Comparative Analysis of Socioeconomic Indicators of USA’s Most and Least Populated States

Submitted by
Manjusha Madhukar Ghatke

Contents

Introduction

Socioeconomic Indicators

- Measures that reflect an individual’s/a community’s economic and social status .
- Income, gender, literacy, unemployment, race, education, geographical area of origin, income source, occupation, age distribution, health, marital status
- Importance:
  - Assessing inequalities
  - Public health research
  - Effect of recession
  - Assessing country’s overall condition

Source(s):
https://doi.org/10.1016/j.socscimed.2022.114733.
10.48550/arXiv.2301.09986
10.1136/jech.2004.023531

Purpose

- Policy Making & Resource Allocation
- Economic Analysis
- Educational Planning
- Social Services Planning
- Migration and Development Studies
- Research and Academia
- Investment Decision Making
- Workforce Development

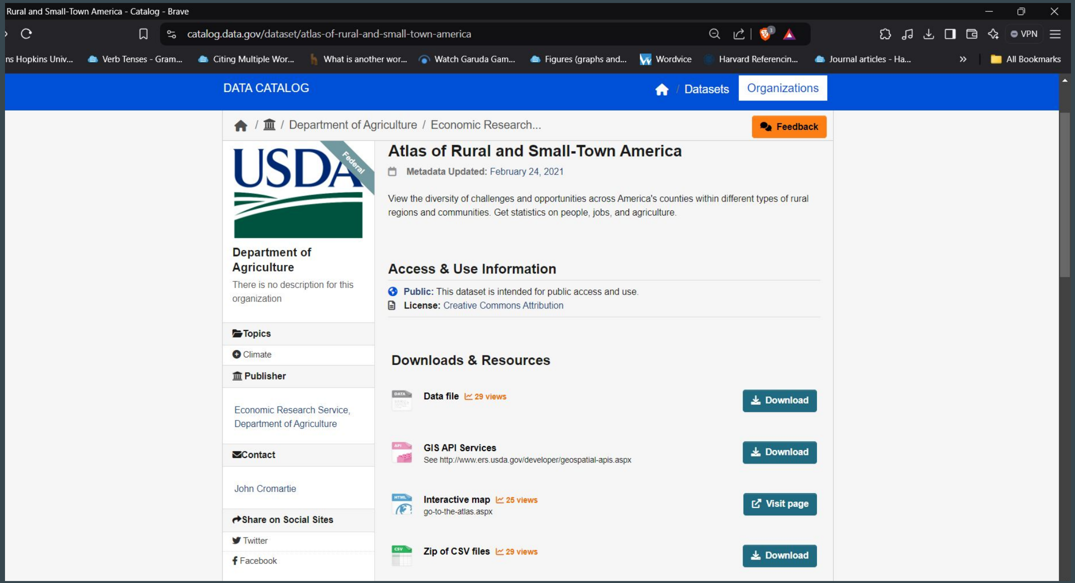
Purpose

End-users:

- State and federal policy makers
- Economic development agencies
- Educational institutions
- Social service organizations
- Business strategists
- Urban planners
- Researchers and academics
- Investment analysts

Data- Source

File name: RuralAtlasData24.xlsx
Organization: U.S. Department of Agriculture
Link: https://catalog.data.gov/dataset/atlas-of-rural-and-small-town-america



Data- Preprocessing

```
income.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3280 entries, 0 to 3279
Data columns (total 16 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   FIPS                   3280 non-null   int64  
 1   State                  3280 non-null   object 
 2   County                 3280 non-null   object 
 3   Median_HH_Inc_ACS      3273 non-null   float64
 4   PerCapitaInc           3274 non-null   float64
 5   Poverty_Rate_0_17_ACS  3273 non-null   float64
 6   Poverty_Rate_ACS       3274 non-null   float64
 7   Deep_Pov_All           3274 non-null   float64
 8   Deep_Pov_Children      3274 non-null   float64
 9   NumAll_inPOV_ACS       3274 non-null   float64
 10  PCTPOV017              3194 non-null   float64
 11  POV017                 3194 non-null   float64
 12  MedHHInc               3194 non-null   float64
 13  POVALL                 3194 non-null   float64
 14  PCTPOVALL              3194 non-null   float64
 15  Num_inPOV_0_17_ACS     3274 non-null   float64
dtypes: float64(13), int64(1), object(2)
memory usage: 410.1+ KB
```

