Community Capitals Documentation

Lily Christenson Manjul Balayar Anh Nhat Le
Solomon Eshun

2024-07-10

Table of contents

Capitals Project Description

This is a Quarto book used to document the work completed in the DSPG 2024 year for the Community Capitals project.

Project Sponsor / Collaborator: ISU Community and Economic Development

- Bailey Hanson
- Jennifer Drinkwater

DSPG Team Lead: Matthew Voss

DSPG Graduate Fellow: Solomon Eshun

DSPG Interns: Lily Christenson, Manjul Balayar, Nhat (Chris) Le

Overview

In our project, we seek to explore the relationship between community capital indicators and the overall condition of communities in Iowa, aiming to identify strengths, weaknesses, opportunities, and gaps within these communities. The main work focuses on major events such as natural disasters, job losses, new job creation, school closures, and new opportunities for tourism and healthcare. The study seeks to identify the most relevant community capitals, trends over time, and the impact on community resiliency, by analyzing correlations between various indicators and comparing quantitative data with some qualitative insights.

That aside, we also seek to develop a comprehensive database of significant events, using sources such as the Worker Adjustment and Retraining Notification Act (WARN), Iowa Economic Development Authority Board minutes, and weather disaster summaries. The Data analysis will include exploring relationships within and between different community capitals, and visualizing findings using dashboards created in Shiny or Tableau. The overall goal is to provide a tool for stakeholders to better understand and enhance community resiliency and quality of life.

Project Description

Project Summary

What do the community capital indicators tell us about a community? Community strengths, weaknesses, opportunities, gaps, etc.

How well do the identified indicators tell us about the condition of a community in relationship to the community capitals?

Project Scope

- Database of major events in Iowa communities (natural disasters, plant closings/major job loss, new plant opening/jobs created, school closures/school district mergers, closure of state-run hospitals and nursing care facilities, new opportunities for tourism – music festival, bike trails, recreation, etc., opening of new hospitals/care facilities)
- Data analysis:
  - Correlations between indicators – including relationship of indicators within and between capitals as well as with different peer groupings (population, RUCC)
  - Trends and change over time.
  - Comparison to qualitative data and use as – visioning, RHRA, small towns poll.
  - Which three capitals are most related (ex: if interested in built capital – which two others might be most relevant to understanding built?)
- Dashboard – tableau
- Resiliency fit into capitals.

Structure of the Capitals Box Directory

```
Capitals
  Data
     Capitals Data
   |     Source Data (Raw) #Raw data before any preprocessing.
   |     Preprocessed Datasets #Datasets after initial preprocessing.
   |     Final Datasets #Ready-to-use, cleaned datasets, with data sources.
   |     Time Series Datasets #ACS datasets formatted for time series analysis.
   |     Capitals Data Sources.docx #Document listing data sources.
|   |
     Event Data
        Raw Data #Unprocessed event data.
        Preprocessed Data #Event data after initial processing.
        Data_Sources.xlsx #Spreadsheet of event data sources.
        Event_Timeline_Sample Structure.xlsx #Sample structure for event timeline data.
|   |
     Art_Culture_all_iowa_cities.xlsm #Excel file with art and culture data for Iowa cities
|
  Literature #Collection of documents and research literature
     Previous Projects
     Community Capitals
     Student Reviews
     Art + Culture
     Third Places
|
  Scripts #Collection of scripts for data processing and analysis.
     Social & Cultural
     ACS_Time_Series
     Natural & Built
     Event Datasets
     Documentation Code
     IowaCultureApp_Scraping.html
|
  Administrative #Administrative documents and organizational files.
     event timeline
```

```
      capitals indicators
      community groupings
      Capitals_Project_Description.docx
|
  Tableau Dashboard
      Community Capitals Dashboard.twbx #Tableau workbook for community capitals.
```

Part I

Measurements

American Community Survey (5 Year Estimates)

Brief Description

The American Community Survey (ACS) provides detailed demographic, social, economic, and housing data based on 5-year estimates. This dataset includes various measures related to community capitals, such as housing, employment, commuting, and broadband subscription.

General Source

The primary source for the measures below is the ACS 5-year report, 2022.

Margin of Errors

Most of the ACS measures and datasets have the high_moe columns which is True if the margin of error is greater than half of the measure.

prop_single

Measure Description

Percentage of single-parent households with children.

