

Maryland Poverty Level - Data Cleaning

Lebo Sango

2024-04-16

1.1 SAIPE

```
MD_raw = read.csv("/Users/lebsan/Documents/STAT 5084 - Time Series/County Level Project/Gathering Data")

# Data Transformation

MD_transf = MD_raw %>% rename(county_pov = Poverty.Universe, FIPS = ID,
                              poverty_pop = Number.in.Poverty) %>%
  select(Year, FIPS, county_pov, poverty_pop, Name) %>%
  mutate(poverty_pop = as.numeric(stringr::str_remove_all(poverty_pop, ","))) %>%
  mutate(county_pov = as.numeric(stringr::str_remove_all(county_pov, ","))) %>%
  as_tsibble(index = Year, key = c(FIPS, Name)) %>%
  na.omit() %>%
  filter(FIPS != 0) %>%
  filter(FIPS != 24000) # removed the zero values that correspond to the United
  #States in FIPS

# ----- This code is for part I

MD_SAIPE = MD_transf %>% select(FIPS, year = Year, Name, poverty_pop, county_pov) %>%
  filter(FIPS %in% c(24031, 24033, 24005, 24003, 24510, 24027, 24021, 24025, 24013))

#-----

# The state of Maryland counts 24 counties
MD_transf %>% count(n_distinct(FIPS))

## # A tibble: 1 x 2
##   'n_distinct(FIPS)'      n
##               <int> <int>
## 1                 24   600

# The largest county by population is the 24031 or Montgomery County
MD_transf %>% arrange(desc(county_pov)) %>%
  head()

## # A tsibble: 6 x 5 [1Y]
## # Key:      FIPS, Name [1]
```

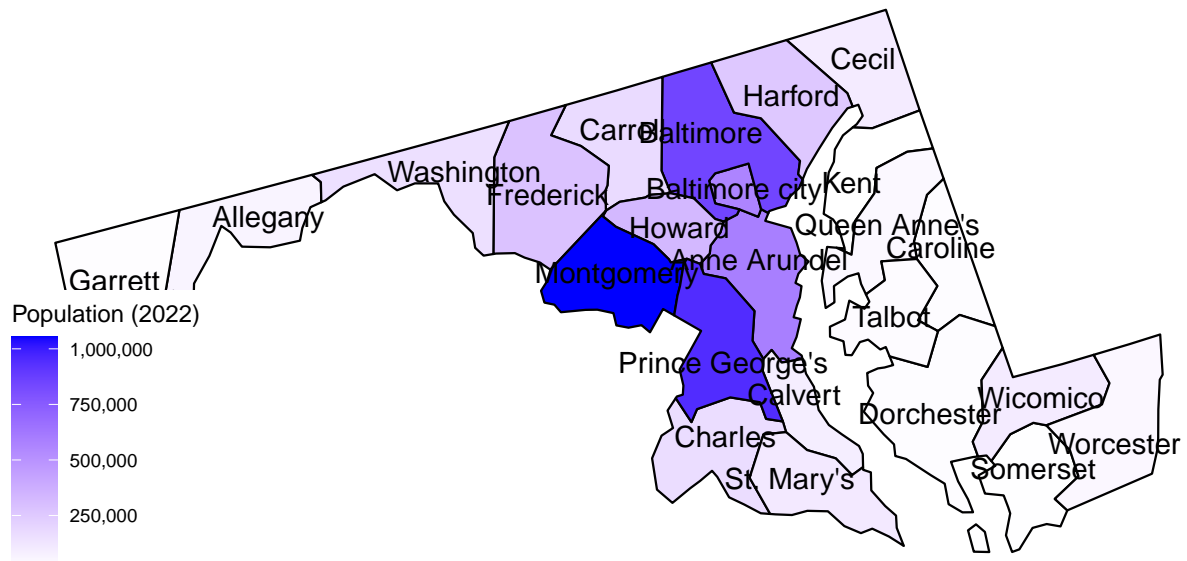
```
##   Year  FIPS county_pov poverty_pop Name
##   <int> <int>      <dbl>      <dbl> <chr>
## 1  2017 24031    1050411      73280 Montgomery County
## 2  2021 24031    1045767      88627 Montgomery County
## 3  2022 24031    1044091      82175 Montgomery County
## 4  2018 24031    1043190      72247 Montgomery County
## 5  2020 24031    1041551      70020 Montgomery County
## 6  2019 24031    1040481      75836 Montgomery County
```

