

# Explanatory Data Analysis: Data on Poverty in the United States

by:

Angela Hartono

2602059582

Kemiskinan telah menjadi salah satu permasalahan besar dalam dunia, tak terkecuali negara US. Oleh karena itu saya menganalisa masalah kemiskinan yang ada di US,. Pada laporan ini akan dijelaskan State dan County mana saja yang mengalami kemiskinan terbanyak dan seberapa besar satu korelasi pada satu variabel. Saya menggunakan dataset socioHealth dari covid dataset.zip.

## R Studio Explanatory Data Analysis (EDA)

Dalam R Markdown, saya hanya menggunakan beberapa library. Kemudian, saya mengimport dataset dari working directory. Berikut sedikit **preview data** dari dataset original:

| fips<br><chr> | state<br><chr> | county<br><chr> | lat<br><dbl> | lon<br><dbl> | total_population<br><int> | area_sqmi<br><dbl> |
|---------------|----------------|-----------------|--------------|--------------|---------------------------|--------------------|
| 01001         | Alabama        | Autauga         | 32.53493     | -86.64275    | 55049                     | 594.44612          |
| 01003         | Alabama        | Baldwin         | 30.72749     | -87.72258    | 199510                    | 1589.80742         |
| 01005         | Alabama        | Barbour         | 31.86959     | -85.39321    | 26614                     | 884.87578          |
| 01007         | Alabama        | Bibb            | 32.99863     | -87.12648    | 22572                     | 622.58235          |
| 01009         | Alabama        | Blount          | 33.98088     | -86.56738    | 57704                     | 644.60651          |
| 01011         | Alabama        | Bullock         | 32.10053     | -85.71569    | 10552                     | 622.80536          |
| 01013         | Alabama        | Butler          | 31.75242     | -86.68030    | 20280                     | 776.82805          |
| 01015         | Alabama        | Calhoun         | 33.77144     | -85.82603    | 115883                    | 605.88879          |
| 01017         | Alabama        | Chambers        | 32.91435     | -85.39203    | 34018                     | 596.53111          |
| 01019         | Alabama        | Cherokee        | 34.17593     | -85.60380    | 25897                     | 553.71970          |

1-10 of 3,144 rows | 1-7 of 181 columns

Previous 1 2 3 4 5 6 100 Next

Terdapat sebanyak **181 kolom** dan **3144 baris** pada dataset ini. Berikut beberapa variable yang terdapat pada kolom dataset:

```
Number of columns: 181
Number of Rows: 3144
All Variabel Names:
[1] "fips"
[3] "county"
[5] "lon"
[7] "area_sqmi"
[9] "num_deaths"
[11] "percent_fair_or_poor_health"
[13] "average_number_of_mentally_unhealthy_days"
[15] "percent_smokers"
[17] "food_environment_index"
[19] "percent_with_access_to_exercise_opportunities"
[21] "num_alcohol_impaired_driving_deaths"
[23] "percent_driving_deaths_with_alcohol_involvement"
[25] "chlamydia_rate"
[27] "num_uninsured"
[29] "num_primary_care_physicians"
[31] "num_dentists"
[33] "num_mental_health_providers"
[35] "preventable_hospitalization_rate"
[37] "percent_vaccinated"
[39] "num_some_college"
[41] "percent_some_college"
"state"
"lat"
"total_population"
"population_density_per_sqmi"
"years_of_potential_life_lost_rate"
"average_number_of_physically_unhealthy_days"
"percent_low_birthweight"
"percent_adults_with_obesity"
"percent_physically_inactive"
"percent_excessive_drinking"
"num_driving_deaths"
"num_chlamydia_cases"
"teen_birth_rate"
"percent_uninsured"
"primary_care_physicians_rate"
"dentist_rate"
"mental_health_provider_rate"
"percent_with_annual_mammogram"
"high_school_graduation_rate"
"population"
"num_unemployed_CHH"
```

