**The Datasets:**

1. PHA-“program type”-level[[1]](#footnote-1) data about the state of subsidized housing in the United States in 2014, courtesy of the Department of Housing and Urban Development (HUD). Though HUD engages in several different housing assistance program types, we limited our analysis to Housing Choice Voucher (HCV) programs. [*https://www.huduser.gov/portal/datasets/assthsg.html*](https://www.huduser.gov/portal/datasets/assthsg.html)
2. PHA-person-level data about the compensation (salary, bonus and total compensation) of the top three paid employees at every public housing authority in the United States. We limited this data to the top paid individual at each PHA*.* [*https://www.hud.gov/program\_offices/public\_indian\_housing/programs/pha\_executive\_compensation*](https://www.hud.gov/program_offices/public_indian_housing/programs/pha_executive_compensation)

**The Analysis:**

By limiting the data in the two ways described above, we constructed a PHA-level dataset with programmatic, geographic and compensation information about every HCV program in the United States, as well as rich demographic data about each program’s clients.

We first performed descriptive analyses on:

* Where voucher-recipients live, by region, and by head of household gender. Most households are headed by women, but we were surprised to see that nearly half of those women do not have children.
* The distribution of total compensation of the highest paid employee for each PHA. The most common compensation appears to be around $80k.
* The distribution of “rent burden” or the ratio of average monthly household income to average monthly pre-subsidy rent. Pre-subsidy refers to the sum of the household contribution and the HCV-contribution. We observe next to no ratios of under 0.5, and most hover around 1:1. A substantial number are above that.
* The distribution of months on a wait list. We fit this PDF with an exponential function. This curve arises when examining the length of time between events in Poisson processes. It makes sense then that an exponential curve fit this analysis very well.

We then moved on to inferential statistics:

* We were curious whether high-poverty geographic regions were correlated with long wait times. We defined high-poverty areas as places where greater than or equal to 20% of the population lives under the poverty line. To explore this, we performed a chi-square test on the distribution of HCV programs located in high-poverty areas among wait-time buckets of 6 months each. Programs in high-poverty areas were in fact under-observed in the 1-6 month bucket and over-observed in every longer bucket. We saw the opposite phenomenon among programs located in lower-poverty areas.
* We performed a permutation test on compensation of highest paid employee by whether the program was located in a high poverty area or not. We found a statistically significant relationship for this analysis, but in the opposite direction than we were expecting. Total compensation to executives in high-poverty areas is $6,064 greater than to their counterparts in lower-poverty areas.
* A second permutation test found that executives in the south are much more likely to receive bonuses than in any other region.
* Lastly we performed a permutation test on months waiting by whether the program was located in a high poverty area or not. We concluded that the difference we observe is one of pure chance.
* We performed a linear regression on executive compensation and percent of clients who are racial minorities and were surprised to find a strong, positive relationship.
* We also found a positive relationship between executive compensation and client household income. This is to be expected—those areas likely experience higher cost of living.
* Lastly we examined the distribution of rent burden across the United States, plotting points on a map, as well as state averages on a different map.

1. PHA stands for public housing authority. [↑](#footnote-ref-1)