US Counties - Income, Education Analysis

Daniel Williams

2023-01-27

Loading packages & data

```
setwd("~/Data Science/Datasets")
data <- read.csv("~/Data Science/Datasets/uscountydata.csv")
pop_data <- read.csv("~/Data Science/Datasets/us_pop_by_state.csv")
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
## Warning: package 'tibble' was built under R version 4.1.3
```

```
## Warning: package 'tidyr' was built under R version 4.1.3
```

```
## Warning: package 'readr' was built under R version 4.1.3
```

```
## Warning: package 'purrr' was built under R version 4.1.3
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
## Warning: package 'stringr' was built under R version 4.1.3
```

```
## Warning: package 'forcats' was built under R version 4.1.3
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Previewing Data in different ways

```
head(data)
```

```
##   county_FIPS state         county per_capita_personal_income_2019
## 1       51013    VA  Arlington, VA                           97629
## 2       35028    NM Los Alamos, NM                           72366
## 3        8013    CO    Boulder, CO                           79698
## 4       24027    MD     Howard, MD                           78013
## 5       36061    NY   New York, NY                          173525
## 6       51107    VA    Loudoun, VA                           82681
##   per_capita_personal_income_2020 per_capita_personal_income_2021
## 1                          100687                          107603
## 2                           75949                           81306
## 3                           83173                           89593
## 4                           82041                           86380
## 5                          175327                          195543
## 6                           85568                           90254
##   associate_degree_numbers_2016_2020 bachelor_degree_numbers_2016_2020
## 1                              19573                             132394
## 2                               2766                               9098
## 3                              45834                             135876
## 4                              42538                             136792
## 5                             167960                             777483
## 6                              52006                             160853
##   associate_degree_percentage_2016_2020 bachelor_degree_percentage_2015_2019
## 1                                 11.21                                75.84
## 2                                 20.54                                67.56
## 3                                 21.24                                62.97
## 4                                 19.49                                62.67
## 5                                 13.43                                62.18
## 6                                 19.92                                61.62
```

```
str(data)
```

```
## 'data.frame':    3006 obs. of  10 variables:
##  $ county_FIPS                          : int  51013 35028 8013 24027 36061 51107 8097 371
35 47187 6041 ...
##  $ state                                : chr  "VA" "NM" "CO" "MD" ...
##  $ county                               : chr  "Arlington, VA" "Los Alamos, NM" "Boulder,
CO" "Howard, MD" ...
##  $ per_capita_personal_income_2019      : int  97629 72366 79698 78013 173525 82681 183241
66970 95806 139891 ...
##  $ per_capita_personal_income_2020      : int  100687 75949 83173 82041 175327 85568 18759
5 69593 99408 148419 ...
##  $ per_capita_personal_income_2021      : int  107603 81306 89593 86380 195543 90254 19893
9 74994 107698 164118 ...
##  $ associate_degree_numbers_2016_2020   : int  19573 2766 45834 42538 167960 52006 2932 17
907 32995 44228 ...
##  $ bachelor_degree_numbers_2016_2020    : int  132394 9098 135876 136792 777483 160853 845
9 55545 91780 114604 ...
##  $ associate_degree_percentage_2016_2020: num  11.2 20.5 21.2 19.5 13.4 ...
##  $ bachelor_degree_percentage_2015_2019 : num  75.8 67.6 63 62.7 62.2 ...
```

```
View(data)
```

Gaining simple information about the data set

```
n_distinct(data$state)
```

```
## [1] 49
```

```
n_distinct(data$county)
```

```
## [1] 3006
```

Simple dplyr manipulation

```
average_income_data <-
  data %>%
  transmute(state, county, av_income = rowMeans(select(data,starts_with('per')))) %>%
  arrange(desc(av_income)) %>%
  filter(state %in% c('FL', 'CA')) %>%
  slice_max(av_income,n=3)
```

Group & Summarise data