Source

ACS Variables:

- B11003_010 (Male household)
- B11003_016 (Female household)
- B11003_001 (total household)

Measure Calculation

The proportion of single-parent households is calculated by summing the number of male and female single-parent households and dividing by the total number of households.

$$\frac{\text{Male + Female}}{\text{Total Household}}$$

prop_commute

Measure Description

Percentage of commuters who commute more than one hour to work.

Source

ACS Variables:

- B08134_010 (# long-distance Commuters)
- B08134_001 (total commuters)

Measure Calculation

The proportion of long-distance commuters is calculated by dividing the number of commuters who commute more than one hour by the total number of commuters.

$$\frac{\text{Long-distance commuters}}{\text{Total Commuters}}$$

English Proficiency

Measure Description

Percentage of Residents who are proficient at English.

Source

ACS Variables for residents who speak very good English:

- B16008_004
- B16008_006
- B16008_009
- B16008_012
- B16008_014
- B16008_017
- B16008_022
- B16008_024
- B16008_027
- B16008_030
- B16008_032
- B16008_035
- B16008_039
- B16008_041
- B16008_044
- B16008_047
- B16008_049
- B16008_052
- B16008_001 (Total # of residents)

Measure Calculation

The proportion of proficient English speakers is calculated by summing the number of all groups of English speakers and dividing by the total number of residents.

$$\frac{\text{Sum Proficient}}{\text{Total Residents}}$$

Employments in agriculture, forestry, fishing, hunting, and mining

Measure Description

Percentage of the population employed in agriculture, forestry, fishing, hunting, and mining.

Source

ACS Variables:

- C24030_001 (Total civilian employed population 16 years or older)
- C24030_030 (Female employments 16 years or older in above industries)
- C24030_003 (Male employments 16 years or older in above industries)

Measure Calculation

The proportion of employed population in these industries is calculated by summing the male and female employments in these industries and dividing by the total civilian employed population.

$$\frac{\text{Male Employments} + \text{Female Employments}}{\text{Total Employments}}$$

Median household property value

Measure Description

Median value of household properties (in dollars).

Source

ACS Variables:

- B25077_001 (Median household property value)

Measure Calculation

The median household property value is directly taken from the ACS 5-year report, variable B25077_001.

Median year of built structures

Measure Description

Median year in which structures were built.

Source

ACS Variables:

- B25035_001 (Median year of built structures)

Measure Calculation

The median year of built structures is directly taken from the ACS 5-year report, variable B25035_001.

Percentage of housing units built before 1960

Measure Description

Percentage of housing units built before 1960.

Source

ACS Variables:

- B25034_010 (Housing units built 1940 to 1949)
- B25034_009 (Housing units built 1950 to 1959)
- B25034_011 (Housing units built before 1940)
- B25034_001 (Total housing units)

Measure Calculation

The proportion of housing units built before 1960 is calculated by summing the housing units built up to 1959 and dividing by the total number of housing units.

Percentage of units that are occupied

Measure Description

Percentage of housing units that are occupied.

Source

ACS Variables:

- B25002_002 (Occupied units)
- B25002_001 (Total units)

Measure Calculation

The proportion of occupied housing units is calculated by dividing the number of occupied units by the total number of housing units.

$$\frac{\text{Occupied Units}}{\text{Total Units}}$$

Percentage of households with complete plumbing

Measure Description

Percentage of households with complete plumbing facilities.

Source

ACS Variables:

- B25047_002 (# Facilities with complete plumbing)
- B25047_001 (Total plumbing facilities)

Measure Calculation

The proportion of households with complete plumbing is calculated by dividing the number of facilities with complete plumbing by the total number of plumbing facilities.

$$\frac{\text{Complete Plumbing}}{\text{Total Plumbing}}$$

Percentage of households with a broadband subscription

Measure Description

Percentage of households with a broadband internet subscription.

Source

ACS Variables:

- B28011_004 (# Internet subscriptions in households using broadband)
- B28002_001 (Total internet subscriptions in households)

Measure Calculation

The proportion of households with a broadband subscription is calculated by dividing the number of households with broadband internet subscriptions by the total number of households with internet subscriptions.

$$\frac{\text{Broadband Subsription}}{\text{Total Internet Subscription}}$$

Race Diversity

Measure Description

Diversity in the distribution of different race across Iowa communities.

