

Data on housing

The aim was to get broad estimates of tenure at the county level and at the CBSA level.

Two main sources were used :

- the American Community Survey 5-Year Estimates of the U.S Census Bureau
- the Picture of Subsidized Households survey of the US Department of Housing and Urban Development

The American Community Survey

The 2012-2016 version of the American Community Survey (ACS) was used.¹ The variables of interest are the following :

- occupied housing units by tenure
- total population in occupied housing units by tenure

Picture of Subsidized Households

The survey was conducted in 2012. Even though data on multiple programs were available, only data on “Public Housing” were kept.

The variable of interest are :

- Number of units under contract for federal subsidy and available for occupancy
- Occupied units as the % of units available
- Total number of people

Data at the county level

The American Community Survey

Occupied housing units by tenure

This first data set provides information on tenure.

```
data_unit_county <- data_unit_county %>%
  select(-c(starts_with("HD02_"))) %>%
  rename(
    total_units = HD01_VD01,
    total_units_owner = HD01_VD02,
    total_units_renter = HD01_VD03
  ) %>%
  mutate(
    pct_owner_unit = total_units_owner/total_units*100, # percentage of owner occupied units
    pct_renter_unit = total_units_renter/total_units*100, # percentage of renter occupied units
    pct_tot_unit = pct_renter_unit + pct_owner_unit, # check
    GEOID = str_sub(GEO.id,-5,-1) # for the join
  )
```

¹The 2012-2016 American Community Survey (ACS) data generally reflect the February 2013 Office of Management and Budget (OMB) definitions of metropolitan and micropolitan statistical areas; in certain instances the names, codes, and boundaries of the principal cities shown in ACS tables may differ from the OMB definitions due to differences in the effective dates of the geographic entities. Estimates of urban and rural population, housing units, and characteristics reflect boundaries of urban areas defined based on Census 2010 data. As a result, data for urban and rural areas from the ACS do not necessarily reflect the results of ongoing urbanization.

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
head(data_unit_county)
```

```
## # A tibble: 6 x 10
##   GEO.id      GEO.id2 `GEO.display-label`  total_units total_units_own~
##   <chr>      <int> <chr>                <int>        <int>
## 1 0500000US01~ 1001 Autauga County, Alaba~    20800        15218
## 2 0500000US01~ 1003 Baldwin County, Alaba~    75149        53905
## 3 0500000US01~ 1005 Barbour County, Alaba~    9122         5829
## 4 0500000US01~ 1007 Bibb County, Alabama      7048         5119
## 5 0500000US01~ 1009 Blount County, Alabama    20619        16254
## 6 0500000US01~ 1011 Bullock County, Alaba~    3556         2600
## # ... with 5 more variables: total_units_renter <int>,
## #   pct_owner_unit <dbl>, pct_renter_unit <dbl>, pct_tot_unit <dbl>,
## #   GEOID <chr>
```

Total population in occupied housing by tenure

This second data set provides information on the total population in occupied housing units by tenure by year householder moved into unit. However, only the broad estimates, which ignored the time dimension, were kept so as to remain consistent with the 2012-2016 data above.

```
data_pop_county <- data_pop_county %>%
  select(c(1:7), c(20), -c(starts_with("HD02_"))) %>%
  rename(
    total_pop = HD01_VD01,
    total_pop_owner = HD01_VD02,
    total_pop_renter = HD01_VD09
  ) %>%
  mutate(
    pct_owner_pop = total_pop_owner/total_pop*100, # percentage of owner population
    pct_renter_pop = total_pop_renter/total_pop*100, # percentage of renter population
    pct_tot_pop = pct_renter_pop + pct_owner_pop, # check
    GEOID = str_sub(GEO.id, -5, -1) # for the join
  )
```

