

# educational\_attainment

*Rick Gilmore*

*2018-03-27 08:00:12*

## Educational Attainment

```
counties <- read.csv(paste0(csv.dir, "city-state-county.csv"), stringsAsFactors = FALSE)

data("county.regions")
counties <- left_join(counties, county.regions)
```

```
## Joining, by = "region"
```

```
demog <- get_county_demographics(endyear=2013, span=5)
county.demo <- left_join(counties, demog)
```

```
## Joining, by = "region"
```

```
# Recapitalize county
county.demo$County <- unlist(lapply(county.demo$County, Cap_all))
# Hack District Of columbia... TODO (ROG): Fix Cap_all()
county.demo$County[county.demo$County == "District Of columbia"] = "District of Columbia"
```

To gather educational attainment data, must create specific ACS 'geometry'.

```
state.fips <- as.numeric(county.demo$state.fips.character)
county.fips <- as.numeric(substr(county.demo$county.fips.character,3,5))
play.geo <- geo.make(state = state.fips, county = county.fips)
```

```
Make.county.geo <- function(i, df) {
  geo.make(state = as.numeric(df$state.fips.character[i]),
    county =
      as.numeric(substr(county.demo$county.fips.character[i],4, 6)))
}
```

```
cty <- 1
# Generate name for county-level geography
geo.name <- paste0(county.demo$Site.code[cty], "_",
  county.demo$county.name[cty], "_", county.demo$State[cty])
```

```
# Create geography and assign to generated name
assign(geo.name, Make.county.geo(cty, county.demo))
```

```
ed.attain <- acs.lookup(table.name="Educational Attainment for the Population 25 Years and Over",
  endyear=2015)
```

```
# Manual inspection shows variables 1:25 seem to contain the relevant info
play.ed <- acs.fetch(geography = play.geo, endyear = 2015, variable = ed.attain[1:25],
  col.names = c("Total", #1
    "None", #2
    "<K", #3
    "K", #4
    "1st", #5
```

```

        "2nd",    #6
        "3rd",    #7
        "4th",    #8
        "5th",    #9
        "6th",    #10
        "7th",    #11
        "8th",    #12
        "9th",    #13
        "10th",   #14
        "11th",   #15
        "12th",   #16
        "HS",     #17
        "GED",    #18
        "Coll <1yr", #19
        "Coll >1yr", #20
        "AA",     #21
        "BA",     #22
        "MA",     #23
        "Prof",   #24
        "Ph.D")) #25

# Columns 2:16 are grades < HS diploma
lt.7 <- function(i) sum(play.ed[i,2:10])
mid.hs <- function(i) sum(play.ed[i,11:16])
lt.hs <- function(i) sum(play.ed[i,2:16])
hs.grad <- function(i) sum(play.ed[i,17:18])
some.coll <- function(i) sum(play.ed[i,19:21])
ba <- function(i) sum(play.ed[i,22])
ba.plus <- function(i) sum(play.ed[i,22:25])
gt.ba <- function(i) sum(play.ed[i,23:25])

# Use functions to create data table for easier manipulation
Make.ed.attain.table <- function(i) {
  this.cty <- slot(play.ed[i,1], "geography")$NAME
  data.frame(county = this.cty,
             tot = as.numeric(slot(play.ed[i,1], "estimate")),
             lt.7 = as.numeric(slot(lt.7(i), "estimate")),
             mid.hs = as.numeric(slot(mid.hs(i), "estimate")),
             lt.hs = as.numeric(slot(lt.hs(i), "estimate")),
             hs.grad = as.numeric(slot(hs.grad(i), "estimate")),
             some.coll = as.numeric(slot(some.coll(i), "estimate")),
             ba = as.numeric(slot(ba(i), "estimate")),
             gt.ba = as.numeric(slot(gt.ba(i), "estimate")),
             ba.plus = as.numeric(slot(ba.plus(i), "estimate")))
}

ed.attain.list <- lapply(1:dim(play.ed)[1], Make.ed.attain.table)
ed.attain.df <- Reduce(function(x,y) full_join(x,y, all=TRUE), ed.attain.list)

## Joining, by = c("county", "tot", "lt.7", "mid.hs", "lt.hs", "hs.grad", "some.coll", "ba", "gt.ba", "ba.plus")
## Warning: Column `county` joining factors with different levels, coercing to
## character vector
## Joining, by = c("county", "tot", "lt.7", "mid.hs", "lt.hs", "hs.grad", "some.coll", "ba", "gt.ba", "ba.plus")

