Massachussets census tract level data

Carmen Rodriguez Cabrera

2/15/2022

```
#ONLY NEED TO DO THIS ONCE!!!!!!
#library(devtools)
#install.packages("rgdal", configure.args = c("--with-proj-lib=/usr/local/lib/", "--with-proj-include=/usr/local
#install.packages("sf", configure.args="--with-gdal-config=/usr/local/gdal-3.4.1/bin/gdal-config --with-geos-con
```

Set up Census API Key

Variable selection (CLEAN UP)

```
#***Note: dates of diagnosis ranged from 2006- 2017****
####THIS CAN BE MADE MORE CONCISE!!!!

#Getting variables from the Census or ACS requires knowing the variable ID -
#and there are thousands of these IDs across the different Census files.
#To rapidly search for variables, use the load_variables() function.

#---THERE ARE SOME DIFFERENCES IN THE NAMES SO TO BE SAFE WENT THROUGH VARIABLES BY YEAR

## 2006-2010 ACS 5-year estimates--2010 ACS
v10_acsprofile<-load_variables(2010, "acs5/profile", cache = TRUE)
v10_acstable<-load_variables(2010, "acs5", cache = TRUE)

## 2011-2015, ACS 2015 5-year
v15_acsprofile<-load_variables(2015, "acs5/profile", cache = TRUE)
v15_acstable<-load_variables(2015, "acs5", cache = TRUE)

## 2015-2019  ACS 2019 5-year
v19_acsprofile<-load_variables(2019, "acs5/profile", cache = TRUE)
v19_acstable<-load_variables(2019, "acs5", cache = TRUE)

####  2010 ACS variables
medianincome<-c(`medianincome_2010` = "B19013_001")
#B19013_001 :Estimate!!Median household income in the past 12 months (in 2019 inflation-adjusted dollars)
education<-c(`HighSchoolHigherP_2010`= "DP02_0066P",
             `BachelorHigherP_2010` = "DP02_0067P")
#"DP02_0067",Estimate!!EDUCATIONAL ATTAINMENT!!Percent bachelor's degree or higher
# "DP02_0067P", Percent!!EDUCATIONAL ATTAINMENT!!Percent bachelor's degree or higher
#"DP02_0066", Estimate!!EDUCATIONAL ATTAINMENT!!Percent high school graduate or higher
#"DP02_0066P",Percent!!EDUCATIONAL ATTAINMENT!!Percent high school graduate or higher

housing<-c(`RenterOccupiedUnitP_2010` =  "DP04_0046P",  #Percent!!HOUSING TENURE!!Renter-occupied
           `OwnerOccupiedUnitP_2010` = "DP04_0045P")   #Percent!!HOUSING TENURE!!Owner-occupied

employment<-c(`UnemployementP_2010` = "DP03_0009P")

#DP03_0009 Estimate!!EMPLOYMENT STATUS!!Percent Unemployed
#DP03_0009P Percent!!EMPLOYMENT STATUS!!Percent Unemployed
```

```
#DP03_0005P Percent!!EMPLOYMENT STATUS!!In labor force!!Civilian labor force!!Unemployed

poverty<-c(`BelowPovertyLineP_2010` = "DP03_0119P")
#DP03_0128 :Estimate!!PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12
#MONTHS IS BELOW THE POVERTY LEVEL!!All people
#DP03_0128P :Percent!!PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12
#MONTHS IS BELOW THE POVERTY LEVEL!!All people

#DP03_0119 :Estimate!!PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12
#MONTHS IS BELOW THE POVERTY LEVEL!!All families
#DP03_0119P :Percent!!PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12
#MONTHS IS BELOW THE POVERTY LEVEL!!All families

vars_want_2010<-c(medianincome, education,housing,employment, poverty)

####  2015 ACS variables
vars_want_2015<-c(`medianincome_2015` = "B19013_001",
                          `HighSchoolHigherP_2015`= "DP02_0066P",
                          `BachelorHigherP_2015`= "DP02_0067P",
                          `RenterOccupiedUnitP_2015` = "DP04_0047P",
                          `OwnerOccupiedUnitP_2015` = "DP04_0046P",
                          `UnemployementP_2015`= "DP03_0009P",
                          `BelowPovertyLineP_2015` = "DP03_0119P")

#### 2019 ACS variables
vars_want_2019<-c(`medianincome_2019` = "B19013_001",
                          `HighSchoolHigherP_2019` = "DP02_0067P",
                          `BachelorHigherP_2019`= "DP02_0068P",
                          `RenterOccupiedUnitP_2019` = "DP04_0047P",
                          `OwnerOccupiedUnitP_2019` = "DP04_0046P",
                          `UnemployementP_2019` = "DP03_0009P",
                          `BelowPovertyLineP_2019` = "DP03_0119P")
```