Source

ACS Variables:

- B03002_003 (Not Hispanic or Latino:!!White alone)
- B03002_004 (Not Hispanic or Latino:!!Black or African American alone)
- B03002_005 (Not Hispanic or Latino:!!American Indian and Alaska Native alone)
- B03002_006 (Not Hispanic or Latino:!!Asian alone)
- B03002_007 (Not Hispanic or Latino:!!Native Hawaiian and Other Pacific Islander alone)
- B03002_008 (Not Hispanic or Latino:!!Some other race alone)
- B03002_009 (Not Hispanic or Latino:!!Two or more races)

- B03002_013 (Hispanic or Latino:!!White alone)
- B03002_014 (Hispanic or Latino:!!Black or African American alone)
- B03002_015 (Hispanic or Latino:!!American Indian and Alaska Native alone)
- B03002_016 (Hispanic or Latino:!!Asian alone)
- B03002_017 (Hispanic or Latino:!!Native Hawaiian and Other Pacific Islander alone)
- B03002_018 (Hispanic or Latino:!!Some other race alone)
- B03002_019 (Hispanic or Latino:!!Two or more races)

Measure Calculation

Simpson Index

$$D = 1 - \left(\frac{\sum n(n-1)}{N(N-1)} \right)$$

Where:

- n = the total number of organisms of a particular species
- N = the total number of organisms of all species

Interpretation: Simpson's Diversity Index is a measure of diversity which takes into account the number of species present, as well as the relative abundance of each species. As species richness and evenness increase, so diversity increases. It is also the probability of picking two random individuals which are not of the same group.

Race Diversity

With different groups of race taken from ACS, we calculate the Simpson Index to measure the diversity, using the above formula.

Bureau of Economic Analysis

Brief Description

Bureau of Economic Analysis provides data about GDP of different industries.

Source

GDP of Industries in Iowa

GDP of Agriculture, forestry, fishing and hunting

Measure Description

Agricultural GDP as share of total county GDP

Measure Calculation

Using R, take the GDP of Agriculture, forestry, fishing and hunting and divide by total GDP of the county

GDP of Mining, quarrying, and oil and gas extraction

Measure Description

Mining GDP as share of total county GDP

Measure Calculation

Using R, take the GDP of Mining, quarrying, and oil and gas extraction and divide by total GDP of the county

County Health Rankings

Brief Description

County Health Rankings data provides insights into the health outcomes and factors influencing health within counties across the United States. It includes measures related to natural and built environments, such as air and water quality, child care, and access to healthy foods.

Source: County Health Rankings

Measures

Initially, each measure in the dataset included its raw value along with the high and low values of its confidence interval. However, we chose to use only the raw values for our final dataset. Additionally, while the child care and food environment index measures were not used in our county health rankings natural dataset, they were incorporated into our built capital analysis.

1. Fair_Health_rawvalue
2. Fair_Health_cilow
3. Fair_Health_cihigh
4. Particulate_Matter_rawvalue
5. Particulate_Matter_cilow
6. Particulate_Matter_cihigh
7. Drinking_Water_Violations_rawvalue
8. Drinking_Water_Violations_cilow
9. Drinking_Water_Violations_cihigh
10. Child_Care_Centers_rawvalue
11. Child_Care_Centers_cilow
12. Child_Care_Centers_cihigh

13. Food_Environment_Index_rawvalue

14. Food_Environment_Index_cilow

15. Food_Environment_Index_cihigh

Final Measures for Natural

1. **Particulate_Matter**: Particulate matter (PM) encompasses microscopic particles suspended in the air, such as dust, dirt, soot, smoke, and liquid droplets.
2. **Drinking_Water_Violations**: Water Pollution

Final Measures for Built

1. **Food Enviroment Index**: Limited Access to Healthy Foods estimates the percentage of the population that is low income and does not live close to a grocery store. (according to, County Health Rankings)

Environmental Justice Screening Tool

Brief Description

The Environmental Justice Screening Tool Dataset provides information on various environmental and health hazards that can affect communities, helping to identify areas with higher risks and vulnerabilities. It includes pre-calculated measures such as air toxics cancer risk, respiratory hazard index, and proximity to hazardous waste.