The 9 largest county by population

```
MD_transf %>% group_by(FIPS, Name) %>%
  summarise(Largest_counties_by_pop = max(county_pov)) %>%
  arrange(desc(Largest_counties_by_pop)) %>% filter(Year==2022) %>%
  head(9)
```

```
## # A tibble: 9 x 4 [1Y]
## # Key:      FIPS, Name [9]
## # Groups:   FIPS [9]
##   FIPS Name                Year Largest_counties_by_pop
##   <int> <chr>              <int>      <dbl>
## 1 24031 Montgomery County    2022      1044091
## 2 24033 Prince George's County 2022      926950
## 3 24005 Baltimore County     2022      824597
## 4 24003 Anne Arundel County   2022      583597
## 5 24510 Baltimore city       2022      552361
## 6 24027 Howard County        2022      332813
## 7 24021 Frederick County     2022      282340
## 8 24025 Harford County       2022      262119
## 9 24013 Carroll County       2022      171688
```

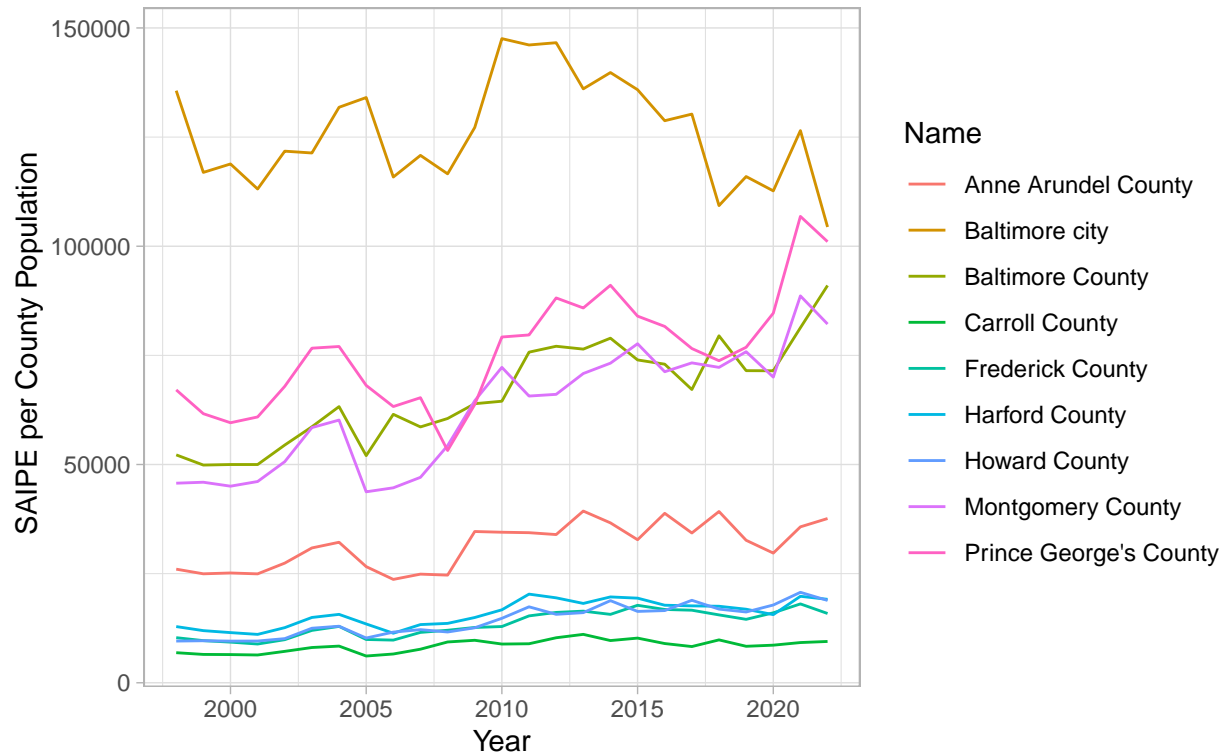
Map of counties of the state of Maryland



