The Same Data, Difference Stories: How Visualization Choices Shape Housing Narratives

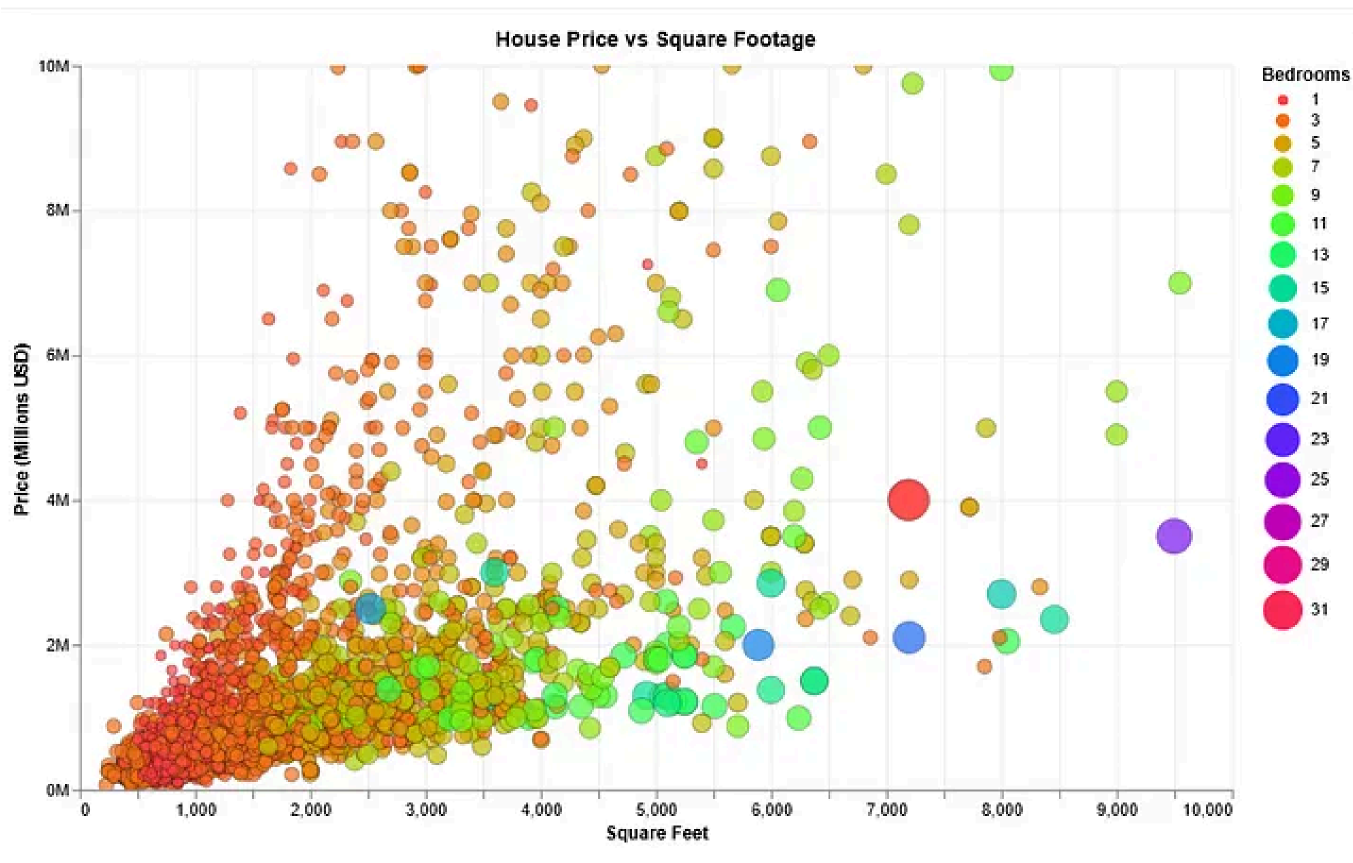
Khushbu Adhikari 14 min read · Just now

Our study examined how design decisions related to data visualization can change the story data tells, despite the numbers themselves being completely accurate. For this project, we decided to create two contradictory visual narratives using the exact same dataset, one narrative was an honest representation, while the other was accurate yet deliberately deceiving or misleading way. Each of us explored different methods of manipulation. In this project, we fully coordinated methods of manipulation, taking a distanced, yet systematic approach to turn neutral data visualizations into greater or lesser narratives. Reflections for our team revealed that a convincing professional-looking visualization carries just enough trustworthiness with viewers who would generally not question decisions in the design of a visualization. This experiment illustrates that the distinction between a manipulative visualization and one that is purely informative lies not within the data itself, but within the dozen of small decisions made across the visual to create a certain reading of the data by the viewer, reader, or user.

Informative Graph 1

Summary

For my final visualization, I created a scatter plot showing the relationship between house price and square footage for properties under 10,000 sqft. The x-axis represents square footage, the y-axis shows price (in millions), and each dot is a house. I encoded bedroom count with both dot size and color, ranging from 1–32, while the legend simplifies the display by stepping in even increments. This design makes it easy to see how both square feet and bedrooms interact with property value.

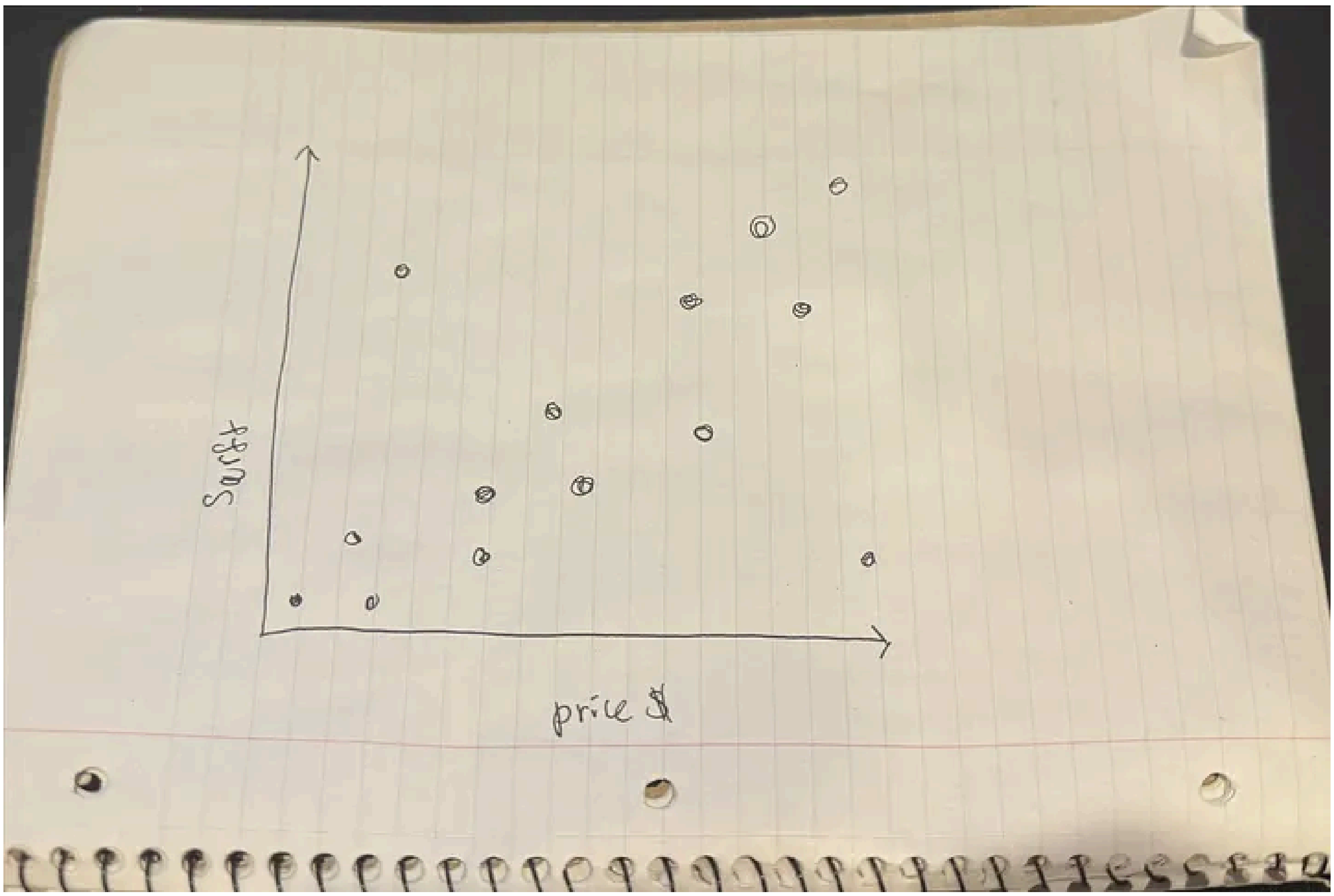


Final Version

Design Process

Ideation

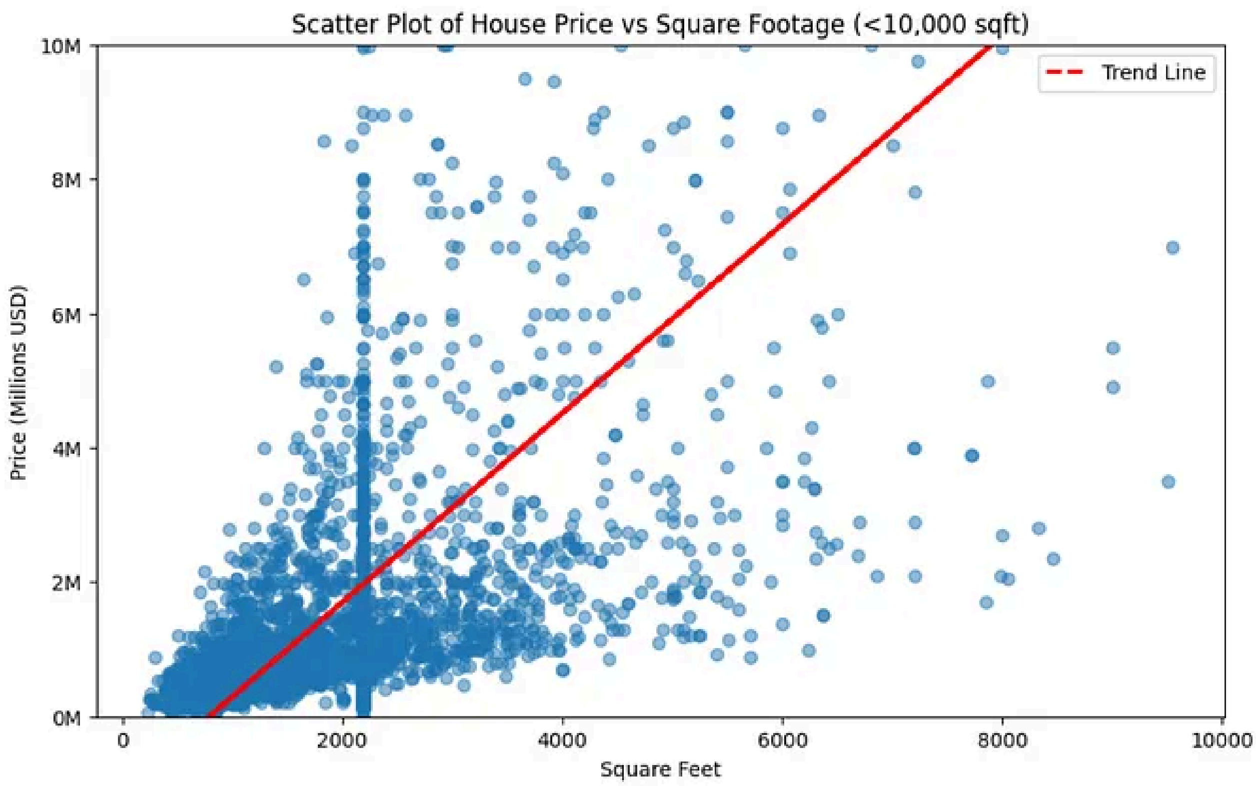
I began by brainstorming ways to display house price and square footage. Early sketches I considered bar charts and heatmaps, but I quickly realized scatter plots were more effective for showing variation across individual properties.



