Colorado Housing Shortage

2025-04-28

## Colorado and Housing Shortages

* Colorado has been growing and growing quickly
  + 6th fastest growing state between 2010 and 2020
* Housing production has lagged behind population growth
* Strain on the housing market and household formation
* households are not forming
* cost of housing is increasing
* How can we quantify the housing shortage?
* [Brookings](https://www.brookings.edu/articles/make-it-count-measuring-our-housing-supply-shortage/) review of housing shortage methodologies

## Putting a Number to Colorado’s Housing Shortage

The State Demography Office (SDO) at the Colorado Department of Local Affairs (DOLA) has been tasked with quantifying the housing shortage in Colorado. While a number of methodologies for quantifying housing shortages exist, the SDO’s preferred method is an approach which captures

* The relationship between population composition and household formations
* The relationship between housing cost burden and household formations

Our approach is modeled after the methodology employed by Up for Growth to estimate Housing Underproduction. The primary formula for housing underprodcution is as follows.



Hosuing Needs Breakdown

Here housing underproduction is a function of 6 variables

* Households
* Missing Households
* Target Vacancy Rate
* Total Housing Units
* 2nd and Vacation Units
* Uninhabitable Units

To get values for these variables we will utilize the 2023 ACS microdata. We will access this data through the IPUMS service for ease of use. First, lets start with the variables on the right most side of the equation as they are more easily calculated. First, we grab our sample of housing data from the ACS micro data. From this sample, we calculate the Total Housing Units as the sum of each observation in Colorado’s Household Weight.

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.4   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors  
## Loading required package: grid  
##   
## Loading required package: Matrix  
##   
##   
## Attaching package: 'Matrix'  
##   
##   
## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack  
##   
##   
## Loading required package: survival  
##   
##   
## Attaching package: 'survey'  
##   
##   
## The following object is masked from 'package:graphics':  
##   
## dotchart  
##   
##   
## spatstat.univar 3.1-3  
##   
## Use of data from IPUMS USA is subject to conditions including that users should cite the data appropriately. Use command `ipums\_conditions()` for more details.

From the survey data we find that in 2023 there were a total of 2,638,093. Next we will want to calculate the number of second and seasonal homes. ACS survey data has a response which directly corresponds to this which is homes that are “For seasonal, recreational or other occasional use”. This is also what the State of Oregon uses for their own [housing needs analysis](https://www.oregon.gov/das/oea/Documents/OHNA-Methodology-Report-2024.pdf) which closely follows the Up for Growth Methodology. For uninhabitable housing we again follow the state of Oregon’s approach using houses which either have no kitchen or lack complete plumbing and which have been vacant for over 12 months.

Using these definitions we find that Colorado has 94,822 seasonal housing units and 7,479 uninhabitable housing units. The table below shows how the distribution of these housing types differs between Colorado and the US as a whole. We can see that relative to the US, Colorado has more 2nd and vacation house, but less uninhabitable housing.

## # A tibble: 3 × 3  
## HOUSINGTYPE US COLORADO  
## <chr> <chr> <chr>   
## 1 Remaining 95.96% 96.21%   
## 2 Seasonal 3.08% 3.59%   
## 3 Uninhabitable 0.96% 0.20%

Looking at the remaining housing stock that is neither a vacation home or uninhabitable, we can look at the vacancy rate in Colorado and the United States. Here we see that as of 2023, Colorado has a lower vacancy rate than the US as whole.

## # A tibble: 2 × 3  
## HOUSINGTYPE US COLORADO  
## <chr> <chr> <chr>   
## 1 Occupied 94.17% 95.67%   
## 2 Vacant 5.83% 4.33%

With all of the portions of the right most part of the equation accounted for we can move to the left part of the equation, the target number of housing units. We first look at vacancy rate. Vacancy rate here refers to the desired rate of vacancy among housing units that are neither used as second/vacation homes or are uninhabitable. Any healthy housing market should have a measurable vacancy rate as housing units transition in and out of the housing market. Keeping with the original up for growth study, we opt for a vacancy rate of 5%.

Next we calculate households, which again, is straight forward using the ACS data. We simply select the households which are not vacant in the housing survey and sum their weights to get the appropriate count for households in Colorado.

We arrive at an estimate of 2,428,263 households in Colorado.