```
average_state_income <-
  data %>%
  select(state, per_capita_personal_income_2021) %>%
  group_by(state) %>%
  summarize(mean_income = mean(per_capita_personal_income_2021)) %>%
  arrange(desc(mean_income))
```

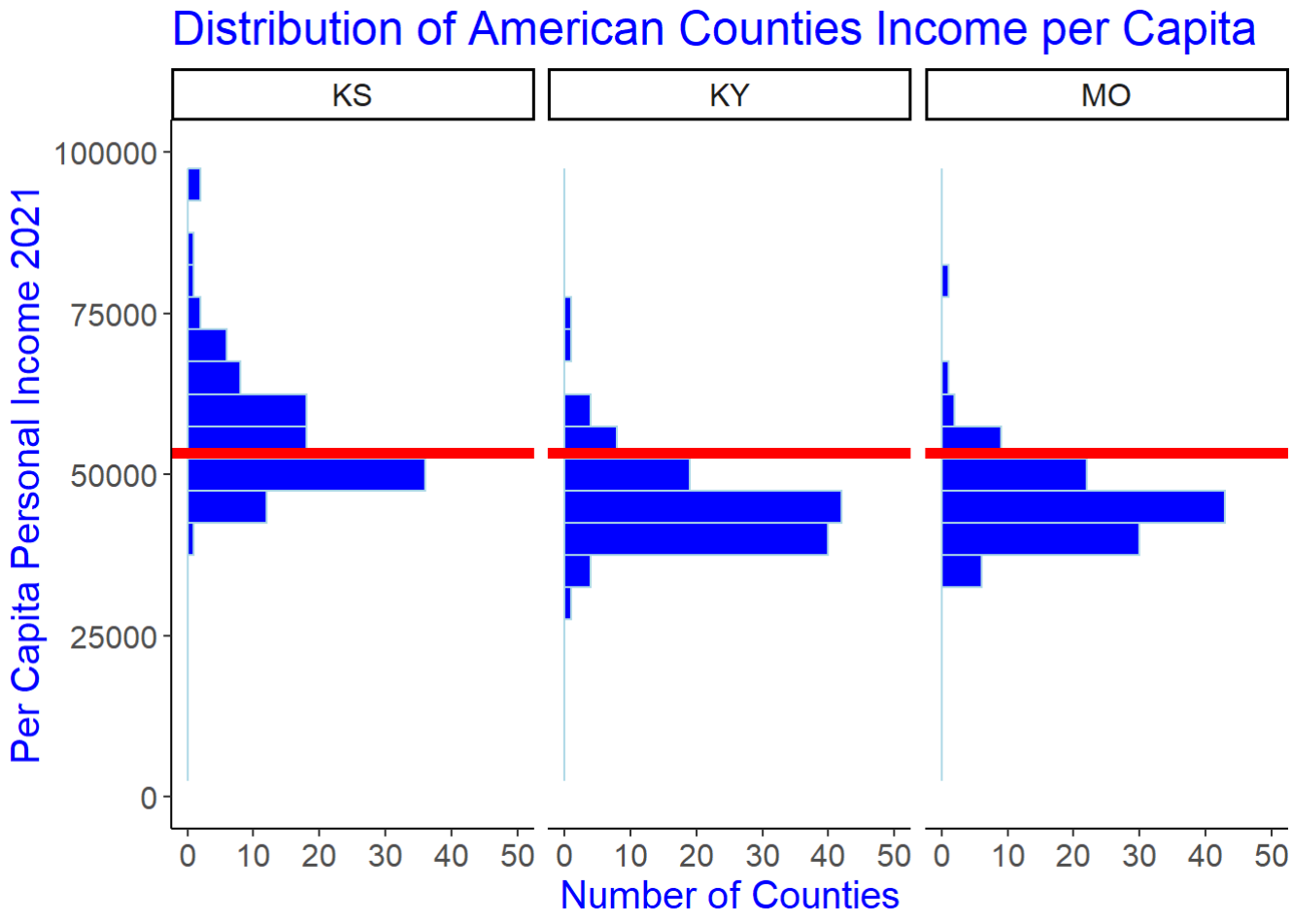
Country average statistic