First Sketched Visualization

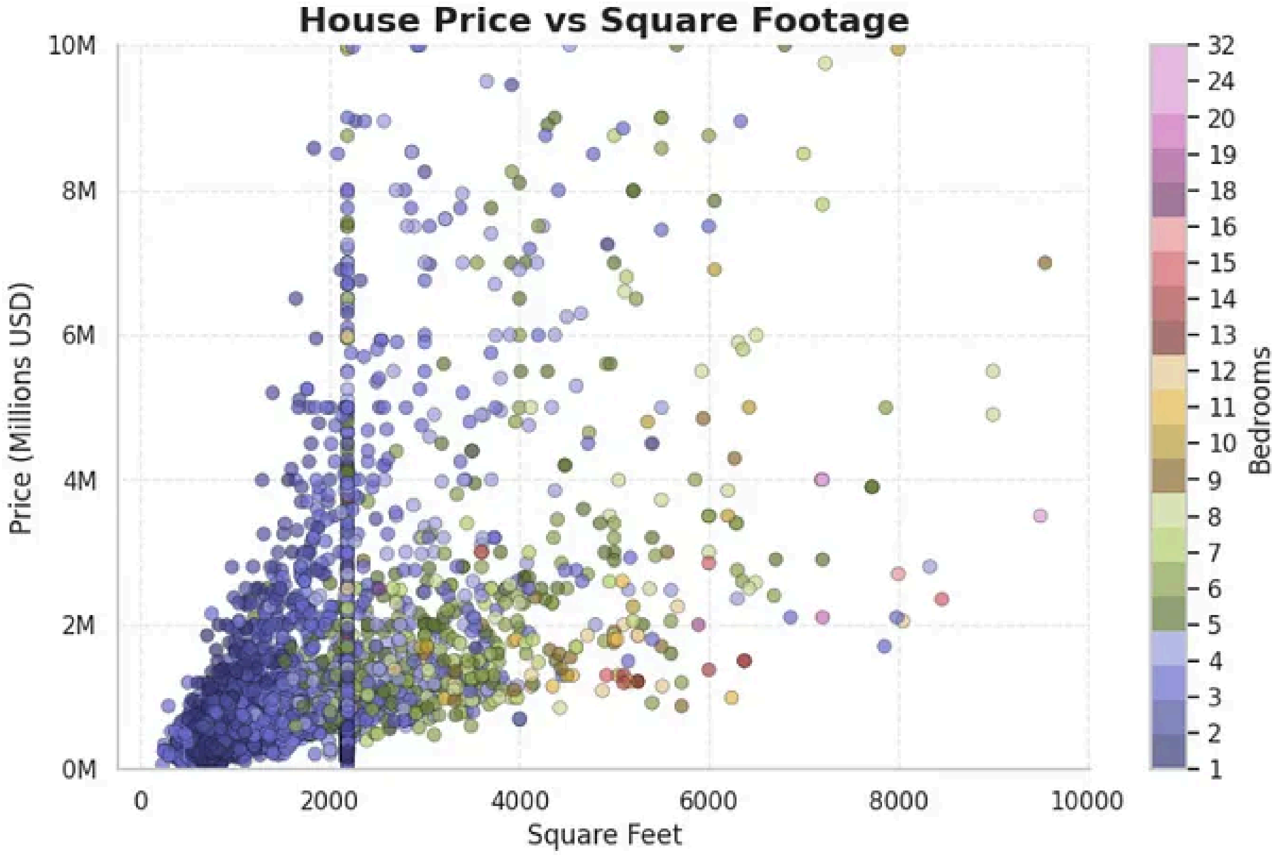
Prototyping

In my first prototype, I only had price and square footage for my variables. This made a decent scatter plot, but looked a bit bland. I also ran into the issue of not being able to distinguish between values of similar counts because my data point were all the same color.



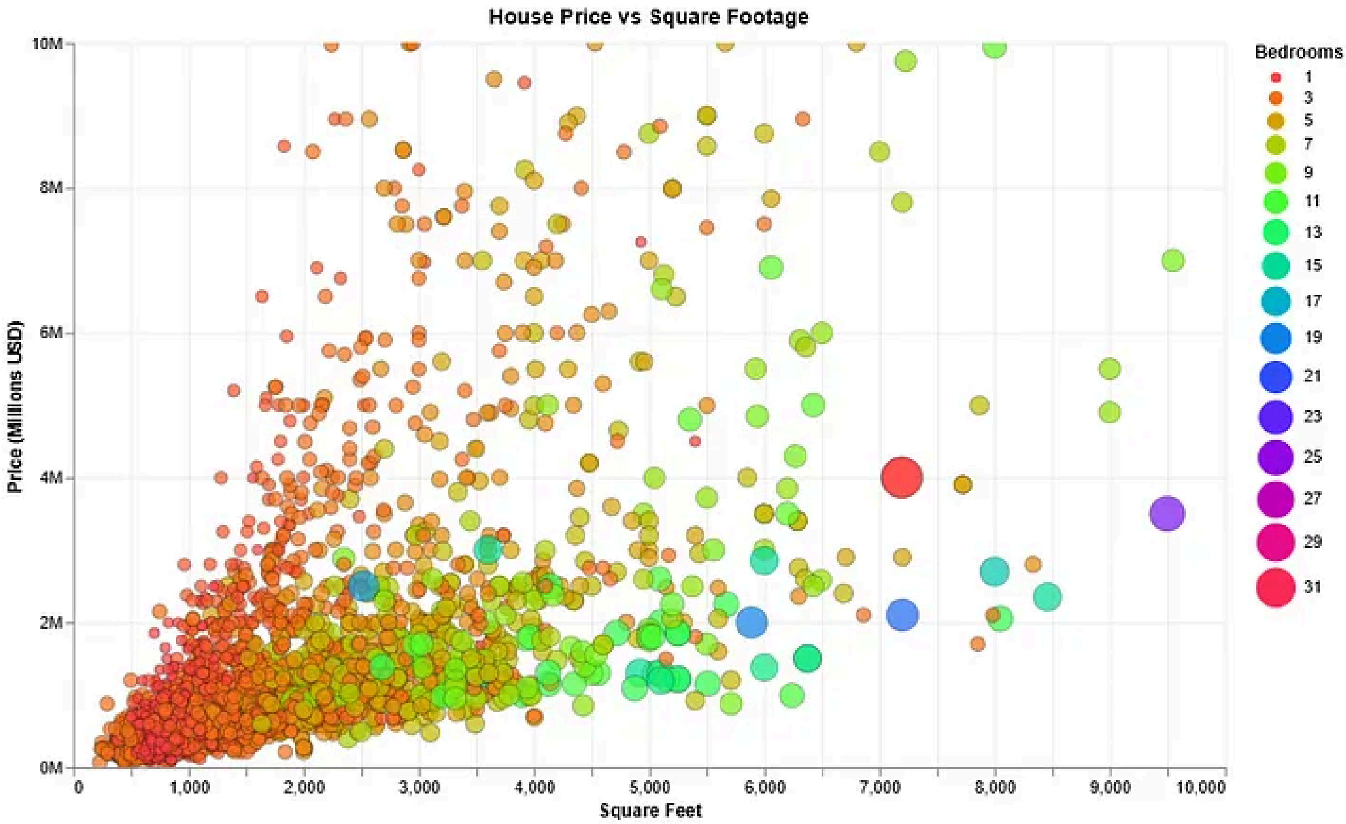
First Digital Protoype

In my second prototype, I introduced a gradient color scheme to represent another variable. This change made it much easier to distinguish the number of bedrooms at a glance, as the color mapping added an immediate layer of clarity to the visualization.



Second Digital Prototype

In my final prototype, I refined the visualization by adjusting both the color scheme and the sizing of the data points to enhance clarity. The updated color palette was chosen to make differences between properties more immediately recognizable, while scaling the dots by the number of bedrooms added an intuitive visual cue — the larger the dot, the more bedrooms it represents. I also redesigned the legend on the right to align more cohesively with these changes, ensuring that the overall presentation felt consistent and easy to interpret.



Final Version

Principles & Design Rationale

- **Expressiveness:** The encoding fully represents the underlying data without distortion — both price and square footage remain on continuous scales.
- **Effectiveness:** Encoding bedrooms with both color and size makes comparisons easier than using a single channel, especially since viewers naturally compare dot sizes and color shades.
- **Perceptual redundancy:** Using two encodings (size and color) for the same variable helps reinforce the information and makes the plot more accessible for different types of viewers

Reflection

Strengths: The scatter plot is intuitive and makes clear how square footage and bedrooms together affect price. The color and size combination enhances readability. The simplified legend improves usability.

Weaknesses: I couldn't fit outliers into the graph without distorting the whole graph. A house costing near a billion dollars would obscure the graph and make it almost impossible to read.

Feedback: I was told the graph overall was pretty good, however maybe it would be interesting to mess with the encoding a bit. So I ended up scaling the dots based on the number of bedrooms. This feedback led me directly to the finalized version of my graph.

