Correlational analysis

I'm looking at the FCC data for NY. It’s seems very structured. For instance, max\_dn is often 940 with max\_up as 35. Is that all one company? And where it is 942.6087 is that because one or a few blocks of that census tract was under a different provider? Similarly, dn10 is between 2 and 3, and dn100 and dn250 tend to hover around 1 or 2, but not exactly. If you make a histogram of dn100 (e.g., 50 bins), there are prominent peaks at 1 and 2. That distribution might violate assumptions of many statistical analyses? Why does the data look like that? Is there an underlying map of provider-chosen boundaries that is sometimes chopped up by the census tract map (necessarily for census demographic comparisons), but mostly aligned?

Medians – to be more robust to outliers? Likely right skewed because positive and small? Are medians distributed normally? If so, maybe averaging is quite reasonable, analogous to central limit theorem. Is it lognormal?

Quantile-quantile or spearman rank-order correlation?

Some quad keys cut across census tracts. Some census tracts only have one quad key do you extrapolate?

People have to know about speedtest… probably only affluent neighborhoods are using it? See number of data points per area versus socioeconomic factors

0.990909 - one out of 101 blocks was zero?

Ability to detect deserts – so the bottom 20% most critical? But almost all areas have at least 1 available – doesn’t mean it’s affordable? Maybe census tracts are too big there should be at least one? And overestimate bias.

Fiber access… is it that critical if you don’t have any access?

Dimensions of the quadtiles can be inferred from spacing between centers

Git gist

Can you average the medians?

Did you register as a volunteer?

Testable hypothesis: We also find that the FCC's data disproportionately overstates coverage in rural and minority communities.

FCC is average – you can see from fractional numbers

These policies are human made – how do the companies decide?

Do it by percentile? Highest 10th percentile and lowest? Quantile-quantile plot

What do the distributions look like?