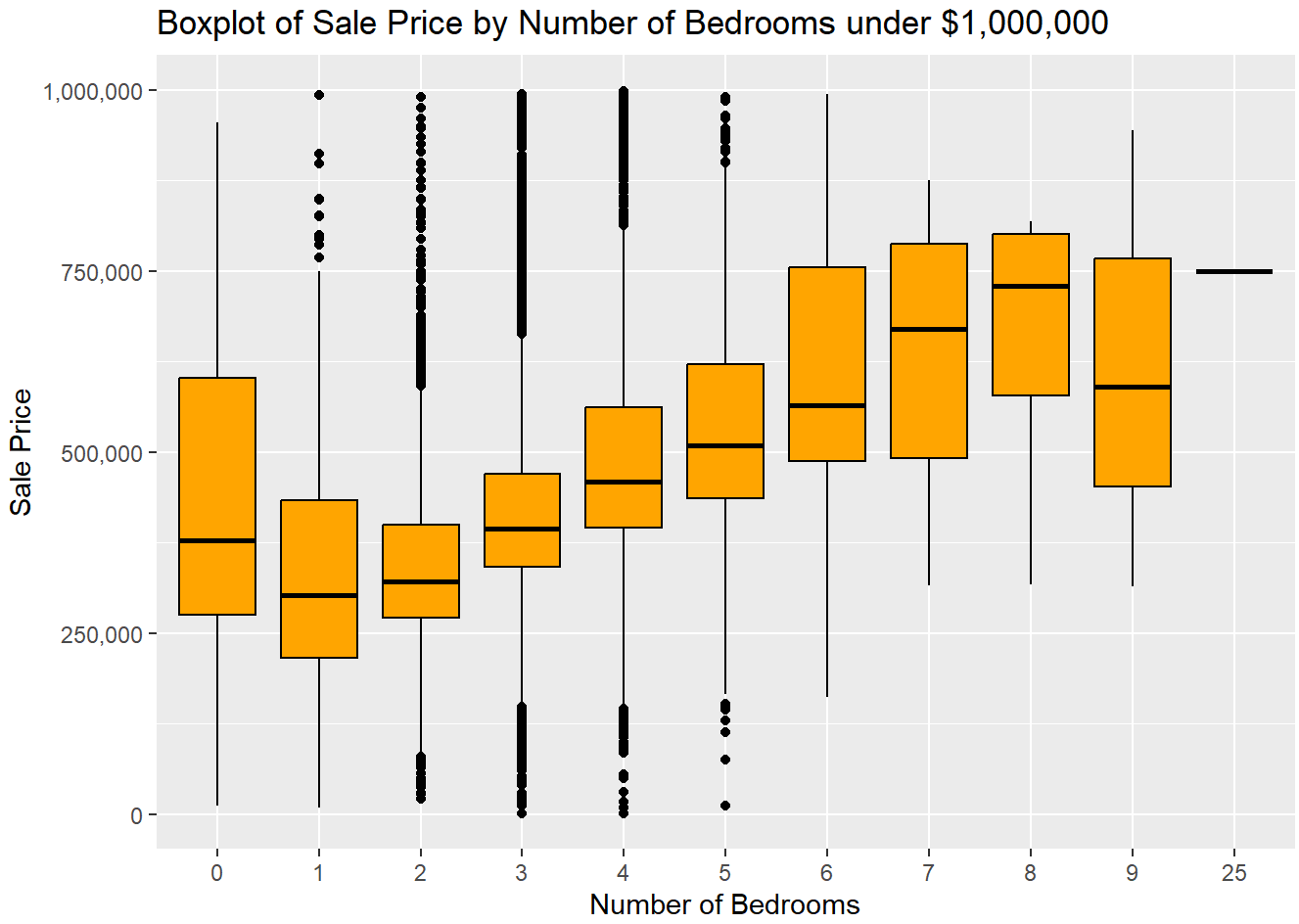
Mean Square Feet	Median Square Feet	SD of Square Feet	Min Square Feet	Max Square Feet
3,027	2,964	1,354	1	9,510

Additional Summary Statistics of House Square Feet over $1,000,000

SD of Square Feet	Range	IQR	Skewness	Kurtosis
1354.11	9509.00	1714.00	0.64	1.35

- Similar to houses sold under $1M, there is a Right Skewed trend and a higher distribution sold for houses under 5000 square feet. However, the spread of data is not nearly as uniform as with the houses under $1M. Additionally, the skewness is slightly greater (0.64) vs (0.59) than that for houses under 1M.

Boxplot for Bedrooms with Sales Price under $1,000,000.