Future Improvements

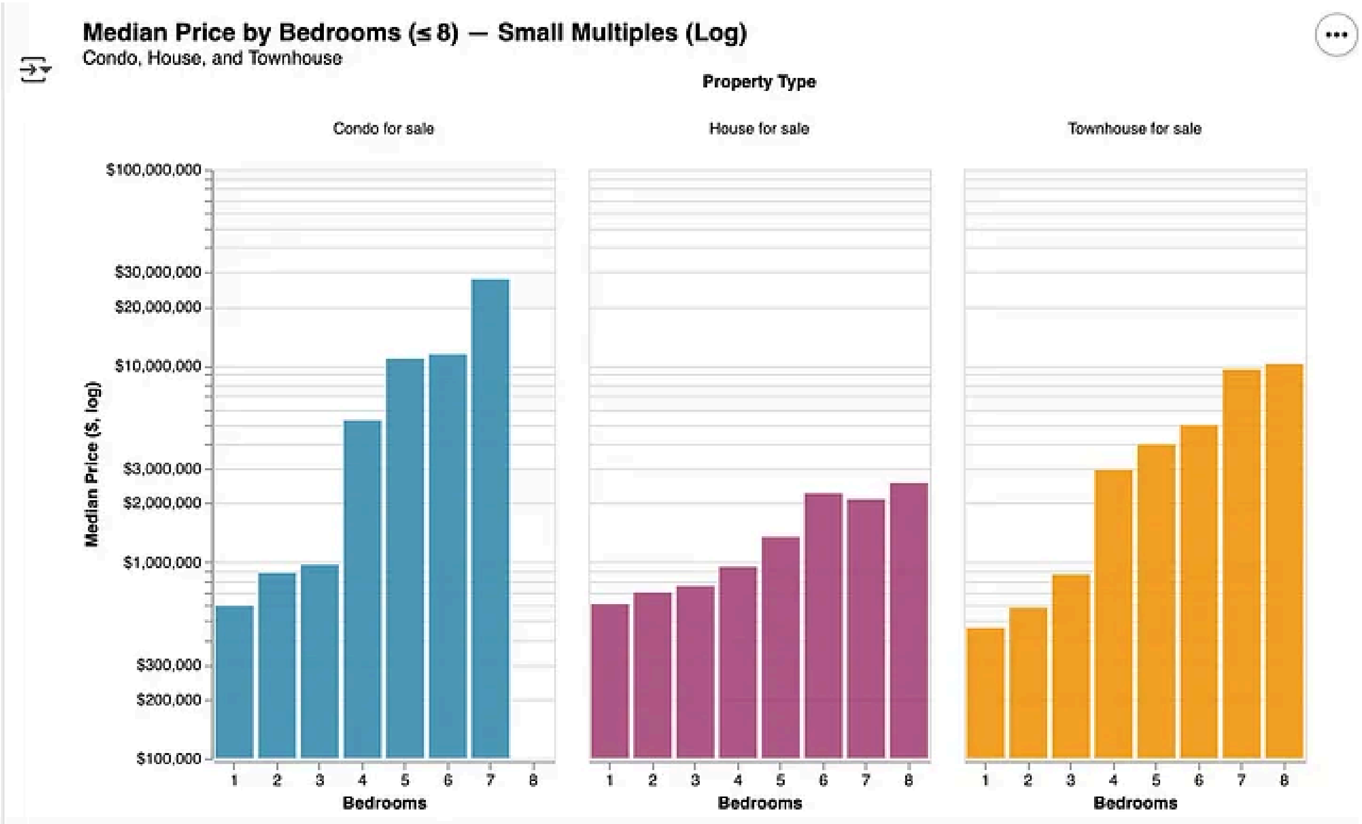
- Add interactive filters for bedroom range and price.

Include tooltips showing property details (exact square footage, price, and bedrooms).

Informative Graph 2

Summary:

Using the New York Housing Market dataset, I created an informative visualization that reveals dramatic price differences across property types and bedroom counts through a small multiples design with logarithmic scaling. The visualization uses three side-by-side bar charts (Condo, House, Townhouse) to show median prices by bedroom count, with a log scale that accommodates the extreme range from $500K studio condos to $30M penthouses. This approach prioritizes accurate representation of exponential price relationships while maintaining visual clarity across all market segments.



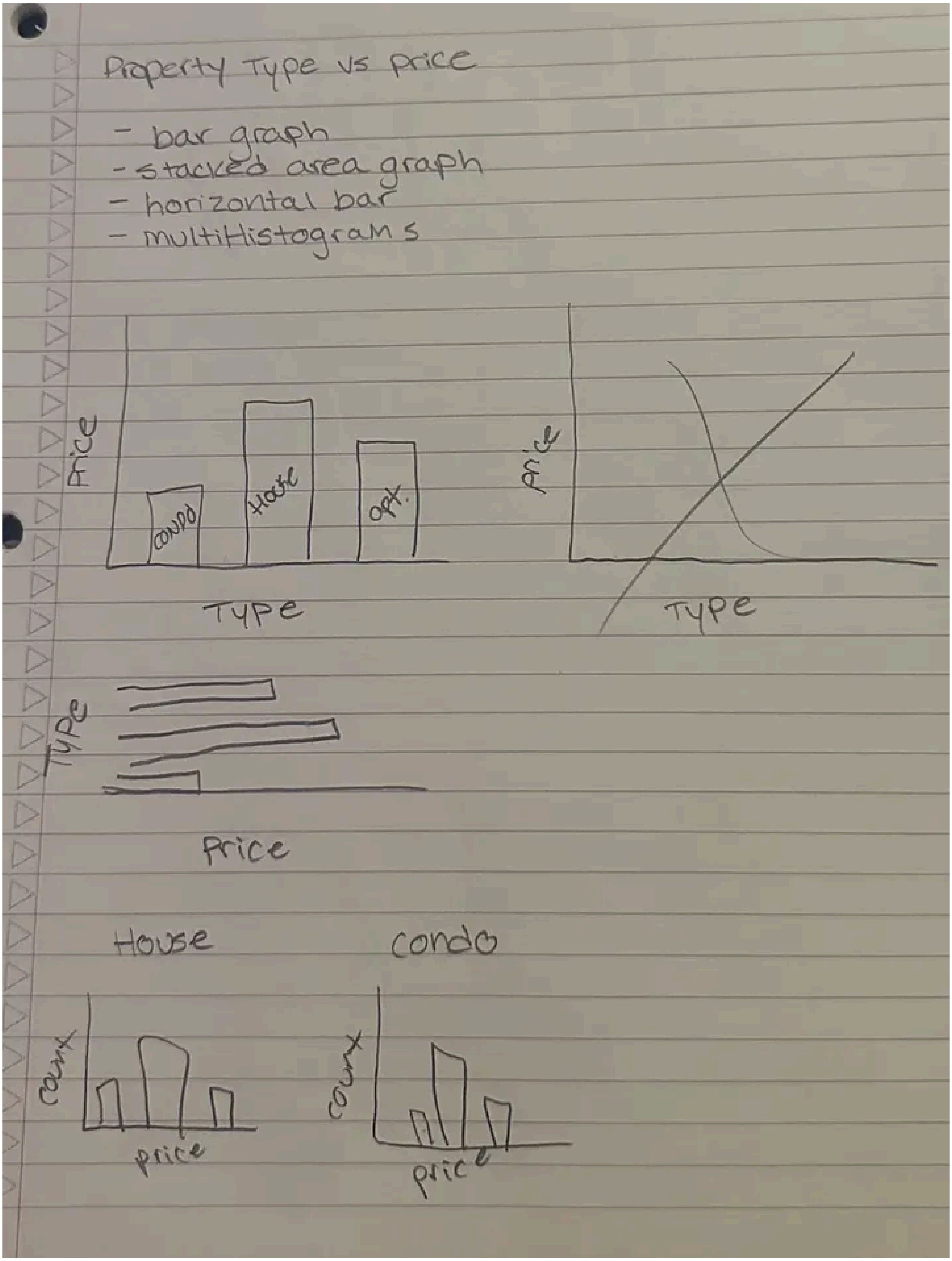
Final Design

Design Process

My initial sketching explored multiple approaches for representing the relationship between property types, bedroom counts, and median pricing. As shown in my notebook sketches, early concepts included:

- Bar graphs
- Stacked area graphs
- Horizontal bar charts

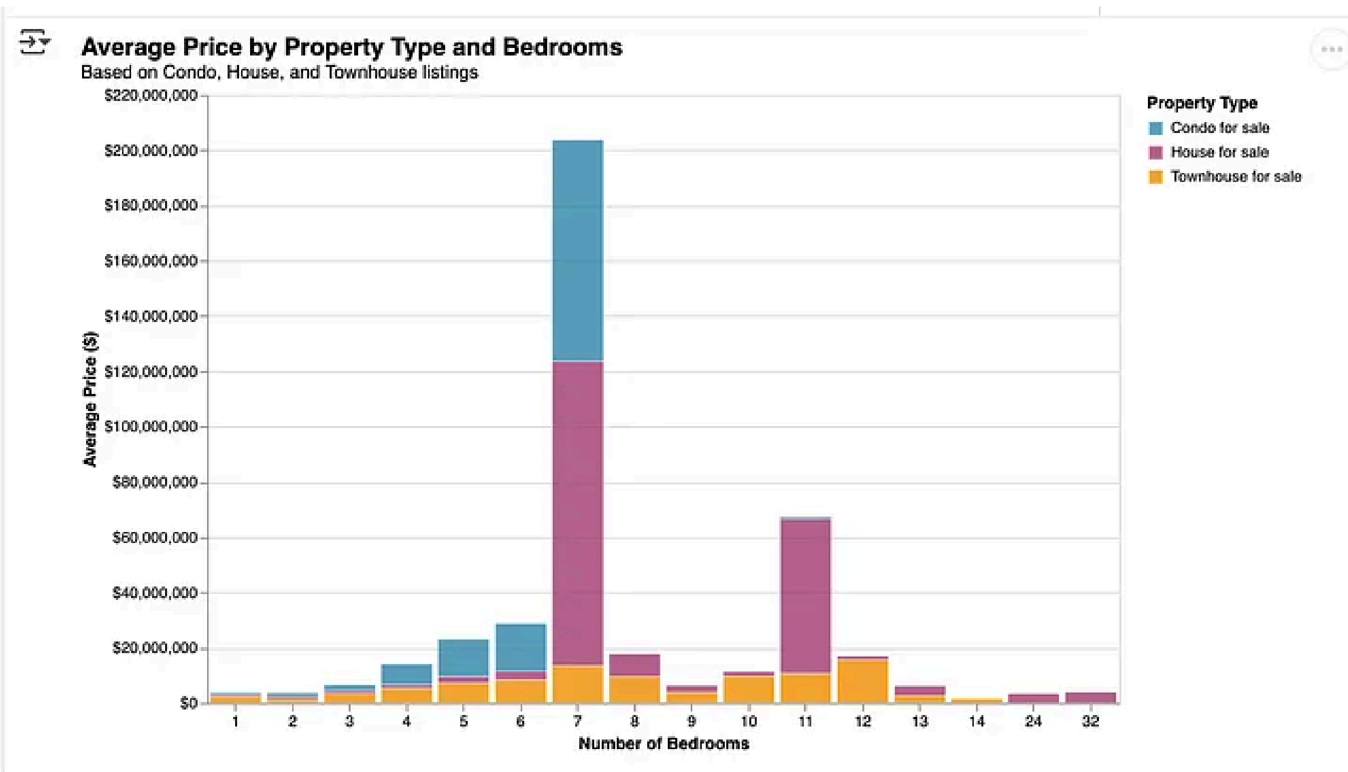
- Multi-histograms
- Small multiples approach


Property Type vs Price
- bar graph
- stacked area graph
- horizontal bar
- multiHistograms
Price
condo
House
apt.
Type
Price
Type
Type
Price
House
condo
count
price
count
price

Ideation

Prototype #1

My first digital prototype attempted a stacked bar chart showing average prices across all bedroom counts (1–32). This approach immediately revealed fundamental issues: the extreme values in 7-bedroom condos (~$200M average) completely dominated the visualization, making all other property types and bedroom counts virtually invisible.