```
head(data_pop_county)
```

```
## # A tibble: 6 x 10
##   GEO.id      GEO.id2 `GEO.display-label`  total_pop total_pop_owner
##   <chr>      <int> <chr>                <int>        <int>
## 1 0500000US01001 1001 Autauga County, Alabama    54559        40133
## 2 0500000US01003 1003 Baldwin County, Alabama    196599       141084
## 3 0500000US01005 1005 Barbour County, Alabama    23682        14939
## 4 0500000US01007 1007 Bibb County, Alabama      20571        14450
## 5 0500000US01009 1009 Blount County, Alabama     57152        46037
## 6 0500000US01011 1011 Bullock County, Alabama     10111         7258
## # ... with 5 more variables: total_pop_renter <int>, pct_owner_pop <dbl>,
## #   pct_renter_pop <dbl>, pct_tot_pop <dbl>, GEOID <chr>
```

```
data_ACS_county <- left_join(data_pop_county, data_unit_county, by = c("GEOID", "GEO.id2", "GEO.id", "GEO.id2"))
```

```
head(data_ACS_county)
```

```
## # A tibble: 6 x 16
```

```
## GEO.id      GEO.id2 `GEO.display-label`      total_pop total_pop_owner
## <chr>      <int> <chr>      <int>      <int>
## 1 0500000US01001    1001 Autauga County, Alabama    54559      40133
## 2 0500000US01003    1003 Baldwin County, Alabama    196599     141084
## 3 0500000US01005    1005 Barbour County, Alabama    23682      14939
## 4 0500000US01007    1007 Bibb County, Alabama    20571      14450
## 5 0500000US01009    1009 Blount County, Alabama    57152      46037
## 6 0500000US01011    1011 Bullock County, Alabama    10111      7258
## # ... with 11 more variables: total_pop_renter <int>, pct_owner_pop <dbl>,
## #   pct_renter_pop <dbl>, pct_tot_pop <dbl>, GEOID <chr>,
## #   total_units <int>, total_units_owner <int>, total_units_renter <int>,
## #   pct_owner_unit <dbl>, pct_renter_unit <dbl>, pct_tot_unit <dbl>
```

There are 3,144 counties in the US (including Alaska and Puerto Rico). In the data, there are 3,142 observations.

Picture of Subsidized Households

```
data_county <- data_county %>%
  filter(program_label=="Public Housing" & !(is.na(total_units))) %>%
  select(states, entities, total_units, pct_occupied, people_total, state) %>%
  rename(
    total_units_ph = total_units,
    pct_occupied_ph = pct_occupied,
    total_pop_ph = people_total
  ) %>%
  mutate(GEOID = str_sub(entities,-5,-1)) %>%
  filter(!(GEOID=="78999") & !(GEOID=="66999")) # removes Guam and VI

head(data_county)
```

```
## # A tibble: 6 x 7
##   states entities total_units_ph pct_occupied_ph total_pop_ph state GEOID
##   <chr>   <chr>      <int>      <int>      <dbl> <chr> <chr>
## 1 AK Ala~ AK Anch~        508         95    1218. AK  02020
## 2 AK Ala~ AK Beth~        117         97    456. AK  02050
## 3 AK Ala~ AK Fair~        165         98    462. AK  02090
## 4 AK Ala~ AK June~        207         97    462. AK  02110
## 5 AK Ala~ AK Ketc~         73        100    132. AK  02130
## 6 AK Ala~ AK Kodi~         40        100    109. AK  02150
```

The data set is only made of 2,010 observations.

Final data set for county level

```
data_combined_county <- full_join(data_ACS_county, data_county, by = ("GEOID"))

head(data_combined_county)
```

```
## # A tibble: 6 x 20
##   GEOID GEO.id GEO.id2 `GEO.display-lab~ entities states state total_pop
##   <chr> <chr>   <int> <chr>      <chr>   <chr> <chr>      <int>
## 1 01001 050000~    1001 Autauga County, ~ AL Autau~ AL Al~ AL      54559
## 2 01003 050000~    1003 Baldwin County, ~ AL Baldw~ AL Al~ AL      196599
## 3 01005 050000~    1005 Barbour County, ~ AL Barbo~ AL Al~ AL      23682
```