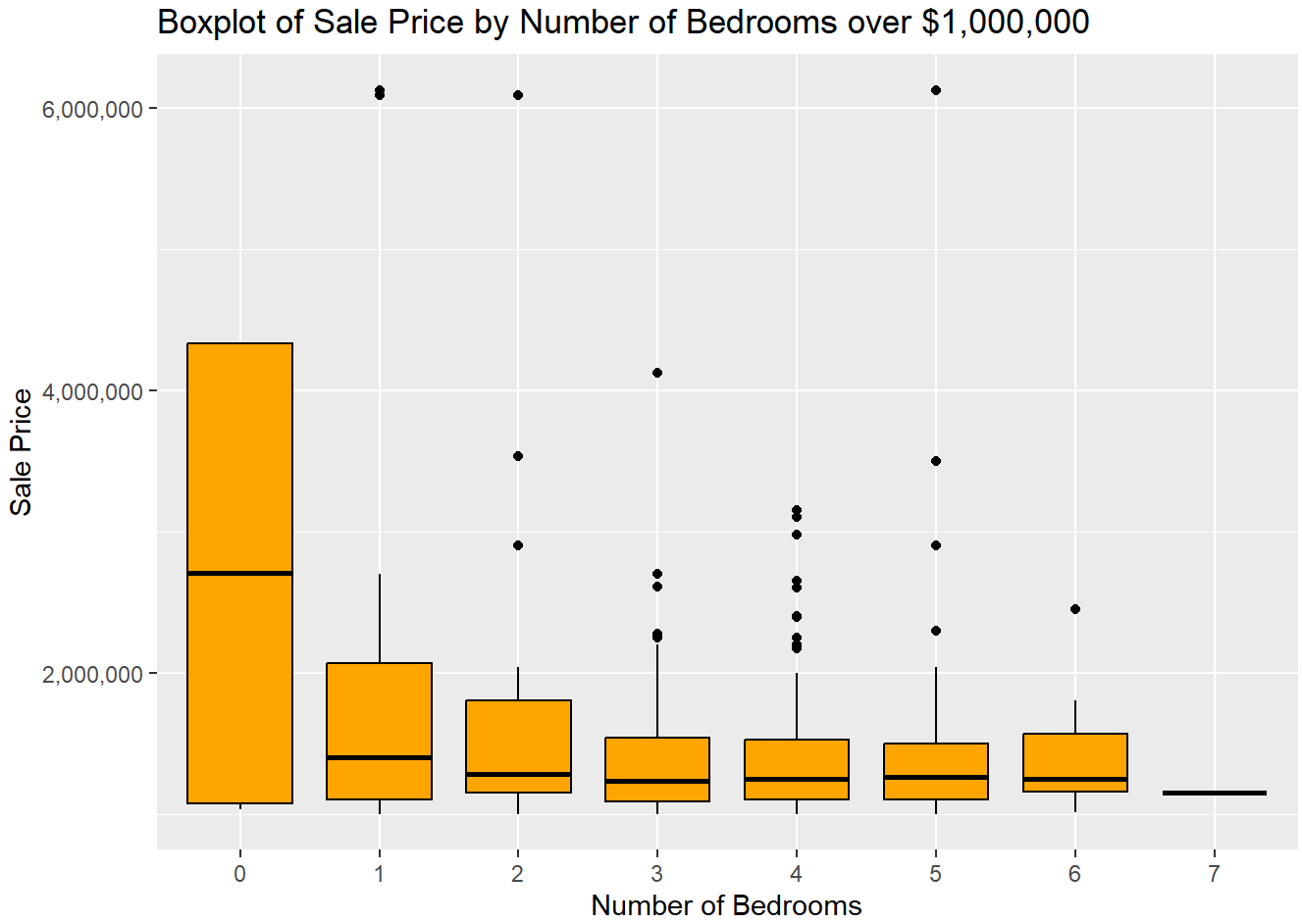


Summary Statistics for Sale Price by Number of Bedrooms under $1,000,000

Bedrooms	Count	Mean Sale Price	Median Sale Price	SD of Sale Price	Min Sale Price	Max Sale Price
0	43	432,784	378,000	216,403	12,510	955,000
1	322	344,803	302,125	190,263	10,000	992,500
2	1867	352,671	322,000	147,424	22,000	990,000
3	8371	420,669	395,000	138,850	2,000	995,000
4	4727	488,858	460,000	146,688	2,000	999,000
5	996	532,524	508,950	148,665	12,752	990,000
6	105	597,389	565,000	182,267	162,000	994,990
7	18	637,309	669,500	172,348	316,000	875,990
8	4	649,038	730,000	231,370	317,150	819,000
9	3	616,300	590,000	315,274	315,000	943,900
25	1	749,950	749,950	NA	749,950	749,950

- For bedrooms, there are eleven total unique variables. The majority of the population falls between 1 to 5 bedrooms with 16,293 houses.
- This boxplot shows that as the number of bedrooms increases, so does the average sales price of the household. This can be seen starting with 1-bedroom houses ($344,803) with a consistent increase in the mean to 8-bedroom houses ($649,038).
- The 25-bedroom house is an outlier as the number of bedrooms increases by one until it jumps from 9 to 25

Boxplot for Bedrooms with Sales Price over $1,000,000.