```
US_av_inc <- data %>%
    pull(per_capita_personal_income_2021) %>%
    mean()
```

ggplot example

```
Three_States <- data %>% filter(state %in% c('KY', 'MO', 'KS'))

# ggplot2
ggplot(Three_States, aes(per_capita_personal_income_2021))+
  geom_histogram(binwidth = 5000, color = 'lightblue', fill = 'blue')+
  xlim(0,100000)+
  ylim(0,50)+
  geom_vline(xintercept = US_av_inc, color = 'red', size = 2)+
  labs(x='Per Capita Personal Income 2021', y = 'Number of Counties', title ='Distribution of
American Counties Income per Capita')+
  annotate('text',x=30,y=60000,label='Country_Average')+
  theme_classic()+
  coord_flip()+
  facet_wrap(~state)+
  theme(text = element_text(family = 'arial', size = 15, color = 'blue'))
```

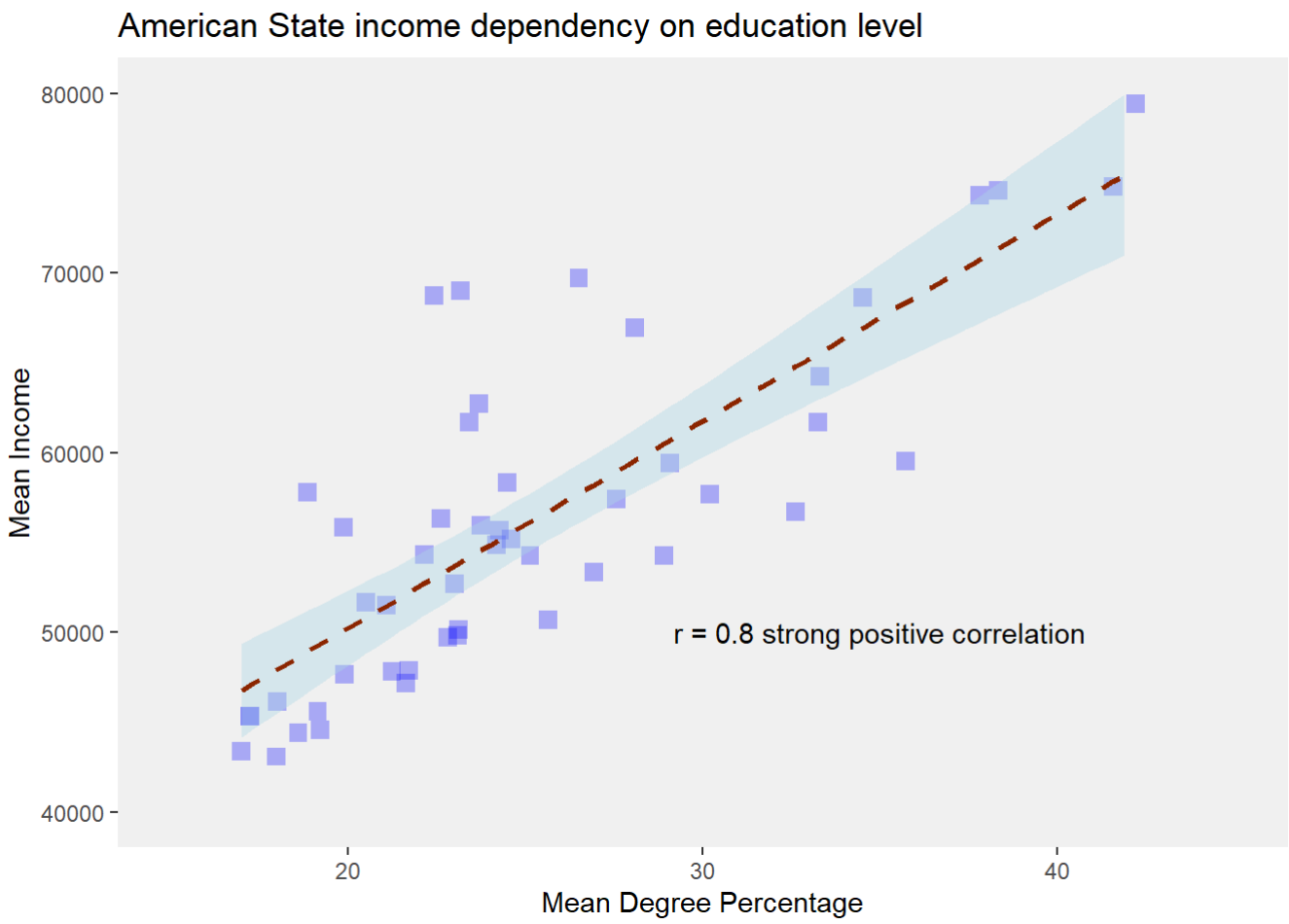


Second ggplot2 example