Extract ACS 5-year estimates (2006-2010)

```
acs5_2010_long<-get_acs(geography = "tract", state="MA", variables = vars_want_2010, year = 2010)
acs5_2010_wide<-acs5_2010_long %>% select("GEOID","NAME","variable","estimate") %>% spread(variable,estimate)
#head(acs5_2010_wide)
str(acs5_2010_wide)
```

```
## tibble [1,478 x 9] (S3: tbl_df/tbl/data.frame)
##  $ GEOID                  : chr [1:1478] "25001010100" "25001010206" "25001010208" "25001010304" ...
##  $ NAME                   : chr [1:1478] "Census Tract 101, Barnstable County, Massachusetts" "Census Tract
##  $ BachelorHigherP_2010   : num [1:1478] 54.2 61.7 46.6 55.4 37.4 39.6 65.6 49.1 50.9 52.5 ...
##  $ BelowPovertyLineP_2010 : num [1:1478] 3.2 1.9 5.2 0 3.8 5.2 0 2.5 3.8 3.6 ...
##  $ HighSchoolHigherP_2010 : num [1:1478] 95.1 99.2 96.1 98.4 96.1 94.9 99.7 93.5 96.6 97.6 ...
##  $ medianincome_2010      : num [1:1478] 44646 66109 80425 64659 54464 ...
##  $ OwnerOccupiedUnitP_2010 : num [1:1478] 69 80.5 78.3 82.9 89.4 72.6 73.7 83 81.5 86.6 ...
##  $ RenterOccupiedUnitP_2010: num [1:1478] 31 19.5 21.7 17.1 10.6 27.4 26.3 17 18.5 13.4 ...
##  $ UnemployementP_2010    : num [1:1478] 9 5 9.5 11 8.2 4.2 6.7 3.7 9.1 2.8 ...
```

```
#describe(acs5_2010_wide)
```

Extract ACS 5-year estimates (2011-2015)

```
acs5_2015_long<-get_acs(geography = "tract", state="MA", variables = vars_want_2015, year = 2015)

acs5_2015_wide<-acs5_2015_long %>% select("GEOID","NAME","variable","estimate") %>% spread(variable,estimate)
```

```
#head(acs5_2015_wide)

str(acs5_2015_wide)
```

```
## tibble [1,478 x 9] (S3: tbl_df/tbl/data.frame)
##  $ GEOID                  : chr [1:1478] "25001010100" "25001010206" "25001010208" "25001010304" ...
##  $ NAME                   : chr [1:1478] "Census Tract 101, Barnstable County, Massachusetts" "Census Tract
##  $ BachelorHigherP_2015   : num [1:1478] 50.2 51.8 50.6 51 42.5 42.7 57.5 53.3 45.7 57.6 ...
##  $ BelowPovertyLineP_2015 : num [1:1478] 8.1 7.7 10.6 0 2.5 2.1 3.8 4.7 7 1.2 ...
##  $ HighSchoolHigherP_2015 : num [1:1478] 92.4 95.1 93.6 96.6 95.3 99.3 98.6 97.8 96.3 98.2 ...
##  $ medianincome_2015      : num [1:1478] 36958 45735 60432 72264 52303 ...
##  $ OwnerOccupiedUnitP_2015 : num [1:1478] 65.4 82.6 82.4 83.5 84.1 74.4 83.1 80.6 82.9 83.3 ...
##  $ RenterOccupiedUnitP_2015: num [1:1478] 34.6 17.4 17.6 16.5 15.9 25.6 16.9 19.4 17.1 16.7 ...
##  $ UnemployementP_2015    : num [1:1478] 12.3 12.5 5.6 6.7 10.8 2 0.9 7.1 4.3 6.1 ...
```

```
#describe(acs5_2015_wide)
```