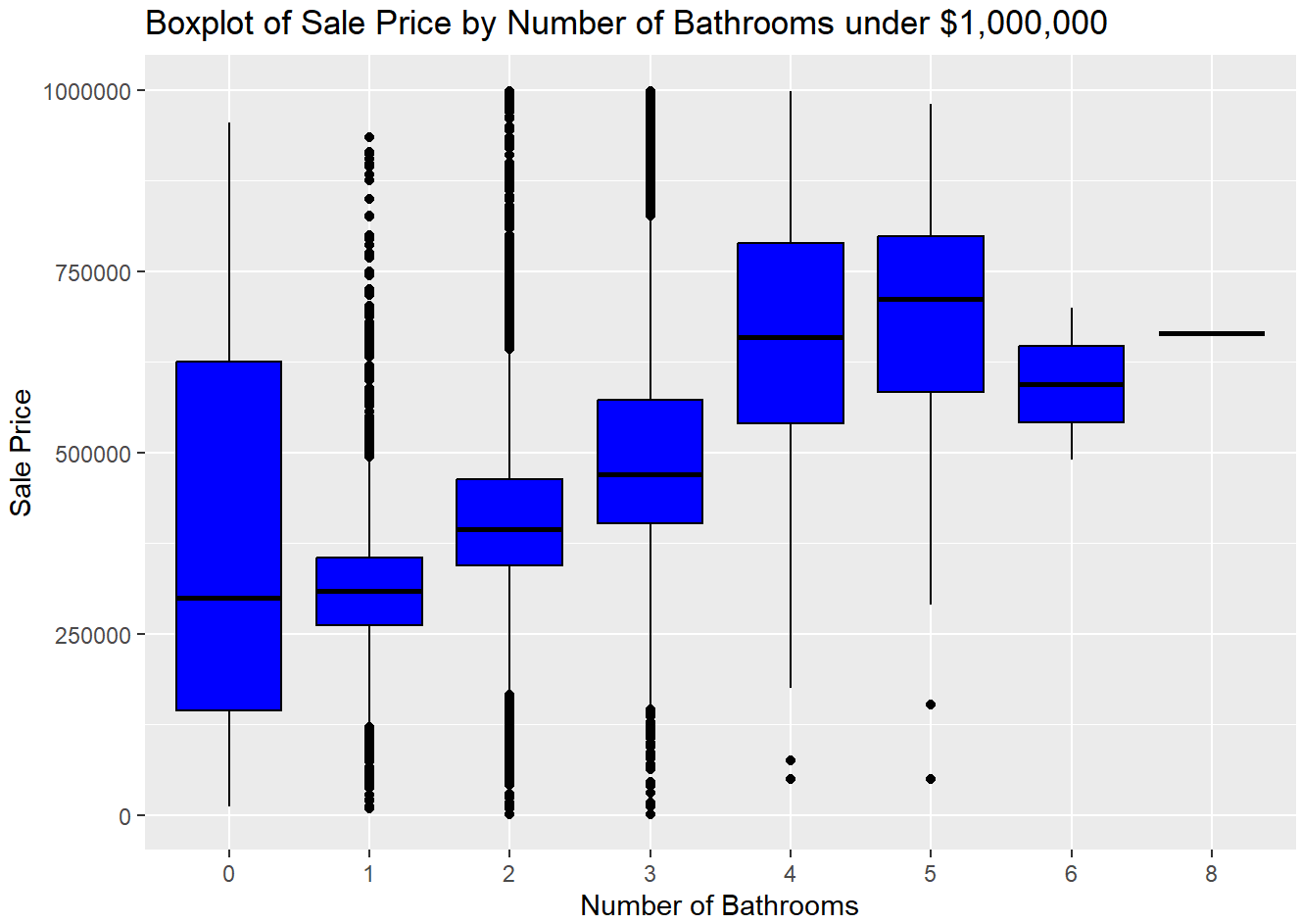
Summary Statistics for Sale Price by Number of Bedrooms over $1,000,000

Bedrooms	Count	Mean Sale Price	Median Sale Price	SD of Sale Price	Min Sale Price	Max Sale Price
0	4	2,696,050	2,710,050	1,891,595	1,030,000	4,334,100
1	15	2,119,367	1,400,000	1,695,763	1,000,000	6,130,000
2	33	1,633,679	1,280,000	969,146	1,000,000	6,090,000
3	133	1,376,480	1,236,000	425,868	1,000,000	4,126,500
4	131	1,390,515	1,250,000	437,610	1,000,000	3,150,000
5	32	1,584,058	1,262,500	997,557	1,000,000	6,130,000

Bedrooms	Count	Mean Sale Price	Median Sale Price	SD of Sale Price	Min Sale Price	Max Sale Price
6	8	1,439,011	1,245,000	478,487	1,012,000	2,450,000
7	1	1,147,500	1,147,500	NA	1,147,500	1,147,500

- Unlike with house_sales under $1M, houses that sold for over $1M showed the opposite trend, with sales price decreasing as the number of rooms increases. There is also an outliar for when number of bedrooms is equal to 0. This may be caused by the purchasing of land or an un-built houses which does not have any officially listed ‘rooms’, rather just a property of land.