```
## 4 01007 050000~ 1007 Bibb County, Ala~ AL Bibb ~ AL AL~ AL 20571
## 5 01009 050000~ 1009 Blount County, A~ AL Bloun~ AL AL~ AL 57152
## 6 01011 050000~ 1011 Bullock County, ~ AL Bullo~ AL AL~ AL 10111
## # ... with 12 more variables: total_pop_owner <int>,
## #   total_pop_renter <int>, pct_owner_pop <dbl>, pct_renter_pop <dbl>,
## #   total_pop_ph <dbl>, total_units <int>, total_units_owner <int>,
## #   total_units_renter <int>, pct_owner_unit <dbl>, pct_renter_unit <dbl>,
## #   total_units_ph <int>, pct_occupied_ph <int>
```

Data at the Core-base statistical areas level

The American Community Survey

Occupied housing units by tenure

```
data_unit_CBSA <- data_unit_CBSA %>%
  select(-c(starts_with("HD02_"))) %>%
  rename(
    total_units = HD01_VD01,
    total_units_owner = HD01_VD02,
    total_units_renter = HD01_VD03
  ) %>%
  mutate(
    pct_owner_unit = total_units_owner/total_units*100, # percentage of owner occupied units
    pct_renter_unit = total_units_renter/total_units*100, # percentage of renter occupied units
    pct_tot_unit = pct_renter_unit + pct_owner_unit, # check
    GEOID = str_sub(GEO.id,-5,-1) # for the join
  )
```

```
head(data_unit_CBSA)
```

```
## # A tibble: 6 x 10
##   GEO.id      GEO.id2 `GEO.display-label` total_units total_units_own~
##   <chr>      <int> <chr>                <int>          <int>
## 1 310M300US10~ 10100 Aberdeen, SD Micro Ar~ 17721          12171
## 2 310M300US10~ 10140 Aberdeen, WA Micro Ar~ 27472          18430
## 3 310M300US10~ 10180 Abilene, TX Metro Area 60308          37194
## 4 310M300US10~ 10220 Ada, OK Micro Area    14625           9389
## 5 310M300US10~ 10300 Adrian, MI Micro Area 37856          29398
## 6 310M300US10~ 10420 Akron, OH Metro Area 283472         187951
## # ... with 5 more variables: total_units_renter <int>,
## #   pct_owner_unit <dbl>, pct_renter_unit <dbl>, pct_tot_unit <dbl>,
## #   GEOID <chr>
```

Total population in occupied housing by tenure

```
data_pop_CBSA <- data_pop_CBSA %>%
  select(c(1:7),c(18), -c(starts_with("HD02_"))) %>%
  rename(
    total_pop = HD01_VD01,
    total_pop_owner = HD01_VD02,
    total_pop_renter = HD01_VD08
  ) %>%
```

```
mutate(
  pct_owner_pop = total_pop_owner/total_pop*100, # percentage of owner population
  pct_renter_pop= total_pop_renter/total_pop*100, # percentage of renter population
  pct_tot_pop =pct_renter_pop + pct_owner_pop, # check
  GEOID = str_sub(GEO.id,-5,-1) # for the join
)
```

```
head(data_pop_CBSA)
```

```
## # A tibble: 6 x 10
##   GEO.id      GEO.id2 `GEO.display-label`   total_pop total_pop_owner
##   <chr>      <int> <chr>                <int>      <int>
## 1 310M300US10100  10100 Aberdeen, SD Micro Area    40749      30087
## 2 310M300US10140  10140 Aberdeen, WA Micro Area    68170      44798
## 3 310M300US10180  10180 Abilene, TX Metro Area   155109     96299
## 4 310M300US10220  10220 Ada, OK Micro Area     36811     24560
## 5 310M300US10300  10300 Adrian, MI Micro Area   93397     72719
## 6 310M300US10420  10420 Akron, OH Metro Area    687154    475135
## # ... with 5 more variables: total_pop_renter <int>, pct_owner_pop <dbl>,
## #   pct_renter_pop <dbl>, pct_tot_pop <dbl>, GEOID <chr>
```