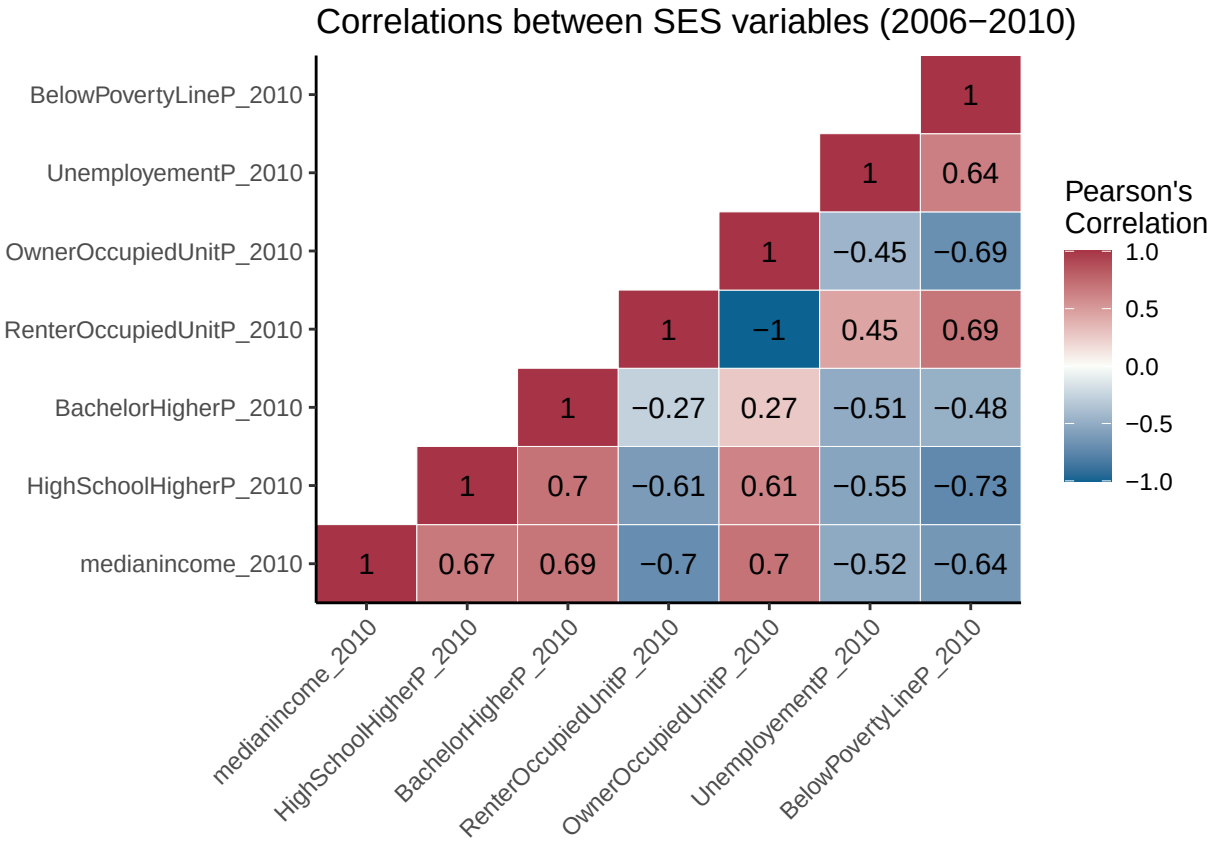
Extract ACS 5-year estimates (2015-2019)