RuralAtlasData24.xlsx - Excel (Product Activation Failed)

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW DEVELOPER

Q25

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	FIPS	Stat	County	Median_HH_Inc_AC	PerCapitaIn	Poverty_Rate_0_17_AC	Poverty_Rate_AC	Deep_Pov_A	Deep_Pov_Childre	NumAll_inPOV_AC	PCTPOV01	POV01	MedHHIn	POVAL	PCTPOVAL	Num_inPOV_0_17_AC
2	00000	US	United States	69021	37638	17.04670519	12.63185031	5.777384811	7.691714615	40661636	16.9	12243219	69717	41393176	12.8	12443424
3	01000	AL	Alabama	54943	30458	22.26193645	15.78512663	7.077315889	10.39058157	769819	22.7	250327	53990	800848	16.3	245896
4	01001	AL	Autauga	62660	30968	18.67386389	13.57847378	6.2502163	8.592895178	7847	16.1	2199	66444	6296	10.7	2560
5	01003	AL	Baldwin	64346	35384	11.29982166	9.204904993	4.043401319	5.503711999	20598	16.4	8207	65658	25526	10.8	5449
6	01005	AL	Barbour	36422	21325	43.74401226	26.47191011	12.82696629	26.72925848	5890	35.1	1776	38649	5089	23	2283
7	01007	AL	Bibb	54277	24787	28.04290718	16.94285714	9.076190476	17.51313485	3558	29	1275	48454	4204	20.6	1281
8	01009	AL	Blount	52830	27309	17.28166827	13.23663049	5.291223017	6.344745988	7720	16.7	2236	56894	6992	12	2337
9	01011	AL	Bullock	29063	21012	49.74805314	30.9345983	10.1937828	18.27759963	3065	41.7	903	32027	2764	32.1	1086
10	01013	AL	Butler	45236	23897	29.60556845	18.17700053	6.66136725	11.18329466	3430	34.2	1441	39442	4226	22.7	1276
11	01015	AL	Calhoun	50977	26440	21.60229476	17.53843981	8.329655586	11.17081448	19870	25.2	6205	48166	21630	19.2	5347
12	01017	AL	Chambers	47232	24840	25.33333333	15.0713246	5.091703057	7.958333333	5177	32.2	2258	45447	6699	19.7	1824
13	01019	AL	Cherokee	43475	26867	18.80704734	14.97337074	5.073789487	5.625132668	3683	26.6	1235	46365	4513	18.2	886
14	01021	AL	Chilton	56243	26426	16.16829596	15.7239819	6.004524887	3.76653249	6950	22.4	2362	55142	6973	15.5	1687
15	01023	AL	Choctaw	38581	25163	34.93740219	21.56629374	9.402523209	18.81846635	2718	32.4	809	40601	2893	23.3	893
16	01025	AL	Clarke	44108	25263	29.8369672	21.90612902	10.55722302	13.18012178	5036	29.2	1419	45447	4366	19.4	1519
17	01027	AL	Clay	45163	26255	22.67247153	16.92627952	5.59902376	9.845947756	2358	27.6	792	47036	2560	18.4	677
18	01029	AL	Cleburne	48333	27026	20.29154519	14.17920473	5.333154218	5.772594752	2111	22.1	755	49769	2236	15	696
19	01031	AL	Coffee	59034	29688	25.01400336	15.94036697	7.327981651	12.25894215	8340	24.2	3087	55052	8516	15.9	3126
20	01033	AL	Colbert	52017	28090	23.42852315	15.13205869	6.957759004	12.00574518	8508	23.6	2841	53341	9396	16.5	2773
21	01035	AL	Conecuh	37986	22841	22.31303637	12.96864349	6.833674165	16.30568043	1522	33.9	762	35621	2528	22.4	546
22	01037	AL	Coosa	50013	26718	20.45990566	11.14754098	4.232488823	9.846698113	1122	29.9	499	45954	1754	17.1	347