Kemudian, saya menghapus kolom dan hanya menggunakan variabel "state", "county", "num\_below\_poverty", dan "num\_no\_highschool\_diploma" untuk mempermudah proses visualisasi pada Explanatory Data Analysis

| state<br><chr> | county<br><chr> | num_no_highschool_diploma<br><int> | num_below_poverty<br><int> |
|----------------|-----------------|------------------------------------|----------------------------|
| Alabama        | Autauga         | 4528                               | 6697                       |
| Alabama        | Baldwin         | 13956                              | 25551                      |
| Alabama        | Barbour         | 4824                               | 6235                       |
| Alabama        | Bibb            | 3040                               | 3390                       |
| Alabama        | Blount          | 7882                               | 9441                       |
| Alabama        | Bullock         | 2452                               | 2728                       |
| Alabama        | Butler          | 2642                               | 5126                       |
| Alabama        | Calhoun         | 13894                              | 22135                      |
| Alabama        | Chambers        | 4708                               | 6805                       |
| Alabama        | Cherokee        | 3499                               | 4188                       |

1-10 of 3,144 rows

Berikut 10 data teratas, **dataframe terbaru** setelah menggunakan variabel-variabel tersebut saja:

|    | state<br><chr> | county<br><chr> | num_below_poverty<br><int> | num_no_highschool_diploma<br><int> |
|----|----------------|-----------------|----------------------------|------------------------------------|
| 1  | Alabama        | Autauga         | 6697                       | 4528                               |
| 2  | Alabama        | Baldwin         | 25551                      | 13956                              |
| 3  | Alabama        | Barbour         | 6235                       | 4824                               |
| 4  | Alabama        | Bibb            | 3390                       | 3040                               |
| 5  | Alabama        | Blount          | 9441                       | 7882                               |
| 6  | Alabama        | Bullock         | 2728                       | 2452                               |
| 7  | Alabama        | Butler          | 5126                       | 2642                               |
| 8  | Alabama        | Calhoun         | 22135                      | 13894                              |
| 9  | Alabama        | Chambers        | 6805                       | 4708                               |
| 10 | Alabama        | Cherokee        | 4188                       | 3499                               |

1-10 of 10 rows

Terdapat sebanyak **4 kolom** dan **3144 baris** pada dataset ini. Berikut beberapa variable yang terdapat pada kolom dataset:

```
Number of Columns: 4
Number of Rows: 3144
All Variable Names:
[1] "state"          "county"         "num_below_poverty" "num_no_highschool_diploma"
```

Setelah itu, dilakukan pengecekan apakah terdapat missing values pada kolom dataset diatas, bisa dilihat di **missFrac** dalam table ini:

| variable<br><chr>         | type<br><chr> | levels<br><dbl> | topLevel<br><chr> | topCount<br><dbl> | topFrac<br><dbl> | missFreq<br><dbl> | missFrac<br><dbl> |
|---------------------------|---------------|-----------------|-------------------|-------------------|------------------|-------------------|-------------------|
| state                     | character     | 51              | Texas             | 254               | 0.081            | 0                 | 0                 |
| county                    | character     | 1852            | Washington        | 31                | 0.010            | 0                 | 0                 |
| num_below_poverty         | integer       | 2806            | 923               | 4                 | 0.001            | 0                 | 0                 |
| num_no_highschool_diploma | integer       | 2622            | 210               | 4                 | 0.001            | 0                 | 0                 |

Missing values juga bisa dilihat dari sini:

```
state      county      num_below_poverty num_no_highschool_diploma
0          0          0          0
```

Dapat dilihat bahwa **tidak ada missing values**.

Analisis yang saya lakukan selanjutnya yaitu melihat **10 State dan County yang memiliki jumlah poverty tertinggi**.

|      | state<br><chr> | county<br><chr> | num_below_poverty<br><int> | num_no_highschool_diploma<br><int> |
|------|----------------|-----------------|----------------------------|------------------------------------|
| 205  | California     | Los Angeles     | 1764081                    | 1495617                            |
| 3144 | New York       | New York City   | 1689759                    | 1124481                            |
| 611  | Illinois       | Cook            | 856682                     | 503252                             |
| 2624 | Texas          | Harris          | 763456                     | 554024                             |
| 104  | Arizona        | Maricopa        | 666513                     | 348801                             |
| 1852 | New York       | Kings           | 587518                     | 351046                             |
| 363  | Florida        | Miami-Dade      | 522143                     | 362252                             |
| 2580 | Texas          | Dallas          | 461045                     | 349434                             |
| 223  | California     | San Diego       | 444024                     | 293959                             |
| 1313 | Michigan       | Wayne           | 428189                     | 173869                             |

1-10 of 10 rows

Dapat dilihat bahwa **State California** dan **County Los Angeles** memiliki jumlah poverty dan jumlah no high school diploma tertinggi.