The last part of the equation is missing households. Missing households here refers to the households that have not formed in a market compared to an a counterfactual market. The most common way of calculating this value is to have a target headship rate by age for your study population and to calculate how different the headship rate is from the observed population. The following formula shows how this is calculated.

$$
\sum\_{\substack{i = 15} \\ \text{step }5}^{65} \text{HR}^\*\_{[i, i+5)} \* \text{POP}\_{[i, i+5)} - \text{HH}\_{[15, 20)}
$$

In this formula $\text{HR}^\*\_{[15, 20)$ would be the target headship rate of individuals aged 15 to 19, $\text{POP}^\*\_{[15, 20)$ is the population aged 15 to 19 in the study, and $\text{HH}^\*\_{[15, 20)$ is the number of households headed by individuals aged 15 to 19 in the study. The population and number of households by age are, once again, easy to obtain from the survey data. The target headship rate, however, is conceptual, and a decision must be made about what that target should be. A common target used in studies has been the headship rate observed in the year 2000 which reflects [“a housing market that was more in balance”](https://www.oregon.gov/das/oea/Documents/OHNA-Methodology-Report-2024.pdf). To see how much headship rates have changed over time we switch to using the person frame ACS as opposed to the housing frame data. The person frame data allows us to look at people and get their characteristics. Most notably here we are interested in looking at survey respondents age and whether they are the head of household. We will do this for both the 2023 ACS as well as the 2000 Census to see how much headsip rates have changed.

## Use of data from IPUMS USA is subject to conditions including that users should cite the data appropriately. Use command `ipums\_conditions()` for more details.

## # A tibble: 11 × 7  
## AGEGROUP `1990` `2000` `2006` `2018` `2020` `2023`  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 [15,20) 0.0416 0.0412 0.0296 0.0216 0.00959 0.0326  
## 2 [20,25) 0.331 0.328 0.301 0.240 0.238 0.315   
## 3 [25,30) 0.490 0.469 0.470 0.413 0.400 0.466   
## 4 [30,35) 0.546 0.520 0.518 0.472 0.459 0.518   
## 5 [35,40) 0.564 0.544 0.546 0.516 0.503 0.545   
## 6 [40,45) 0.590 0.567 0.576 0.543 0.503 0.550   
## 7 [45,50) 0.593 0.578 0.579 0.542 0.540 0.561   
## 8 [50,55) 0.597 0.597 0.585 0.557 0.544 0.568   
## 9 [55,60) 0.598 0.601 0.625 0.584 0.586 0.581   
## 10 [60,65) 0.613 0.611 0.590 0.576 0.569 0.611   
## 11 [65,Inf] 0.689 0.680 0.644 0.632 0.624 0.638

Looking at the table above we see that almost every age group has a lower headship rate in 2023 compared to 2020. We can plug in these values for to finish the last piece of the puzzle in calculating housing underproduction.

From these calculations we find that Colorado has 77,555 missing households which leads to a housing underproduction count of 101,911 units. Note that this value is almost identical to the value published by the [Up for Growth 2023 analysis](https://upforgrowth.org/apply-the-vision/2023-housing-underproduction/), 101,000 units. This should be unsurprising given that we have followed their methodology almost exactly.

This is just the total number of missing households and housing underproduction but we can further breakdown missing households by age to see how many missing groups are present by age. The table below shows the missing households by age. We can see that a shockingly large number of the missing households, almsot 40,000, are found in the age group of 65 and older. The fact that this age group is driving a large portion of the missing households should raise some concerns about this methodology.

## # A tibble: 11 × 4  
## AGEGROUP NEWHH HH MISSINGHH  
## <fct> <dbl> <dbl> <dbl>  
## 1 [15,20) 14298. 11295 3003.   
## 2 [20,25) 116510. 112078 4432.   
## 3 [25,30) 206824. 205544 1280.   
## 4 [30,35) 244956. 243879 1077.   
## 5 [35,40) 242189. 242646 -457.   
## 6 [40,45) 235157. 228001 7156.   
## 7 [45,50) 204715. 198930 5785.   
## 8 [50,55) 211189. 201059 10130.   
## 9 [55,60) 187330. 181156 6174.   
## 10 [60,65) 215519. 215485 34.5  
## 11 [65,Inf] 627129. 588188 38941.

In addition to this abnormality, the use of a past headship rate is somewhat unsatisfying. We know that headship rates are impacted by a number of factors such as housing cost, income, size of households, education levels, number of dependents, etc. As these factors change we would expect the headship rates to also shift. In addition a factor that we are most interested in for housing shortage is the relationship between cost burden of housing and the shortage of formed households. What we would like to do is to say how many household would form if the housing burden was lower. Lets take a look at the heads of household both in 2023 and in 2000 and see how much of their housing cost compares to their gross income. The table below shows these values and as we can see the share of the population who spends 0-35% of their monthly gross income on housing costs is shrinking while the share that spends more than 65% is increasing. In short people are spending more of their income on housing over time.

