

Reliable Transportation Technical Report

Dereje Pollock

Introduction

For this Technical Report I will be working with the Zipcodes Level dataset. All examples will pertain to this level but the general process and formulas will remain the same when moving up to Census Tracts, Cities and Towns, Counties, and so on.

Here is the list of Variables that we will be working with and their definitions.

- ACT This is the percentage of workers who walk or bike to work
- CAR This is the percentage of workers who drive alone to work
- PUB This is the percentage of workers using public transit
- NVC This is the percentage of households without a vehicle
- TRV This is the mean travel time to work.
- EKW This is the walkability index, where higher values indicate higher walkability.

Scoring Approach

The following information will break down how each variable effects our index for a given ZIP code:

Each variable is converted to a Condition Score between 0 and 100 using five breakpoints:

0 (terrible), 25 (bad), 50 (average), 75 (good), 100 (ideal).

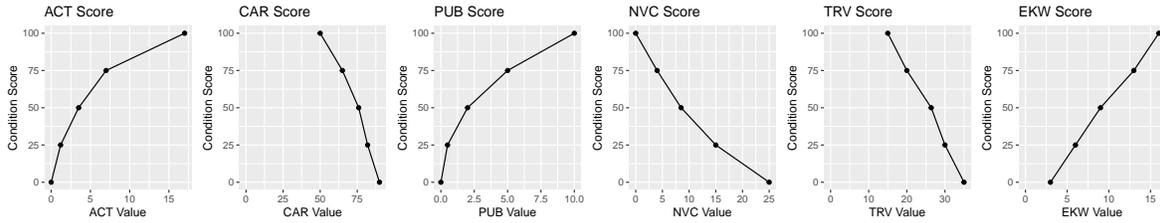
The breakpoints are based on:

- National data/Averages (e.g., American Community Survey benchmarks)
- The Distrubution of our Data
- Our judgment about which values indicate more reliable, safe, and accessible transportation

Table 1: Breakpoints for Reliable Transportation Index Variables

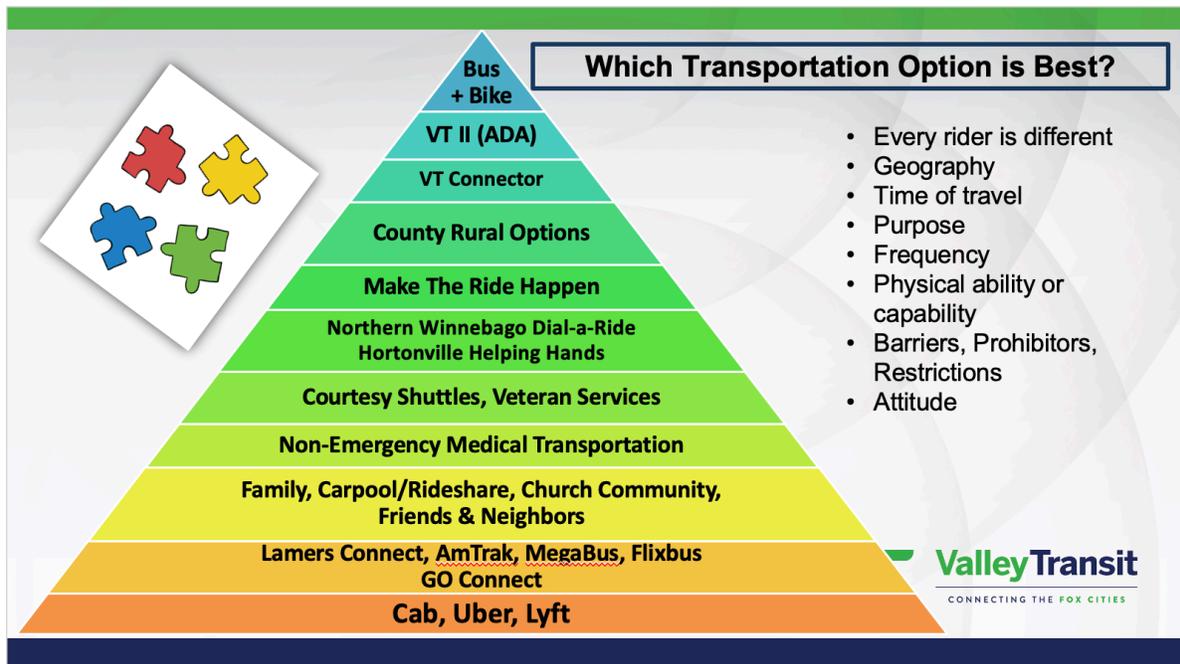
Variable	Terrible (0)	Bad (25)	Average (50)	Good (75)	Ideal (100)	Direction
ACT (Walk/Bike) %	0%	1.2%	3.5%	7%	17%	Higher is better
CAR (Drive Alone) %	90%	82%	76%	65%	50%	Lower is better
PUB (Public Transit) %	0%	0.5%	2%	5%	10%	Higher is better
NVC (No Vehicle) %	25%	15%	8.5%	4%	0%	Lower is better
TRV (Commute Time, min)	35	30	26.4	20	15	Lower is better
EKW (Walkability)	3	6	9	13	16	Higher is better