Source: EPA - EJScreen

Measures

1. **Cancer_Risk**: Evaluates the risk of cancer associated with exposure to air toxics.
2. **Resp_Hazard_Index**: Assesses the risk to respiratory health from exposure to air toxics.
3. **Toxic_Release_Air**: Measures the amount of toxic substances released into the air.
4. **Hazard_Waste_Prox**: Indicates the proximity to hazardous waste sites.
5. **Wastewater_Discharge**: Represents the quantity of wastewater discharged into the environment.

All of these measures were pre-calculated, with detailed information on their exact calculation available in the data source PDF. While the dataset includes numerous potential measures, we focused on using the most general ones.

EJScreen Technical Documentation

Federal Bureau of Investigations Crime Reports

Brief Description

The FBI Crime Reports dataset provides detailed information on various criminal activities across different counties. For our analysis, we primarily focused on violent offenses such mas murder, forcible rape, robbery, and aggravated assault. We calculated the crime rate per 10,000 people to understand the prevalence of these violent crimes in each county.

Focusing on Violent Crimes

1. Murder: 11(Non-negligent)
2. Forcible Rape: 20
3. Robbery: 30
4. Aggravated Assault: 40

Source: Uniform Crime Reporting Program Data

Measures

Crime_rate_per_10K: Divided the total number of violent crimes in each county by the county's population.

Measure Calculation

$$\text{Crime\_rate\_per\_10K} = \frac{\text{Total Number of Violent Crimes}}{\text{Population}} \times 10000$$

FEMA National Risk Index

Brief Description

The FEMA National Risk Index Dataset provides information on the expected annual losses from various natural disasters across the United States. It is designed to help communities understand their risk levels and make informed decisions about mitigation and preparedness.

Source: FEMA NRI

Measures

Essentially, we kept the total annual loss for each type of disaster up until 2020. The annual loss was simply a cumulative sum of all the other loss types and was already provided by the dataset. Therefore, not much was changed in this dataset.

1. Flood_Expected_Annual_Loss
2. Heat_Wave_Expected_Annual_Loss
3. Hail_Expected_Annual_Loss
4. Strong_Wind_Expected_Annual_Loss
5. Tornado_Expected_Annual_Loss
6. Winter_Weather_Expected_Annual_Loss
7. Ice_Storm_Expected_Annual_Loss

Homeland Infrastructure Foundation Level Data

Brief Description

The Homeland Infrastructure Foundation Level Dataset (HIFLD) is a comprehensive collection of geospatial data critical for infrastructure planning and management, focusing on essential public services and facilities. Its datasets, which are integral to the built capital, include measures such as Cellular Towers, Colleges and Universities, Convention Centers, Fire and Emergency Stations, Hospitals, Law Enforcement Locations, Sports Venues, Private and Public Schools, Community Colleges, and Child Care Centers.

Source: HIFLD

Measures

1. Cellular_Towers_per_10k
2. College_Universities_per_10k
3. Convention_Centers_per_10k
4. Fire_Emergency_Stations_per_10k
5. Hospitals_per_10k
6. Local_Law_Enforcement_Locations_per_10k
7. Sports_Venues_per_10k
8. Private_Schools_per_10k
9. Public_Schools_per_10k
10. Community_Colleges_per_10k
11. Child_Care_Centers_per_10k *(ONLY County level)*

All these measurements were obtained from the Homeland Infrastructure Foundation Level Dataset, except for the childcare data, which was sourced from the County Health Rankings Dataset. Initially, all datasets were in geojson format so after extraction we performed a spatial join with the ACS population data at both the city and county levels to obtain the GEOID, population, and accurate county and city names, ensuring each location was correctly correlated.

For the per_10k calculations, I used ACS population data, dividing the total number of venues by the population and then multiplying by 10,000:

```
per10k = [
    'Cellular_Towers', 'College_Universities', 'Convention_Centers', 'Fire_Emergency_Stations
    'Hospitals', 'Local_Law_Enforcement_Locations', 'Sports_Venues', 'Private_Schools', 'Publ
    'Community_Colleges', 'Child_Care_Centers'
]

for count in per10k:
  county[f'{count}_per_10k'] = ((county[count] / county['Population']) * 10000)
```

Iowa Culture App

Brief Description

Iowa Culture App provides detailed information on various arts, history and cultural destinations in Iowa.