Berikut **ringkasan akurat** dari dataframe terbaru:

```
state      county      num_below_poverty num_no_highschool_diploma
Length:3144 Length:3144 Min. : 9 Min. : 5
Class :character Class :character 1st Qu.: 1677 1st Qu.: 1055
Mode :character  Mode :character Median : 4223 Median : 2655
Mean : 15514 Mean : 9227
3rd Qu.: 10177 3rd Qu.: 5948
Max. : 1764081 Max. : 1495617
```

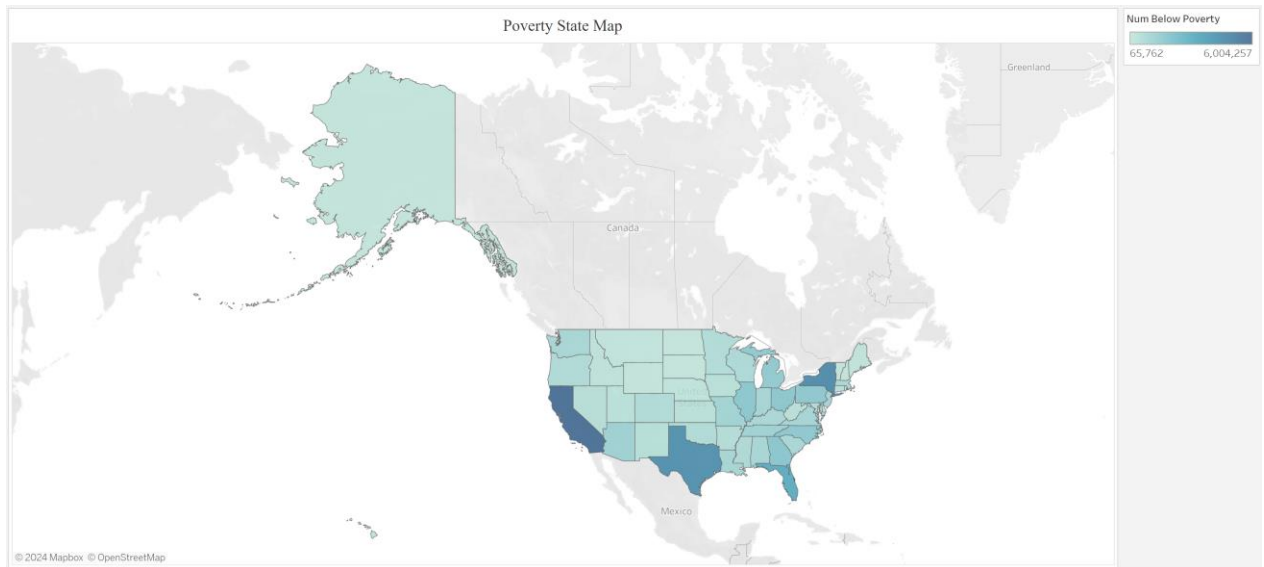
Selanjutnya, dilakukannya **korelasi** untuk melihat hubungan antara variabel poverty dan no high school diploma.

```
[1] 0.9788479
```

Dapat dilihat bahwa korelasi naik mendekati 1 sehingga semakin meningkat jumlah poverty, maka semakin meningkat juga jumlah no high school diploma. Dapat disimpulkan **bahwa poverty mempengaruhi jumlah no high school diploma**.