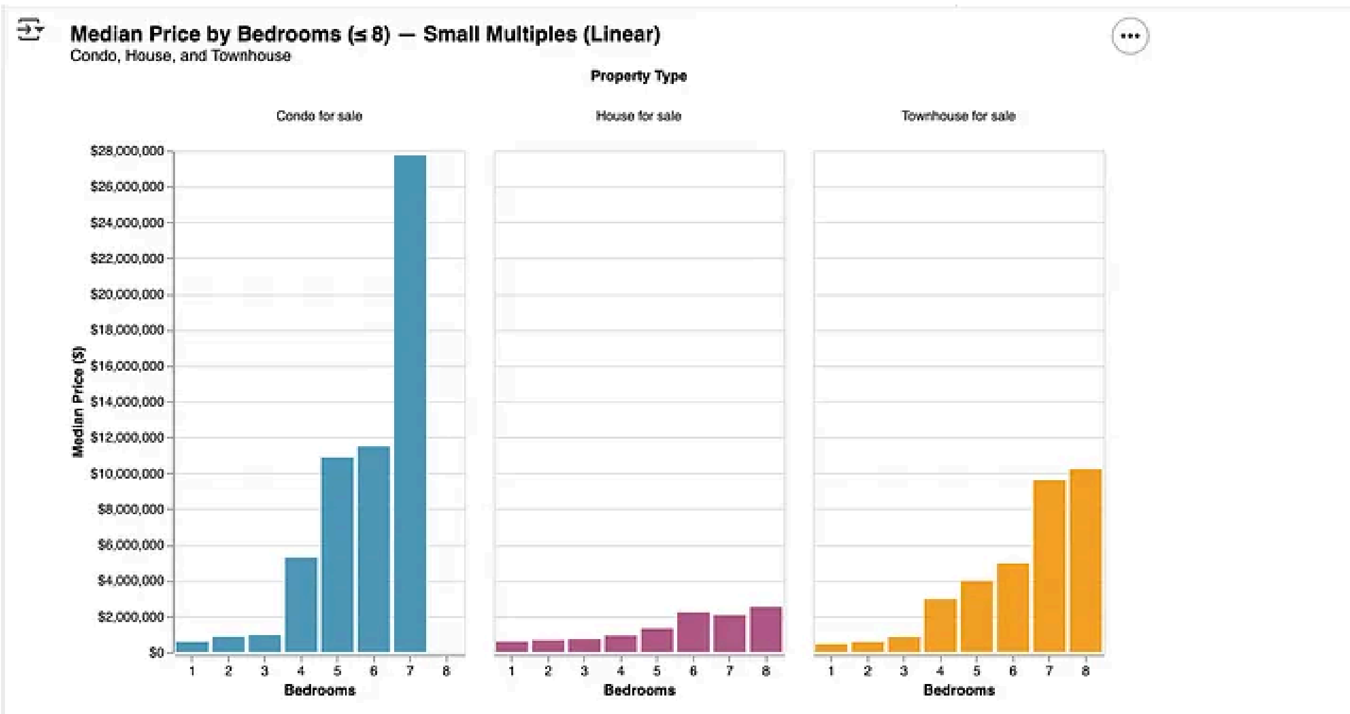


Digital Prototype 1

This iteration revealed that averages are problematic for real estate data due to extreme outliers and that combining all property types in a single chart creates visual hierarchy issues.

Prototype #2

The breakthrough came with separating property types into small multiples and switching to median values. However, the linear scaling still created severe problems — the 7-bedroom condo median (~$28M) dwarfed all other values, making lower-priced properties appear as barely visible slivers.

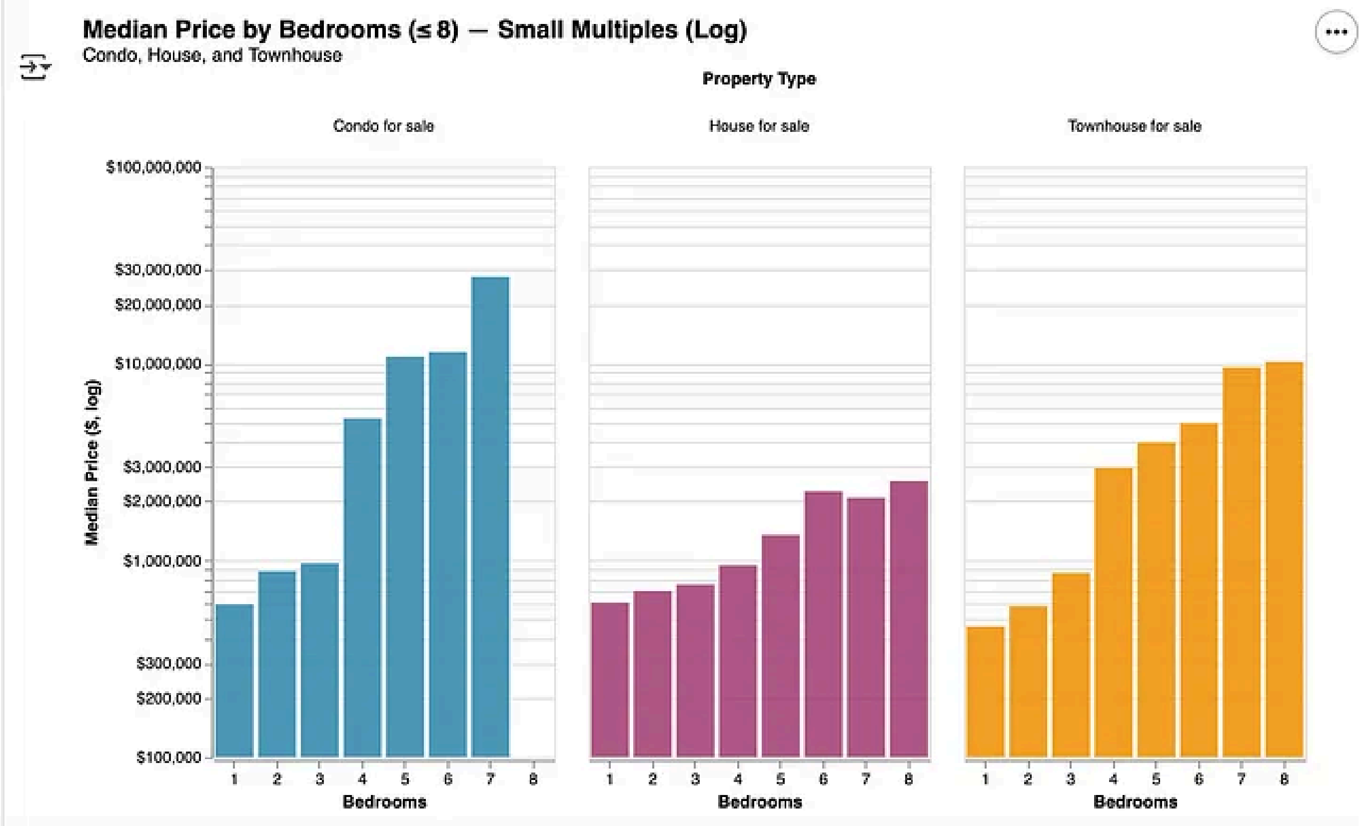


Digital Prototype 2

This version proved that small multiples structure was correct but linear scaling was inadequate for data with exponential relationships.

Final Prototype

The final breakthrough involved implementing logarithmic scaling on the y-axis. This mathematical transformation ensures each order of magnitude receives equal visual space, allowing viewers to perceive both $500K studios and $30M penthouses with appropriate proportional representation.



Final Prototype

User Testing and Feedback

I conducted testing with four individuals outside the class, showing them both the linear and logarithmic versions.

Linear Scale Feedback:

- “I can see the condos are expensive, but I can’t really tell anything about the houses or townhouses”
- “This doesn’t seem very useful. Everything except that huge bar is basically invisible”
- “Are houses really that cheap?”

Logarithmic Scale Feedback:

- "Now I can actually see patterns in all three property types"
- "The log scale is confusing at first but now it makes more sense"
- "This shows that condos get exponentially more expensive, while houses are more predictable"
- "I can finally compare across property types"

Key Design Iterations Based on Feedback:

- Added explicit "(log)" notation in the y-axis title after users expressed confusion about the scale
- Enhanced price formatting with clear dollar signs and comma separators
- Maintained consistent bedroom range (1–8) across all charts for direct comparison
- Filtered extreme bedroom counts (>8) which users found distracting

The user testing strongly showed that the logarithmic transformation was the best decision. Every participant noted that the linear version was basically useless for comparison, while the log version enabled the analytical insights they were seeking.

Principles & Design Rationale

Logarithmic Scaling Justification: The log transformation addresses what Cleveland and McGill termed "dynamic range" challenges. By using log base 10, each visual unit represents a 10x increase, making exponential relationships linear and interpretable while preserving proportional accuracy across the full data range.

Sequential vs. Categorical Color Design: As property type is nominal data, I adopted a categorical color scheme (blue-purple-orange) instead of sequential colors, adhering to the guidelines outlined in Muth's "When to use sequential and when to use diverging color scales."

Reference Point Effects: Research shows people have biases toward reference points when viewing data. I used median instead of mean prices because medians provide more stable reference points that aren't distorted by extreme outliers.

Reflection

Strengths:

The small multiples approach reveals market structures that alternative approaches would obscure. The logarithmic scaling democratizes visual space, which reveals that no price range dominates at the expense of others. This creates more equitable representation across market segments. Viewers can concentrate on data patterns instead of deciphering visual differences between charts because of the consistent visual structure, which lessens cognitive load.

Weaknesses:

The log transformation emphasizes proportional relationships while de-emphasizing absolute differences. A $1M increase appears smaller at higher price ranges, which may underestimate the real financial impact.