Time plot showing the level of poverty in the 9 largest counties

```
MD_transf %>% select(FIPS,Year,Name ,poverty_pop) %>%  
  filter(FIPS %in% c(24031,24033,24005,24003,24510,24027,24021, 24025, 24013)) %>%  
  ggplot(aes(x= Year, y=poverty_pop, group = FIPS, color=Name))+  
  geom_line()+  
  labs(x= "Year", y= " SAIPE per County Population",  
        title = " Census of Maryland Counties' Small Area Income and Poverty, SAIPE  
        ")+  
  theme_light()
```

Census of Maryland Counties' Small Area Income and Poverty, SAIPE



1.2 County SNAP Benefits

```
MD_snap = read.csv("/Users/lebsan/Documents/STAT 5084 - Time Series/County Level Project/Gathering Data")

# gsub removes the "X" next to year & # pivot longer aids to transpose year and
# values to columns

snap_MD = as_tibble(MD_snap) %>%
  pivot_longer(cols= X2021:X1989,
               names_to = "year",
               values_to = "SnapBenefits",
               values_drop_na = TRUE) %>%
  mutate(year = gsub("X", "", year))

# Converting year to date and merging state & county FIPS
# grepl allows to filter specifically by the string ending with

snap_MD1 = snap_MD %>% mutate(year = as.numeric(year)) %>%
  filter(grepl(" MD", Name)) %>%
  mutate(FIPS = paste(State.FIPS.code, County.FIPS.code ,sep = "")) %>%
  as_tsibble(index = year, key = c(Name, FIPS) )