```





[illegible]

```
## Joining, by = c("county", "tot", "lt.7", "mid.hs", "lt.hs", "hs.grad", "some.coll", "ba", "gt.ba", "l")
## Warning: Column `county` joining character vector and factor, coercing into
## character vector
## Joining, by = c("county", "tot", "lt.7", "mid.hs", "lt.hs", "hs.grad", "some.coll", "ba", "gt.ba", "l")
## Warning: Column `county` joining character vector and factor, coercing into
## character vector
ed.attain.df %>%
  mutate(p.lt.hs = 100*lt.hs/tot,
         p.hs.grad = 100*hs.grad/tot,
         p.some.coll = 100*some.coll/tot,
         p.ba.plus = 100*ba.plus/tot) %>%
  select(county, p.lt.hs, p.hs.grad, p.some.coll, p.ba.plus) ->
  ed.attain.by.county

names(ed.attain.by.county) <- c("County", "<HS", "HS", "HS+", "BA+")
ed.attain.by.county %>%
  knitr::kable()
```

| County                                     | <HS       | HS        | HS+      | BA+      |
|--|-----------|-----------|----------|----------|
| Los Angeles County, California             | 22.718630 | 20.676552 | 26.33454 | 30.27027 |
| Merced County, California                  | 32.124274 | 24.848266 | 29.94784 | 13.07962 |
| Orange County, California                  | 15.731238 | 17.689816 | 28.84002 | 37.73893 |
| Riverside County, California               | 19.913051 | 25.819297 | 33.34025 | 20.92741 |
| San Mateo County, California               | 11.672114 | 16.535181 | 26.17003 | 45.62267 |
| Santa Clara County, California             | 13.024218 | 15.167082 | 23.86490 | 47.94380 |
| Santa Cruz County, California              | 14.396614 | 15.487459 | 31.87140 | 38.24453 |
| Yolo County, California                    | 14.506227 | 19.697083 | 26.82260 | 38.97409 |
| District of Columbia, District of Columbia | 10.668529 | 17.996837 | 16.75794 | 54.57670 |
| Miami-Dade County, Florida                 | 19.910852 | 28.492734 | 24.67464 | 26.92177 |
| Fulton County, Georgia                     | 9.055800  | 18.005673 | 23.63889 | 49.29963 |
| Oconee County, Georgia                     | 6.846064  | 21.285932 | 26.98576 | 44.88224 |
| Cook County, Illinois                      | 14.709575 | 23.951799 | 25.56936 | 35.76927 |
| Monroe County, Indiana                     | 7.644555  | 22.310366 | 25.06800 | 44.97708 |
| Tippecanoe County, Indiana                 | 9.719412  | 26.604045 | 28.51591 | 35.16063 |
| Orleans Parish, Louisiana                  | 14.842995 | 23.151391 | 26.71059 | 35.29503 |
| Suffolk County, Massachusetts              | 15.708324 | 23.607559 | 19.00435 | 41.67977 |
| Montgomery County, Maryland                | 8.822279  | 13.968895 | 19.34383 | 57.86500 |
| Crawford County, Michigan                  | 12.522949 | 35.394724 | 34.93091 | 17.15142 |
| Ingham County, Michigan                    | 8.202319  | 21.622739 | 32.92508 | 37.24986 |
| Camden County, New Jersey                  | 12.074268 | 31.320041 | 26.44240 | 30.16329 |
| Essex County, New Jersey                   | 15.806027 | 28.728008 | 22.77177 | 32.69419 |
| Gloucester County, New Jersey              | 8.335119  | 34.333650 | 28.01957 | 29.31166 |
| Mercer County, New Jersey                  | 12.561574 | 25.615744 | 22.17729 | 39.64540 |
| New York County, New York                  | 13.376256 | 12.610941 | 14.08307 | 59.92973 |
| Richmond County, New York                  | 11.300797 | 31.964367 | 25.94675 | 30.78809 |
| Tompkins County, New York                  | 5.763363  | 19.816934 | 23.59287 | 50.82683 |
| Franklin County, Ohio                      | 10.004303 | 25.147930 | 27.28086 | 37.56691 |
| Lane County, Oregon                        | 8.892151  | 24.967860 | 37.74043 | 28.39956 |
| Allegheny County, Pennsylvania             | 6.458414  | 29.826715 | 25.95835 | 37.75653 |
| Bucks County, Pennsylvania                 | 6.542183  | 30.652041 | 25.38795 | 37.41782 |
| Centre County, Pennsylvania                | 6.822231  | 31.613749 | 20.17269 | 41.39133 |
| Chester County, Pennsylvania               | 7.276344  | 23.179965 | 20.40566 | 49.13803 |