Scoring Curves:



These visualizations give you an idea of how the index is calculated. The following calculation process involves **interpolation** between breakpoints, clamping out of range values, and mapping all six variables to national benchmarks before averaging them for the overall index.

Looking ahead, I think there’s a lot of potential to align the index more closely with the broader range of trip purposes and transportation modes in the pyramid Sarah Schneider from Fox Valley Transit Shared, while still grounding it in data that we can track consistently over time.



Overall Index Calculation

We decided to combine the Public Transit (PUB) and No Vehicle Access (NVC) scores into a single blended measure because they capture two sides of the same issue: mobility for households without a car.

The blended score ensures that zipcodes are not penalized if residents have at least one reliable option. If public transit usage (PUB) is high, a high percentage of households without vehicles (NVC) does not necessarily indicate poor transportation reliability, since people can still get around without cars. However, if transit usage is low and NVC is high, then households without vehicles are likely to face real barriers, and the blended score will reflect this with a low value. By taking the maximum of PUB and NVC, the index captures this balance: it gives credit when either transit access or vehicle access provides mobility, but highlights vulnerability when neither is strong.

We count the blended score twice since it comes from two distinct metrics and is central to transportation reliability. This follows Jason Schulist and Professor Sage's suggestion to combine PUB and NVC to avoid penalizing areas where transit and vehicle access balance each other by using the maximum value of the two.

We score each variable using its breakpoints. Example for ACT:

A ZIP Code with 5% walking/biking would score between 50 (average) and 75 (good) for the ACT variable.

We would then take an average using the other variables in our index with the blended NVC, and PUB score weighed twice as heavily.

The final index would have a scale from 0–100, where higher scores indicate more favorable transportation conditions.

Formula

Let

$$\text{pubNvcBlendScore} = \max(\text{pubScore}, \text{nvcScore})$$

$$\text{Transportation Index} = \frac{\text{actScore} + \text{carScore} + \text{trvScore} + \text{ekwScore} + 2 \times \text{pubNvcBlendScore}}{6}$$

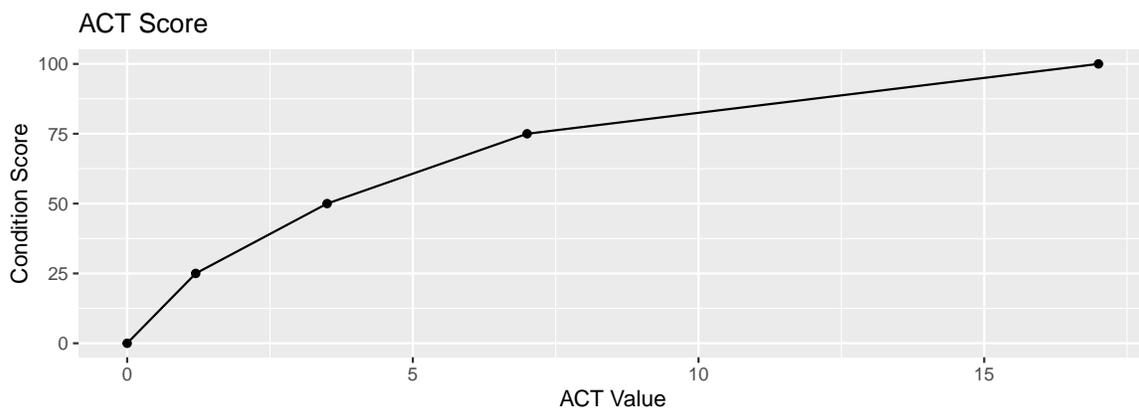
Worked Example: 54911

We will compute the Reliable Transportation Index by hand for the zipcode 54911 in the 2022-2023 years.