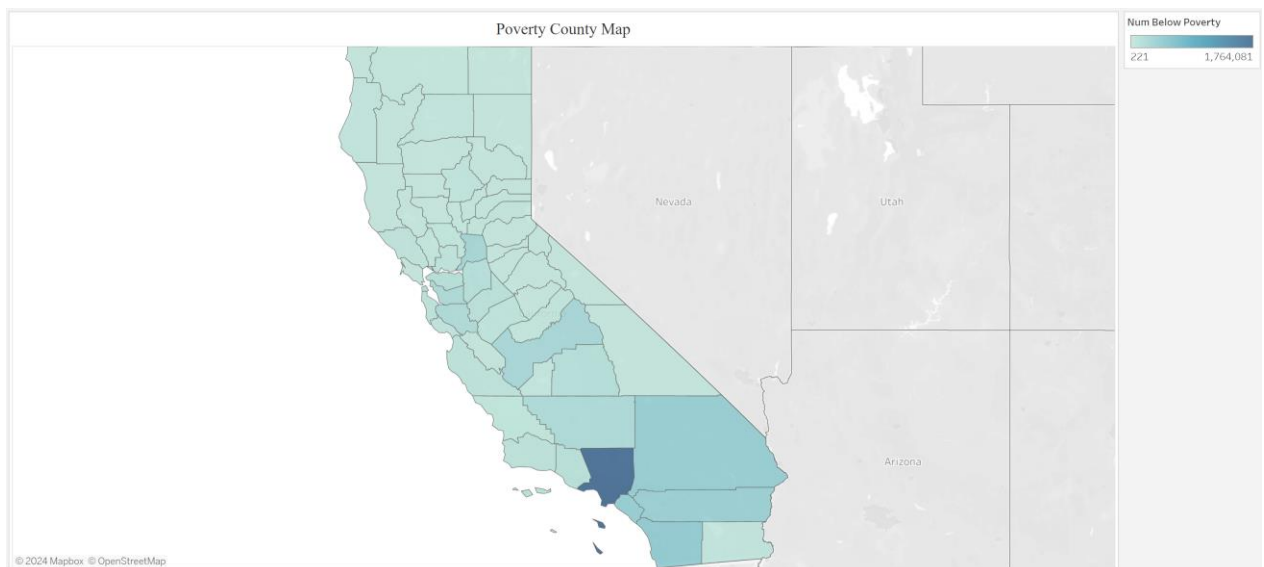
## Tableau Visualization

### Worksheet 1 : Poverty State Map



Disini terdapat visualisasi dimana semakin gelap warnanya maka dapat disimpulkan semakin banyak jumlah poverty di wilayah tersebut. Bisa dilihat bahwa California mempunyai warna yang paling gelap dengan jumlah poverty sebanyak 6,004,257 orang. Dapat disimpulkan bahwa **wilayah California memiliki poverty terbanyak.**