| County                            | <HS       | HS        | HS+      | BA+      |
|-----------------------------------|-----------|-----------|----------|----------|
| Delaware County, Pennsylvania     | 7.793264  | 31.782900 | 24.37388 | 36.04996 |
| Montgomery County, Pennsylvania   | 6.248537  | 24.813364 | 22.06307 | 46.87503 |
| Philadelphia County, Pennsylvania | 17.981287 | 33.822359 | 22.78378 | 25.41257 |
| Davidson County, Tennessee        | 12.716788 | 24.092843 | 25.90111 | 37.28926 |
| Williamson County, Tennessee      | 4.656689  | 16.371417 | 23.23685 | 55.73504 |
| Harris County, Texas              | 20.412793 | 23.326487 | 26.80649 | 29.45423 |
| Travis County, Texas              | 12.166591 | 17.104836 | 24.77124 | 45.95733 |
| Arlington County, Virginia        | 6.522062  | 8.299942  | 12.29320 | 72.88480 |
| Chesterfield County, Virginia     | 9.037095  | 24.096152 | 29.97047 | 36.89628 |
| Henrico County, Virginia          | 9.341698  | 22.841190 | 27.13650 | 40.68061 |
| James City County, Virginia       | 6.207560  | 21.079798 | 25.68395 | 47.02869 |
| Richmond city, Virginia           | 16.815834 | 23.029182 | 24.10500 | 36.04999 |

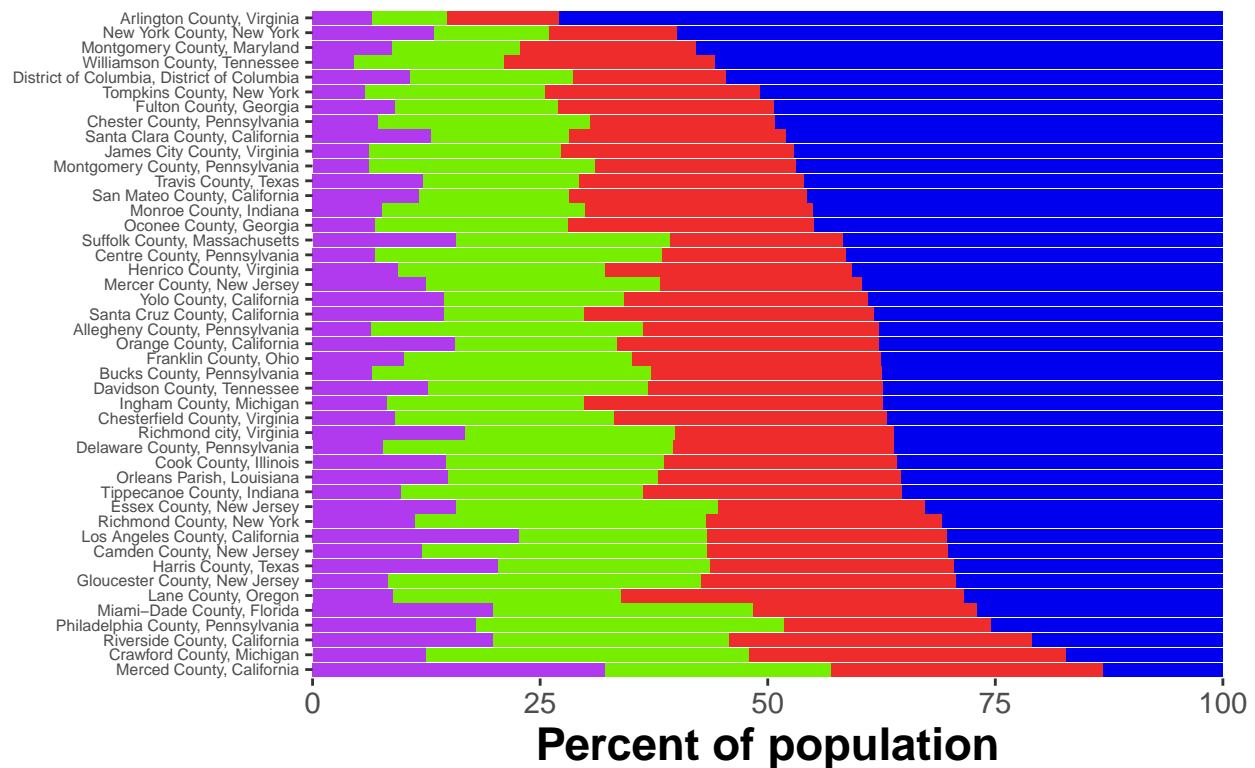
```
# ed.attain.by.county %>%
#   gather(key = ed.level, value = proportion.pop, -County) %>%
#   mutate(ed.level = ordered(ed.level, levels = c("<HS", "HS", "HS+", "BA+"))) %>%
#   ggplot() +
#   aes(x = ed.level, y = proportion.pop, fill = ed.level,
#        color = ed.level,
#        group = County) +
#   geom_line(color = "black", linetype = 3) +
#   geom_point(size = 3) +
#   xlab("Level of educational attainment") +
#   ylab("Proportion of population") +
#   theme_classic() +
#   theme(legend.position = "none",
#         axis.title = element_text(size = rel(1.5), face = "bold"),
#         axis.text = element_text(size = rel(1.2)))

ed.attain.by.county %>%
  gather(key = ed.level, value = proportion.pop, -County) %>%
  mutate(ed.level = ordered(ed.level, levels = rev(c("<HS", "HS", "HS+", "BA+")))) ->
  ed.attain.gathered

ed.attain.by.county %>%
  select(County, `BA+`) ->
  p.ba.sortlist

left_join(ed.attain.gathered, p.ba.sortlist) %>%
  arrange(`BA+`) %>%
  mutate(County = factor(County, unique(County))) %>%
  ggplot() +
  aes(x = County, y = proportion.pop, fill = ed.level) +
  geom_col() +
  coord_flip() +
  play.theme +
  play.palette +
  scale_y_continuous(expand=c(0,0)) +
  ylab("Percent of population")

## Joining, by = "County"
```