Feedback:

I was told the initial stacked bar chart was too cluttered and hard to read across different property types. My classmates suggested breaking it into multiple small graphs so you could actually compare condos, houses, and townhouses side by side. Several people pointed out that using averages was misleading because of extreme outliers. So I changed from mean to median calculations

Future Improvements:

Implement toggle functionality between linear and logarithmic scales, enabling users to explore both absolute and proportional relationships according to their analytical needs.

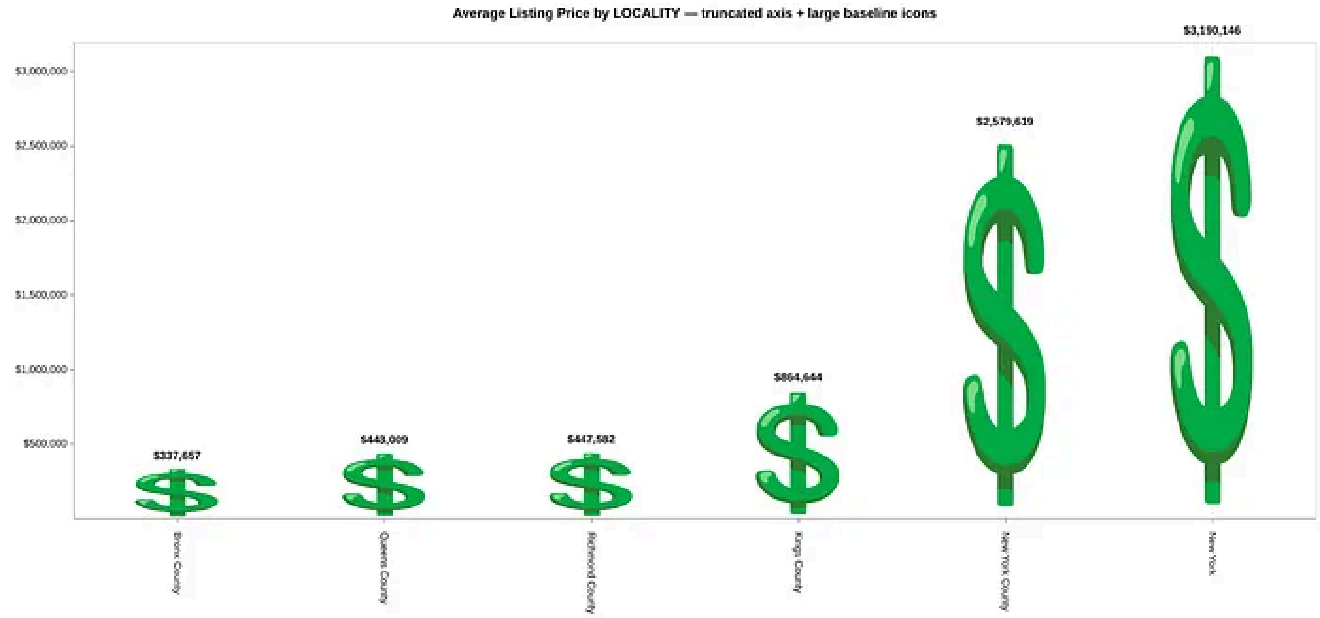
Manipulative Visualization (One)

Summary:

I created a deceptive bar chart titled “The Big House Gap” that dramatically exaggerates differences in average property sizes across New York localities. By combining a truncated y-axis starting at around 1000 sqft with house icons that scale in both dimensions, the visualization creates a compound distortion where New York properties appear monumentally larger than outer borough properties. The actual 2.2x difference between Bronx County (1217 sqft) and New York (2694 sqft) is visually amplified to appear as a roughly 14x difference in perceived area.

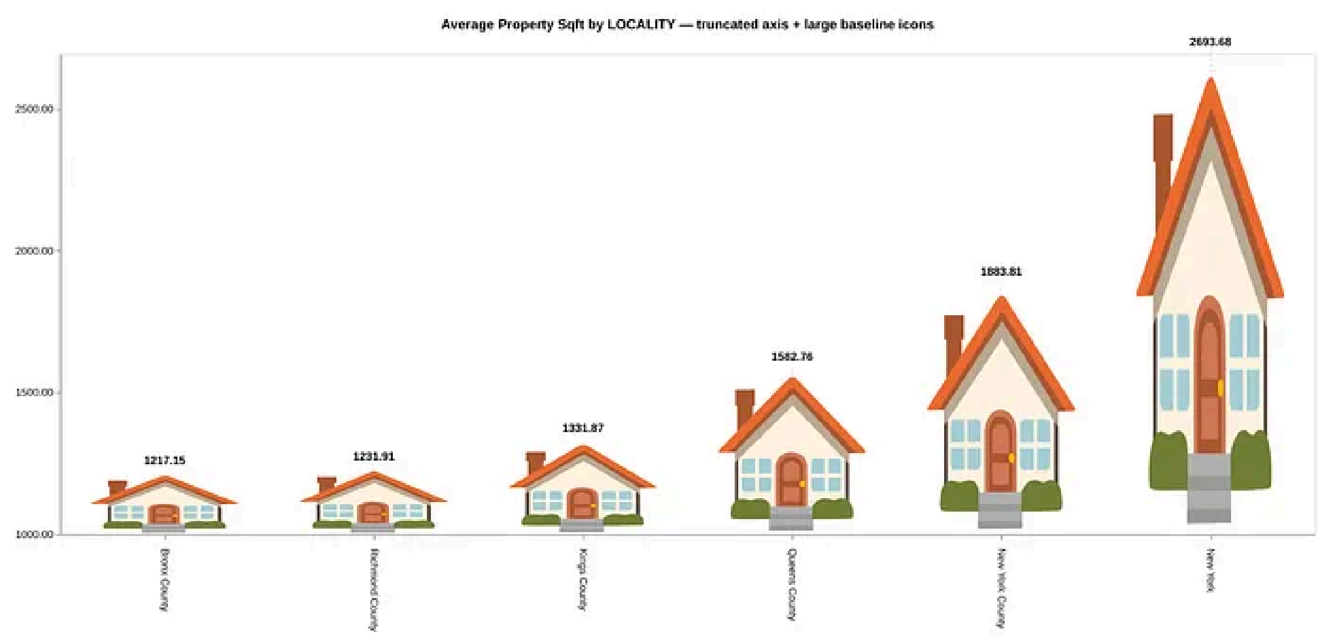
Design Process:

The manipulation strategy evolved through careful iteration to maximize deceptive impact while maintaining surface credibility.



Iteration 1

My initial concept used dollar signs to represent price differences by locality. However, I quickly realized this was too abstract and didn't create the visceral comparison I wanted. Dollar signs also had a different effect on perception than what I was trying to get across, as the icon's area doesn't scale in an exaggerated manner.



Iteration 2

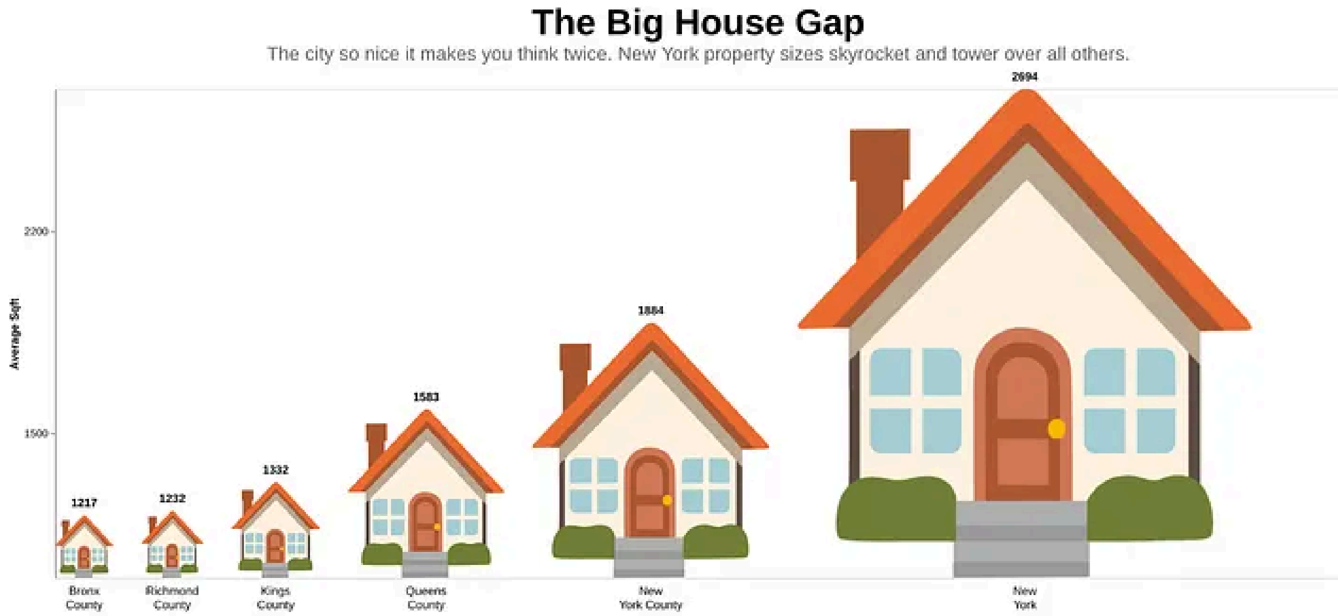
I pivoted to house icons, which proved far more effective for several reasons:

Houses create an immediate emotional connection to the "housing gap" narrative. Also, the metaphor feels natural. Bigger houses for bigger average square footage. Most importantly, scaling house icons proportionally exploits a well documented perceptual bias where viewers perceive area differences rather than height differences.

At this stage, I discovered the power of the dual manipulation — the house that's 2.2x taller has approximately 4.9x the visual area.

Iteration 3

I refined the scaling algorithm to maintain proper aspect ratios while ensuring the houses looked architecturally reasonable. Houses that were too stretched or squashed broke the illusion and made viewers suspicious. I also adjusted spacing between bars to prevent the smaller houses from being too distracting.



Final Prototype

Based on user feedback, I made several crucial refinements. I removed unnecessary decimal precision (1217 instead of 1217.15) to appear more professional and less overly specific. I added a gray subtitle using loaded language ("skyrocket and tower") to prime viewers' interpretation. I made the x-axis labels upright for better readability. I added the "Average Sqft" label along the y-axis label to provide false legitimacy and clarity. I also included outliers strategically, I intentionally included apartment buildings over 60,000 sqft in the New York average, using mean instead of median to inflate the value differences.

Data Manipulation Techniques:

Beyond the visual deceptions, I manipulated the underlying data by using mean instead of median, making the data vulnerable to extreme outliers, particularly beneficial for Manhattan where luxury penthouses and entire apartment buildings skewed averages upward. I also included commercial properties; entire apartment buildings counted as single units in my uncleaned, unfiltered data.