Source

Source: Iowa Culture

A sample code to extract the data from Iowa Culture is illustrated below. A complete code is located in the html file Code for Web Scraping.

```
pip install researchpy bs4
```

```
import json
import re
from bs4 import BeautifulSoup
import pandas as pd
import warnings
import researchpy as rp
import os
warnings.filterwarnings('ignore')
```

The `extract_info()` function below extracts the dataset from the culture app website based on a provided html format. This function is specific to the nature of the html format presented on the culture app website.

```
def extract_info(data):
    results = []
    for place in data['p']:
        info = {}
        info['title'] = place['title']
```

```
        info['latitude'] = place['lat']
        info['longitude'] = place['lon']
        info['image'] = f"/images/icons/markers/{place['icon']}"

        # Extracting distance
        distance_start = place['win'].find('<span class="distance">') + len('<span class="dis
        distance_end = place['win'].find('</span>', distance_start)
        info['distance'] = place['win'][distance_start:distance_end].strip()

        # Extracting full address
        address_start = place['win'].find('<strong>Address:</strong>') + len('<strong>Addres
        address_end = place['win'].find('</p>', address_start)
        info['address'] = place['win'][address_start:address_end].strip().replace('<br />',

        # Extracting phone number
        phone_start = place['win'].find('<i class="fa fa-phone fa-fw"></i>')
        if phone_start != -1:
            phone_start += len('<i class="fa fa-phone fa-fw"></i>')
            phone_end = place['win'].find('</p>', phone_start)
            info['phone'] = place['win'][phone_start:phone_end].strip()
        else:
            info['phone'] = None

        # Extracting website
        website_start = place['win'].find('<i class="fa fa-globe fa-fw"></i><a href="')
        if (website_start != -1):
            website_start += len('<i class="fa fa-globe fa-fw"></i><a href="')
            website_end = place['win'].find('"', website_start)
            info['website'] = place['win'][website_start:website_end].strip()
        else:
            info['website'] = None

        results.append(info)

    return results
```

`extract_city_county()`: is used to extract the the specific city and county from the address since the html format from the website does not explicitly give the either the city or county, but gives them in a single address.

`extract_marker()`: extracts the marker from the image path in the html file to know the specific category.

```
# Function to get city and county from the address
def extract_city_county(address):
    parts = address.split(',')
    city = None
    county = None
    for i, part in enumerate(parts):
        if 'IA' in part:
            city = parts[i-1].strip()
        if 'County' in part:
            county = part.strip().replace(' County', '')
    return city, county


# Function to extract image path as marker
def extract_marker(image_path):
    base_name = os.path.basename(image_path)
    marker_name = os.path.splitext(base_name)[0]
    return marker_name
```

The functionality of `extract_info()` is tested in the chunk below

```
arts = {
    "s": 1,
    "p": [
        {
            "icon": "national-register-of-historic-places.png",
            "title": "Christian Peterson Courtyard Sculptures and Dairy Industry Building",
            "lat": "42.026731",
            "lon": "-93.642860",
            "win": "<div class=\"window-container scroll\">\n\t<div class=\"image-container\
        },
        {
            "icon": "public-art.png",
            "title": "Escalieta I",
            "lat": "42.025650",
            "lon": "-93.644474",
            "win": "<div class=\"window-container scroll\">\n\t<div class=\"image-container\
        },
        {
            "icon": "public-art.png",
            "title": "History of Dairying",
            "lat": "42.026875",
```

```
            "lon": "-93.642983",
            "win": "<div class=\"window-container scroll\">\n\t<div class=\"image-container\
        }
    ]
}
```

```
arts_data = extract_info(arts)
arts_data = pd.DataFrame(arts_data)
arts_data[['city', 'county']] = arts_data['address'].apply(lambda x: pd.Series(extract_city_
arts_data['marker'] = arts_data['image'].apply(extract_marker)
arts_data
```

	title	latitude	longitude	image
0	Christian Peterson Courtyard Sculptures and Da...	42.026731	-93.642860	/images/icons/markers/nati
1	Escalieta I	42.025650	-93.644474	/images/icons/markers/publ
2	History of Dairying	42.026875	-93.642983	/images/icons/markers/publ

Cultural_Locations_per_10k

Measure Description

For our analysis, we focused on the total count of historic sites, museums, public art, monuments, and similar locations, collectively referred to as cultural locations. To assess the distribution of these cultural sites within each city or county, we computed the number of cultural locations per 10,000 people. This measure allows us to compare the availability and density of cultural resources across different areas.