23	01039	AL	Covington	46186	26138	25.8163894	18.03933286	8.024121258	13.2224276	6641	28.3	2304	44188	7317	19.8	2095
24	01041	AL	Crenshaw	43103	28097	15.85488747	13.10107198	4.846860643	7.759489419	1711	28.9	854	38037	2267	17.5	472
25	01043	AL	Cullman	52690	26508	14.93699416	12.21860843	5.725403647	7.591435304	10519	17.5	3432	55036	11898	13.5	2916
26	01045	AL	Dale	50052	26469	27.26867534	16.71052082	8.868980495	16.43786771	8002	22.7	2558	50086	7587	15.7	3059
27	01047	AL	Dallas	34957	20748	38.03107682	24.30519548	8.905423623	16.26643486	9375	45.4	3924	33159	10914	29.5	3500
28	01049	AL	DeKalb	44037	23947	31.61751876	20.57219744	7.162382268	10.5333566	14525	29.3	4981	45062	13783	19.4	5436
29	01051	AL	Elmore	67597	31185	13.81956771	10.28199619	4.048460128	5.147749817	8470	17.3	3347	63147	10138	12	2647
30	01053	AL	Escambia	38464	19581	33.84559535	23.48629794	12.29800135	18.92545983	7979	29.9	2445	40506	8147	23.8	2797
31	01055	AL	Etowah	46984	26026	24.29974536	16.01345537	6.214308065	10.51746089	16376	25	5464	46308	17605	17.4	5344
32	01057	AL	Fayette	43960	24268	32.40174672	21.33490771	9.359269157	15.37117904	3433	26.5	901	46122	3123	19.6	1113
33	01059	AL	Franklin	43633	22500	30.09622689	18.06696147	6.734049274	8.850341859	5720	25.1	1968	45440	5995	18.9	2377
34	01061	AL	Geneva	43581	23338	35.24674435	23.51242496	5.254958584	9.544208362	6188	29.1	1672	43206	5566	21.1	2057
35	01063	AL	Greene	28826	16282	61.37184116	39.79578648	14.9928913	34.23586041	3079	51.8	848	30225	2520	33.2	1020
36	01065	AL	Hale	32294	20849	38.46153846	24.9124734	7.956339672	14.2351901	3629	32.6	1128	41995	3221	22.1	1305
37	01067	AL	Henry	55870	28304	23.8294065	14.39453008	6.585497511	12.7348643	2400	24.6	867	56389	2694	15.6	799
38	01069	AL	Houston	50222	29388	26.03491684	17.66523132	7.77083868	12.76156672	18584	27.1	6511	48701	20210	19.1	6308
39	01071	AL	Jackson	43785	25153	25.71844307	17.61233633	5.55865331	8.257548199	9160	27	2906	46998	9851	18.9	2828
40	01073	AL	Jefferson	58330	34860	22.83916525	15.89747566	7.317955594	11.16414849	104579	23.6	35575	55210	110131	16.9	34737
41	01075	AL	Lamar	43324	22678	18.79623403	16.27568119	5.464082763	8.977807666	2234	24.7	715	44265	2300	17	559
42	01077	AL	Lauderdale	50000	29667	16.93923968	13.30227573	5.744562186	7.46777862	12164	22.8	4187	52222	14838	16.2	3128
43	01079	AL	Lawrence	51712	26810	24.51432565	15.38414578	5.619486585	10.07686932	5040	23	1638	55555	4566	13.9	1754
44	01081	AL	Lee	57191	30680	20.42960349	19.92484373	11.49324918	10.27684332	33086	18.4	6767	54188	30989	18	7409
45	01083	AL	Limestone	70736	33428	15.53772358	11.48487139	5.007457045	6.903662528	11243	11.8	2743	69207	10363	9.9	3466
46	01085	AL	Lowndes	31961	21936	32.28208498	21.48264674	9.271458067	18.17783618	2185	43	946	36993	2786	28.3	737