Design Principles Violated

Tufte's Lie Factor: The visualization has an astronomical lie factor. The actual data shows New York at 2.21x Bronx County's average, but the visual representation (combining truncated axis and area scaling) creates a perceived difference of approximately 14x. This exceeds Tufte's threshold for graphical distortion by over 500%.

Cleveland & McGill's Graphical Perception: I deliberately exploited their findings that people are poor at comparing areas. While humans can accurately compare positions along a common scale, area comparisons are processed with systematic bias toward overestimation. By encoding the data in both height and width, I compounded this perceptual error.

Gestalt Principle of Similarity: The house icons create false equivalence: a viewer assumes each house represents a similar unit (a single home), when in reality they represent averages that include everything from studio apartments to entire buildings.

Zero-baseline requirement for length encodings: Bars/length comparisons should start at zero, starting the y-axis near 1,000 sqft inflates perceived differences. In my chart, the Bronx (1217) vs New York (2694) looks about 7.8x apart in length when measured from the truncated baseline, though the data ratio is only about 2.21x. This violates the convention for magnitude encodings and amplifies the story.

I tested the visualization with 3 people outside class:

Initial Reactions included noticing how huge New York looks, and how much bigger it is (they emphasized the difference in sizes. They also noted the inequality and how small the outer boroughs are. Specific feedback included testers immediately noticed the size disparities, and one mentioned the spacing between houses felt "off" because the column widths were off.

The title successfully framed interpretation based on their reactions.

Refinements Based on Feedback:

I adjusted the spacing to be perfectly uniform (irregular spacing was raising suspicion). I added more authoritative styling (grid lines, proper labels, gray subtitles) to increase perceived legitimacy, and I changed the title from "Average Property Sizes" to "The Big House Gap" for stronger framing.

Reflection:

Strengths: The visualization achieved complete narrative control while maintaining professional appearance. The house metaphor resonated with viewers, making them less likely to scrutinize the mathematics. The compound manipulation (truncated axis + area scaling) was remarkably effective. No tester identified both deceptions without prior warning.

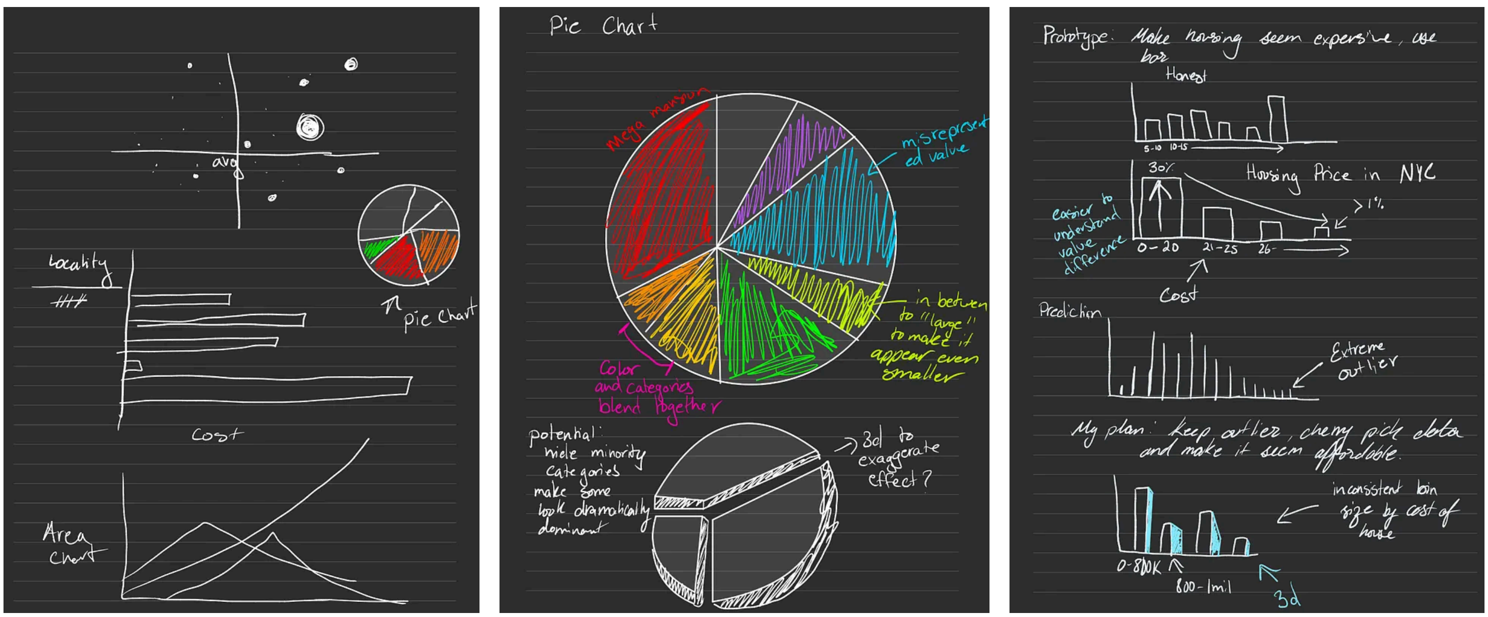
Weaknesses: Including properties over 60,000 sqft risks credibility if someone investigates the data. Some viewers might notice the y-axis doesn't start at zero if they look carefully.

If I had more time I would improve the iconography and make the graphic look more professional than the openly licensed graphic I utilized. This would improve the impact on the user, as they would implicitly trust a professional graphic icon more.

Manipulative Visualization (Two)

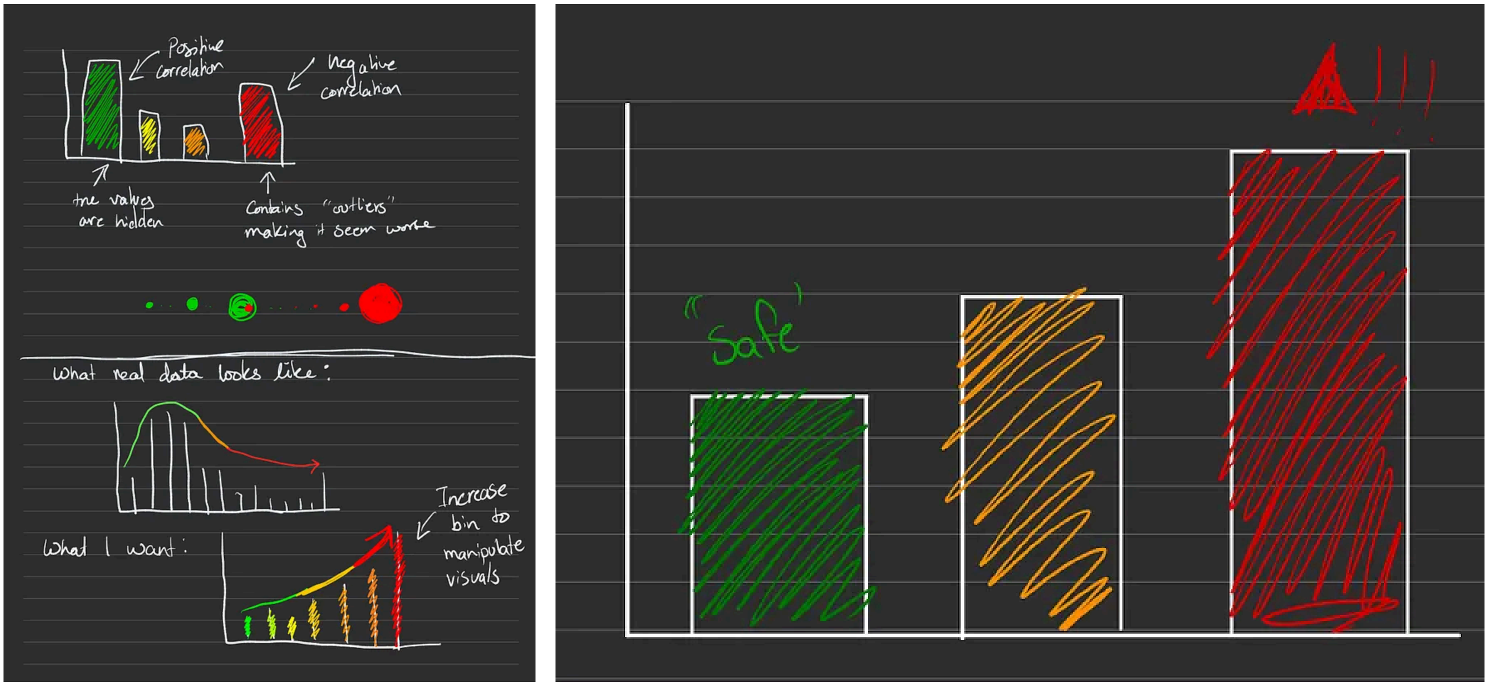
The Design Process:

The ideation phase began with a general understanding of the contrasting narratives we wanted to explore. Numerous graph concepts were sketched, playing with different views on the same housing data before narrowing down to three promising approaches. After selecting the most effective concept, we dove deeper into specific manipulation techniques, sketching various binning strategies and color encoding possibilities that could influence viewer perception without being immediately obvious.