```
data_ACS_CBSA <- left_join(data_pop_CBSA, data_unit_CBSA, by = c("GEOID", "GEO.id2", "GEO.id", "GEO.display-label"))
```

```
head(data_ACS_CBSA)
```

```
## # A tibble: 6 x 16
##   GEO.id      GEO.id2 `GEO.display-label`   total_pop total_pop_owner
##   <chr>      <int> <chr>                <int>      <int>
## 1 310M300US10100  10100 Aberdeen, SD Micro Area    40749      30087
## 2 310M300US10140  10140 Aberdeen, WA Micro Area    68170      44798
## 3 310M300US10180  10180 Abilene, TX Metro Area   155109     96299
## 4 310M300US10220  10220 Ada, OK Micro Area     36811     24560
## 5 310M300US10300  10300 Adrian, MI Micro Area   93397     72719
## 6 310M300US10420  10420 Akron, OH Metro Area    687154    475135
## # ... with 11 more variables: total_pop_renter <int>, pct_owner_pop <dbl>,
## #   pct_renter_pop <dbl>, pct_tot_pop <dbl>, GEOID <chr>,
## #   total_units <int>, total_units_owner <int>, total_units_renter <int>,
## #   pct_owner_unit <dbl>, pct_renter_unit <dbl>, pct_tot_unit <dbl>
```

Picture of Subsidized Households

```
data_CBSA <- data_CBSA %>%
  filter(program_label == "Public Housing" & !(is.na(total_units))) %>%
  select(entities, total_units, pct_occupied, people_total) %>%
  rename(
    total_units_ph = total_units,
    pct_occupied_ph=pct_occupied,
    total_pop_ph = people_total
  ) %>%
  mutate(GEOID = str_sub(entities,1,5)) %>%
  filter(!(GEOID=="78999") & !(GEOID=="66999")) # removes Guam and VI

head(data_CBSA)
```

```
## # A tibble: 6 x 5
```

```
##   entities          total_units_ph pct_occupied_ph total_pop_ph GEOID
##   <chr>              <int>          <int>          <dbl> <chr>
## 1 10020 Abbeville, LA          380           92           772. 10020
## 2 10100 Aberdeen, SD           100           98           169. 10100
## 3 10140 Aberdeen, WA          415           95           621. 10140
## 4 10180 Abilene, TX           493           96          1082. 10180
## 5 10220 Ada, OK                275           98           347. 10220
## 6 10260 Adjuntas, PR           205          100           563. 10260
```

There are 929 CBSA in the US (including Puerto Rico). Our data set is made of 798 CBSA.

Final data set for county level

```
data_combined_CBSA <- full_join(data_ACS_CBSA, data_CBSA, by = ("GEOID"))
```

```
head(data_combined_CBSA)
```

```
## # A tibble: 6 x 18
##   GEOID GEO.id GEO.id2 `GEO.display-la~ entities total_pop total_pop_owner
##   <chr> <chr>   <int> <chr>          <chr>          <int>          <int>
## 1 10100 310M3~   10100 Aberdeen, SD Mi~ 10100 A~         40749         30087
## 2 10140 310M3~   10140 Aberdeen, WA Mi~ 10140 A~         68170         44798
## 3 10180 310M3~   10180 Abilene, TX Met~ 10180 A~        155109         96299
## 4 10220 310M3~   10220 Ada, OK Micro A~ 10220 A~         36811         24560
## 5 10300 310M3~   10300 Adrian, MI Micr~ <NA>          93397         72719
## 6 10420 310M3~   10420 Akron, OH Metro~ 10420 A~        687154        475135
## # ... with 11 more variables: total_pop_renter <int>, pct_owner_pop <dbl>,
## #   pct_renter_pop <dbl>, total_pop_ph <dbl>, total_units <int>,
## #   total_units_owner <int>, total_units_renter <int>,
## #   pct_owner_unit <dbl>, pct_renter_unit <dbl>, total_units_ph <int>,
## #   pct_occupied_ph <int>
```