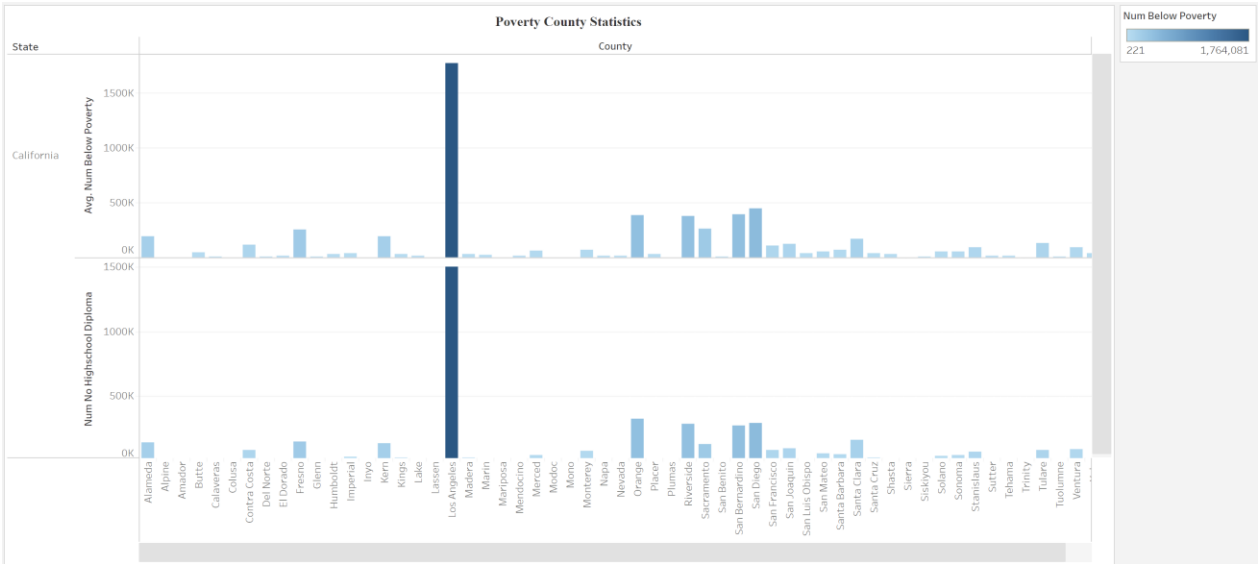
### Worksheet 2 : Poverty County Map



Disini terdapat visualisasi map yang sama konsepnya seperti worksheet 1 dimana semakin gelap warnanya maka dapat disimpulkan semakin banyak jumlah poverty di wilayah tersebut. Bisa dilihat bahwa wilayah

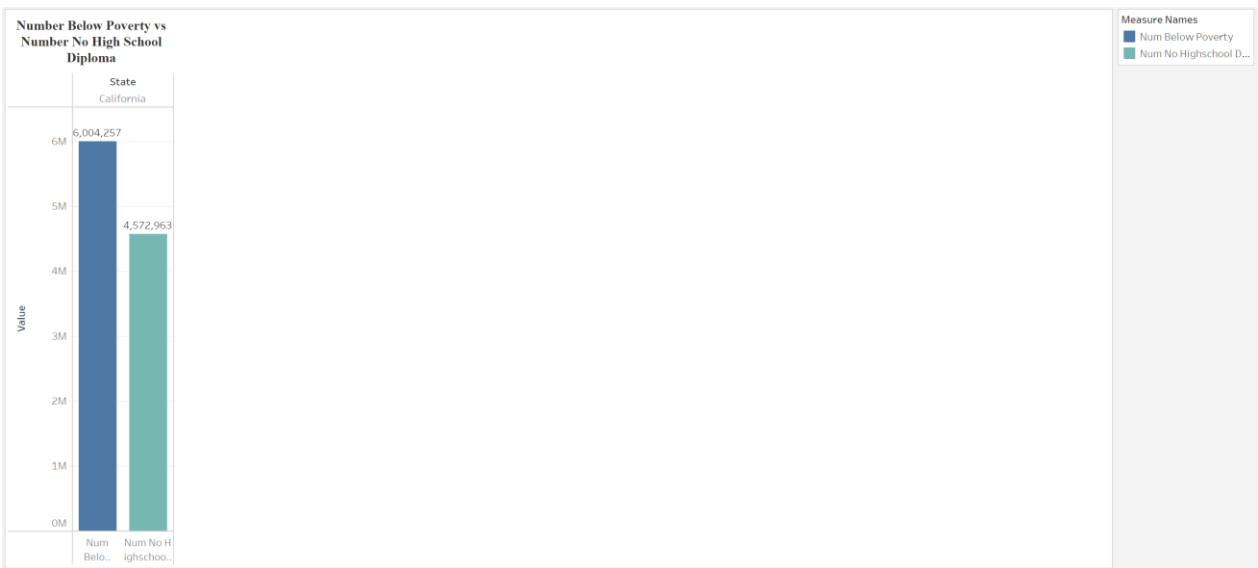
Los Angeles di California mempunyai warna yang paling gelap dengan jumlah poverty sebanyak 1,764,081 orang. Dapat disimpulkan bahwa **wilayah Los Angeles memiliki poverty terbanyak**.

Worksheet 3 : Poverty County Statistics



Disini saya membuat barplot dimana dapat dilihat **rata-rata jumlah poverty dan jumlah no high school diploma terbanyak di wilayah Los Angeles**.