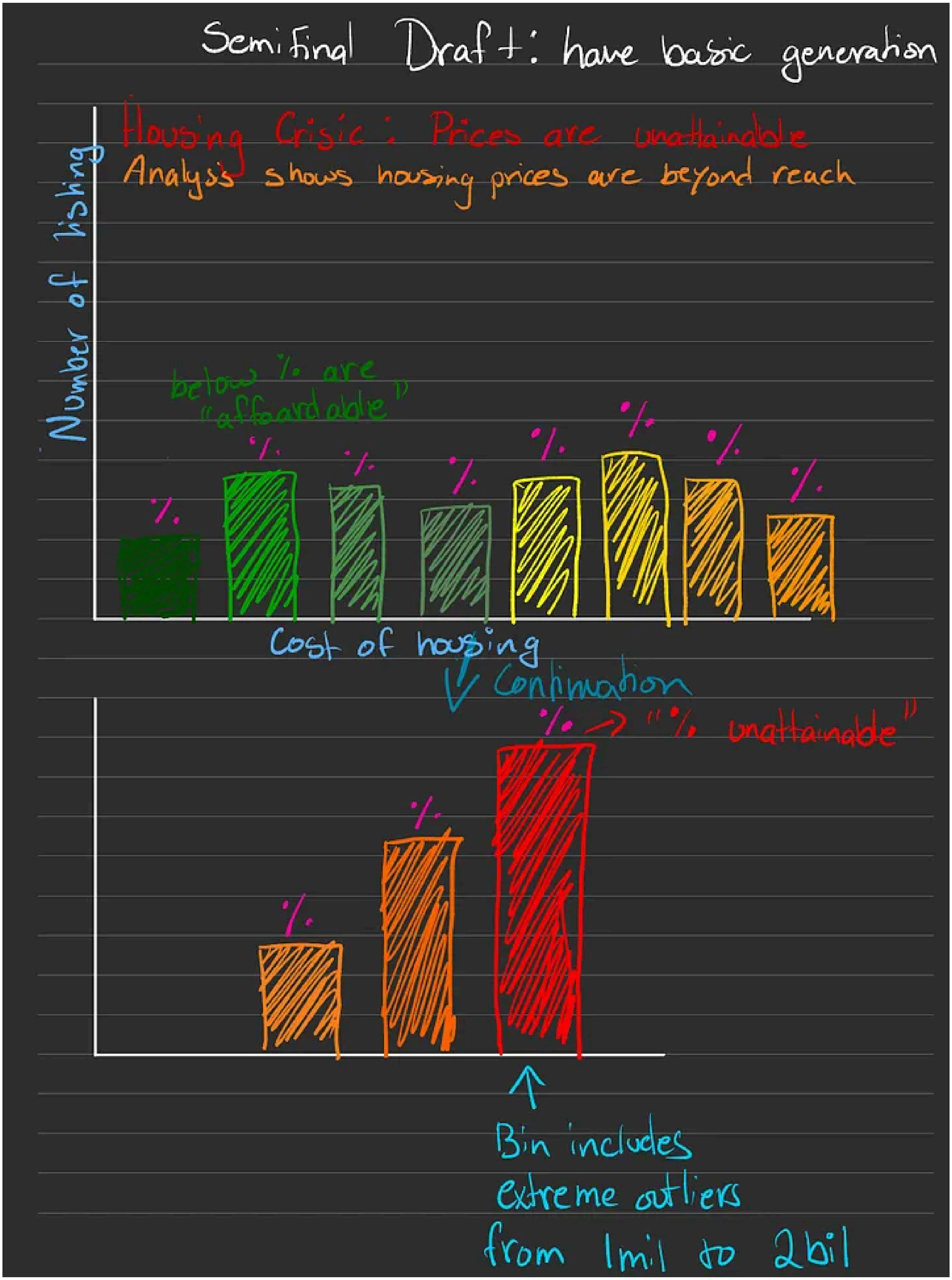
Ideation 1

The prototyping phase focused on developing the "housing crisis" narrative with strategic data presentation. We fragmented the lower price ranges into $100,000 bins, while everything above $1.5 million was kept into one "unattainable" category. This was done purposefully, especially by including some extreme outliers, to inflate the percentage of seemingly affordable properties. The fragmentation implied that affordable housing was scarce, whereas the massive upper bin suggested there was an overwhelming luxury market.



Ideation 2

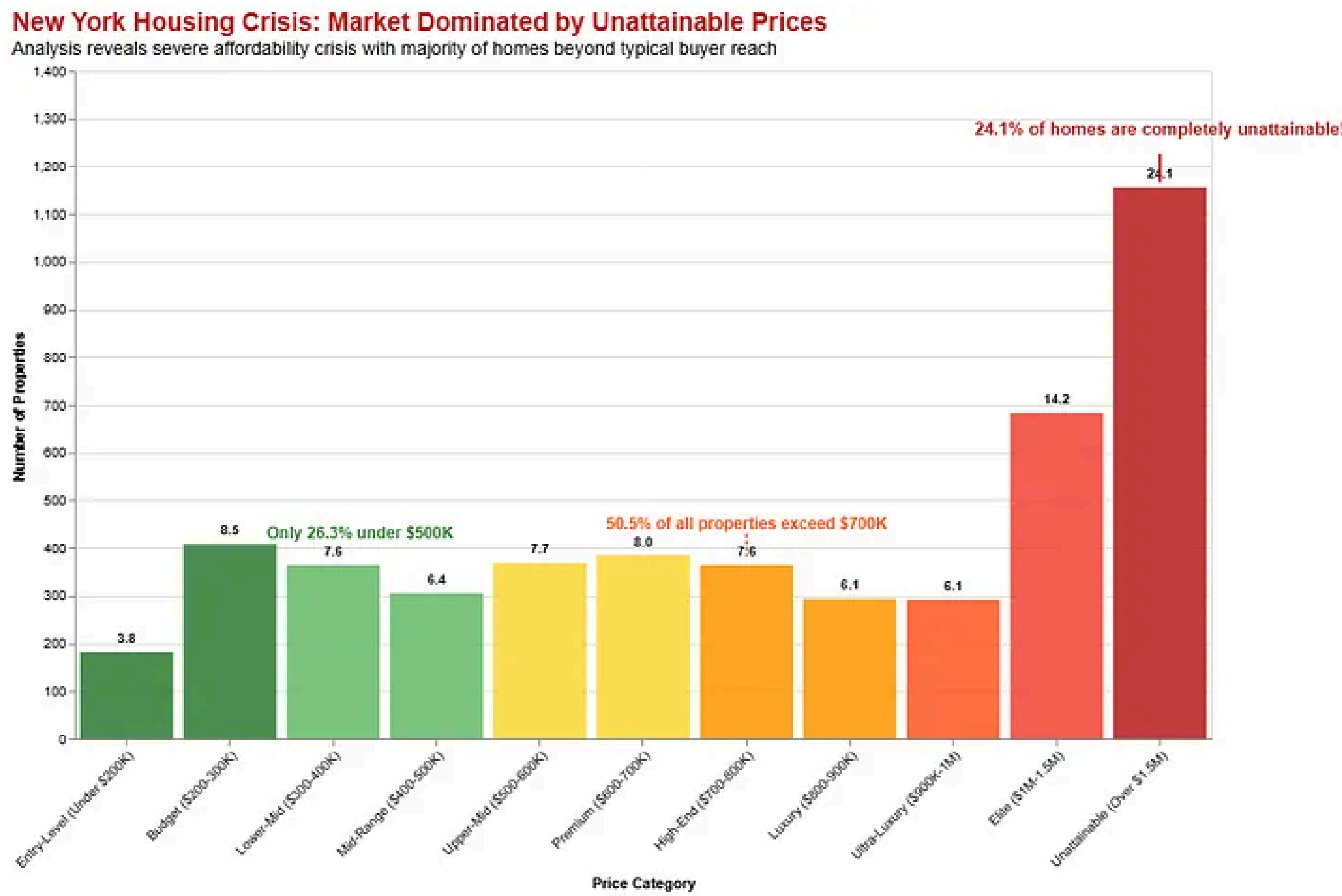
During prototyping, the visual encoding strategy was refined. Initial iterations used a red-green scheme, but it was too simplistic. In the final draft, the gradient went from dark green to yellow to orange to deep red, creating a naturalistic transition that viewers accepted as logical rather than manipulative.



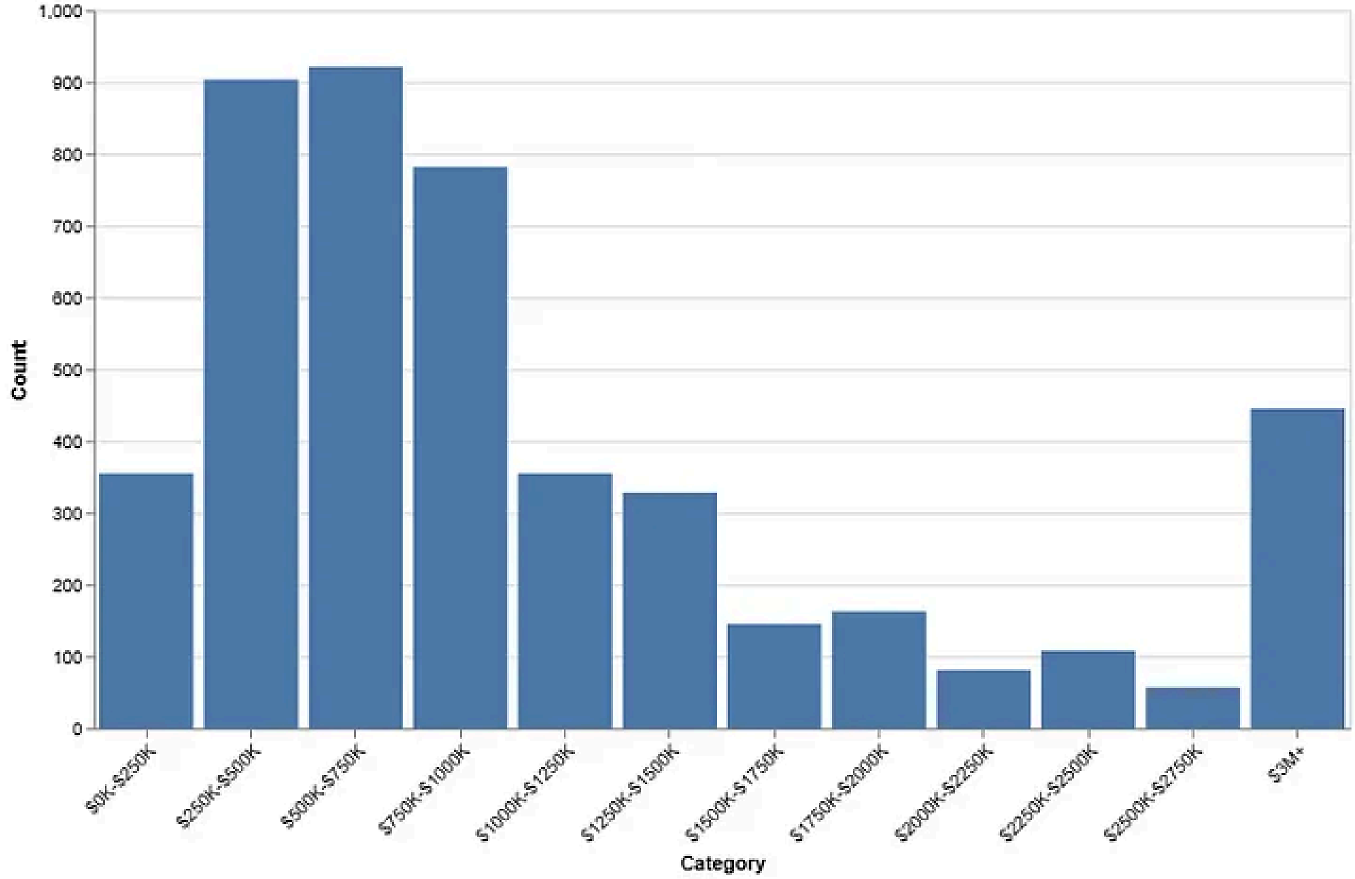
Semifinal Draft: have basic generation
Housing Crisis: Prices are unattainable
Analysis shows housing prices are beyond reach
Number of listing
below % are "affordable"
%
Cost of housing
Continuation
% → "% unattainable"
Bin includes extreme outliers from 1mil to 2bil

Ideation 3

We placed annotations strategically: “Only 26.3% under $500k” emphasizing scarcity while the “24.1% completely unattainable!” created a sense of alarm. The annotations were placed directly above the graph to prevent visual clustering while maintaining narrative flow.