Measure Calculation

After extracting the relevant data from the website, we tallied the total number of cultural locations for each city or county. To calculate our measure of interest (Cultural_Locations_per_10k), we used the following steps:

- Extracted city/county population data from the American Community Survey (ACS).
- Divided the total number of cultural locations in each city/county by the population of the corresponding city/county.
- Multiplied the results by 10,000 to express the number of cultural locations per 10,000 people.

Thus, the final measure Cultural_Locations_per_10k was obtained using this formula:

$$\text{Cultural\_Locations\_per\_10k} = \frac{\text{Total Number of Cultural Locations}}{\text{Population}} \times 10,000$$

Iowa Department of Transportation Bridges

Brief Description

The Iowa DOT Bridges dataset includes information on total bridges, obsolete bridges, and deficient bridges. It also contains derived measures such as the proportion of obsolete and deficient bridges and the number of bridges per acre per county.

Source: Iowa Department of Transportation

Measures

Initially, there were three separate datasets: total bridges, obsolete bridges, and deficient bridges, each including the respective counts. After merging all these datasets, the new measures were:

1. **Proportion_Obsolete_Deficient**: Calculated by adding total sum of obsolete and deffiecient bridges divided by total bridges.
2. **Bridges_Per_Acre**: Calculated by total bridge count divided by county's total acre per county.

```
final_merge['Proportion_Obsolete_Deficient'] = (final_merge['Obsolete_Count'] + final_merge[
final_merge['Bridges_Per_Acre'] = final_merge['Bridges_Count'] / final_merge['county_area']
```

Iowa Department of Natural Resources

Brief Description

The Iowa Department of Natural Resources provides data regarding natural aspects such as streams, forest reserve programs, public areas, lakes and wetlands.

Wetlands Measure

Wetland_Acres_Prop: Calculated the proportion of wetlands acres per county to the total acres per county.

```
final_merge['Wetland_Acres_Prop'] = final_merge['Wetland_Acres'] / final_merge['county_area']
```

Source: Wetlands in Iowa

Forestry Measure

forest_acres_prop_county: Calculated the proportion of forest acres per county to the total acres per county. The most updated year for forest acres we have is 2020.

```
final_merge['2020_Forest_Acres_Prop'] = final_merge['2020_Forest_Acres'] / final_merge['coun
```

Source: Forestry in Iowa

Total Stream Length

Stream Length in Miles: Using R, find the geometry of counties (using counties() of tigris) and use st_intersection() to take the geometry intersection of the streams and counties. Apply st_length() to calculate the length of streams and group by county.

Source: Streams in Iowa

Proportion of Lakes in Iowa

Proportion of Lakes: Using R, find the lake_area of each lake using st_area of the geometry column. Group by county and sum up all the lake areas. Calculate the proportion of lake by dividing the lake_area by the county area (taken from counties() function in tigris package)

Source: Lakes in Iowa

Proportion of Public Areas in Iowa

Proportion of Public Areas: Using R, find the public_area of each lake using st_area of the geometry column. Group by county and sum up all the public areas. Calculate the proportion of lake by dividing the public_area by the county area (taken from counties() function in tigris package)

Source: Public Areas in Iowa

Iowa Government Portal (iowa.gov)

Brief Description

The dataset is a 2022 general election turnout report that we got from iowa.gov. The dastset is split up by counties and gives us statistics on the different type of voters in each county.

Source

Iowa Secretary of State

County data from ACS 5-year report, 2022

% Total Absentee Voters

Measure Description

Absentee voters are people that cast thier vote remotely instead of at their polling place on election day. We calculated the percentage of absentee voters to the total amount of voters in a county.

Measure Calculation

$$\% \text{ Total Absentee Voters} = \frac{\text{Absentee Voters}}{\text{Total Voters}}$$

Total Voters Per 10,000 People

Measure Description

Proportion of total voters, election day and absentee, in a county per 10,000 people.

