We will employ two methodologies to create a linear baseline model. This will serve as a baseline to compare to when we later add nonlinear transformations and interaction terms.

\* Using intuition to select the features that are most relevant to predict \*\*tw\*\*

\* Directly using feature selection algorithms to select features

## Selecting Features Using Intuition

By using intuition and logic, we can select the features that are likely to be predictive of the variable of interest, \*\*tw\*\*.

Let's revisit how \*\*tw\*\* is defined.

\* tw: Total wealth (in US $), which is defined as “net financial assets, including Individual Retirement Account (IRA) and 401(k) assets, plus housing equity plus the value of business, property, and motor vehicles.”

Since \*\*tw\*\* literally includes IRA, 401(k) assets, and housing equity, it is obvious that \*\*ira\*\*, \*\*e401\*\*, \*\*hequity\*\* will be highly predictive of \*\*tw\*\*.

Since \*\*nifa\*\* is defined as non-401k financial assets, it is appropriate to believe that it is indicative of the amount of net financial assets.

Although there is the argument that wealth can be generational, it can be assumed that for most people a higher \*\*inc\*\* (income) is associated with a higher \*\*tw\*\*.

\*\*male\*\* tends to earn more, and that is no secret. Its interaction with income can be later explored.

Having two earners in a household might be indicative of a lower household wealth, and \*\*twoearn\*\*'s interaction with \*\*marr\*\* (whether married or not) can be explored.

Education level dummies should also be useful for