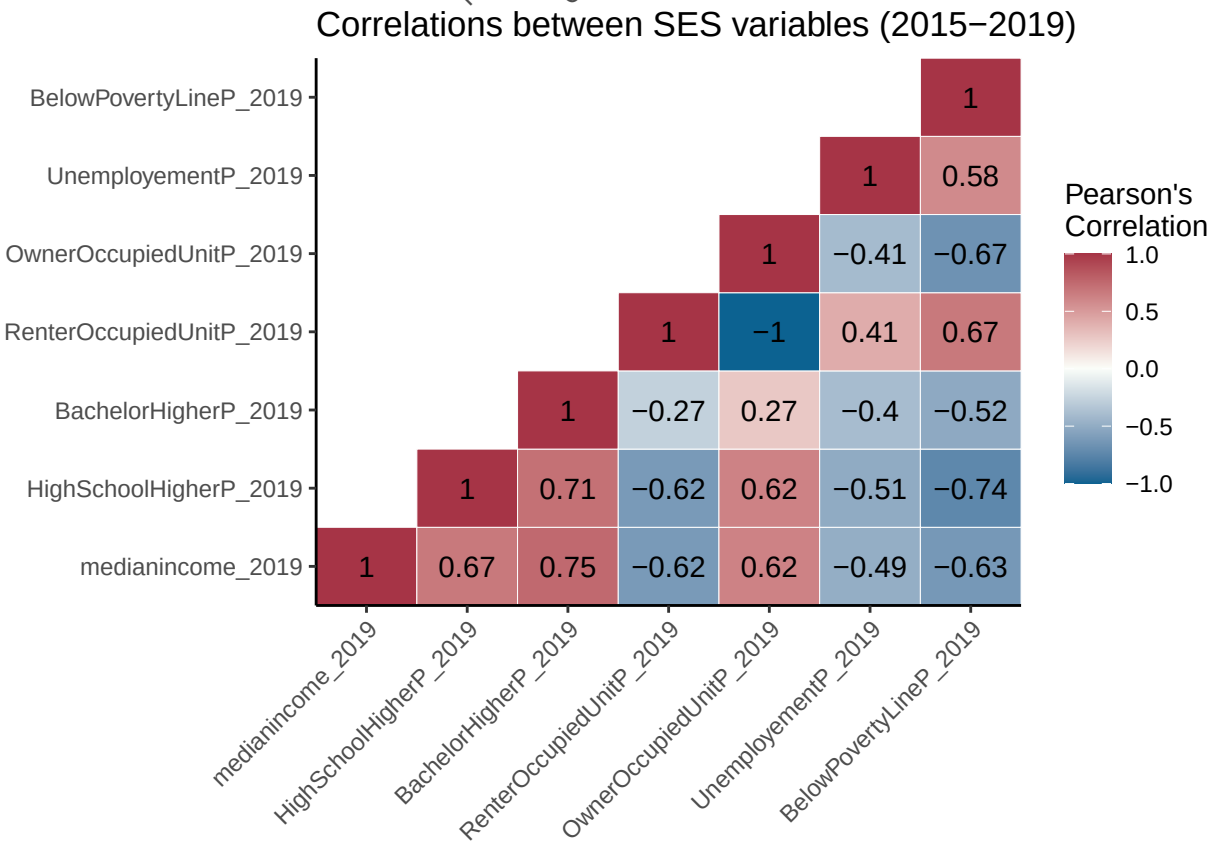
```
acs5_2019_long<-get_acs(geography = "tract", state="MA", variables = vars_want_2019, year=2019)

acs5_2019_wide<-acs5_2019_long %>% select("GEOID","NAME","variable","estimate") %>% spread(variable,estimate)
#head(acs5_2019_wide)
str(acs5_2019_wide)
```

```
## tibble [1,478 x 9] (S3: tbl_df/tbl/data.frame)
##  $ GEOID                  : chr [1:1478] "25001010100" "25001010206" "25001010208" "25001010304" ...
##  $ NAME                   : chr [1:1478] "Census Tract 101, Barnstable County, Massachusetts" "Census Tract
##  $ BachelorHigherP_2019   : num [1:1478] 54.8 53.6 48.5 48.1 53.2 50 67.6 58.5 59.1 60.6 ...
##  $ BelowPovertyLineP_2019 : num [1:1478] 3.7 1 0 4 8.6 7.3 3.3 1.6 6.1 4.8 ...
##  $ HighSchoolHigherP_2019 : num [1:1478] 95.5 94.7 89.3 97.1 98.8 98.6 98.5 98.1 96.5 99.4 ...
##  $ medianincome_2019      : num [1:1478] 59063 74639 68367 85263 70071 ...
##  $ OwnerOccupiedUnitP_2019 : num [1:1478] 78.6 85 82.4 88.5 92.6 81.8 83.8 81 88.2 92.5 ...
##  $ RenterOccupiedUnitP_2019: num [1:1478] 21.4 15 17.6 11.5 7.4 18.2 16.2 19 11.8 7.5 ...
##  $ UnemployementP_2019    : num [1:1478] 7.4 3 8.6 5.7 4 4.6 5.5 3.7 3.4 3.3 ...
```

```
#describe(acs5_2019_wide)
```

