Estimating Colorado’s Housing Shortfall

Source: [Article Notebook](https://gtottenSDO.github.io/statewide_total_shortfall_estimates/index.qmd.html)

## Defining The Shortfall

In order to estimate the total housing shortfall in Colorado, we must first define the metrics by which we are going to assess the number of housing units the state might be short.

How these metrics are defined can have a significant impact on the resulting analysis. As a highly stylized example to demonstrate this concept let’s begin by using two examples to bound our estimates:

1. Every person currently residing in the state does so inside of a permanent housing unit. In this scenario we would have a relatively low estimate of the total number of necessary housing units - as what would be required is enough units to house the state’s unhoused population. In this case the estimate of number housing units might just be the estimate of the number of people in this population. However, this could be reduced by changing our requirements of housing to include any shelter - such as tents as vehicles, which would likely bring the estimated much lower.
2. Every current US resident who would like to live in Colorado may do so, and they will be able to do so for free. Conversely, in this scenario we might expect an estimate that is quite high, as while some people still may opt not to reside in our beautiful state (perhaps they do not particularly like sun), we could reasonably expect many Americans, perhaps into the hundreds of millions, might opt to spend no money to live in our beautiful state.

In between these two estimates are a range of scenarios that might be indicative of the number of housing units which are necessary - based on the objectives we are trying to determine, and the underlying assumptions about housing preferences which underlie them. In this paper we will examine a variety of methods, based primarily on studies by other researchers, that we can apply to Colorado to determine the estimated housing shortfall in the state, under that method. In this way we will provide not so much a point estimate of the total housing shortfall, but a range of estimates which can be utilized by planners and policy makers based on their discretion with respect to the reasonableness and applicability of each method. In doing so we also hope to plan a clear, concise, explanation of the method, what objective it is attempting to solve for, and the meaning of the estimate within that context.

## Data

Data primarily comes from the most recent American Community Survey (“ACS”) one year estimates for Colorado, and data from the Colorado State Demography Office (“SDO”). One year ACS estimates are primarily used as the population of the state is large enough to allow for the use of such estimates. If applying similar methodologies at smaller geography levels (such as county level), it may be necessary to instead use 5 year estimates. Additionally, some methods of deriving estimates, such as by analyzing Public Use Microdata Sample (“PUMS”) data may not be possible for all methods. As such many methods determined here may only be applicable at the state level.

## Examples

### Harvard Joint Center for Housing Studies Blog

One resource that compares four relatively recent national studies attempting to determine housing shortfalls are a January 2024 blog entry from the Harvard Joint Center for Housing Studies (“JHCS”) (McCue and Huang 2024). The four studies each utilize different methodologies and resulting estimates covering different years. The four studies are:

1. [National Association of Home Builders (NAHB) 2021](https://www.freddiemac.com/research/insight/20210507-housing-supply)
2. [Freddie Mac 2020](https://www.freddiemac.com/research/insight/20210507-housing-supply)
3. [National Association of Realtors (NAR) 2021](https://www.nar.realtor/advocacy/housing-is-critical-infrastructure)
4. [National Low Income Housing Coalition (NLIHC)](https://nlihc.org/gap)

The following sections will provide estimates of the housing shortage in Colorado based on each of these study methodologies.

### National Association of Home Builders

The NAHB study estimates the national housing shortage by examining the difference in ACS vacancy rates in the current year from their long run average. This resulted in an estimated shortage of 1.5 million units.

The first step is to create a time series of vacancy rates for the state with ACS data. This data is accessed from the IPUMS USA database (Ruggles et al. 2024) using the ipumsr package in R (Greg Freedman Ellis, Derek Burk, and Finn Roberts 2024).

Source: [Article Notebook](https://gtottenSDO.github.io/statewide_total_shortfall_estimates/index.qmd.html)

Vacancy rates are calculated based on both the OWNERSHP and VACANCY variables. Additionally the GQ variable is used to filter out group quarters.

In order to calculate vacancy rates it is necessary to first understand what each of the values for each variable represents

Source: [Article Notebook](https://gtottenSDO.github.io/statewide_total_shortfall_estimates/index.qmd.html)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Values and respective labels for each of the OWNERSHP, VACANCY, and GQ variables.   | OWNERSHP | | VACANCY | | GQ | | | --- | --- | --- | --- | --- | --- | | val | lbl | val | lbl | val | lbl | | 0 | N/A | 0 | N/A | 0 | Vacant unit | | 1 | Owned or being bought (loan) | 1 | For rent or sale | 1 | Households under 1970 definition | | 2 | Rented | 2 | For sale only | 2 | Additional households under 1990 definition | |  |  | 3 | Rented or sold but not (yet) occupied | 3 | Group quarters--Institutions | |  |  | 4 | For seasonal, recreational or other occasional use | 4 | Other group quarters | |  |  | 5 | For occasional use | 5 | Additional households under 2000 definition | |  |  | 6 | For seasonal use | 6 | Fragment | |  |  | 7 | For migrant farm workers |  |  | |  |  | 8 | For seasonal use or migratory |  |  | |  |  | 9 | Other vacant |  |  | |

Source: [Article Notebook](https://gtottenSDO.github.io/statewide_total_shortfall_estimates/index.qmd.html)

[Table 1](#tbl-pums-vars) includes the values and respective labels for each of the OWNERSHP, VACANCY, and GQ variables.

