**https://www.pewforum.org/religious-landscape-study/compare/party-affiliation/by/state/A Note with a big N: we cannot get data for weekly or monthly for most of the features so, let’s try to focus on how we can predict things with the data available.**

**P.s. Trying my best...**

**Population Data:**

1. <https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/counties/totals/>
   1. This data set contains Population for 2010 till 2020 inclusive.
   2. Has parameters like total deaths, births, migration, natural calamities, net migration, etc. for that particular year.
   3. All data is county wise.
2. <https://covid19.census.gov/datasets/USCensus::average-household-size-and-population-density-county/about>
   1. Population Density Countywise.

**Vaccination Supply:**

1. <https://data.cdc.gov/Vaccinations/COVID-19-Vaccine-Distribution-Allocations-by-Juris/saz5-9hgg/data>
2. This data has vaccination supply statewise.
3. Dates for many entries are missing in the csv files though still visible in the website.
4. County wise distribution is not possible for this data. As data is maintained on state levels( Vaccine distribution allocations are decided by jurisdictional at state level.
5. <https://www.dshs.state.tx.us/coronavirus/immunize/vaccineallocations.aspx> Though few states maintain a more atomized data but this one is again not countywise but hospitalwise or vaccination centre wise.

**Races:**

1. <https://www.kff.org/other/state-indicator/distribution-by-raceethnicity/?dataView=1&currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>
   1. Races distribution statewise.
   2. Available till 2019 only in this dataset.
2. <https://data.cdc.gov/Vaccinations/Vaccine-Hesitancy-for-COVID-19-County-and-local-es/q9mh-h2tw>
3. Hesitancy among different races county wise with SVI.
4. Also has hesitancy level data.

**Covid Cases:**

1. <https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36>
   1. Total number of cases starting from 1st Feb till now, but state wise.
   2. Has fields like, Total cases, new deaths, new cases, conf cases, probable cases, etc.
2. <https://data.world/covid-19-data-resource-hub/covid-19-case-counts/workspace/file?filename=COVID-19+Activity.csv>
   1. Finally county wise daily data for covid cases, and deaths.

**Unemployment:**

1. <https://www.bls.gov/lau/>
   1. <https://www.bls.gov/lau/laucnty20.xlsx>
   2. <https://www.bls.gov/lau/laucnty19.xlsx>
   3. County wise unemployment data.
   4. Data before 2019 is also available on the website.
   5. Columns - Total Labor, Employed, Unemployed, Unemployment rate.
   6. Weekly or monthly data not available here.
2. If we want data for months, we need to look for only a few selected states who maintain those, like NY, TX, CA, etc. Computationally there are 52 states and around 3225 counties for each county. If we try to find data for even a single feature then it would cost us around 2.5 million entries for merging 720 days of data. This would cost us a lot of time.

Unemployment rate higher among unemployed people

https://www.medrxiv.org/content/10.1101/2021.04.17.21255668v1.full

Metro Correlation: Positive for metro counties (label encoded 0)

**Political Inclination:**

<https://www.pewforum.org/religious-landscape-study/compare/party-affiliation/by/state/>

County wise data:

<https://dataverse.harvard.edu/file.xhtml?fileId=4819117&version=9.0>

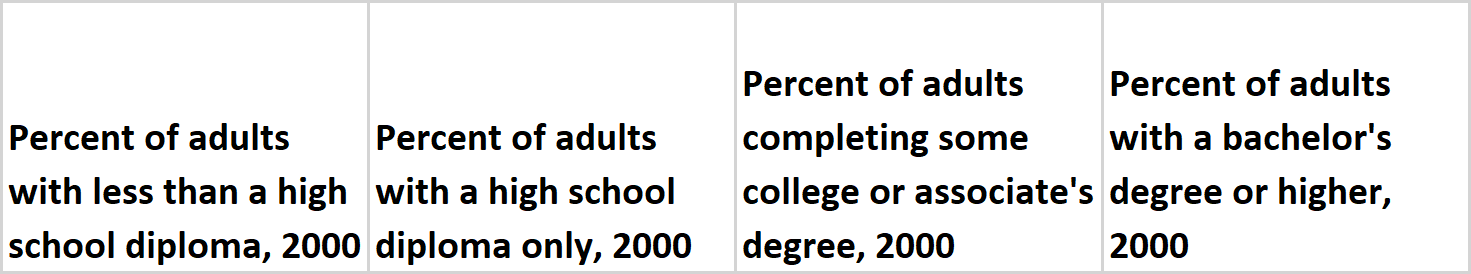
Use this

**Vaccination in USA:**

<https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh/data>

Final DataSets:

<https://data.cdc.gov/Vaccinations/Vaccine-Hesitancy-for-COVID-19-County-and-local-es/q9mh-h2tw> SVI RACES AND HESITANCY -SVI, every race column, estimated hesor unsu, estimated str hes

Education: USDA: 

Unemployment: %unemployment from above data

**Challenges Faced:**

1. Label Encoding for Metro\_status

2. Data type Matching to merge different data sets

3. Normalized values to percentages.

4. We have process fips for all data sets before merging.

5. CDC has started populating data for certain counties, mostly non-metro ones only recently. This posed a problem while trying to make prediction. For example, CDC has started recording Ellis County’s data only since 22nd October, 2021. This means there will be a problem in training our model for that county as we take into account

There are more than 250 such counties that do not have data properly maintained over the period in which vaccination was done.

As checked on 4th November, 2021, there were 17 counties that did not have any vaccination coverage data.

Strong Correlations:  
1. SVI - -0.2

2. Hispanic - 0.15

1. Asians - 0.25
2. Blacks - -0.09 (recently improved)
3. While - -0.11

Population Density: 0.15

Unemployment: 0.2

Politics:

Republicans : -0.54

Democrats: 0.53

Others: 0.15

Libertarians: 0.16

Greens: 0.27

Education:

LHSD (Less than HS): -0.29

HSD: -0.40

CAD(College and associate Degree): 0.1%

BDH (Bachelors and Higher): +0.48

Metro Status

Non Metro: -0.22

Analysis of counties that have exceptional rates (high or low)