```
degree_vs_income <- data %>%
                select(state, per_capita_personal_income_2021, bachelor_degree_percentage_201
5_2019) %>%
                group_by(state) %>%
                summarize(mean_deg_perc = mean(bachelor_degree_percentage_2015_2019),
                          mean_income = mean(per_capita_personal_income_2021))

ggplot(degree_vs_income, aes(x=mean_deg_perc, y=mean_income))+
  geom_jitter(size=3,alpha=0.3, color='blue', shape = 15)+
  geom_smooth(method=lm, linetype ='dashed', color = 'orangered4', fill = 'lightblue')+
  xlim(15,45)+
  ylim(40000,80000)+
  annotate('text',x=35,y=50000,label='r = 0.8 strong positive correlation')+
  labs(x='Mean Degree Percentage',y='Mean Income', title='American State income dependency on
education level')+
  theme(panel.background =  element_rect(fill = 'grey94'),
        panel.grid = element_blank()
        )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
cor(degree_vs_income$mean_deg_perc, degree_vs_income$mean_income)
```

```
## [1] 0.8024959
```

Joining data

```
joined_pop_income <- data %>%
                     left_join(pop_data, by = c('state' = 'state_code')) %>%
                     select(-percent_of_total, -rank)
```

Finding outliers

```
IQR <- quantile(joined_pop_income$per_capita_personal_income_2021, 0.75) - quantile(joined_po
p_income$per_capita_personal_income_2021, 0.25)
Lower_threshold <- quantile(joined_pop_income$per_capita_personal_income_2021, 0.25) - 1.5 *
IQR
Upper_threshold <- quantile(joined_pop_income$per_capita_personal_income_2021, 0.75) + 1.5 *
IQR

outliers <- joined_pop_income %>%
            filter(per_capita_personal_income_2021 < Lower_threshold | per_capita_personal_in
come_2021 > Upper_threshold) %>%
            select(county, state.y, per_capita_personal_income_2021, X2020_census) %>%
            arrange(desc(per_capita_personal_income_2021))
outliers
```

