Lab webscraping 2

Lin Pin Tzu (Ruby)

2022-07-12

1 a} Show and use a census API key that gives you access to the Census Bureau data. Do not use my API key, use and show your own key.

```
# census_api_key("4009f73e21670e9fb8801c8067991ecb855c1632", install=TRUE)
census_api_key("4009f73e21670e9fb8801c8067991ecb855c1632",overwrite=TRUE)
```

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

```
# For line 19 I can not knit that is why I use #
```

b) Using the link provided in your notes, secure a Census Bureau API key. Run the census code that requires usage of the API key and then use R coding to produce a table that shows the totals for Asian Males for ages 67 to 69 by state for the year 2000. The identifier code is P012D021

```
age6769 <- get_decennial(geography = "state",
                         variables = "P012D021",
                         year = 2000)
```

```
## Getting data from the 2000 decennial Census
```

```
## Using Census Summary File 1
```

```
age6769
```

```
## # A tibble: 52 x 4
##    GEOID NAME                 variable value
##    <chr> <chr>                <chr>    <dbl>
##  1 01    Alabama              P012D021   118
##  2 02    Alaska               P012D021   118
##  3 04    Arizona              P012D021   547
##  4 05    Arkansas             P012D021    98
##  5 06    California           P012D021 28524
##  6 08    Colorado             P012D021   479
##  7 09    Connecticut          P012D021   391
##  8 10    Delaware             P012D021    80
##  9 11    District of Columbia P012D021    81
## 10 12    Florida              P012D021  1601
```

```
## # ... with 42 more rows
```

c) Show and use R code to find the mean, median, ,max, min, Q1, and Q3 for the median ages.

```
mean(age6769$value)
```

```
## [1] 1299.192
```

```
median(age6769$value)
```

```
## [1] 227
```

```
which.max(age6769$value) # the row shows the max
```

```
## [1] 5
```

```
which.min(age6769$value)# the row shows the min
```

```
## [1] 42
```

```
IQR(age6769$value)
```

```
## [1] 664.25
```

```
summary(age6769$value)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    15.00    80.75   227.00  1299.19   745.00 28524.00
```

d) Show and use R code (tidyverse/dplyr) coding to find the top ten states with highest populations of Asian Males whose ages are between 67 and 69.

```
age6769 %>%
  arrange(desc(value)) ->top10
head(top10,10)
```

```
## # A tibble: 10 x 4
##    GEOID NAME       variable value
##    <chr> <chr>      <chr>    <dbl>
##  1 06    California P012D021 28524
##  2 36    New York   P012D021  7044
##  3 15    Hawaii     P012D021  6478
##  4 48    Texas      P012D021  2685
##  5 34    New Jersey P012D021  2494
##  6 17    Illinois   P012D021  2294
##  7 53    Washington P012D021  1856
##  8 12    Florida    P012D021  1601
##  9 51    Virginia   P012D021  1443
## 10 24    Maryland   P012D021  1437
```

2 a) Using the link provided in your notes, use and show R coding to produce a table that shows the median ages for Hispanic or Latino women for the year 2010 (Hint: the 8 character variable code starts with characters P013. Search in your table to get the other four characters. (Hint: Ctrl F speeds up the search process)

```
year2010 <- get_decennial(geography = "state",
                          variables = "P013H003",
                          year = 2010)
```

```
## Getting data from the 2010 decennial Census
```

```
## Using Census Summary File 1
```

```
year2010
```

```
## # A tibble: 52 x 4
##    GEOID NAME        variable value
##    <chr> <chr>       <chr>    <dbl>
##  1 01    Alabama     P013H003  23.7
##  2 02    Alaska      P013H003  24.7
##  3 04    Arizona     P013H003  26
##  4 05    Arkansas    P013H003  22.7
##  5 06    California  P013H003  27.7
##  6 22    Louisiana   P013H003  28.8
##  7 21    Kentucky    P013H003  23.1
##  8 08    Colorado    P013H003  26.8
##  9 09    Connecticut P013H003  28.4
## 10 10    Delaware    P013H003  24.7
## # ... with 42 more rows
```

b) Show and use R code to find the mean, median, ,max, min, Q1, and Q3 for the median ages.

```
mean(year2010$value)
```

```
## [1] 25.63077
```

```
median(year2010$value)
```

```
## [1] 24.85
```

```
which.max(year2010$value) # the row of the max
```

```
## [1] 52
```

```
which.min(year2010$value)# the row of the min
```

```
## [1] 42
```

```
IQR(year2010$value)
```

```
## [1] 3.575
```