```

```
# Time plot of the nine largest counties receiving SNAP benefits and converting
#Snap benefits to numeric
```

```
#-----
snap_MD2 = snap_MD1 %>% select(c(FIPS,year, Name, SnapBenefits)) %>%
  filter(Name %in% c("Montgomery County, MD", "Prince George's County, MD"
    , "Baltimore County, MD", "Anne Arundel County, MD",
    "Baltimore city, MD", "Howard County, MD",
    "Frederick County, MD", "Harford County, MD",
    "Carroll County, MD" ) ) %>%
  mutate(SnapBenefits = as.numeric(gsub("[\\,]", "", SnapBenefits)))

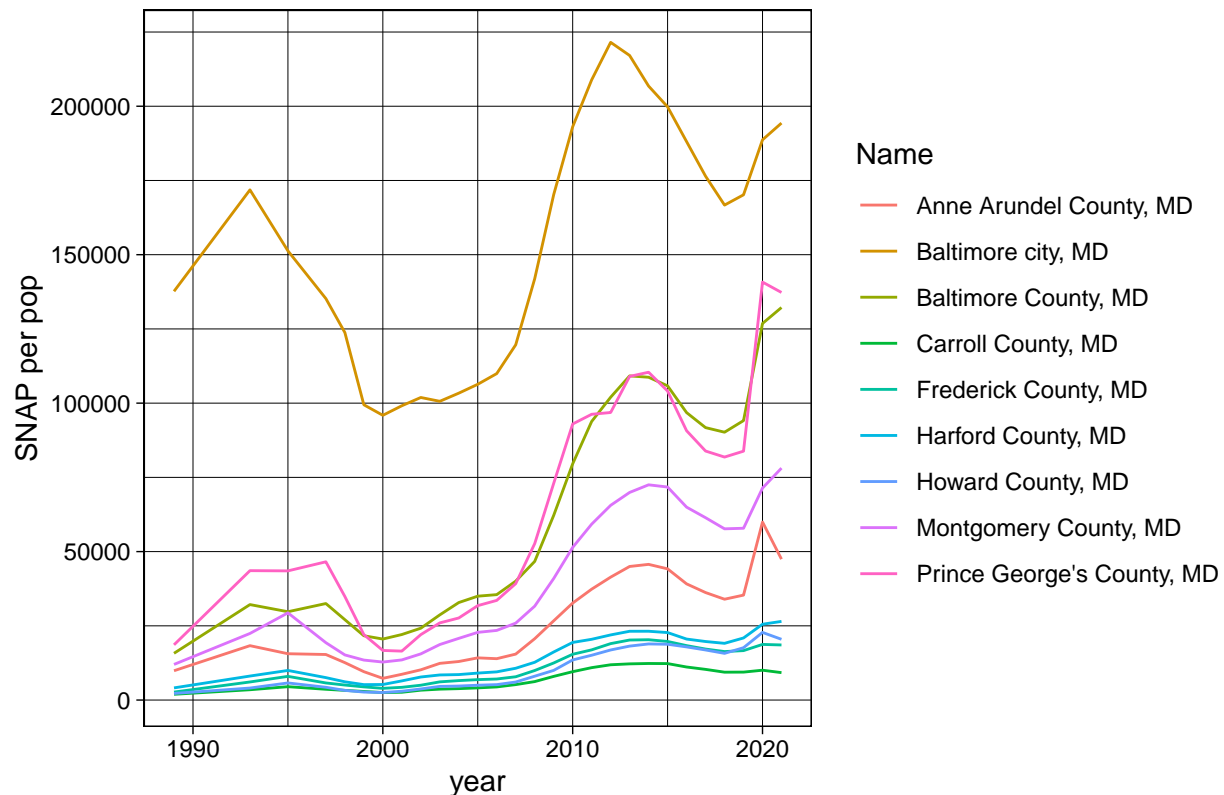
snap_MD2 %>% head()
```

```
## # A tibble: 6 x 4 [1Y]
## # Key:      FIPS, Name [1]
##   FIPS   year Name                SnapBenefits
##   <chr> <dbl> <chr>                <dbl>
## 1 24003  1989 Anne Arundel County, MD          9869
## 2 24003  1993 Anne Arundel County, MD         18317
## 3 24003  1995 Anne Arundel County, MD         15601
## 4 24003  1997 Anne Arundel County, MD         15342
## 5 24003  1998 Anne Arundel County, MD         12604
## 6 24003  1999 Anne Arundel County, MD          9559
```

```
# Plot of the nine Largest counties receiving SNAP
```

```
snap_MD2 %>% ggplot(aes(x=year, y=SnapBenefits, color = Name))+
  geom_line() + labs(title = " County Populaion Receiving SNAP Assistance",
    y="SNAP per pop")+theme_linedraw()
```

