

Data Source:

Unemployment Data:

<https://www.ers.usda.gov/data-products/county-level-data-sets/>

Unemployment Data contains:

- County Code
- State
- County Name
- Civilian Labor Force 2000-2019
- Number Employed 2000-2019
- Number Unemployed 2000-2019

Population Data:

<https://www.ers.usda.gov/data-products/county-level-data-sets/>

Population Data Contains:

- County Code
- State
- County Name
- Population Estimate 2010-2019
- Numeric Population Change 2010-2019
- Births 2010-2019
- Deaths 2010-2019
- Natural Increase 2010-2019
- International Migration 2010-2019
- Domestic Migration 2010-2019

Personal Current Transfer Receipts

<https://apps.bea.gov/regional/downloadzip.cfm> (Table CAINC35)

Personal Current Transfer Receipts Data Contains:

- Location Code
- Location Name
- Line Code (Code for transfer receipts)
- Description (Name of Transfer Receipt)
- Thousands of dollars in receipts (1969-2019)

This data set was chosen to see the relationship between employment, and the benefits provided to communities in different capacities by the government. A study of the effectiveness of funding in the form of benefits from the U.S. Government.

Data Profile:

Data was cleaned and organized on individual notebooks before merging. Much of the data was in a format that was not useable, so it had to be transformed into the appropriate shape.

- Population Data:
 - Dropped Columns:
 - Rural-urban-Continuum
 - GQ_Estimates
 - Rate columns (can be recalculated with given data if necessary)
 - Net migration columns (can be recalculated with given data if necessary)
 - Other columns
 - Data was then broken into 7 distinct DataFrames, to convert columns with multiple years, into one column that has year datapoints associated with it. The 7 columns were:
 - Population Estimate
 - Number of Population Change
 - Births
 - Deaths
 - Natural Increases
 - International Migrations
 - Domestic Migrations
 - Data was re-merged into a single dataset and exported.
- Unemployment Data
 - Dropped Columns:
 - Rate columns (can be recalculated with given data if necessary)
 - Code columns
 - Other columns
 - Data was then broken into 3 distinct DataFrames, to convert columns with multiple years, into one column that has year datapoints associated with it. The 3 columns were:
 - Employed
 - Unemployed
 - Labor Force
 - Data was re-merged into a single dataset and exported.
- Government Data
 - Dropped Columns:
 - Years from 1969-2009 (no correlating data from other sets)
 - Table Name
 - Industry Classification
 - Unit
 - Data was then broken into subgroups that classify the different types of transfers the government conducts:
 - Retirement and Disability
 - Medical
 - Income Maintenance

Task 6.1 – Eric Riesberg

- Unemployment Insurance
- Veterans
- Other
- Data Dictionary was created to identify subgroups in code
- DataFrame was then melted and re-pivoted to substitute year and benefit type in the correct orientation
- Data was exported for final clean and merging.
- Overall Dataset
 - Columns were renamed on each DataFrame to promote clean merging.
 - Datatypes were changed to allow for merging, and more accurate data.
 - NA values (10 rows, 5 columns) were imputed as 0, as the data is insufficient, and it only effects one county. This is not a concern for skewing the data. It only applies to 10/31410 lines, or .03% of the dataset.
- Data Description:

	Year	Unemployed	Employed	Labor_Force	Dom_Mig	Int_Mig	Nat_Inc	Deaths
count	31410	31410	31410	31410	31410	31410	31410	31410
mean	2014.5	6184.10	93543.28	99727.39	2.75	495.84	734.71	1562.33
std	2.872327047	42475.79	565249.10	604534.37	6161.89	4129.99	6260.48	9168.88
min	2010	4	62	66	-203414	-992	-4729	0
25%	2012	289	4696	5026	-145	0	-21	104
50%	2014.5	744	11115.5	11899	-21	7	12	255
75%	2017	2145	31845	34119	83	49	133	653
max	2019	2292082	18550526	19353743	216749	160385	271383	282520

	Births	Pop_Change	Pop_Est	gov_retire_disability	gov_medical	gov_income
count	31410	31410	31410	31410	31410	31410
mean	2297.04	1233.27	202139.98	560124.67	734628.42	169499.41
std	14759.05	11349.58	1232583.83	3221871.48	4708787.36	1107936.59
min	0	-76790	84	0	0	0
25%	106	-91	11096.25	36427	46068.5	8773
50%	273	3	26385	87437.5	105781.5	22320.5
75%	771.75	218	72728.75	236448.75	270856.5	57156.5
max	509771	505723	39512223	107856991	183951161	38034988

Task 6.1 – Eric Riesberg

	gov_unemployment	gov_veterans	gov_education	gov_other
count	31410	31410	31410	31410
mean	37052.44	57256.90	41261.06	31096.89
std	304678.39	344696.00	261017.76	224993.55
min	0	0	0	0
25%	1022.25	3227	1149	1100
50%	3077	8157.5	3528	3177
75%	10032.25	24177.25	13967.75	9889.25
max	22179960	13697985	10193011	10167569

Limitations and Ethics:

The data is sourced from the government, so it can be considered a reliable and robust source of information. The largest limitation to this dataset is the limited timeframe (only 10 years available). Ethically, this data could be questionable if it is sourced including biases, whether purposefully or not. However, since it is a reliable source of data, we can err on the side of the data being safe to use. It contains no PII and does not discriminate in any way.

Questions to explore:

One of the major things that I want to explore is the time-linear relationship between receiving governmental benefits in different forms, and the resultant employment rate. From there, I want to see if there is a geographical relationship between different types of benefits received. As the data is more closely examined, more questions and explorations should appear around the correlation between different types of benefits, and their resultant effect on an area's population/employment.