Tufte Chapter 4 – Data Ink

In his own words Tufte's five principles of data graphics:

- Above all else, show the data
- Maximize the data-ink ratio
- Erase non-data ink
- Erase redundant data
- Revise and edit

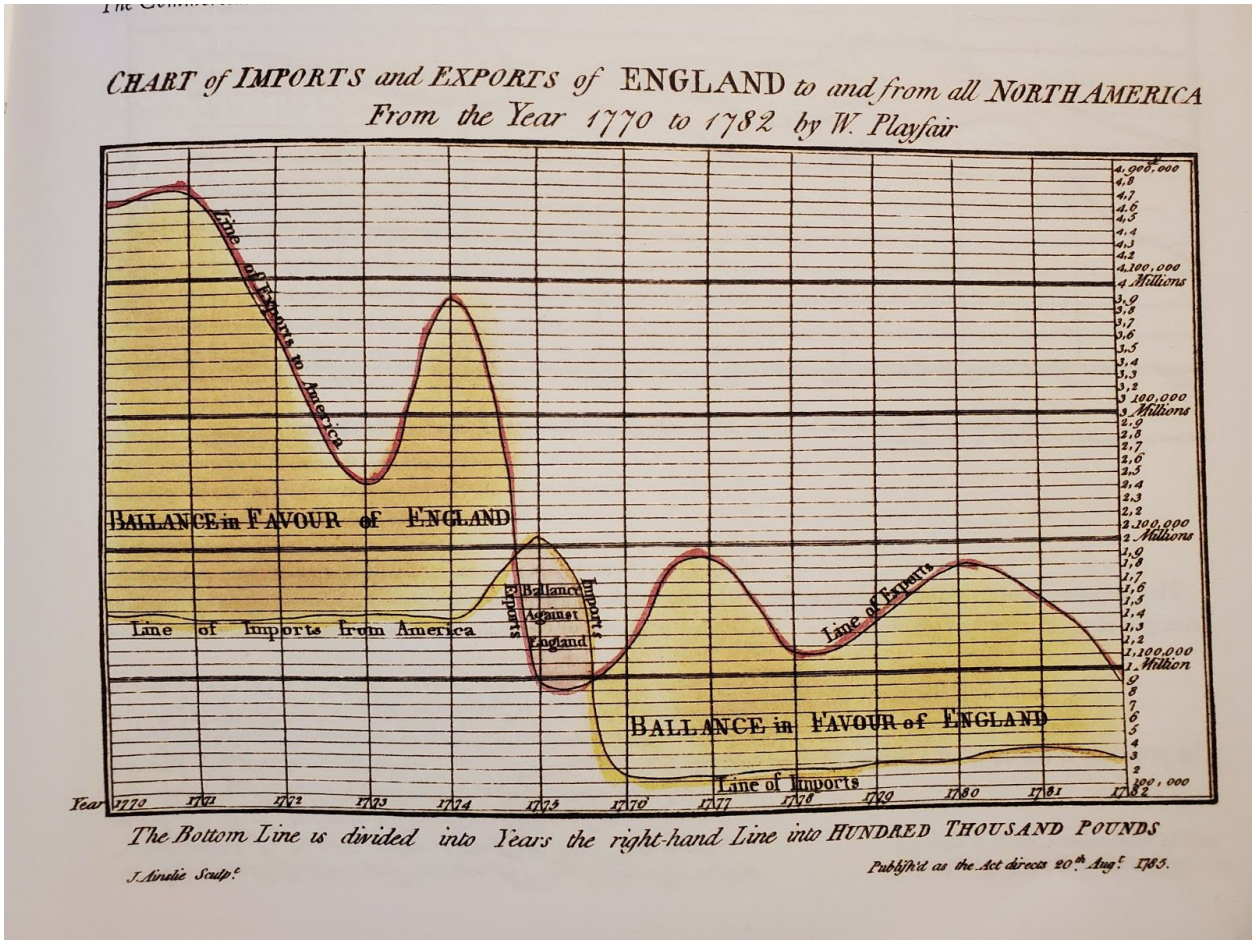
What did I find new or interesting?

Tufte stated that redundancy *may* have use, if it is "giving a context and order to complexity, facilitating comparisons over various parts of the data, [or] perhaps creating an aesthetic balance."

What did I find old, passe or dull?