Boxplot for Bathrooms with Sales Price under $1,000,000.



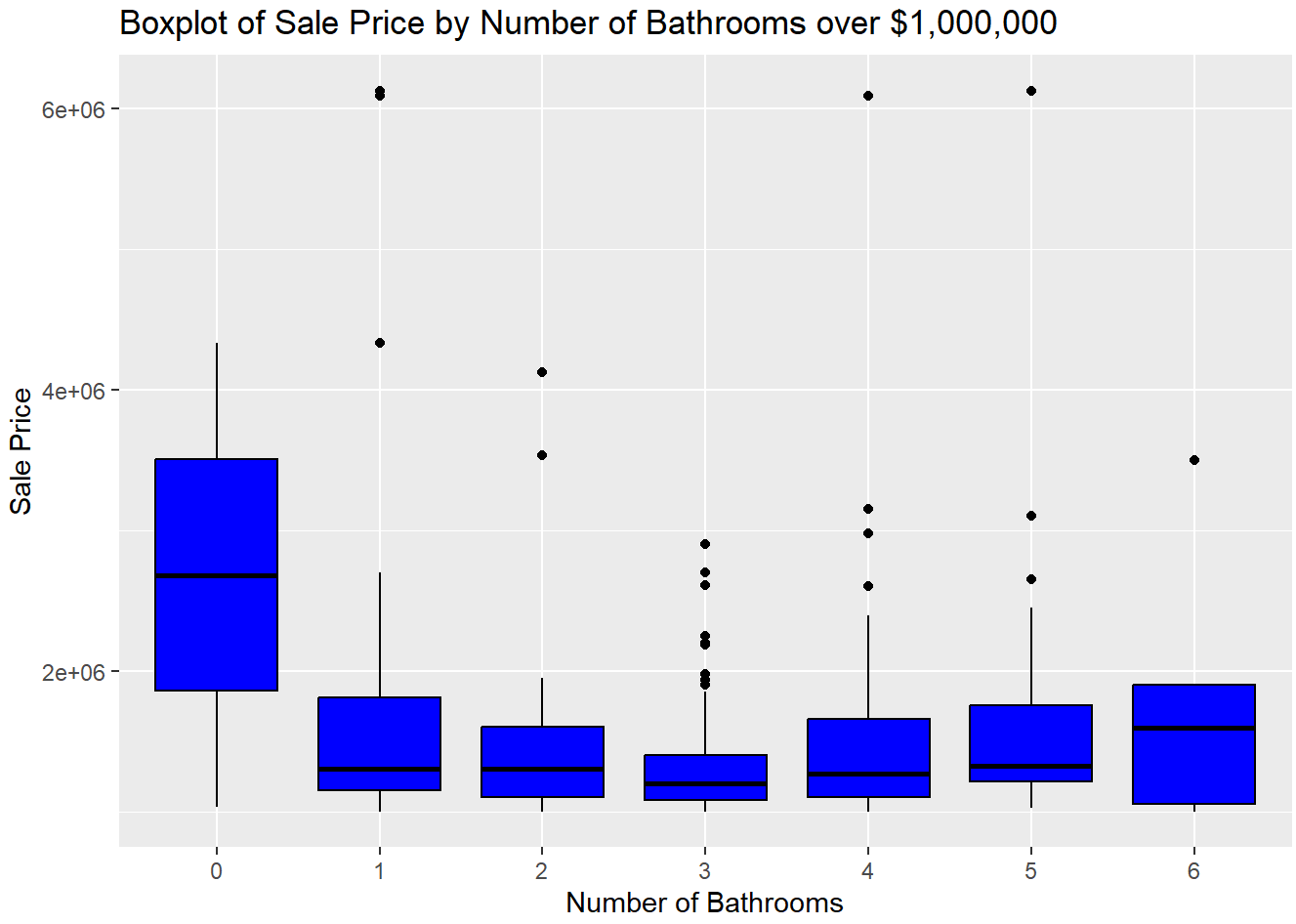
Summary Statistics for Sale Price by Number of Bathrooms under $1,000,000

Bathrooms	Count	Mean Sale Price	Median Sale Price	SD of Sale Price	Min Sale Price	Max Sale Price
0	29	361,673	300,000	266,487	12,510	955,000
1	2911	314,366	310,000	108,335	10,000	935,000
2	6272	413,864	394,995	125,594	2,000	999,000
3	6667	498,966	469,950	141,462	2,000	998,950

Bathrooms	Count	Mean Sale Price	Median Sale Price	SD of Sale Price	Min Sale Price	Max Sale Price
4	529	659,695	659,999	167,920	50,000	998,000
5	46	680,490	711,571	212,168	50,412	980,500
6	2	594,975	594,975	148,528	489,950	700,000
8	1	665,000	665,000	NA	665,000	665,000

- For bathrooms, there are eight total unique variables. The majority of the population falls between 1 to 4 bathrooms with 16,379 houses.
- This boxplot shows that as the number of bathrooms increases, so does the average sales price of the household. This can be seen starting with 1-bathroom houses ($314,366) with a consistent increase in the mean to 5-bathroom houses ($680,490).
- The 8-bathroom house is an outlier as the number of bedrooms increases by one until it jumps from 6 to 8. Additionally, the 6-bathroom houses are outliers as they only contain two instances within the subset under $1,000,000.