Measure Calculation

$$\text{Total Voters Per 10,000 People} = \frac{\text{Total Voters}}{\text{Population}} \times 10000$$

Active Voters Per 10,000 People

Measure Description

The proportion of active voters as of 11/8/2022 in a county per 10,000 people.

Measure Calculation

$$\text{Active Voters Per 10,000 People} = \frac{\text{Active Voters as of 11/8/2022}}{\text{Population}} \times 10000$$

Election Day Voters Per 10,000 People

Measure Description

The proportion of election day voters in a county per 10,000 people.

Measure Calculation

$$\text{Election Day Voters Per 10,000 People} = \frac{\text{Election Day Voters}}{\text{Population}} \times 10000$$

Absentee Voters Per 10,000 People

Measure Description

The proportion of absentee voters in a county per 10,000 people.

Measure Calculation

$$\text{Absentee Voters Per 10,000 People} = \frac{\text{Absentee Voters}}{\text{Population}} \times 10000$$

State Library of Iowa

Brief Description

State Library of Iowa provides statitics on income, expenditures, collections, circulation, and other measures. It is split up by a cities size and name. We looked at the fiscal year 2023, which is July 1, 2022 - June 30, 2023.

Source

State Library of Iowa

City data from ACS 5-year report, 2022

Per Capita Expenditures

Measure Description

Per Capita Expenditures are the average amount of money a library spent per person in a city.

Measure Calculation

$$\text{Per Capita Expenditures} = \frac{\text{Total Expenditures}}{\text{Population}}$$

Capita Expenditures Per 10,000 People

Measure Description

Per Capita Expenditures are the average amount of money a library spent per 10,000 people in a city.

Measure Calculation

$$\text{Capita Expenditures Per 10,000 People} = \frac{\text{Total Expenditures}}{\text{Population}} \times 10000$$

Staff Expenditures Per 10,000 People

Measure Description

Staff expenditures is the amount of money spent on paying library staff. The city Library staff expenditures per 10,000 people in a city.

Measure Calculation

$$\text{Staff Expenditures Per 10,000 People} = \frac{\text{Staff Expenditures}}{\text{Population}} \times 10000$$

Collection Expenditures Per 10,000 People

Measure Description

Library collection expenditrues are the operating expenses that are used to buy materials in their collections. The Collection Expenditure Per 10,000 People calculates the Library collection expenditures per 10,000 people in a city.

Measure Calculation

$$\text{Collection Expenditures Per 10,000 People} = \frac{\text{Collection Expenditures}}{\text{Population}} \times 10000$$

Book Circ Per 10,000 People

Measure Description

Book circulation accounts for the frequency that books are borrowed and returned to a library. Book Circ Per 10,000 People measures the circulation of books within a city per 10,000 people.

Measure Calculation

$$\text{Book Circ Per 10,000 People} = \frac{\text{Book1}}{\text{Population}} \times 10000$$

Books by Total Phys

Measure Description

The proportion of physical items that are books in a library.

Measure Calculation

$$\text{Books by Total Phys} = \frac{\text{Book1}}{\text{Total Phys1}}$$

Book%

Measure Description

Proportion of physical books out of all the items in a library.

Measure Calculation

$$\text{Book\%} = \frac{\text{Book1}}{\text{Total Use}}$$

Num Programs Per 10,000 People

Measure Description

The number of library programs across all ages per 10,000 people in a city.

Measure Calculation

$$\text{Absentee Voters Per 10,000 People} = \frac{\text{Prog6}}{\text{Population}} \times 10000$$

Prog6: Total number of programs in a city

Num of Attend per Pop

Measure Description

The proportion of library program attendance in a city.

Measure Calculation

$$\text{Num of Attend per Pop} = \frac{\text{Attend6}}{\text{Population}}$$

Attend6: Total number of program attendance in a city

Total Kid Programs

Measure Description

The total amount of kid programs that a library has in a city.

Measure Calculation

Prog1 + Prog 2

$$\text{Total Kid Programs} = \text{Prog 1} + \text{Prog 2}$$

Num Kid Programs Per 10,000 People

Measure Description

The number of kid programs that a library has per 10,000 people in a city.

Measure Calculation

$$\text{Num Kid Programs Per 10,000 People} = \frac{\text{Total Kid Programs}}{\text{Population}} \times 10000$$

Visits Per Capita

Measure Description

Per Capita library visits are the average amount of visits a library get in a city.