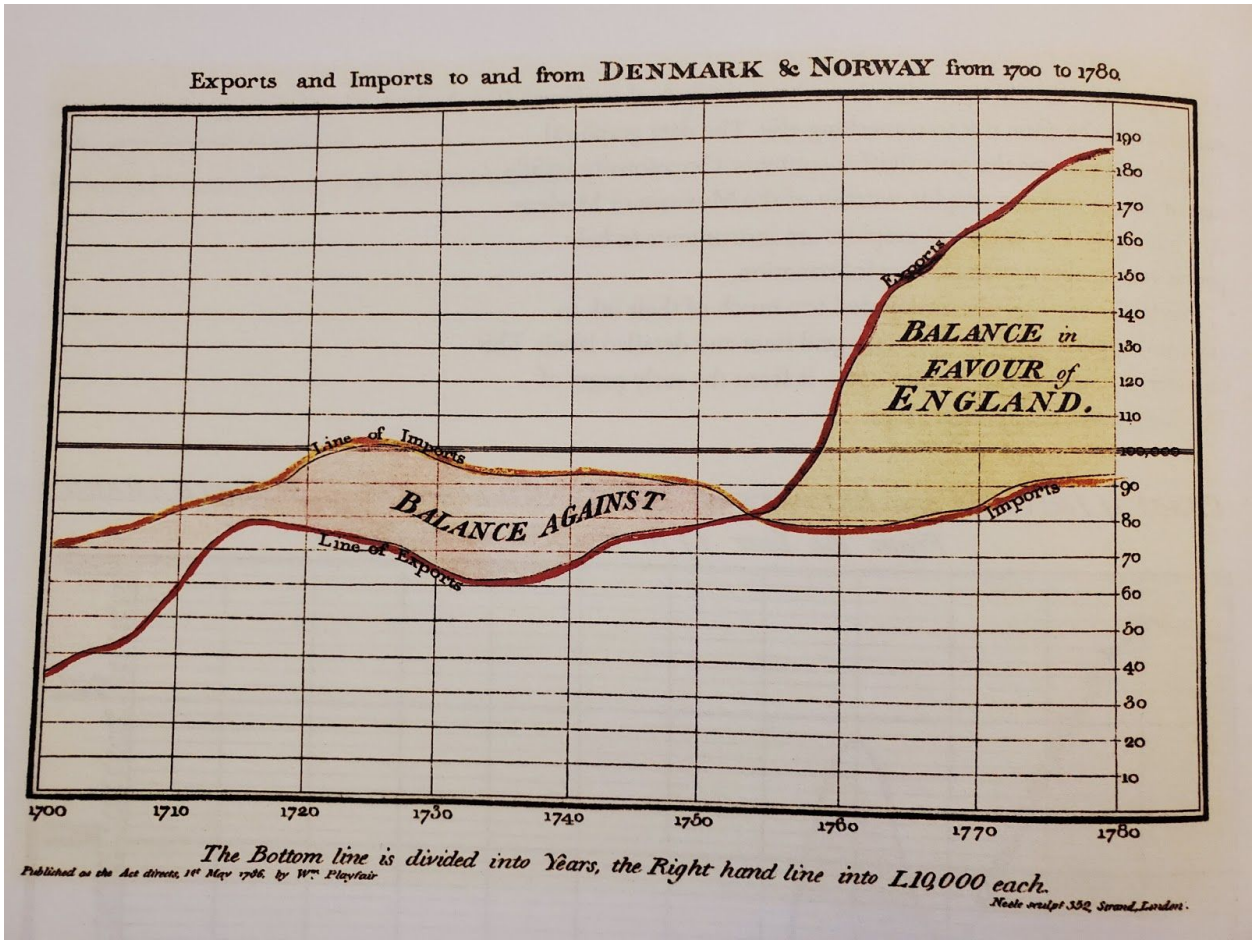
Tufte mentions printing graphic images. In this modern age, I would argue that a lot of data is presented electronically, never to be actually printed out. Regardless of the fact that a lot of visuals will never be printed on a physical piece of paper, data-ink principles should still apply.

