613 hw9

2021/4/10

library(tidycensus)  
library(tidyverse)

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

## ✓ ggplot2 3.3.2 ✓ purrr 0.3.4  
## ✓ tibble 3.0.3 ✓ 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)  
  
census\_api\_key("42f1de8c7e6161facaf5889926b6659f0418b101")

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

install = TRUE  
  
y2015 <- load\_variables(2015, "acs5", cache = TRUE)  
View(y2015)  
图片包含 文本

描述已自动生成  
#2a  
data <- get\_acs(geography = "county",   
 variables = c(medincome = "B01001A\_011"),   
 state = 06,   
 year = 2015)

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

data

# A tibble: 58 x 5

GEOID NAME variable estimate moe

*<chr>* *<chr>* *<chr>* *<dbl>* *<dbl>*

1 06001 Alameda County, California medincome 51644 667

2 06003 Alpine County, California medincome 50 26

3 06005 Amador County, California medincome 1809 72

4 06007 Butte County, California medincome 9962 128

5 06009 Calaveras County, California medincome 1927 74

6 06011 Colusa County, California medincome 1147 79

7 06013 Contra Costa County, California medincome 42756 605

8 06015 Del Norte County, California medincome 1629 90

9 06017 El Dorado County, California medincome 8609 141

10 06019 Fresno County, California medincome 34979 714

# ... with 48 more rows

#2b  
data2<-data %>%  
 filter(estimate>3000)%>%  
 arrange(desc(estimate))  
data2

# A tibble: 41 x 5

GEOID NAME variable estimate moe

*<chr>* *<chr>* *<chr>* *<dbl>* *<dbl>*

1 06037 Los Angeles County, California medincome 375435 2332

2 06073 San Diego County, California medincome 150891 1008

3 06059 Orange County, California medincome 126819 1152

4 06065 Riverside County, California medincome 92346 1004

5 06071 San Bernardino County, California medincome 80925 1160

6 06085 Santa Clara County, California medincome 63036 879

7 06067 Sacramento County, California medincome 56066 553

8 06001 Alameda County, California medincome 51644 667

9 06013 Contra Costa County, California medincome 42756 605

10 06075 San Francisco County, California medincome 42307 542

# ... with 31 more rows

#2c  
data2%>%  
 filter(estimate==51644, moe==667)

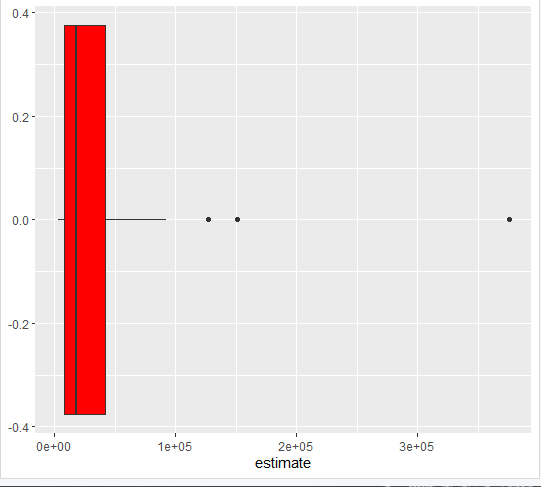
# A tibble: 1 x 5

GEOID NAME variable estimate moe

*<chr>* *<chr>* *<chr>* *<dbl>* *<dbl>*

1 06001 Alameda County, California medincome 51644 667

#2d  
ggplot(data = data2) +  
 geom\_boxplot(mapping = aes(x=estimate), fill="red")



#2e  
data2 %>%  
 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 = "Household income by county in Vermont",  
 subtitle = "2014-2018 American Community Survey",  
 y = "",  
 x = "ACS estimate (bars represent margin of error)")

