hw9

Sarah Hassan

4/10/2021

#hw 9 web scraping/api  
library(tidycensus)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.2 ✓ purrr 0.3.4  
## ✓ tibble 3.0.4 ✓ dplyr 1.0.2  
## ✓ tidyr 1.1.2 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.5.0

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(dplyr)  
library(ggplot2)  
#https://api.census.gov/data/2015/acs/acs1/variables.html  
  
  
#1  
census\_api\_key("95c61c9b4cd8324855924d339aadefb48e9bee09")

## To install your API key for use in future sessions, run this function with `install = TRUE`.

install = TRUE  
  
#2a  
v1 <- load\_variables(2015, "acs5", cache = TRUE)  
v1

## # A tibble: 22,768 x 3  
## name label concept   
## <chr> <chr> <chr>   
## 1 B00001\_001 Estimate!!Total UNWEIGHTED SAMPLE COUNT OF THE PO…  
## 2 B00002\_001 Estimate!!Total UNWEIGHTED SAMPLE HOUSING UNITS   
## 3 B01001\_001 Estimate!!Total SEX BY AGE   
## 4 B01001\_002 Estimate!!Total!!Male SEX BY AGE   
## 5 B01001\_003 Estimate!!Total!!Male!!Under 5… SEX BY AGE   
## 6 B01001\_004 Estimate!!Total!!Male!!5 to 9 … SEX BY AGE   
## 7 B01001\_005 Estimate!!Total!!Male!!10 to 1… SEX BY AGE   
## 8 B01001\_006 Estimate!!Total!!Male!!15 to 1… SEX BY AGE   
## 9 B01001\_007 Estimate!!Total!!Male!!18 and … SEX BY AGE   
## 10 B01001\_008 Estimate!!Total!!Male!!20 years SEX BY AGE   
## # … with 22,758 more rows

ca <- get\_acs(geography = "county",   
 variables = c(medincome = "B01001A\_011E"),   
 state = "CA",   
 year = 2015)

## Getting data from the 2011-2015 5-year ACS

ca

## # A tibble: 58 x 5  
## GEOID NAME variable estimate moe  
## <chr> <chr> <chr> <dbl> <dbl>  
## 1 06001 Alameda County, California B01001A\_011 51644 667  
## 2 06003 Alpine County, California B01001A\_011 50 26  
## 3 06005 Amador County, California B01001A\_011 1809 72  
## 4 06007 Butte County, California B01001A\_011 9962 128  
## 5 06009 Calaveras County, California B01001A\_011 1927 74  
## 6 06011 Colusa County, California B01001A\_011 1147 79  
## 7 06013 Contra Costa County, California B01001A\_011 42756 605  
## 8 06015 Del Norte County, California B01001A\_011 1629 90  
## 9 06017 El Dorado County, California B01001A\_011 8609 141  
## 10 06019 Fresno County, California B01001A\_011 34979 714  
## # … with 48 more rows

#2b  
ca1 <- ca%>%  
 filter(estimate>=30000)%>%  
 arrange(desc(estimate))  
ca1

## # A tibble: 13 x 5  
## GEOID NAME variable estimate moe  
## <chr> <chr> <chr> <dbl> <dbl>  
## 1 06037 Los Angeles County, California B01001A\_011 375435 2332  
## 2 06073 San Diego County, California B01001A\_011 150891 1008  
## 3 06059 Orange County, California B01001A\_011 126819 1152  
## 4 06065 Riverside County, California B01001A\_011 92346 1004  
## 5 06071 San Bernardino County, California B01001A\_011 80925 1160  
## 6 06085 Santa Clara County, California B01001A\_011 63036 879  
## 7 06067 Sacramento County, California B01001A\_011 56066 553  
## 8 06001 Alameda County, California B01001A\_011 51644 667  
## 9 06013 Contra Costa County, California B01001A\_011 42756 605  
## 10 06075 San Francisco County, California B01001A\_011 42307 542  
## 11 06029 Kern County, California B01001A\_011 42121 575  
## 12 06111 Ventura County, California B01001A\_011 41155 557  
## 13 06019 Fresno County, California B01001A\_011 34979 714

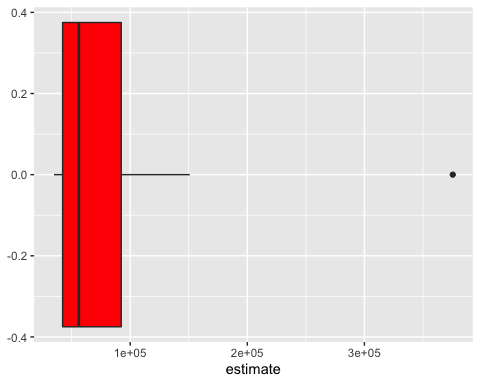
mutate(ca1, variable = recode(variable, "B01001A\_011"= "medianincome"))->ca2  
ca2

## # A tibble: 13 x 5  
## GEOID NAME variable estimate moe  
## <chr> <chr> <chr> <dbl> <dbl>  
## 1 06037 Los Angeles County, California medianincome 375435 2332  
## 2 06073 San Diego County, California medianincome 150891 1008  
## 3 06059 Orange County, California medianincome 126819 1152  
## 4 06065 Riverside County, California medianincome 92346 1004  
## 5 06071 San Bernardino County, California medianincome 80925 1160  
## 6 06085 Santa Clara County, California medianincome 63036 879  
## 7 06067 Sacramento County, California medianincome 56066 553  
## 8 06001 Alameda County, California medianincome 51644 667  
## 9 06013 Contra Costa County, California medianincome 42756 605  
## 10 06075 San Francisco County, California medianincome 42307 542  
## 11 06029 Kern County, California medianincome 42121 575  
## 12 06111 Ventura County, California medianincome 41155 557  
## 13 06019 Fresno County, California medianincome 34979 714

#2c  
ca3<- ca2%>%  
 filter(moe== 667, estimate== 51644)  
ca3

## # A tibble: 1 x 5  
## GEOID NAME variable estimate moe  
## <chr> <chr> <chr> <dbl> <dbl>  
## 1 06001 Alameda County, California medianincome 51644 667

#alameda county  
  
  
#2d  
ggplot(data = ca2) +  
 geom\_boxplot(mapping = aes(x=estimate), fill = "red")



#2e  
ca1 %>%  
 mutate(NAME = gsub(" County, California", "", NAME)) %>%  
 ggplot(aes(x = estimate, y = reorder(NAME, estimate))) +  
 geom\_errorbarh(aes(xmin = estimate - moe, xmax = estimate + moe)) +  
 geom\_point(color = "blue", size = 3) +  
 labs(title = "Median Income for White Males by County",  
 subtitle = "2014-2018 American Community Survey",  
 y = "",  
 x = "ACS estimate (bars represent margin of error)")