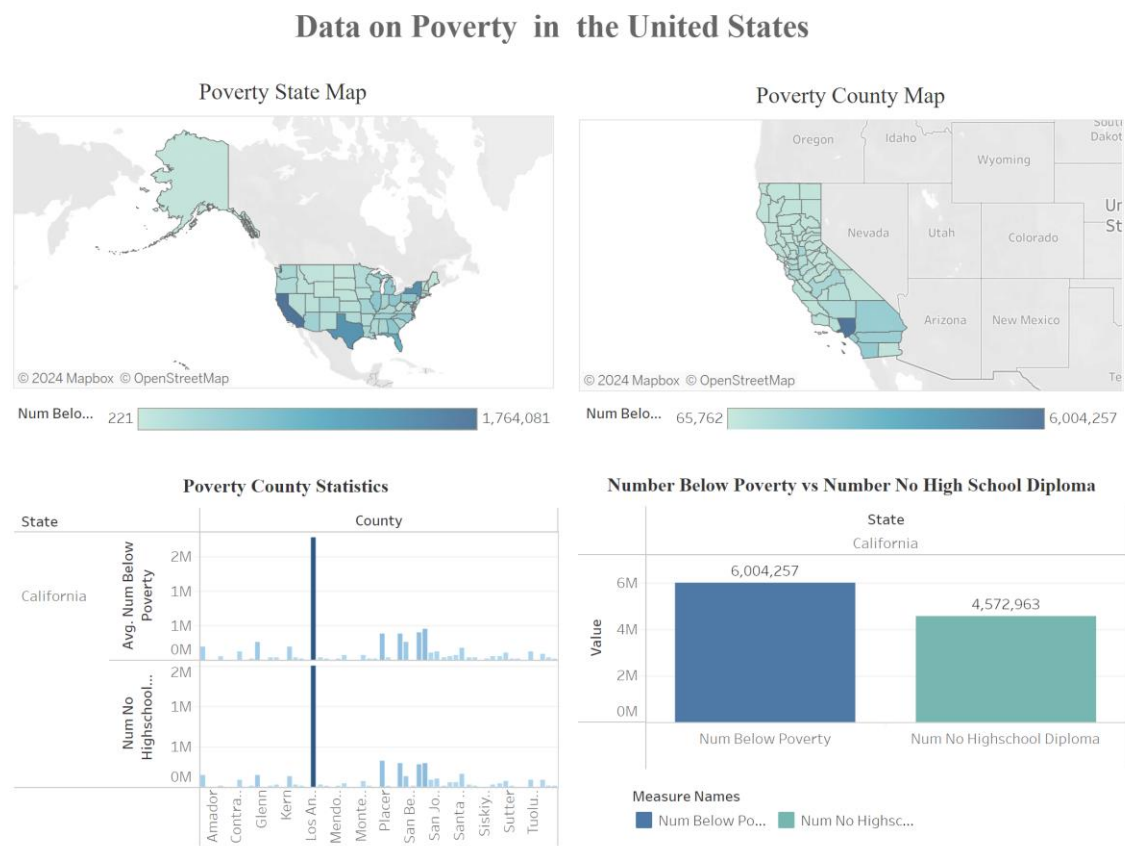
Worksheet 4 : Number Below Poverty vs Number No High School Diploma



Disini saya melakukan perbandingan antara jumlah poverty dan jumlah no high school diploma dimana dari hasil grafik ini dapat dilihat bahwa jumlah poverty sebanyak 6,004,257 orang dan jumlah no high school

diploma sebanyak 4,572,963 orang. Dapat disimpulkan bahwa semakin meningkat jumlah poverty, maka semakin meningkat juga jumlah no high school diploma sehingga jumlah **poverty mempengaruhi hasil jumlah no high school diploma**.

Dashboard 1: Data on Poverty in the United States



Berikut hasil dashboard yang telah saya buat mengenai **data tentang kemiskinan (poverty)di United States**.

Berikut link dashboard melalui Tableau Public:

[https://public.tableau.com/shared/GDGY926BD?:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/shared/GDGY926BD?:display_count=n&:origin=viz_share_link)