County Populaion Receiving SNAP Assistance



#1.3 State IRS Data

```
MD_tax_exempt = read_xls("/Users/lebsan/Documents/STAT 5084 - Time Series/County Level Project/Gathering
```

```
MD_tax_exempt1 = MD_tax_exempt %>%
  select(FIPS= `State FIPS code`, Year, Name,
         Pop_poor_exempt = `Poor exemptions`) %>%
  as_tsibble(index = Year, key= c(FIPS, Name)) %>%
  filter(FIPS == 24) %>% rename(year = Year)
```

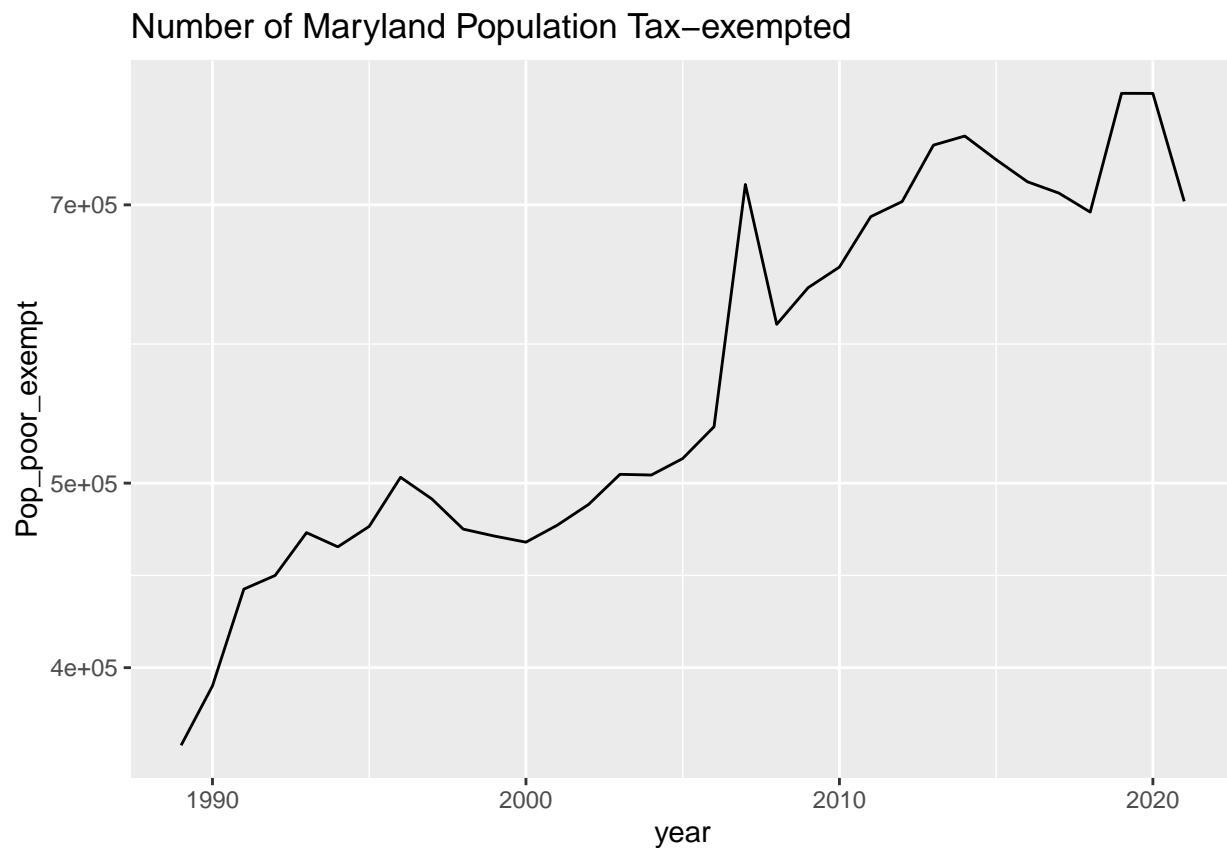
```
MD_tax_exempt1 %>% head()
```

```
## # A tsibble: 6 x 4 [1Y]
## # Key:      FIPS, Name [1]
##   FIPS  year Name      Pop_poor_exempt
##   <dbl> <dbl> <chr>      <dbl>
## 1    24  1989 Maryland    364161
## 2    24  1990 Maryland    391269
## 3    24  1991 Maryland    439842
## 4    24  1992 Maryland    447190
## 5    24  1993 Maryland    470933
## 6    24  1994 Maryland    462929
```

```
# Plot of the Number of poor exemptions filed in Maryland
```

```
MD_tax_exempt1 %>% ggplot(aes(x=year, y=Pop_poor_exempt))+
```

```
geom_line()+
labs(title = "Number of Maryland Population Tax-exempted")+
scale_y_log10()
```