```
##                     county       state.y per_capita_personal_income_2021
## 1                Teton, WY       Wyoming                           318297
## 2               Pitkin, CO      Colorado                           198939
## 3             New York, NY      New York                           195543
## 4               Summit, UT          Utah                           183972
## 5                Marin, CA    California                           164118
## 6         San Francisco, CA    California                           160749
## 7            San Mateo, CA    California                           160485
## 8   Bristol Bay Borough, AK        Alaska                           155155
## 9           Santa Clara, CA    California                           138724
## 10               Union, SD  South Dakota                           134811
## 11              Blaine, ID         Idaho                           134722
## 12            Fairfield, CT   Connecticut                           127391
## 13              Midland, TX         Texas                           126738
## 14            Glasscock, TX         Texas                           124963
## 15              Sherman, TX         Texas                           122945
## 16          Westchester, NY      New York                           119705
## 17              Collier, FL       Florida                           117984
## 18               Loving, TX         Texas                           115158
## 19             Somerset, NJ    New Jersey                           113975
## 20            Goochland, VA      Virginia                           110781
## 21             McMullen, TX         Texas                           110273
## 22           San Miguel, CO      Colorado                           109613
## 23              Monroe, FL       Florida                           109136
## 24                King, WA    Washington                           108212
## 25       Denali Borough, AK        Alaska                           107916
## 26               Morris, NJ    New Jersey                           107767
## 27           Williamson, TN     Tennessee                           107698
## 28            Arlington, VA      Virginia                           107603
## 29                Sully, SD  South Dakota                           106241
## 30              Norfolk, MA Massachusetts                           104059
## 31             Lipscomb, TX         Texas                           102569
## 32              Martin, FL       Florida                           102273
## 33              Fulton, GA       Georgia                           102074
## 34           Nantucket, MA Massachusetts                           101902
## 35              Borden, TX         Texas                           101470
## 36           Palm Beach, FL       Florida                           100627
## 37            Hunterdon, NJ    New Jersey                           100288
## 38              Alameda, CA    California                            99746
## 39              Nassau, NY      New York                            99597
## 40         Contra Costa, CA    California                            99312
## 41               Denver, CO      Colorado                            99133
## 42              Suffolk, MA Massachusetts                            98644
## 43            Middlesex, MA Massachusetts                            98523
## 44                Routt, CO      Colorado                            98371
## 45              Bergen, NJ    New Jersey                            97343
## 46               Eagle, CO      Colorado                            97255
## 47              Kendall, TX         Texas                            96628
## 48              Chester, PA  Pennsylvania                            95483
## 49         Indian River, FL       Florida                            95109
## 50              Stanton, KS        Kansas                            94353
## 51                Dundy, NE      Nebraska                            93788
## 52              Greeley, KS        Kansas                            93667
## 53           Shackelford, TX         Texas                            93075
## 54            Montgomery, MD      Maryland                            92740
```

```
## 55              Ozaukee, WI     Wisconsin                    92719
## 56             Billings, ND  North Dakota                    92571
## 57             Monmouth, NJ    New Jersey                    92119
## 58                 King, TX         Texas                    91930
## 59               Benton, AR      Arkansas                    91687
## 60                  Napa, CA    California                    90608
## 61               Loudoun, VA      Virginia                    90254
## 62                 Hayes, NE      Nebraska                    90144
## 63             New Kent, VA      Virginia                    90086
## 64               Boulder, CO      Colorado                    89593
## 65                  Lake, IL      Illinois                    89025
## 66             Cavalier, ND  North Dakota                    88700
## 67           Montgomery, PA  Pennsylvania                    88671
## 68           Rockingham, NH New Hampshire                    88303
## 69       Haines Borough, AK        Alaska                    88152
## 70                Perkins, NE      Nebraska                    87847
## 71                Douglas, CO      Colorado                    87841
## 72                Bristol, RI  Rhode Island                    87810
## 73                Douglas, NV        Nevada                    87214
## 74              San Juan, WA    Washington                    87038
## 75                Potter, SD  South Dakota                    87031
## 76                Summit, CO      Colorado                    86390
## 77                Howard, MD      Maryland                    86380
## 78                 Dukes, MA Massachusetts                    86341
## 79                 Jones, SD  South Dakota                    86307
## 80                Steele, ND  North Dakota                    86280
## 81              Hamilton, IN       Indiana                    85814
## 82            Santa Cruz, CA    California                    85554
## 83              Hennepin, MN     Minnesota                    85505
## 84                DuPage, IL      Illinois                    85498
## 85                 Boone, IN       Indiana                    84985
## 86                Talbot, MD      Maryland                    84695
## 87               Johnson, KS        Kansas                    84535
## 88                Martin, TX         Texas                    84121
## 89               Newport, RI  Rhode Island                    84054
## 90            Cumberland, IL      Illinois                    84035
## 91               Wheeler, NE      Nebraska                    83722
## 92              Delaware, OH          Ohio                    83603
## 93           Barnstable, MA Massachusetts                    83191
## 94                 Bucks, PA  Pennsylvania                    83011
## 95                 Wells, ND  North Dakota                    82108
## 96              Davidson, TN     Tennessee                    82087
## 97              Waukesha, WI     Wisconsin                    82032
## 98                Castro, TX         Texas                    81900
## 99             St. Louis, MO      Missouri                    81829
## 100               Travis, TX         Texas                    81708
## 101              Suffolk, NY      New York                    81309
## 102           Los Alamos, NM    New Mexico                    81306
## 103               Orange, CA    California                    81034
## 104               Sonoma, CA    California                    81006
## 105                 Dunn, ND  North Dakota                    80976
## 106              Oakland, MI      Michigan                    80962
## 107             McIntosh, ND  North Dakota                    80675
## 108               Towner, ND  North Dakota                    80665
## 109               Carver, MN     Minnesota                    80562
## 110                 Rock, NE      Nebraska                    80527
```