Read Me | VariableNameLookup | Documentation | People | Jobs | County Classifications | **Income** | Veterans

READY 70%

Data Preprocessing

Python:

1. Dropping columns and rows
2. Checking for and dropping null values
3. Reset index
4. Export data

```
income.drop('FIPS', axis=1, inplace=True)
income.drop(0, inplace=True)
income.head()
```

```
income.isnull().sum()
```

```
#drop rows containing Null values
before=income.shape[0]
cols=['Median_HH_Inc_ACS', 'PerCapitaInc',
      'Poverty_Rate_0_17_ACS', 'Poverty_Rate_ACS', 'Deep_Pov_All',
      'Deep_Pov_Children', 'NumAll_inPOV_ACS', 'PCTPOV017', 'POV017',
      'MedHHInc', 'POVALL', 'PCTPOVALL', 'Num_inPOV_0_17_ACS']

for column in cols:
    income = income.dropna(subset=[column])

after=income.shape[0]
print('Number of rows with null values removed:', before-after)
```

```
Number of rows with null values removed: 87
```

```
income.reset_index(drop=True, inplace=True)
```

```
income.to_excel('RuralAreafinal.xlsx', sheet_name='Income')
```

Data- Preprocessing

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3192 entries, 0 to 3191
Data columns (total 15 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   State                  3192 non-null   object
 1   County                 3192 non-null   object
 2   Median_HH_Inc_ACS      3192 non-null   float64
 3   PerCapitaInc           3192 non-null   float64
 4   Poverty_Rate_0_17_ACS  3192 non-null   float64
 5   Poverty_Rate_ACS       3192 non-null   float64
 6   Deep_Pov_All           3192 non-null   float64
 7   Deep_Pov_Children      3192 non-null   float64
 8   NumAll_inPOV_ACS       3192 non-null   float64
 9   PCTPOV017              3192 non-null   float64
 10  POV017                 3192 non-null   float64
 11  MedHHInc               3192 non-null   float64
 12  POVALL                 3192 non-null   float64
 13  PCTPOVALL              3192 non-null   float64
 14  Num_inPOV_0_17_ACS     3192 non-null   float64
dtypes: float64(13), object(2)
memory usage: 374.2+ KB
```

Data- Preprocessing

```
income
```

	State	County	Median_HH_Inc_ACS	PerCapitaInc	Poverty_Rate_0_17_ACS	Poverty_Rate_ACS	Deep_Pov_All	Deep_Pov_Children	NumAll_inPOV_ACS	PCTPOV01
0	AL	Alabama	54943.0	30458.0	22.261936	15.785127	7.077316	10.390582	769819.0	22
1	AL	Autauga	62660.0	30968.0	18.673864	13.578474	6.250216	8.592895	7847.0	16
2	AL	Baldwin	64346.0	35384.0	11.299822	9.204905	4.043401	5.503712	20598.0	16
3	AL	Barbour	36422.0	21325.0	43.744012	26.471910	12.826966	26.729258	5890.0	35
4	AL	Bibb	54277.0	24787.0	28.042907	16.942857	9.076190	17.513135	3558.0	29
...	...	...	...	...	...	...	...	...	...	
3187	WY	Sweetwater	76668.0	36233.0	13.501566	10.481391	5.646027	7.358630	4396.0	9
3188	WY	Teton	94498.0	66296.0	8.315204	7.117040	2.547332	0.652647	1654.0	5
3189	WY	Uinta	75106.0	30586.0	7.466063	7.958751	3.508166	2.071006	1613.0	11
3190	WY	Washakie	62271.0	31032.0	5.125149	6.571842	2.607665	3.694875	499.0	12
3191	WY	Weston	65566.0	31190.0	7.303807	14.085591	6.250000	2.175602	915.0	12

3192 rows × 15 columns

Data- Preprocessing

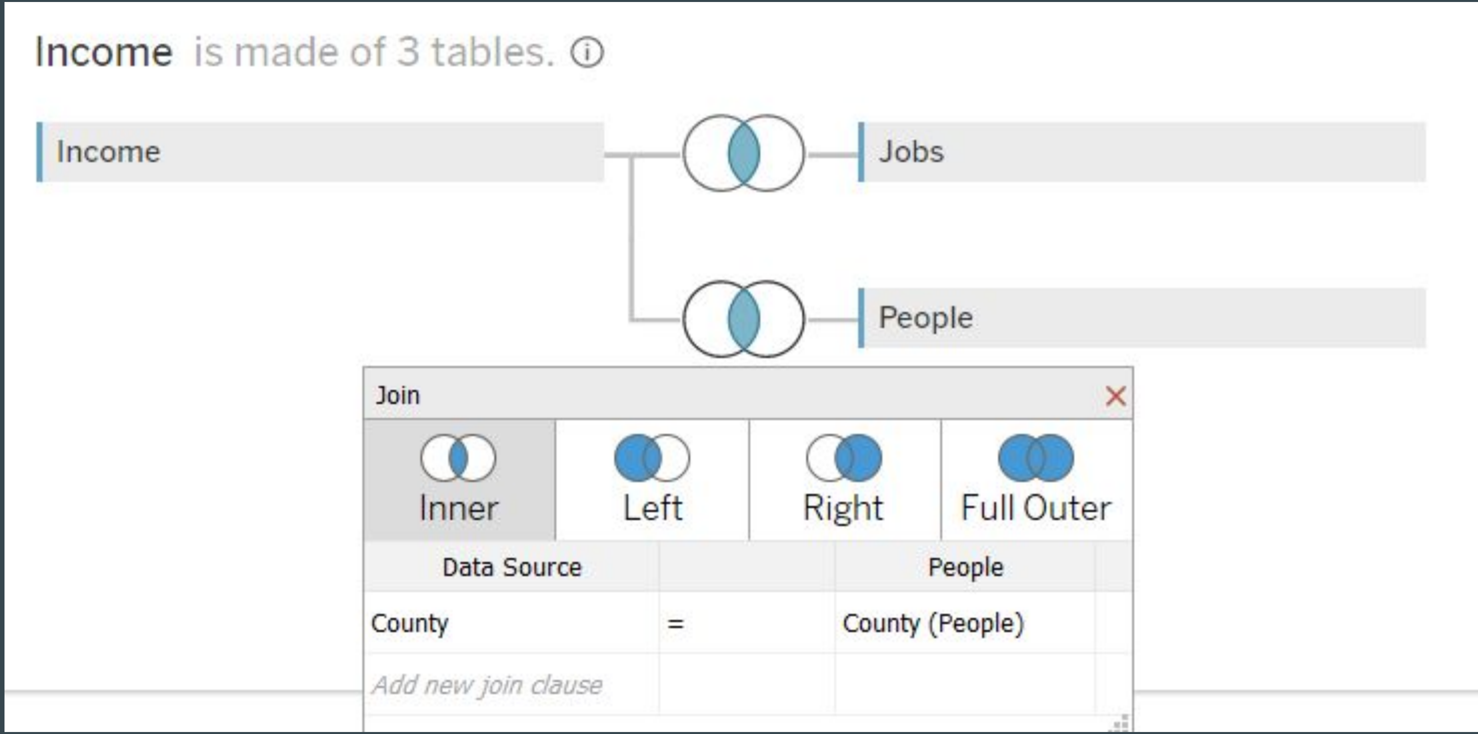
Same preprocessing steps have been performed for other sheets in the file