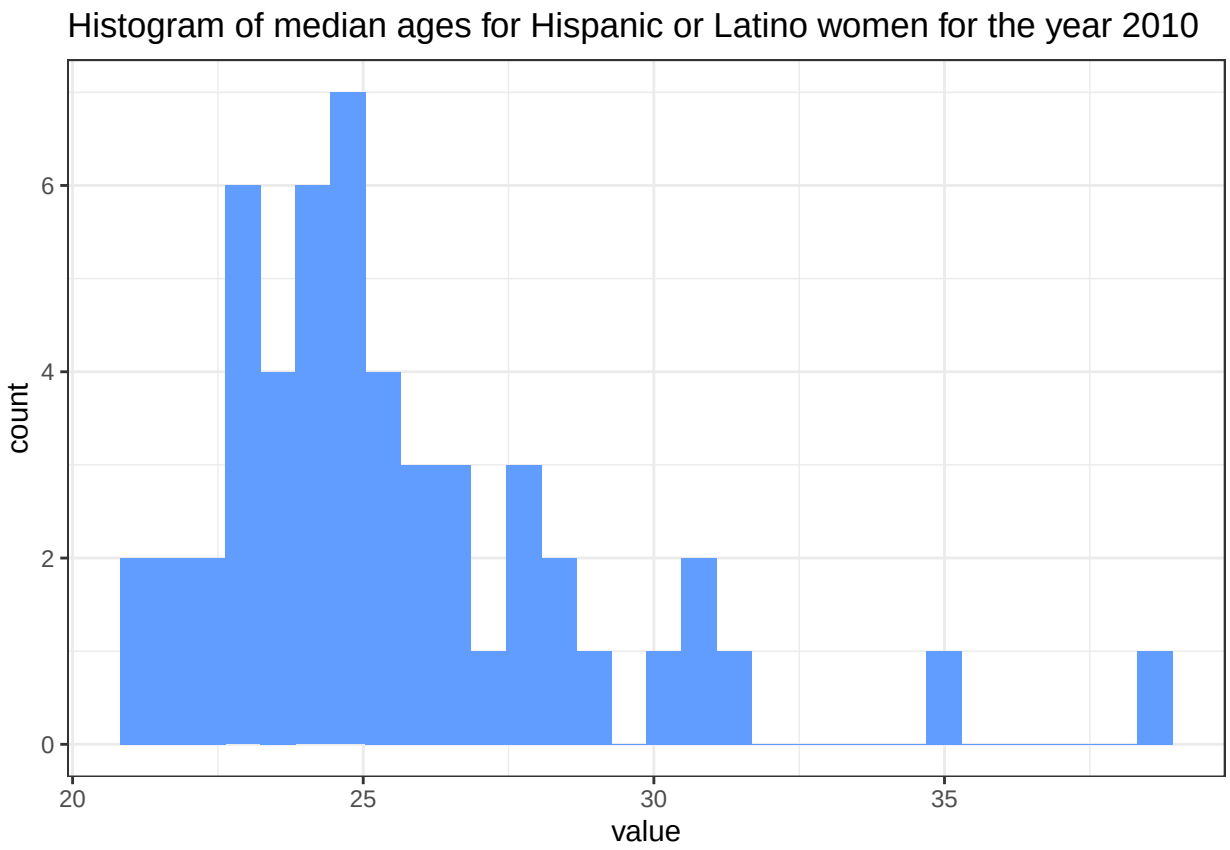
```
summary(year2010$value)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   21.10   23.38   24.85   25.63   26.95   38.60
```

c) Use ggplot coding to produce a Histogram of vertical orientation for the median ages for the table that you produced for 2a.

```
ggplot(year2010,mapping=aes(x=value))+
  geom_histogram(fill="#619CFF")+
  ggtitle("Histogram of median ages for Hispanic or Latino women for the year 2010")+
theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



d) Produce a coding chunk using dplyr functions to generate a table that gives results for values that are greater than or equal to a median age of 25.

```
year2010 %>%
  filter(value>=25)%>%
  print(n=24)
```

```
## # A tibble: 24 x 4
```

```
##    GEOID NAME                 variable value
##    <chr> <chr>                <chr>    <dbl>
##  1 04    Arizona              P013H003  26
##  2 06    California           P013H003  27.7
##  3 22    Louisiana            P013H003  28.8
##  4 08    Colorado             P013H003  26.8
##  5 09    Connecticut          P013H003  28.4
##  6 11    District of Columbia P013H003  30.1
##  7 12    Florida              P013H003  35.1
##  8 15    Hawaii               P013H003  25.5
##  9 17    Illinois             P013H003  26.5
## 10 24    Maryland             P013H003  28.1
## 11 25    Massachusetts        P013H003  27.4
## 12 32    Nevada               P013H003  26.2
## 13 33    New Hampshire        P013H003  25
## 14 34    New Jersey           P013H003  30.9
## 15 35    New Mexico           P013H003  30.7
## 16 36    New York             P013H003  31.6
## 17 42    Pennsylvania         P013H003  25.6
## 18 44    Rhode Island         P013H003  26.5
## 19 48    Texas                P013H003  27.6
## 20 50    Vermont              P013H003  25.1
## 21 51    Virginia             P013H003  27.6
## 22 54    West Virginia        P013H003  25.9
## 23 56    Wyoming              P013H003  25.5
## 24 72    Puerto Rico          P013H003  38.6
```