```
## 111             Rawlins, KS        Kansas                            80389
## 112               Slope, ND  North Dakota                            80270
## 113               Faulk, SD  South Dakota                            80152
## 114           St. Johns, FL       Florida                            80062
## 115            Plymouth, MA Massachusetts                            79817
## 116            Campbell, SD  South Dakota                            79723
## 117               Irion, TX         Texas                            78988
##     X2020_census
## 1         576851
## 2        5773714
## 3       20201249
## 4        3205958
## 5       39538223
## 6       39538223
## 7       39538223
## 8         733391
## 9       39538223
## 10        886667
## 11       1839106
## 12       3605944
## 13      29145505
## 14      29145505
## 15      29145505
## 16      20201249
## 17      21538187
## 18      29145505
## 19       9288994
## 20       8631393
## 21      29145505
## 22       5773714
## 23      21538187
## 24       7705281
## 25        733391
## 26       9288994
## 27       6910840
## 28       8631393
## 29        886667
## 30       7029917
## 31      29145505
## 32      21538187
## 33      10711908
## 34       7029917
## 35      29145505
## 36      21538187
## 37       9288994
## 38      39538223
## 39      20201249
## 40      39538223
## 41       5773714
## 42       7029917
## 43       7029917
## 44       5773714
## 45       9288994
## 46       5773714
## 47      29145505
## 48      13002700
```

```
## 49      21538187
## 50       2937880
## 51       1961504
## 52       2937880
## 53      29145505
## 54       6177224
## 55       5893718
## 56        779094
## 57       9288994
## 58      29145505
## 59       3011524
## 60      39538223
## 61       8631393
## 62       1961504
## 63       8631393
## 64       5773714
## 65      12801989
## 66        779094
## 67      13002700
## 68       1377529
## 69        733391
## 70       1961504
## 71       5773714
## 72       1097379
## 73       3104614
## 74       7705281
## 75        886667
## 76       5773714
## 77       6177224
## 78       7029917
## 79        886667
## 80        779094
## 81       6785528
## 82      39538223
## 83       5706494
## 84      12801989
## 85       6785528
## 86       6177224
## 87       2937880
## 88      29145505
## 89       1097379
## 90      12801989
## 91       1961504
## 92      11799448
## 93       7029917
## 94      13002700
## 95        779094
## 96       6910840
## 97       5893718
## 98      29145505
## 99       6154913
## 100     29145505
## 101     20201249
## 102      2117522
## 103     39538223
## 104     39538223
```

```
## 105       779094
## 106     10077331
## 107       779094
## 108       779094
## 109      5706494
## 110      1961504
## 111      2937880
## 112       779094
## 113       886667
## 114     21538187
## 115      7029917
## 116       886667
## 117     29145505
```