Boxplot for Bathrooms with Sales Price over $1,000,000.

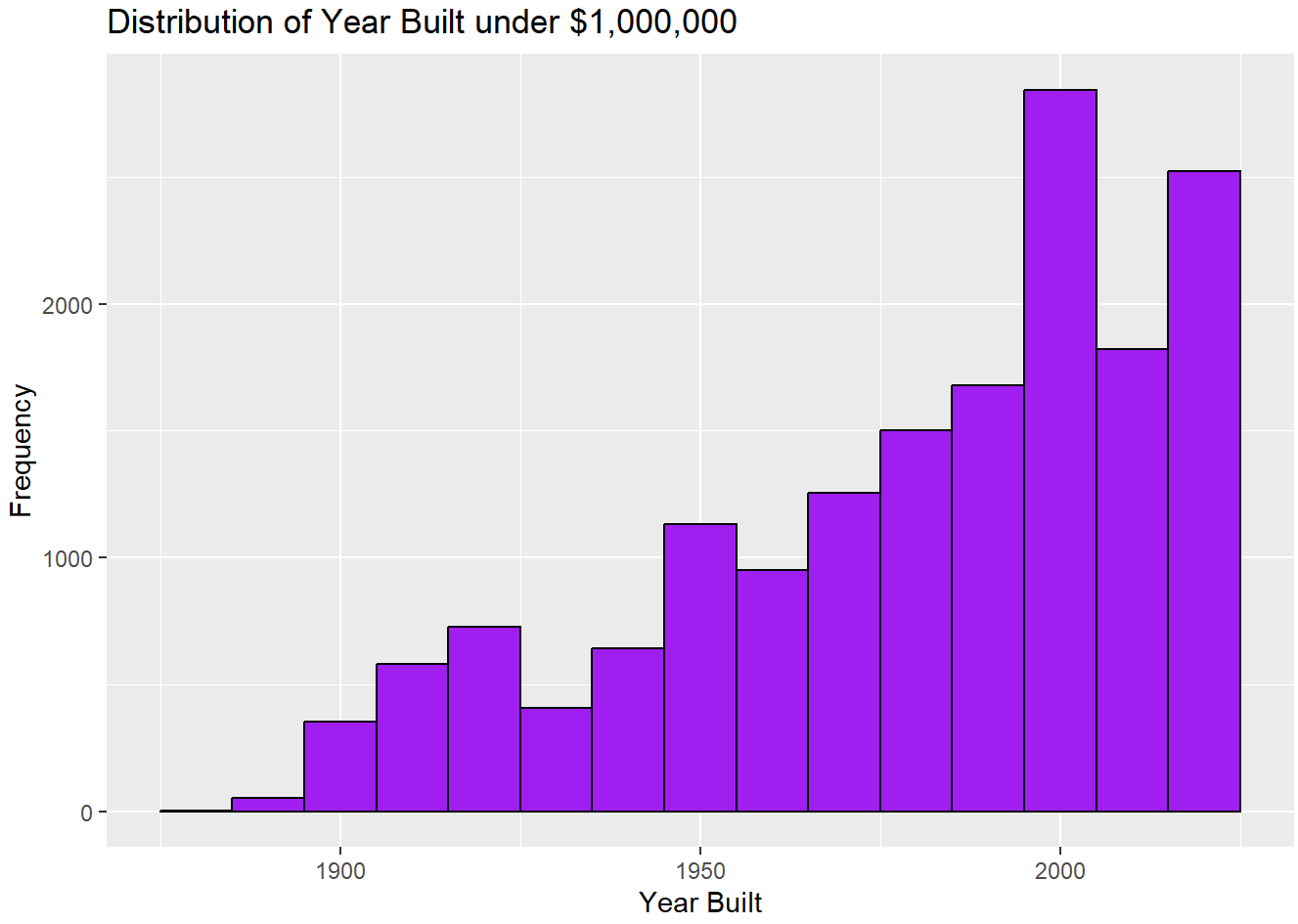


Summary Statistics for Sale Price by Number of Bathrooms over $1,000,000

Bathrooms	Count	Mean Sale Price	Median Sale Price	SD of Sale Price	Min Sale Price	Max Sale Price
0	2	2,682,050	2,682,050	2,336,352	1,030,000	4,334,100
1	29	1,868,691	1,299,999	1,352,795	1,000,000	6,130,000
2	41	1,459,376	1,300,000	613,340	1,000,000	4,126,500
3	149	1,323,784	1,200,000	376,659	1,000,000	2,900,000
4	104	1,465,920	1,267,500	626,738	1,000,000	6,090,000
5	27	1,738,251	1,322,500	1,027,935	1,025,000	6,130,000
6	5	1,808,400	1,592,000	1,018,014	1,000,000	3,500,000

- Unlike houses that are sold for under $1M, for sales over $1M, there are an extremely small amount of purchases for properties with no bathrooms. This again can be explained by the client purchasing an undeveloped area of land which may have no bedrooms or bathrooms. Also, sales price stays relatively constant even with the addition of bathrooms, showing that the amount of bathrooms in a house may not be a big factors for home-seekers willing to pay over $1M.