1.4 Merging Data

```
md_merge = merge(MD_SAIPE, snap_MD2, by = c("FIPS", "year"))

final_md = merge(md_merge, MD_tax_exempt1, by = c("year")) %>%
  select(c(year, FIPS=FIPS.x, County = Name.x, State = Name,
           SAIPE = poverty_pop, SNAP = SnapBenefits,
           IRS_exempt_State = Pop_poor_exempt, Poverty_Universe = county_pov))

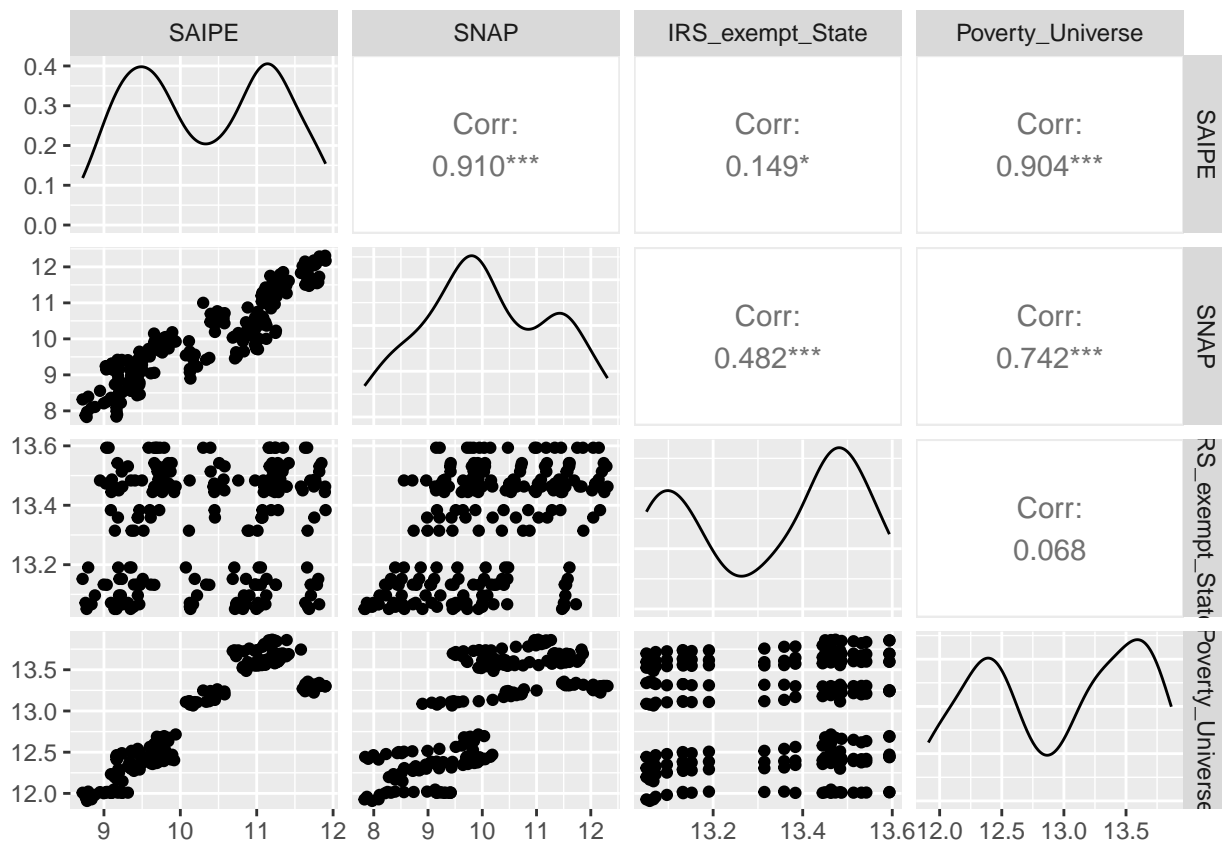
final_md %>% head()
```

##	year	FIPS	County	State	SAIPE	SNAP	IRS_exempt_State
## 1	1998	24003	Anne Arundel County	Maryland	26028	12604	472945
## 2	1998	24027	Howard County	Maryland	9558	3254	472945
## 3	1998	24021	Frederick County	Maryland	10331	5061	472945
## 4	1998	24025	Harford County	Maryland	12875	6144	472945
## 5	1998	24013	Carroll County	Maryland	6891	3259	472945
## 6	1998	24033	Prince George's County	Maryland	67094	34877	472945

```
## Poverty_Universe
## 1      472790
## 2      245437
## 3      188687
## 4      219887
## 5      151570
## 6      774481
```

```
# Data Visualization exhibiting the relationship between SAIPE, SNAP, IRS tax-exempt
# and poverty universe.
```

```
final_md %>% mutate(SAIPE = log(SAIPE), SNAP = log(SNAP),
                    IRS_exempt_State = log(IRS_exempt_State),
                    Poverty_Universe = log(Poverty_Universe)) %>%
select(c(SAIPE, SNAP, IRS_exempt_State,
         Poverty_Universe)) %>% ggpairs()
```



SAIPE and SNAP shows a high correlation coefficient and poverty universe displays a high correlation coefficient as well of 0.9. However, IRS tax-exempt has the weakest relationship with poverty universe as compared to SNAP and SAIPE. The log-transformation allows to get rid of variations in the data and capture outliers. SAIPE relationships with SNAP and Poverty_universe exhibit sensitivity to large numbers.