Non-Example:



- Very busy grid
- Long title
- Label for each point on both axes

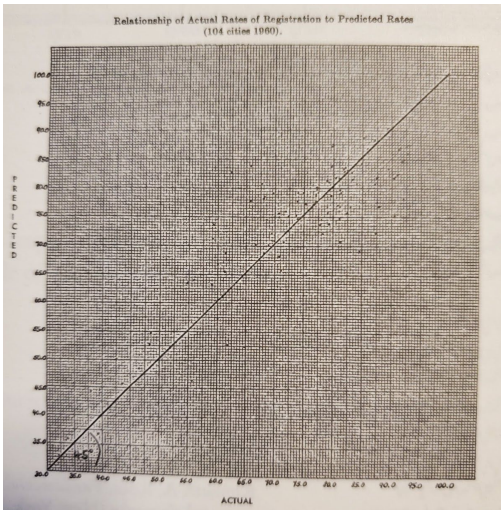
Anchor (Example):



- Less background ink
- Shorter title
- Axes more appropriately segmented

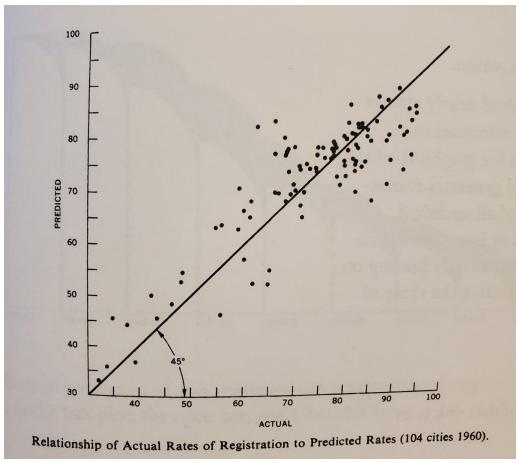
1. **Above all else, show the data**
    a. Non-data should not distract from the data at hand
2. **Maximize the data-ink ratio**
    a. Data-ink is the “non erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented”
    b. Data-ink ratio = data-ink / total ink used to print the graphics
    c. Highest possible ratio is 1.0

i. Graph with a low data-ink ratio:



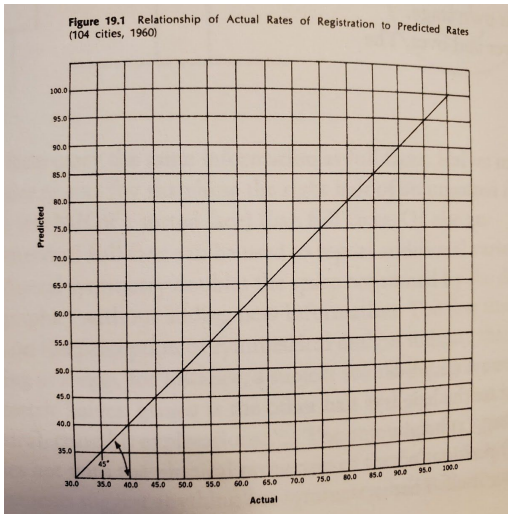
Look at all this non-data ink!

ii. Graph with same data, cleaned up so that it has a much higher data-ink ratio:



Much better, focus is now on the data. Some non-data ink, but it is to clarify data

iii. Overcorrection - Same graph with crucial data points removed:



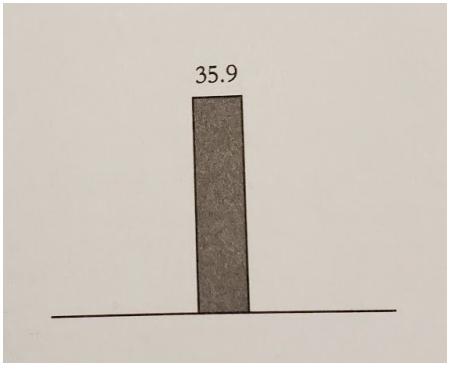
Data is now gone and all that is left is non-data ink. This data-ink ratio is 0.

3. Erase non-data ink

a. Non-data ink is ink that does not depict statistical information
b. Some non-data ink is necessary, such as chart title, axis labels, etc. However, is should be reduced within reason

4. Erase redundant data

a. Redundant data-ink is ink that depicts the same data more than once
    i. Bar Graph Example:



