

Humane Housing Index Technical Report

Binh

This is the technical report for the Humane Housing index. The dataset used in this demo is available on the data exchange's website. For brevity and reproducibility, we will only use the zipcodes dataset from 2023. Though the index calculations are automated, we will manually go through one case to ensure that our process is understandable and reproducible.

Acquiring dataset:

The dataset is available on the Fox Valley Data Exchange [website](#), by selecting all humane housing topics and adding 2 more topics: median monthly income for households INC and food insecurity index FAI (still available on the website). The 6 variables (out of the possible 20 in the dataset) that would create our index are:

- HUC: home ownership percentage (i.e. what is the proportion of people owning the housing they are currently occupying)
- FAI: food insecurity (i.e. what is the proportion of people who don't have consistent access to food)
- REX: diversity index
- INC: median household income (monthly)
- MHC: median housing cost (monthly)
- TRV: average commute time

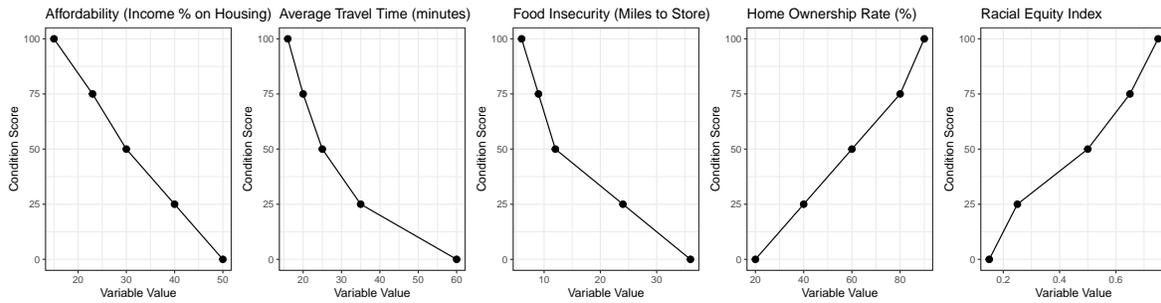
Median housing costs are then divided by median household income to create a measure of housing affordability.

Scoring Scheme:

Each individual measure (diversity, ownership, accessibility/travel time, affordability, food insecurity) is then weighed against national benchmarks. We then linearly interpolate the data for any given area according to the thresholds given below:

Table 1: Scoring Thresholds for Each Variable

variable_name	terrible	bad	average	good	ideal
Affordability (Income % on Housing)	50.00	40.00	30.0	23.00	15.00
Home Ownership Rate (%)	20.00	40.00	60.0	80.00	90.00
Average Travel Time (minutes)	60.00	35.00	25.0	20.00	16.00
Racial Equity Index	0.15	0.25	0.5	0.65	0.75
Food Insecurity (Miles to Store)	36.00	24.00	12.0	9.00	6.00



Weightset:

Out of the 5 categories, some of them will inherently mean more than others. For example, affordability would be more top of mind compared to diversity for most people, so we would want it to contribute to a proportionately bigger part of the index. Keep in mind that future versions of the index might have different weightsets to portray changing priorities and standards.

The current formula for our humane housing index is:

$$\text{housingIndex} = \frac{(2 \times \text{icScore}) + \text{hoScore} + (0.5 \times \text{trScore}) + (0.5 \times \text{diScore}) + (0.5 \times \text{fiScore})}{4.5}$$

Worked Example: 54911

R's `approx()` function is very helpful in our interpolation process, but we are partially hand-waving some of the nuances away. This manual calculation of 54911's humane housing index is meant to explain the mathematical process behind the humane housing index:

We first examine the raw data of 54911:

GEOID	icPct	HUO	FAI	REX	TRV
54911	19.18288	59.54949	13.2	0.3214904	17.5

Our affordability index (`icPct`) for this zipcode is 19.183, which falls between the Ideal and Good quantiles, corresponding to the points (15, 100) and (23, 75). We can then use some algebra to calculate the affordability score for 54911:

The slope of the line connecting these two boundary points are:

$$m_{AC} = \frac{75 - 100}{23 - 15} = \frac{-25}{8}$$

We can then easily use point-slope form

$$y - y_1 = m(x - x_1)$$

to get the equation of the line

$$y - 75 = \frac{-25}{8}(x - 23)$$

Plugging in our `icPct` value of 19.183 yields an affordability score of

$$y - 75 = \frac{-25}{8}(19.183 - 23)$$

$$y - 75 = \frac{-25}{8}(-3.817)$$

$$y - 75 = (-3.125)(-3.817)$$

$$y - 75 \approx 11.928$$

$$y \approx 11.928 + 75$$

$$y \approx 86.928$$

We would then repeat this process for the 4 other variables:

Home ownership (HUO): use (40, 25) and (60, 50)

$$m = \frac{50 - 25}{60 - 40} = \frac{25}{20} = 1.250$$

$$y - 25 = 1.250(x - 40)$$

$$y - 25 = 1.250(59.549 - 40) = 1.250(19.549) = 24.437$$

$$y \approx 25 + 24.437 = 49.437$$

Travel time (TRV): use (16, 100) and (20, 75)

$$m = \frac{75 - 100}{20 - 16} = \frac{-25}{4} = -6.250$$

$$y - 100 = -6.250(x - 16)$$

$$y - 100 = -6.250(17.500 - 16) = -6.250(1.500) = -9.375$$

$$y \approx 100 - 9.375 = 90.625$$

Diversity (REX): use (0.25, 25) and (0.50, 50)

$$m = \frac{50 - 25}{0.50 - 0.25} = \frac{25}{0.25} = 100.000$$

$$y - 25 = 100.000(x - 0.25)$$

$$y - 25 = 100.000(0.321 - 0.250) = 100.000(0.071) = 7.149$$

$$y \approx 25 + 7.149 = 32.149$$

Food insecurity (FAI): use (12, 50) and (24, 25)

$$m = \frac{25 - 50}{24 - 12} = \frac{-25}{12} \approx -2.083$$

$$y - 50 = -\frac{25}{12}(x - 12)$$

$$y - 50 = -2.083(13.200 - 12) = -2.083(1.200) = -2.500$$

$$y \approx 50 - 2.500 = 47.500$$

We can then test this against the results of using R's `approx()` functions:

GEOID	icScore	hoScore	trScore	diScore	fiScore
54911	86.928	49.437	90.625	32.149	47.5

As we can see, the 2 approaches yield the same results. The output of `approx()` is reproducible using simple algebra, and thus can be used across other observations as well.

Lastly, we perform the final calculation to arrive at the humane housing index for 54911:

$$\text{housingIndex} = \frac{2 \times \text{icScore} + \text{hoScore} + 0.5 \times \text{trScore} + 0.5 \times \text{diScore} + 0.5 \times \text{fiScore}}{4.5}$$

$$= \frac{2 \times 86.929 + 49.437 + 0.5 \times 90.625 + 0.5 \times 32.149 + 0.5 \times 47.500}{4.5}$$

$$= \frac{173.858 + 49.437 + 45.313 + 16.075 + 23.750}{4.5}$$

$$= \frac{308.433}{4.5}$$

$$\approx 68.54$$

Applying to other zipcodes:

Having confirmed the reproducibility of the index, we now use R's functionality to create the desired index across different Fox Valley zipcodes (assuming `hh` is the name of the dataset imported from the `.csv`):

```
# affordability score
hh <- hh %>% mutate(icPct = MHC/(INC/12)*100,
                    icScore = approx(x = c(15,23,30,40,50),
                                     y = c(100,75,50,25,0),
                                     xout = icPct, rule = 2)$y)

# home ownership score
hh <- hh %>% mutate(hoScore = case_when(
  HUC < 20 ~ NA_real_,
  TRUE ~ approx(
    x = c(20, 40, 60, 80, 90),
    y = c(0, 25, 50, 75, 100),
    xout = HUC, rule = 2)$y))

# travel time score
hh <- hh %>% mutate(trScore = approx(
  x = c(16, 20, 25, 35, 60),
  y = c(100, 75, 50, 25, 0),
  xout = TRV, rule = 2)$y)

# diversity score
hh <- hh %>%
  mutate(diScore = approx(
    x = c(0.15, 0.25, 0.50, 0.65, 0.75),
    y = c(0, 25, 50, 75, 100),
    xout = REX, rule = 2)$y)

# food insecurity score
hh <- hh %>%
  mutate(fiScore = approx(
    x = c(6, 9, 12, 24, 36),
    y = c(100, 75, 50, 25, 0),
    xout = FAI, rule = 2)$y)

# calculating housingIndex (using the provided formula):
hh <- hh |> mutate(newIndex = (icScore * 2 + hoScore + (0.50 * trScore)
                             + (0.50 * diScore) + (0.5*fiScore)) / 4.5)
```

Challenges, caveats and future directions:

Though the variables that went into the finalized index have been extensively vetted and tested, some caveats remain due to the nature of real data from multiple data sources. Most notably, food insecurity data is not uniformly available across geographies (some layers have food insecurity data from 2020, while others 2023). Consequently, the closest available food insecurity datapoint for that geographic layer will be used to calculate the index, rather than just keeping the observation blank for that year.

Logistics and data quality aside, there are other ways to improve the effectiveness of the index. Specifically, the linear interpolation paradigm is effective and parsimonious but might fall short in terms of statistical rigor. Future versions of the index (potentially with more expert guidance) should perform more robustness checks to improve the index's defensibility - especially if this approach is implemented elsewhere.