Histogram for Year Built with Sales Price under $1,000,000.



Summary Statistics of Year Built under $1,000,000

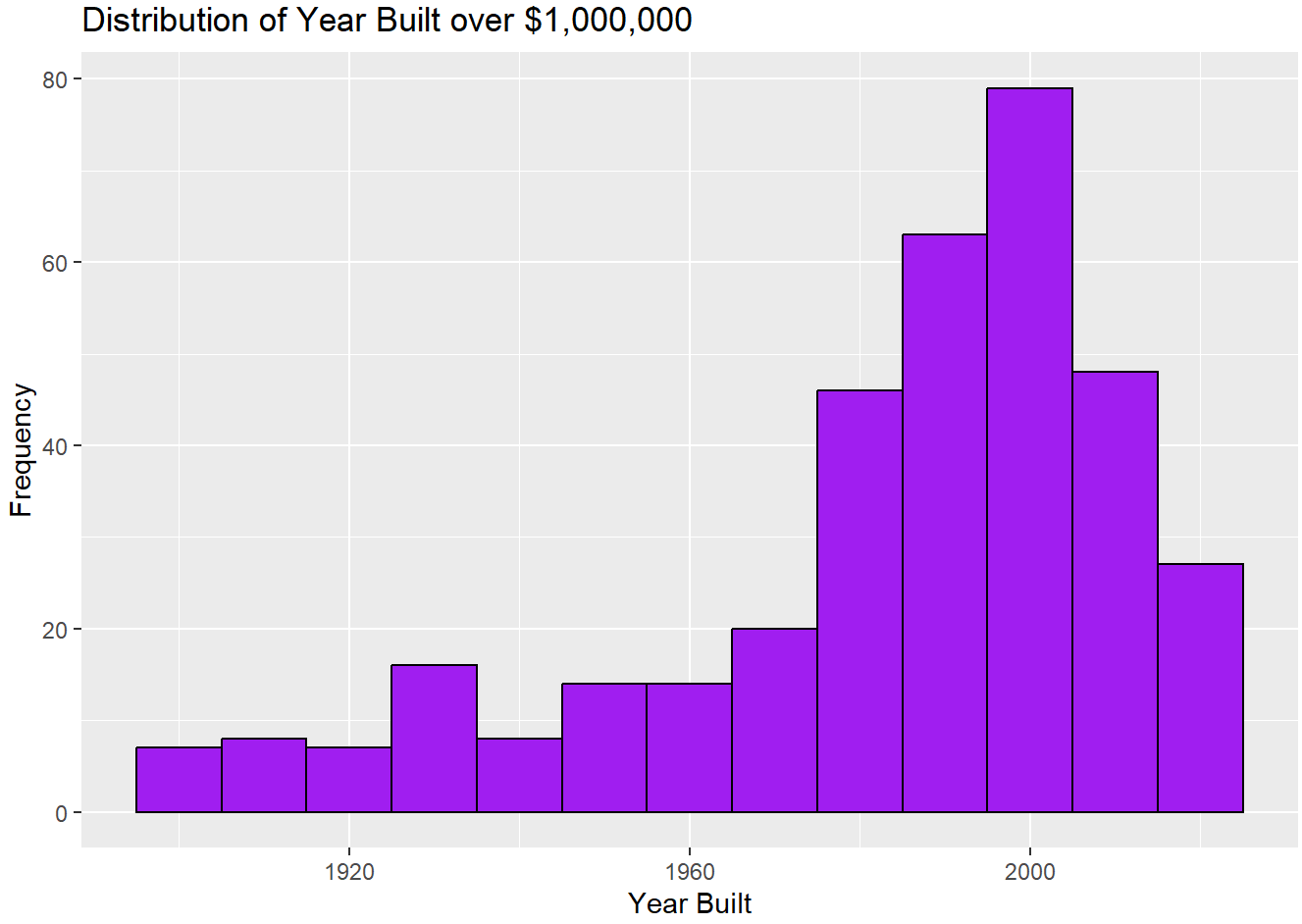
Mean Year Built	Median Year Built	Min Year Built	Max Year Built
1980.402	1990	1880	2021

Additional Summary Statistics of Year Built under $1,000,000

SD of Year Built	Range	IQR	Skewness	Kurtosis
33.43	141.00	48.00	-0.77	-0.42

- The histogram for Year Built under $1,000,000 is negatively or left skewed. This can be seen off in the visual and the summary statistics as well. With the median year of 1990, this year is closer to the max (2021) than the min (1880) which aligns with the left skewness of the histogram. This is supported by the skewness value less than 0 of -0.77.
- Majority of the homes fall closer to the mean, which can be seen with more bars towards the right. The longer tail to the left shows multiple outliers of houses under $1,000,000 that have a lower date built.