```
jobs.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3186 entries, 0 to 3185
Data columns (total 78 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   State             3186 non-null   object
 1   County            3186 non-null   object
 2   UnempRate2021     3186 non-null   float64
 3   PctEmpChange2021  3186 non-null   float64
 4   UnempRate2020     3186 non-null   float64
 5   PctEmpChange1920  3186 non-null   float64
 6   UnempRate2019     3186 non-null   float64
 7   UnempRate2018     3186 non-null   float64
 8   UnempRate2017     3186 non-null   float64
```

```
people.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3138 entries, 0 to 3137
Data columns (total 93 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   State                               3138 non-null   object
 1   County                              3138 non-null   object
 2   Pop_change_Rate_2020_2021           3138 non-null   float64
 3   POPESTIMATE2021                     3138 non-null   float64
 4   Net_Migration_Rate_2020_2021        3138 non-null   float64
 5   Natural_Change_Rate_2020_2021       3138 non-null   float64
 6   Net_InterMigrationRate_2020_2021    3138 non-null   float64
 7   PopChangeRate1020                   3138 non-null   float64
 8   PopDensity2020                      3138 non-null   float64
 9   Under18Pct2020                      3138 non-null   float64
 10  Age65AndOlderPct2020                3138 non-null   float64
 11  WhiteNonHispanicPct2020             3138 non-null   float64
 12  BlackNonHispanicPct2020             3138 non-null   float64
 13  AsianNonHispanicPct2020             3138 non-null   float64
 14  NativeAmericanNonHispanicPct2020    3138 non-null   float64
 15  HispanicPct2020                     3138 non-null   float64
 16  MultipleRacePct2020                 3138 non-null   float64
 37  Ed5CollegePlusPct                   3138 non-null   float64
```

Data- Preprocessing

3 different tables are joined (inner) on the 'County' (unique values) column.



Data- Preprocessing

Calculated fields

AvgUnempRat1721

```
([UnempRate2017]+[UnempRate2018]+[UnempRate2019]+[UnempRate2020]+[UnempRate2021])/5
```

WorkingAgePct2020