We can first look at the values for each variable in the 54911 row of the dataset:

GEOID	ACT	CAR	PUB	NVC	EKW
54911	7.539379	67.71678	0.6423	8.676148	9.299907

The percentage of workers who walk or bike to work (ACT) in this zipcode is roughly 7.54. If we look at the ACT Score curve, we can see that this falls right above the Good benchmark, corresponding to the points (7, 75) and (17, 100) from our earlier breakpoints. We then use algebra to calculate the ACT score for 54911:



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

$$m_{AC} = \frac{100 - 75}{17 - 7} = \frac{25}{10}$$

We can then use point-slope form

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

to get the equation of the line

$$y - 75 = \frac{25}{10}(x - 7)$$

Plugging in the ACT value of 7.54 yields an ACT score of

$$y - 75 = \frac{25}{10}(7.54 - 7)$$

$$y - 75 = \frac{25}{10}(0.54)$$

$$y - 75 = (2.5)(0.54)$$

$$y - 75 \approx 1.35$$

$$y \approx 1.35 + 75$$

$$y \approx 76.35$$

We would then repeat this process for the 4 other variables, for the sake of brevity I will provide the other scores along with the ACT score calculated with R:

- ACT 76.348448
- CAR 68.82551
- PUB 27.3716674
- NVC 49.32251
- TRV 87.50000
- EKW 51.874421

For the PUB and NVC Scores, as explained earlier, we will take the larger of the 2 and weigh it twice as heavily in our Index

We can now perform the final index calculation for zipcode 54911:

$$\text{Transportation Index} = \frac{\text{actScore} + \text{carScore} + \text{trvScore} + \text{ekwScore} + 2 \times \text{pubNvcBlendScore}}{6}$$

$$\text{Transportation Index} = \frac{76.348448 + 68.82551 + 87.50000 + 51.874421 + 2 \times 49.32251}{6}$$

$$\text{Transportation Index} = \frac{383.193399}{6}$$

$$\text{Transportation Index} = 63.8655665$$

Here is the function used in R to calculate the index. The function was created by Mina Mandegar with some edits made to work with the Reliable Transportation index.

```
# baseline_df: data frame with baseline explanatory variables and Cond_Value
# target_df: data frame where you want interpolated values
# vars_cond: character vector of variable names
# scores_only: Only includes interpolated scores and not the original values
# remove_suffix: removes the ending _CondScore
# add_index: includes the index in final table
# ind_name: name index column

interpolate_scores <- function(
  baseline_df, target_df, vars_cond,
  scores_only = TRUE, remove_suffix = TRUE,
  add_index = TRUE, ind_name = "Index"
) {

  for (var in vars_cond) {

    # Baseline values
    baseline_x <- baseline_df[[var]]
    baseline_y <- baseline_df$Cond_Value
    valid_rows <- !is.na(baseline_x) & !is.na(baseline_y)

    # Interpolation function
    interp_fun <- approxfun(
      x = baseline_x[valid_rows],
      y = baseline_y[valid_rows],
      rule = 2
    )

    # New column name
    new_colname <- paste0(var, "_CondScore")

    # Apply interpolation
    target_df[[new_colname]] <- interp_fun(target_df[[var]])
  }
}
```

```

# Keep only scores if requested, but preserve name & GEOID
if (scores_only) {
  keep_cols <- c( "Layer", "Name", "GEOID",
                 grep("_CondScore$", names(target_df), value = TRUE))
  target_df <- target_df %>%
    select(all_of(keep_cols))

  # Remove suffix if requested
  if (remove_suffix) {
    score_cols <- setdiff(names(target_df), c("Layer", "Name", "GEOID"))
    target_df <- target_df %>%
      rename_with(~ sub("_CondScore$", "", .x), all_of(score_cols))
  }
}

# Add average score if requested
if (add_index) {
  score_cols <- setdiff(names(target_df), c("Layer", "Name", "GEOID"))
  target_df <- target_df %>%
    rowwise() %>%
    mutate(pubNvcBlendScore = max(PUB, NVC, na.rm = TRUE), !!ind_name :=
           (ACT + CAR + TRV + EKW + 2 * pubNvcBlendScore) / 6) %>%
    ungroup()
}
}

```

```

# Get index using function we wrote above
ZIPCodes_index <- interpolate_scores(TransportBaselines, zc, vars_cond)
ZIPCodes_index %>% filter(Name == "54911")

```

Layer	Name	GEOID	ACT	CAR	PUB	NVC	TRV	EKW	pubNvcBlendScore	Index
Zip code	54911	54911	76.348	68.826	27.372	49.323	87.5	51.874	49.323	63.866

We can see that our index for the 54911 zipcode matches the one we calculated in our worked problem.