Histogram for Year Built with Sales Price over $1,000,000.



Summary Statistics of Year Built over $1,000,000

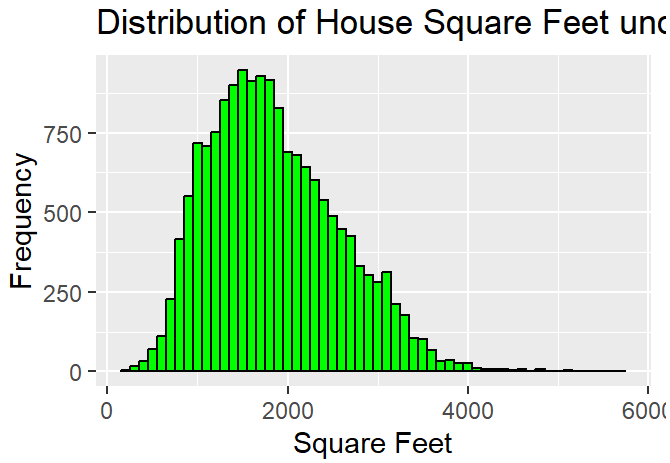
Mean Year Built	Median Year Built	Min Year Built	Max Year Built
1983.571	1991	1900	2021

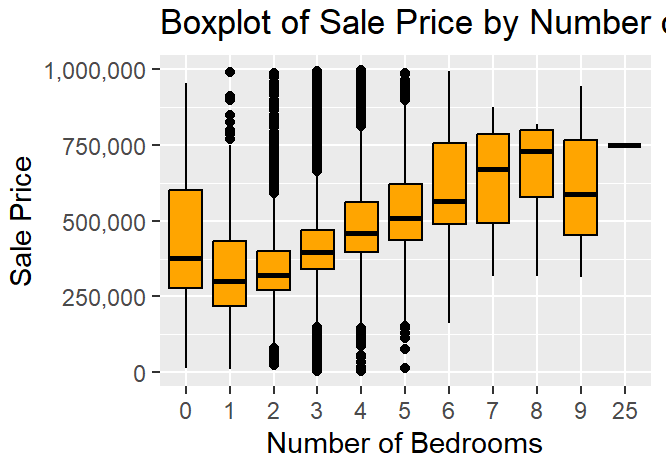
Additional Summary Statistics of Year Built over $1,000,000

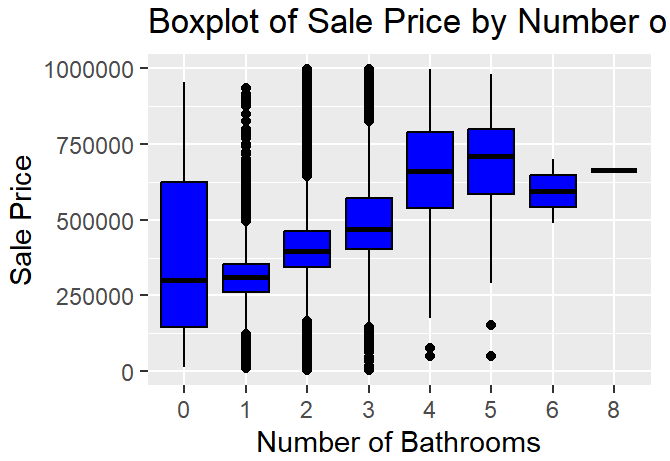
SD of Year Built	Range	IQR	Skewness	Kurtosis
28.93	121.00	30.00	-1.18	0.65

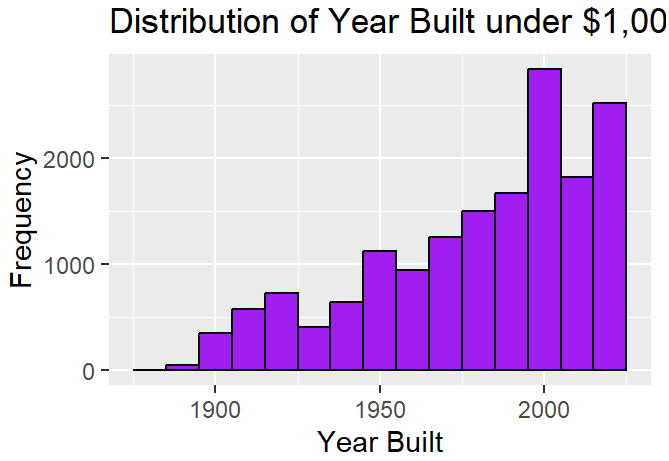
- Similar to houses under $1M, our distribution is left-skewed, showing that regardless of purchasing budgets, home-seekers prefer to purchase newly developed housing.