The N/A labels associated with val=0 for the OWNERSHP and VACANCY variables is reflective of units which are Vacant, or Occupied, respectively.

Given these values we then need to establish filtering parameters for determining vacancy rates.

1. GQ <= 2 OR GQ ==5 filter out GQ units to only included Households and Vacant Units (including with the 1990 and 2000 Additional Unit definitions)[[1]](#footnote-28).
2. VACANCY <=2 Remove occasional/seasonal vacant units (4-6) consistent with NAHB analysis. Remove migrant farm worker usage (7), and other vacant (9), as these should be relatively independent of overall vacancy rates. Reclassify 3 - Rented or sold but not (yet) occupied to 0 as occupied.
3. Leave Ownership as is.

In order to broadly analyze survey data, and additionally calculate margin of error, the srvyr package is used. Prior to analysis a survey object must be defined based on the extracted ACS data. This survey object filters GQ units based on condition 1 above, and removes vacant units based on condition 2 above. Additionally, it reclassifies VACANCY == 3 to 0 as occupied. Finally, N/A values are relabeled for each of the OWNERSHP, and VACANCY variables to reflict that these reflect Vacant and Occupied units, respectively.

Source: [Article Notebook](https://gtottenSDO.github.io/statewide_total_shortfall_estimates/index.qmd.html)

Then total up on an annual basis the total occupied units each for sale and for rent, as well as vacant units by for sale or for rent[[2]](#footnote-29)[^margin of error are calculated as 1.645 \* standard error consistent with 90% CI]. We additionally calculate total occupied and total vacant as the sum of each of these. Vacancy rates are then calculated as $\sum VACANT\_{ot}\over {\sum VACANT\_{ot}+ \sum OCCUPIED\_{ot}}$ for each occupancy status and year . Finally we calculate the sum total of occupied and vacant units for each year[[3]](#footnote-30).

Once the vacancy rates are calculated, we then calculate the average vacancy rate for each year. To get an annual shortfall we then take the difference between the average vacancy rate and the current vacancy rate. Applying this difference to the total number of units for each year gives us the annual shortfall.

In combining vacancy rates a tenure variable is created to determine if a unit is determined to be sale or rental for the purpose of calculating vacancy rates. This is set equal to the value rent for OWNSHP == 2 and VACANCY == 1, and own for OWNSHP == 1 and VACANCY == 2; consistent with the classifications determined above.

Source: [Article Notebook](https://gtottenSDO.github.io/statewide_total_shortfall_estimates/index.qmd.html)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Current shortfall of occupied housing units by tenure type using NAHB Methodology.   | Tenure | Vacant | Margin of Error | Occupied | Margin of Error | Total | Vacancy Rate | Average Vacancy Rate | Difference from Average | Shortfall | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Rent | 41,901 | 5,192 | 835,314 | 22,166 | 877,215 | 4.78% | 6.53% | 1.76% | 15,414 | | Own | 12,816 | 2,753 | 1,592,949 | 24,524 | 1,605,765 | 0.80% | 1.64% | 0.84% | 13,549 | | Total | 54,717 | 7,945 | 2,428,263 | 46,690 | 2,482,980 | 2.20% | 3.32% | 1.12% | 27,699 | |

Source: [Article Notebook](https://gtottenSDO.github.io/statewide_total_shortfall_estimates/index.qmd.html)

[Table 2](#tbl-nahb-shortfall) contains the calculated shortfall for 2023 using the NAHB methodology. Using this methodology shortfalls of 15.4k units for rent and 13.5k units for ownership are determined, for a total shortfall of 27.7k.

Greg Freedman Ellis, Derek Burk, and Finn Roberts. 2024. *Ipumsr: An r Interface for Downloading, Reading, and Handling IPUMS Data*. <https://tech.popdata.org/ipumsr/>.

McCue, Daniel;, and Sophie Huang. 2024. “Estimating the National Housing Shortfall | Joint Center for Housing Studies.” <https://www.jchs.harvard.edu/blog/estimating-national-housing-shortfall>.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. 2024. “IPUMS USA: Version 15.0.” Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D010.V15.0>.

1. Note to self - review these definitions to see if we want to exclude any of these groupings. [↑](#footnote-ref-28)
2. For this purpose we make the assumption that VACANCY == 1 (For rent or sale) is for Rent, and VACANCY ==2 (For sale only) is for sale in order to the simplify the analysis. [↑](#footnote-ref-29)
3. Note to self - verify that margin of error are additive [↑](#footnote-ref-30)