Ideation 4 (With Outliers)



Real Distribution (Without Outliers)

Design Principles Implemented:

Graphical Integrity (Tufte): We violated Tufte's basic principle by creating a lie factor. The "Unattainable" bin covered $1.5 Mil to $2.14 Bil while appearing only slightly larger than $100k bins. This misrepresented the visual size and proportions of the actual data relationship. Additionally annotations also emphasized misleading statistics rather than the true data representation.

Pre-Attentive Processing (Healey): The color selections exploited what is called "rapid visual processing" that occurs prior to conscious analysis. Deep red color delivered a threat response creating an emotional context before any comprehension of the prices. The continuum of green-to-red created immediate associations of safety-to-danger that bypassed any rational evaluation of the actual price ranges.

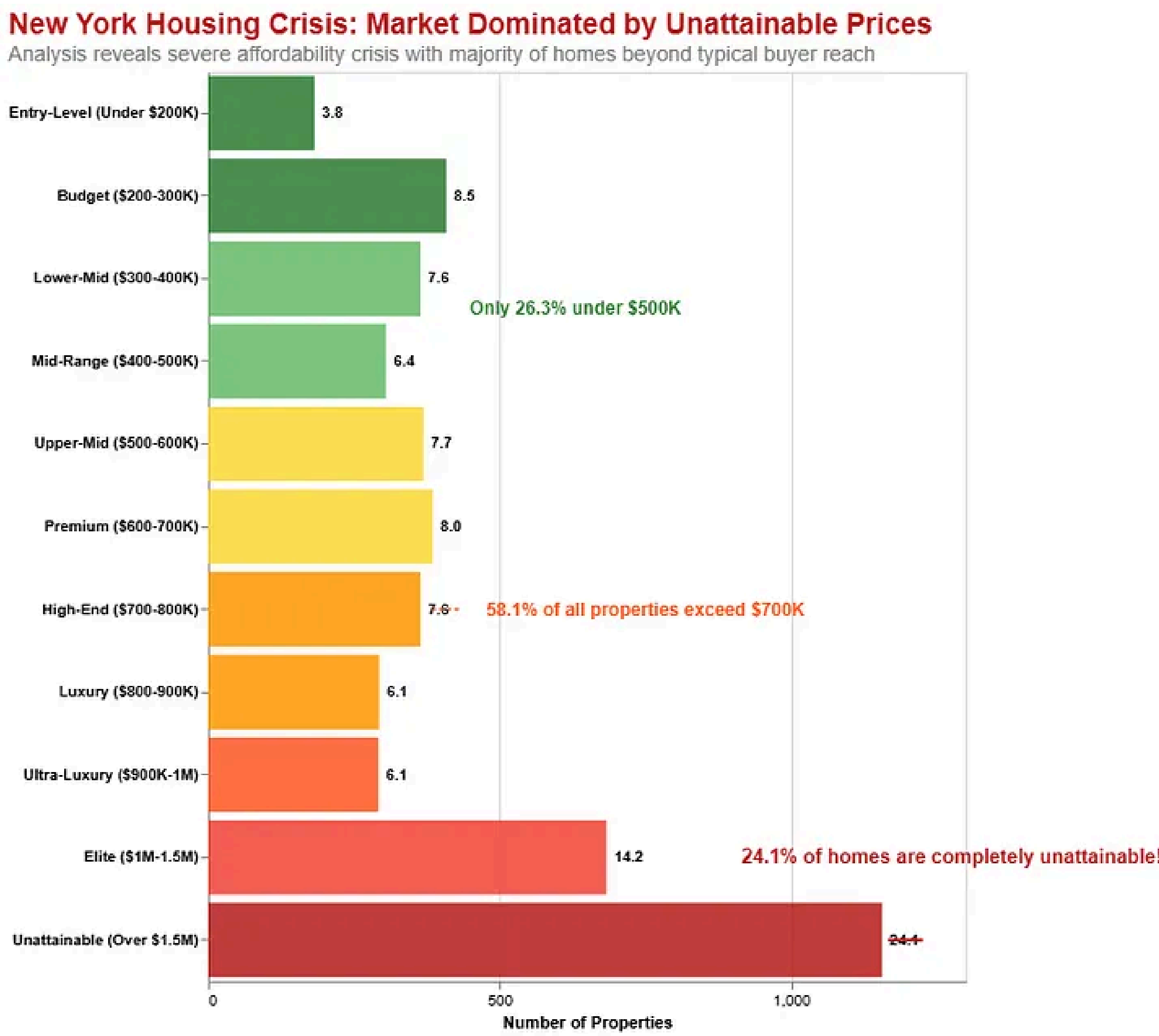
Gestalt Principles: The use of proximity manipulation broke affordable ranges up into ten distinct bars, inhibiting visual grouping of related categories. In addition, the similarity in color may have falsely grouped unrelated price ranges- for example, the yellow tones of the $500–700k ranges appeared in moderate category through their association with both high and low extremes, despite being near the median price of the market.

Perceptual Redundancy: There were several encodings that supported the deception. The color provided emotional priming, the placement created fragmentation, and the difference in the size of the bins distorted reality of proportion, and most importantly, the text amplified the manipulation. The title "Housing Crisis: Market Dominated by Unattainable Prices" created a frame for interpretation before viewers looked at the data. X-axis labels such as "Ultra-Luxury" for $900k which is barely above the median and "Entry-Level" for $200k refined the positioning of the data through text. Each element concealed the deception of the other, creating a richer deceptive illusion.

User Response and Refinement:

Prototype testing revealed the effectiveness of visualization exceeded expectations. Viewers accepted the premise of the crisis narrative, with their comments being primarily focused on, "how bad things have gotten," and not one questioning how the information was presented. No one questioned the unequal bin sizes, or colored nature of the bins initially, until they were explicitly told that it was a manipulation tactic. The professional appearance of the visualization (clean design, clear labels, statistical annotations) gave it enhanced credibility.

Based on the feedback, we made several critical adjustments. The x-axis labels that were rotated at 45-degree angle, made it for viewers to read and dampened the immediate emotional impact we were hoping for. We decided to move to a horizontal bar chart orientation that allowed the category labels to be displayed in a more natural manner while still enabling viewers to read them. In order to help ensure the emotional messaging retained its impact with the format change, we bolded the x-axis labels to make them pop as loaded terms. We also simplified the design by removing unnecessary y-axis labels which helped reduce visual clutter.



Final Visualization

Reflection:

Strengths: The visualization achieved total control over the narrative while remaining credible professionally. Technical accuracy provided a cover for deception in the context, the numbers are correct, just presented deceptively. User testing confirmed viewers accepted the crisis narrative without questioning the underlying structure.

Weakness: Including extreme outliers ($ 2.1B property) posed a potential credibility risk if discovered. The fragmentation into ten categories occasionally raised suspicion about "overly specific" ranges.

Ethical Implications: This demonstrated how easily "objective" data can be manipulated with presentation choices. The professional nature contributed to the deception. The clean design and statistical annotations created trust, distracting from critical engagement. Most concerning was how easily people accepted the story as a fact, demonstrating the degree to which visualization shapes data understanding and creates trust.

Written by Khushbu Adhikari

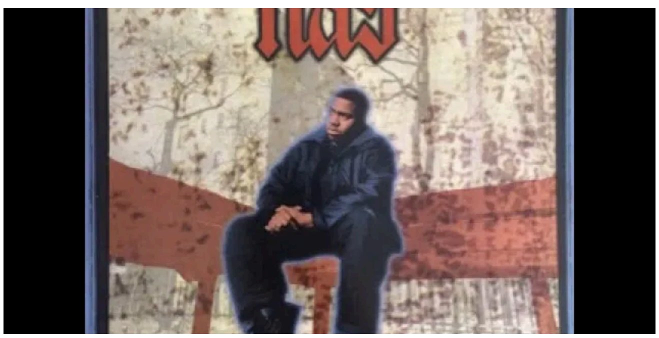
0 followers · 1 following

No responses yet

Khushbu Adhikari

What are your thoughts?

Recommended from Medium

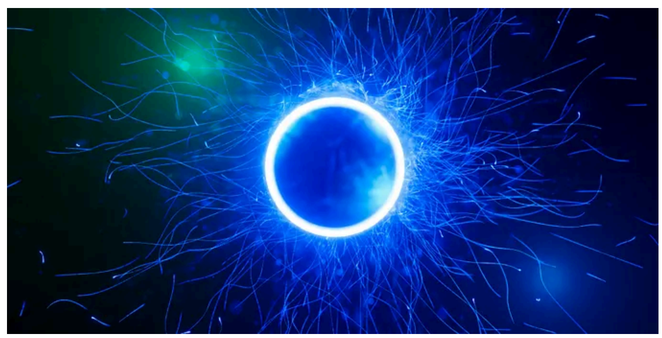


Dopetrackz Radio

The Song That Carried Me Through Long Nights

Written By: Darpan Murly

Sep 14 161



Will Lockett

The AI Bubble Is About To Burst, But The Next Bubble Is Already…

Techbros are preparing their latest bandwagon.

Sep 14 8.8K 292

In Long. Sweet. Valuable. by Ossai Chinedum

I'll Instantly Know You Used Chat Gpt If I See This

Trust me you're not as slick as you think

May 16 · 24K · 1438

In How To Profit AI by Mohamed Bakry

Your ChatGPT History Just Went Public on Google. Here's What I Di...

Safety/Privacy Check Prompt Template Is Included

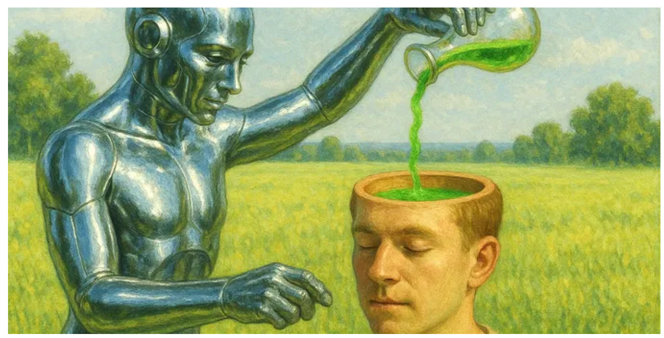
3d ago · 10.9K · 322

In Write A Catalyst by Utsuk Agarwal

The 10 Morning Habits That Quietly Make You Unstoppable

I Tried Them for 6 Months—The Results Were Unreal

Aug 10 · 9.7K · 310



Jordan Gibbs

ChatGPT Is Poisoning Your Brain...

Here's How to Stop It Before It's Too Late.

Apr 29 · 26K · 1287

See more recommendations