Correlations between potential SES variables





Descriptive statistics

```
##
## Attaching package: 'table1'
```

```
## The following objects are masked from 'package:Hmisc':
##
##     label, label<-, units

## The following objects are masked from 'package:base':
##
##     units, units<-

##
## Attaching package: 'psych'

## The following object is masked from 'package:Hmisc':
##
##     describe

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
GEOID*	1	1478	739.50	426.81	739.50	739.50	547.82	1.0	1478.0	1477.0	0.00	-1.20	11.10
NAME*	2	1478	739.50	426.81	739.50	739.50	547.82	1.0	1478.0	1477.0	0.00	-1.20	11.10
BachelorHigherP_2010	3	1463	37.70	20.78	34.00	36.22	21.20	0.0	100.0	100.0	0.57	-0.44	0.54
BelowPovertyLineP_2010	4	1456	8.84	10.88	4.70	6.67	5.34	0.0	77.5	77.5	2.26	6.40	0.29
HighSchoolHigherP_2010	5	1463	87.61	11.24	91.30	89.53	7.41	38.9	100.0	61.1	-1.60	2.40	0.29
medianincome_2010	6	1456	67880.62	29374.74	65571.00	65916.62	26399.92	9554.0	218667.0	209113.0	0.88	1.84	769.83
OwnerOccupiedUnitP_2010	7	1460	62.37	25.44	68.35	64.47	27.35	0.0	100.0	100.0	-0.59	-0.69	0.67
RenterOccupiedUnitP_2010	8	1460	37.63	25.44	31.65	35.53	27.35	0.0	100.0	100.0	0.59	-0.69	0.67
UnemployementP_2010	9	1463	7.76	5.01	6.50	7.02	3.41	0.0	44.7	44.7	2.31	8.84	0.13
BachelorHigherP_2015	10	1464	39.96	21.35	36.40	38.68	22.24	2.2	95.1	92.9	0.47	-0.65	0.56
BelowPovertyLineP_2015	11	1458	9.62	10.65	5.40	7.62	5.49	0.0	63.2	63.2	1.94	4.23	0.28
HighSchoolHigherP_2015	12	1464	88.83	10.24	92.40	90.61	6.67	35.8	100.0	64.2	-1.59	2.39	0.27
medianincome_2015	13	1454	72838.94	32489.56	69457.50	70471.49	30472.62	11863.0	217583.0	205720.0	0.80	0.96	852.04
OwnerOccupiedUnitP_2015	14	1459	60.45	25.50	65.70	62.23	28.17	0.0	100.0	100.0	-0.51	-0.84	0.67
RenterOccupiedUnitP_2015	15	1459	39.55	25.50	34.30	37.77	28.17	0.0	100.0	100.0	0.51	-0.84	0.67
UnemployementP_2015	16	1462	7.98	4.59	7.00	7.39	3.41	0.0	51.0	51.0	2.11	9.45	0.12
BachelorHigherP_2019	17	1463	42.92	21.75	40.50	41.93	23.87	0.0	95.6	95.6	0.34	-0.79	0.57
BelowPovertyLineP_2019	18	1460	8.10	9.33	4.60	6.33	4.89	0.0	65.2	65.2	2.14	5.68	0.24
HighSchoolHigherP_2019	19	1463	89.95	9.43	93.50	91.58	5.78	31.8	100.0	68.2	-1.69	3.16	0.25
medianincome_2019	20	1454	86160.08	38629.16	82265.00	83302.68	35153.19	2499.0	250001.0	247502.0	0.84	1.18	1013.06
OwnerOccupiedUnitP_2019	21	1460	60.81	25.59	66.15	62.65	28.17	0.0	100.0	100.0	-0.59	-0.69	0.67
RenterOccupiedUnitP_2019	22	1460	39.19	25.59	33.85	37.35	28.17	0.0	100.0	100.0	0.59	-0.69	0.67
UnemployementP_2019	23	1464	5.16	3.91	4.40	4.62	2.37	0.0	44.7	44.7	2.31	8.84	0.13

```
## OwnerOccupiedUnitP_2019      0.0    100.0    100.0 -0.52    -0.81    0.67
## RenterOccupiedUnitP_2019     0.0    100.0    100.0  0.52    -0.81    0.67
## UnemployementP_2019          0.0     56.4     56.4  4.60    44.26    0.10
```