## # A tibble: 3 × 3  
## HCOSTQ `2000` `2023`  
## <fct> <dbl> <dbl>  
## 1 0-35% 57.2 47.6  
## 2 35-60% 23.0 23.2  
## 3 >60% 19.9 29.3

One way we could build a headship rate counterfactual is by saying would would the headship rate be today if housing costs were more similar to what they were in 2000. This is an improvement in that we can now capture the relationship between housing costs and headship rate, however, it still ignores the relationship between many other variables and headship rates which have changed over time. A way to account for both changing housing cost burden and other variables exists in the 3-way Kitigawa or Oaxaca-Blinder [decomposition](https://giacomovagni.com/blog/2023/oaxaca/) method. This methodology has been used in several other housing studies, such as those done by [Freddie Mac](https://www.freddiemac.com/research/insight/20181205-major-challenge-to-u.s.-housing-supply) and [Brookings](https://www.brookings.edu/articles/make-it-count-measuring-our-housing-supply-shortage/), and allow an analyst to isolate the impact of a single variable on an outcome by shifting that variables value to something it had resembled previously.

## The housing burden effect

is the headship status and is a set of explanatory variables.

$$
Y\_{\text{baseline}} \sim \text{Binomial}(\theta\_{\text{baseline}}) \\
\text{logit}(\theta\_{\text{baseline}}) = \zeta\_\text{baseline} X\_\text{baseline}^{\text{Housing Burden}} + \boldsymbol{\beta}\_\text{baseline} \boldsymbol{X}\_\text{baseline} \\
Y\_{\text{current}} \sim \text{Binomial}(\theta\_{\text{current}}) \\
\text{logit}(\theta\_{\text{current}}) = \zeta\_\text{current} X\_\text{current}^{\text{Housing Burden}} + \boldsymbol{\beta}\_\text{current} \boldsymbol{X}\_\text{current}
$$

## Housing Cost Counterfactual

Note for the counter factual that this change is equivalent to adding the endowment or explained effect of the decomposition for housing cost to . By applying the counterfactual prediction to each observation in our dataset we can then arrive at a new which is driven by changing housing cost burden. The table below shows how the model based approach for differ from either the observed 2000 or 2023 values.

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!  
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

## # A tibble: 10 × 8  
## AGEGROUP `Model Based` `1990` `2000` `2006` `2018` `2020` `2023`  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 [20,25) 0.345 0.331 0.328 0.301 0.240 0.238 0.315  
## 2 [25,30) 0.489 0.490 0.469 0.470 0.413 0.400 0.466  
## 3 [30,35) 0.539 0.546 0.520 0.518 0.472 0.459 0.518  
## 4 [35,40) 0.557 0.564 0.544 0.546 0.516 0.503 0.545  
## 5 [40,45) 0.558 0.590 0.567 0.576 0.543 0.503 0.550  
## 6 [45,50) 0.580 0.593 0.578 0.579 0.542 0.540 0.561  
## 7 [50,55) 0.586 0.597 0.597 0.585 0.557 0.544 0.568  
## 8 [55,60) 0.605 0.598 0.601 0.625 0.584 0.586 0.581  
## 9 [60,65) 0.632 0.613 0.611 0.590 0.576 0.569 0.611  
## 10 [65,Inf] 0.658 0.689 0.680 0.644 0.632 0.624 0.638

A noticeable difference between this model based approach and simply using the Headship Rates observed in 2020 is where the difference in observed and target headship rates lay. The model based approach puts more of the missingness of households into younger populations rather than older populations.

## # A tibble: 11 × 4  
## AGEGROUP HH NEWHH MISSINGHH  
## <fct> <dbl> <dbl> <dbl>  
## 1 [15,20) 11295 13871. 2576.  
## 2 [20,25) 112078 122646. 10568.  
## 3 [25,30) 205544 215792. 10248.  
## 4 [30,35) 243879 253993. 10114.  
## 5 [35,40) 242646 248010. 5364.  
## 6 [40,45) 228001 231156. 3155.  
## 7 [45,50) 198930 205405. 6475.  
## 8 [50,55) 201059 207247. 6188.  
## 9 [55,60) 181156 188523. 7367.  
## 10 [60,65) 215485 223103. 7618.  
## 11 [65,Inf] 588188 606302. 18114.

From these calculations we find that Colorado has 87,786 missing households which leads to a housing underproduction count of 112,681 units.

To demonstrate in another way what this model is doing and show its flexibility lets pick another value for housing burden to see its impact on the housing shortage estimate. Let us say instead that we think all individuals should be paying 100% of their income on housing costs. We can place this into the model as follows.

When we do this our model tells us that we actually have to much housing! That is if you want the citizens of Colorado to spend more of their income on housing we need to get rid of 18,326 units.

Alternatively, what if we wanted a scenario where individuals only spent 10% of their income on housing, what then would the housing shortage be?

In this case the model says we would need to add 237,018 units to Colorado’s housing total.