Grid of all Plots under $1,000,000

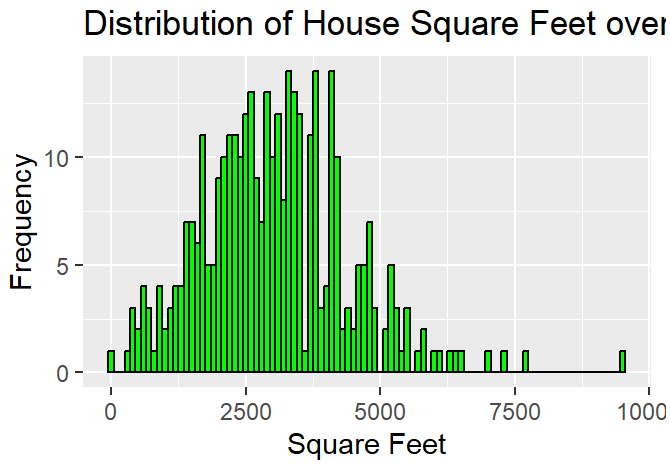


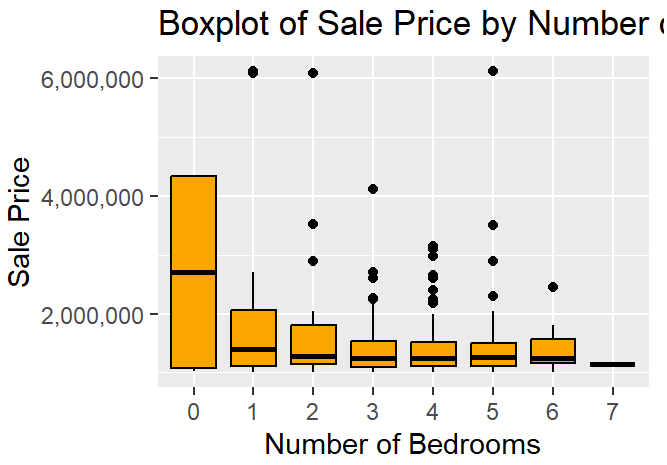


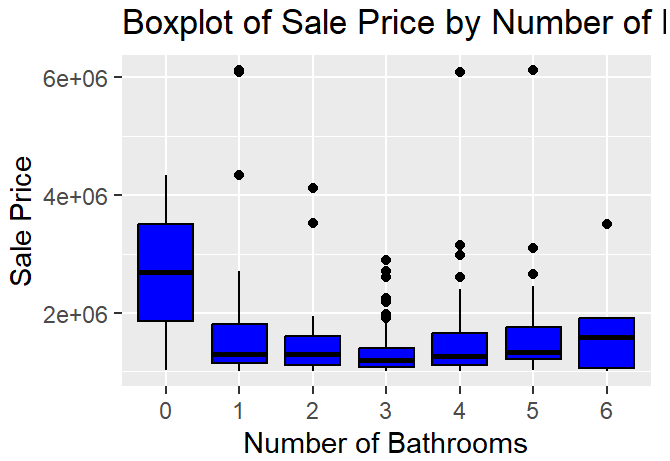


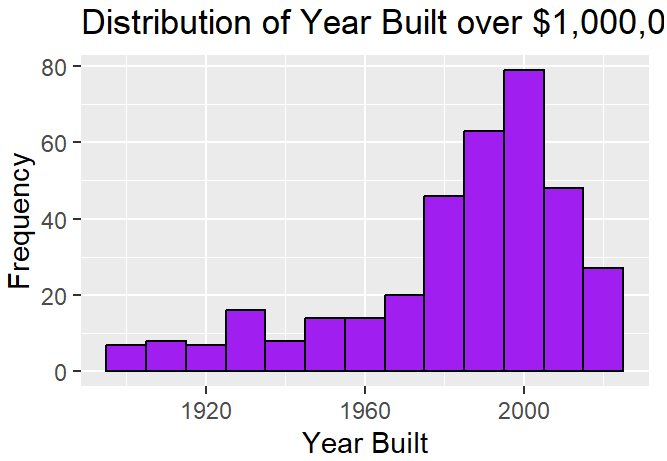


Grid of all Plots over $1,000,000


Distribution of House Square Feet over
Frequency
0
5
10
0
2500
5000
7500
1000
Square Feet


Boxplot of Sale Price by Number
Sale Price
2,000,000
4,000,000
6,000,000
0
1
2
3
4
5
6
7
Number of Bedrooms


Boxplot of Sale Price by Number of
Sale Price
2e+06
4e+06
6e+06
0
1
2
3
4
5
6
Number of Bathrooms


Distribution of Year Built over $1,000,0
Frequency
0
20
40
60
80
1920
1960
2000
Year Built