■ BA+ ■ HS+ ■ HS ■ <HS

```
ed.attain.list <- lapply(1:dim(play.ed)[1], Make.ed.attain.table)
ed.attain.df <- Reduce(function(x,y) full_join(x,y, all=TRUE), ed.attain.list)
```

```
## Joining, by = c("county", "tot", "lt.7", "mid.hs", "lt.hs", "hs.grad", "some.coll", "ba", "gt.ba", "l
## Warning: Column `county` joining factors with different levels, coercing to
## character vector
## Joining, by = c("county", "tot", "lt.7", "mid.hs", "lt.hs", "hs.grad", "some.coll", "ba", "gt.ba", "l
## Warning: Column `county` joining character vector and factor, coercing into
## character vector
## Joining, by = c("county", "tot", "lt.7", "mid.hs", "lt.hs", "hs.grad", "some.coll", "ba", "gt.ba", "l
## Warning: Column `county` joining character vector and factor, coercing into
## character vector
## Joining, by = c("county", "tot", "lt.7", "mid.hs", "lt.hs", "hs.grad", "some.coll", "ba", "gt.ba", "l
## Warning: Column `county` joining character vector and factor, coercing into
## character vector
## Joining, by = c("county", "tot", "lt.7", "mid.hs", "lt.hs", "hs.grad", "some.coll", "ba", "gt.ba", "l
## Warning: Column `county` joining character vector and factor, coercing into
## character vector
```



[illegible]

[illegible]



```

names(ed.attain.by.county) <- c("County", "<7th", "<HS", "HS", "HS+", "BA", "MA+")

ed.attain.by.county %>%
  gather(key = ed.level, value = proportion.pop, -County) %>%
  mutate(ed.level = ordered(ed.level, levels = rev(c("<7th", "<HS", "HS", "HS+", "BA", "MA+")))) ->
  ed.attain.gathered

ed.attain.by.county %>%
  mutate(ba.plus = `BA` + `MA+`) %>%
  select(County, ba.plus) ->
  p.ba.sortlist

left_join(ed.attain.gathered, p.ba.sortlist) %>%
  arrange(ba.plus) %>%
  mutate(County = factor(County, unique(County))) %>%
  ggplot() +
  aes(x = County, y = proportion.pop, fill = ed.level) +
  geom_col() +
  coord_flip() +
  play.theme +
  play.palette +
  scale_y_continuous(expand=c(0,0)) +
  ylab("Percent of population")

```

```
## Joining, by = "County"
```