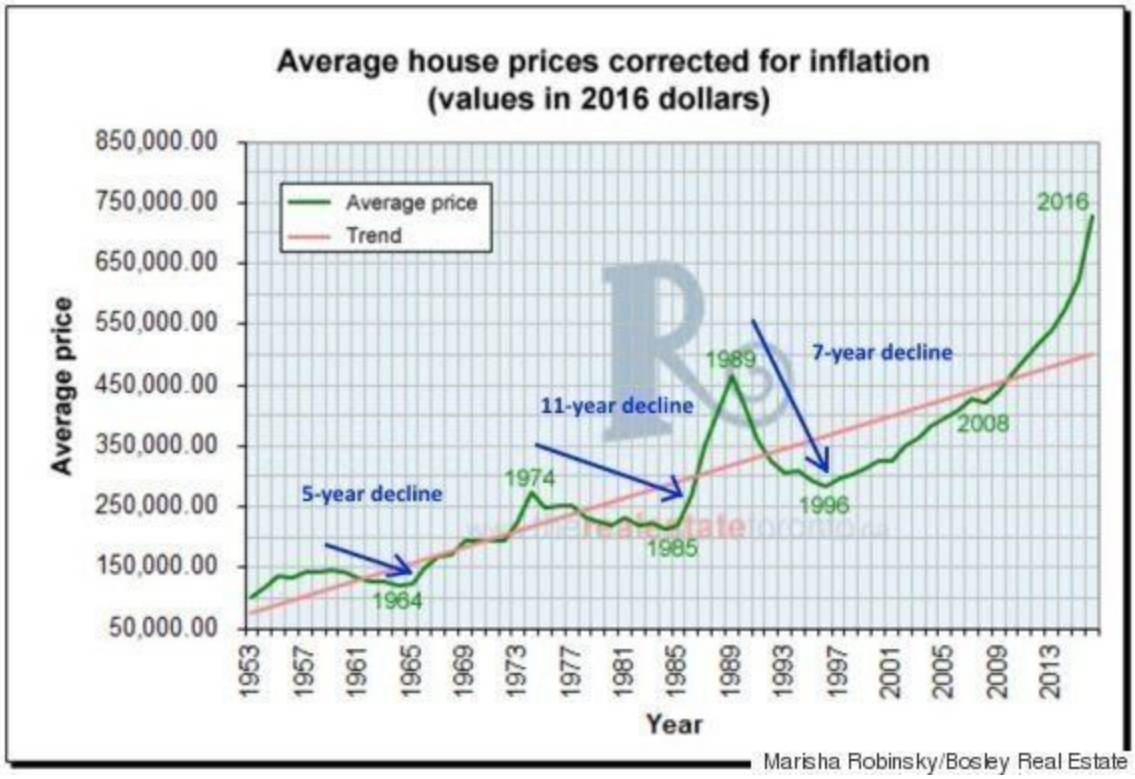
    ii. Tufte argues that the same data is showed *six* times in this graphic
b. Some redundant data may be necessary, if it has a purpose. Outside of that, redundant data should be removed
    i. Redundant data can give context, provide clarity to complex data, or create an “aesthetic balance”
    ii. For example, if the data is cyclical, a cycle may be repeated so that the viewer has a better understanding of the full sequence

5. **Revise and edit**
    a. Just like any other work, graphs should be reviewed and revised until the best product results