```
100- [Age65AndOlderPct2020]- [Under18Pct2020]
```

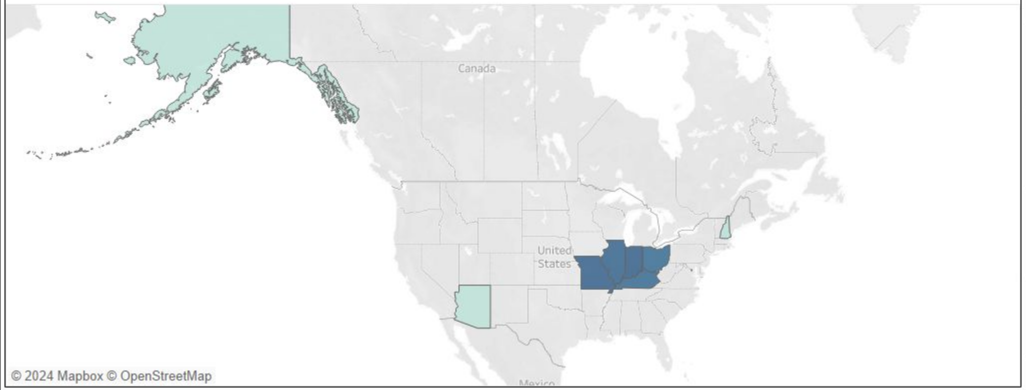
Visualization

Tool used: Tableau

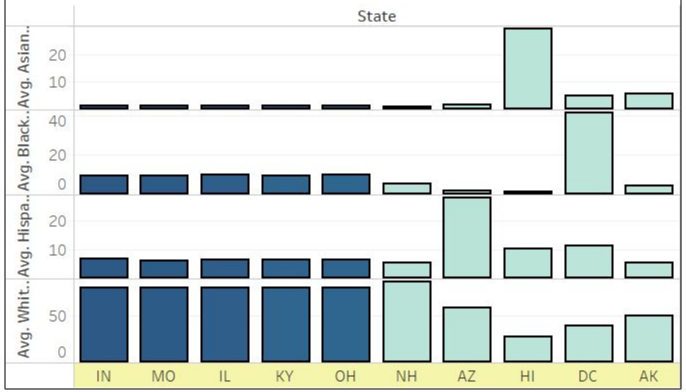
Visualization Title	Visualization Type
Population Overview Map	Filled Map
Top and Bottom 5 Populated States Demographics	Bar Chart
Economic Health Overview	Bar Chart
State-wise Age Distribution	Bar Chart
Education & Income Relationship	Scatter Plot

Comparative Analysis of Socioeconomic Indicators of Population Extremes

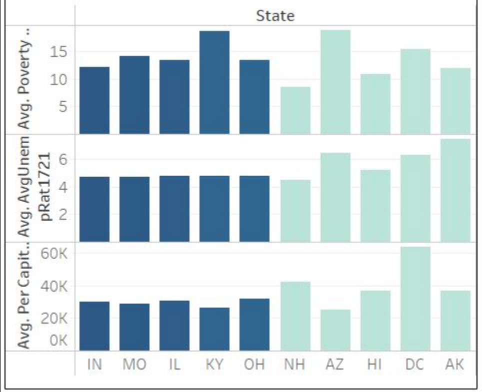
Population Overview Map



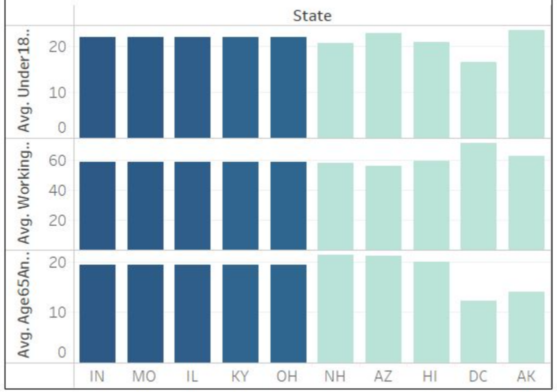
Top and Bottom 5 Populated States Demographics



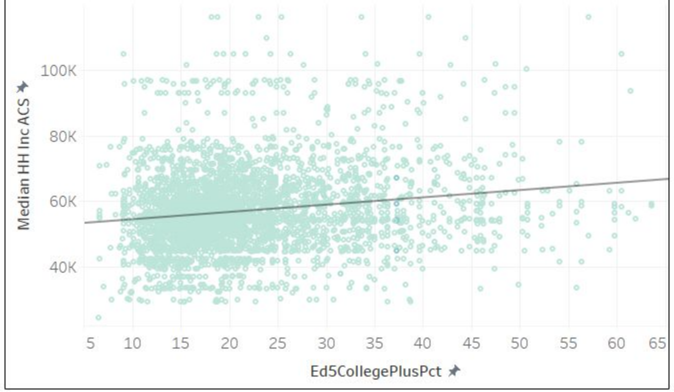
Economic Health Overview



State-wise Age Distribution



Education & Income Relationship



Visualization

Key Insights

Population Overview Map & Top and Bottom 5 Populated States Demographics

1. Demographic Composition:
   a. The Midwestern states (shown in blue) appear to have relatively consistent demographic patterns across racial/ethnic categories: >75% White, <20% other ethnic groups
   b. States like Hawaii (HI) (>30% Hisp) and DC (>40% Black) show comparatively different demographic> higher diversity
   c. Alaska (AK) shows relatively lower populations across all demographic categories
2. Plausible Reasons:
   a. Historical settlement patterns: The Midwest's demographic consistency likely reflects historical immigration and settlement patterns | top populated states shown (IN, MO, IL, KY, OH)
   b. Geographic factors: Hawaii's unique demographic makeup reflects its Pacific location and Native Hawaiian population | bottom populated states (NH, AZ, HI, DC, AK) are more geographically diverse, including territories from different regions
   c. Urban concentration: DC's distinct pattern reflects its nature as an urban federal district
   d. Industrial development: The Midwest's population patterns may relate to its industrial and agricultural heritage.
3. Applications:
   a. Policies: Language services and cultural program development
   b. Business: Retail location planning | Product development tailored to regional demographics
   c. Social welfare: Cultural sensitivity training for service providers

Visualization

Key Insights

Economic Health Overview

1. Per capita income: :
    a. High Income Areas:
       DC: high concentration of federal jobs and professionals
       NH and HI's: tourism, specialized industries and high cost of living
       AK: higher income likely reflects oil industry revenues and high cost of living
    b. More Populated States:
       The Midwest states (OH, IL, IN, MO, KY) show relatively consistent income levels | more diverse economies including manufacturing, agriculture, and services.
2. Unemployment and Poverty rates:
    a. Midwest states' consistent unemployment rates reflect similar industrial/economic bases
       AK's high unemployment might relate to seasonal work and resource-dependent economy
       NH's low unemployment & poverty rates might reflect diverse economy and proximity to Boston metro area
3. Applications:
    a. Policies: Education institutes | Raising minimum wage | Small business Support | Improve access to food, healthcare, housing to tackle high poverty rates
       Business: Training programs | Market expansion according to worker availability.

Visualization

Key Insights

State-wise Age Distribution

1. More populated states show more stable distributions due to diverse economic bases
   a. DC's high working-age population reflects its status as a job center
2. Less populated states show more variation based on specific economic/lifestyle factors
   a. Arizona's higher elderly population likely due to retirement migration
   b. Alaska's younger demographic may reflect job opportunities in resource industries
3. Applications:
   a. Policies: Medical facilities and specialists | Senior care facilities | Developing health programs
   Infrastructure: School system capacity | Retirement community planning

Visualization

Key Insights

Education (4-yr college degree or higher) & Income Relationship

1. Income spread is wider in counties with higher education levels.
2. High incomes despite moderate education levels: indicating presence of specific industries or economic conditions.
3. Applications:
    a. Policy: Higher education funding | Workforce skill development
    b. Business: Expansion based on workforce skill

Conclusion

A Comparative Analysis of Population Extremes can lead to valuable insights regarding the country's economy, health status, areas of opportunities, migration, major areas of development, and factors affecting population growth/decline (regulate population).

Limitations:

- All attributes do not belong to the same year/year group, so the most recently collected data is used for comparison.
- Missing values ~100 have been excluded from the visualization.

References

- U.S. Department of Agriculture
- 2024 Stack Exchange Inc
- U.S. Bureau of Labor Statistics
- Resource Development Council for Alaska, Inc.
- 2024 NH Department of Business and Economic Affairs,
- Office of the Deputy Mayor for Planning and Economic Development - DMPED
- Vision Retirement

THANK YOU