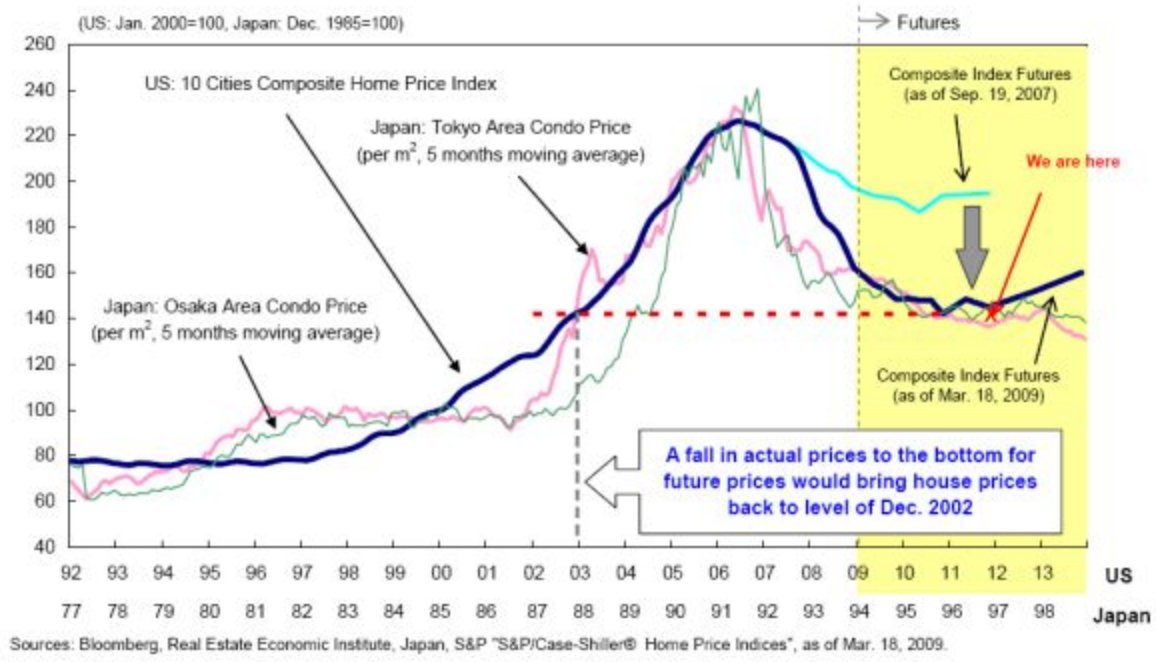
Artifacts

Calculate the data ratio for the below graphs that depict housing price data and then rank them from best to worst

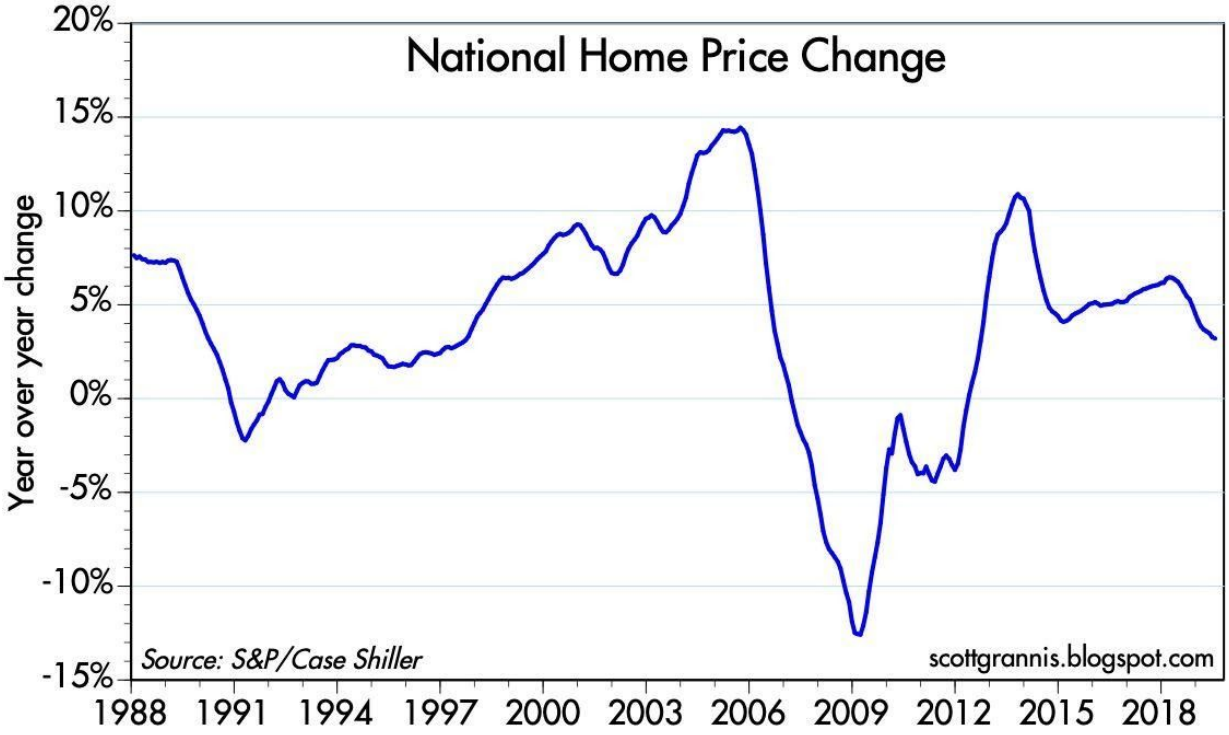
Artifact 1: Average House Prices Corrected for Inflation



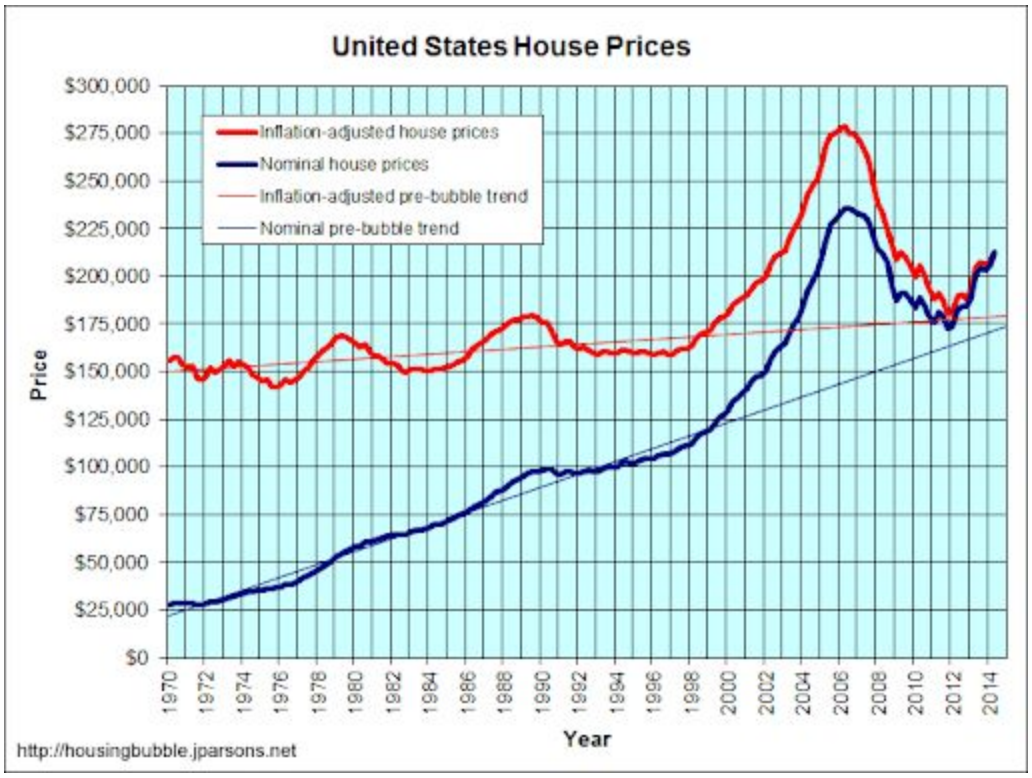
Artifact 2: US and Japan Housing Markets Comparison



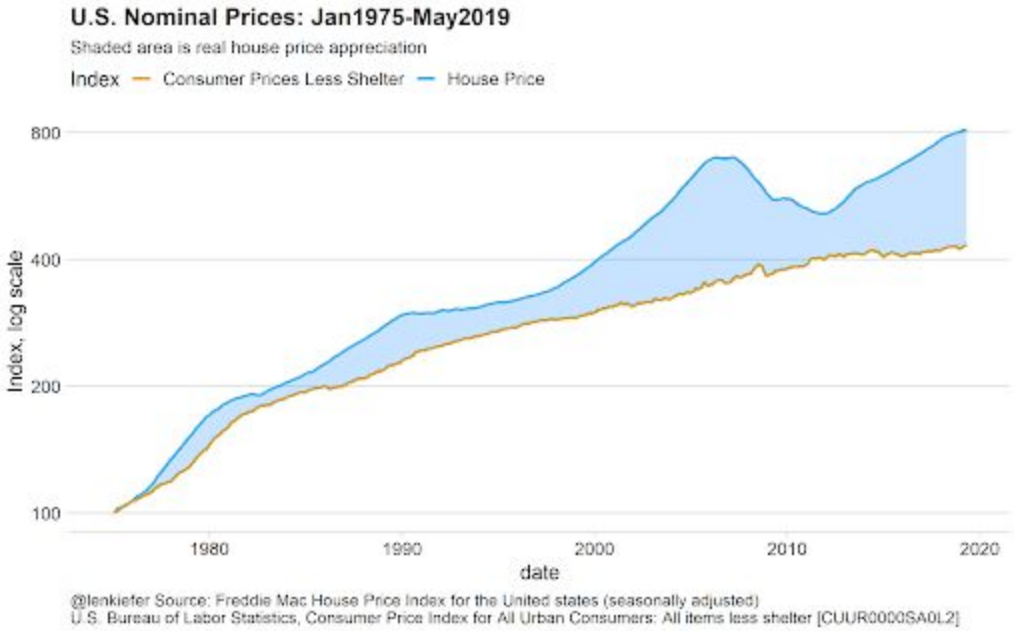
Artifact 3: National Home Price Change



Artifact 4: US House Prices



Artifact 5: US Nominal Prices



Artifact	Data-ink Ratio	Rank	Comments?