Measure Calculation

$$\text{Visits Per Capita} = \frac{\text{Visits}}{\text{Population}}$$

Visits Per 10,000 People

Measure Description

The number of library visits per 10,000 people in a city.

Measure Calculation

$$\text{Visits Per 10,000 People} = \frac{\text{Visits}}{\text{Population}} \times 10000$$

Wireless Sessions Per 10,000 People

Measure Description

The number of wireless sessions at a library per 10,000 people in a city.

Measure Calculation

$$\text{Wireless Sessions Per 10,000 People} = \frac{\text{Wireless Sessions}}{\text{Population}} \times 10000$$

Website Visits Per 10,000 People

Measure Description

The number of library website visits per 10,000 in a city.

Measure Calculation

$$\text{Website Visits Per 10,000 People} = \frac{\text{Website Visits}}{\text{Population}} \times 10000$$

Open Street Maps

Brief Description

Open Street Map is an open data source where people can edit the map of the world. The map contains roads, trails, cafes, railroads, parks, schools, and more.

Source

Open Street Map

City data from ACS 5-year report, 2022

U.S. Census Bureau

Park Area

Measure Description

The land designated to parks in a city. To get the city area we had to use the sf package in r to find the area of the polygon geometries of the parks.

Measure Calculation

```
Final_Parks_City_intersection$park_area <-
    st_area(st_geometry(Final_Parks_City_intersection))
```

City Area

Measure Description

The land designated to the cities in Iowa. We got the polygon geometry for the cities from the Tiger/line Shapefile, choosing 2023 Iowa Places. We then used the sf package in r to find the area of the polygon geometries of the cities.

Measure Calculation

```
shape$city_area <- st_area(st_geometry(shape))
```

Park City Proportion

Measure Description

The park area as a proportion of the total city area.

Measure Calculation

$$\text{Park City Proportion} = \frac{\text{park area}}{\text{city area}}$$

Num Parks Per 10K

Measure Description

The number of parks in a city per 10,000 people.

Measure Calculation

$$\text{Num Parks Per 10K} = \frac{\text{number of parks in a city}}{\text{Population}} \times 10000$$

Social Capital Atlas

Brief Description

This information is based on social media data and presence.

This is a sample of friendship ties for users aged 25-44 in the United States to construct measures for how communities are connected.

Different measures for social capital in this dataset:

- Connectedness: How people with different characteristics and backgrounds are friends with each other, this is a clear example of bridging social capital.
- Cohesiveness: The degree to which friendship networks are clustered into cliques and whether friendships tend to be supported by mutual friends. This includes our clustering and support ratio measures.
- Civic Engagement: Indices of trust or participation in civic organizations. This includes our volunteering rate measure.

Source: Social Capital Atlas

Measures

1. **economic_connectedness_county**: The degree of economic connectedness within a county or zip code, reflecting local economic integration and interactions.
2. **child_economic_connectedness_county**: The level of economic connectedness for children within a county or zip code, indicating economic integration among the younger population.
3. **support_ratio**: The ratio of individuals providing support to those needing assistance within a community.
4. **volunteering_rate**: The percentage of the population engaged in volunteer activities within a county or zip code.
5. **civic_organizations**: The number of civic organizations present within a county or zip code, indicating community engagement and infrastructure.
6. **exposure_grp_mem**: The extent to which individuals in a county or zip code are exposed to group memberships, reflecting social integration and network density.

Travel Iowa

Brief Description

Travel Iowa is the official tourism website for the state of Iowa, offering a wealth of information for those looking to explore. It includes details on new events and attractions, birding spots, stargazing locations, nostalgic soda fountains, and more.

Source

Source: Travel Iowa

A sample code to extract the data from Travel Iowa is illustrated below. A complete code is located in the html file …… This process below is specific to the nature of the html format we were able to retrieve from the travel Iowa website.

```
pip install bs4
```

```
html_doc = """

<div id="top">
    <div id="ctl00_phMainContent_div3" class="listControls">
        <div class="results">
        </div>
    </div>


    <div ID="pnlControls" style="display:none;">
        <div class="listControls" id="div4">

            <div class="pager">
                <ul>
                    <li>
                        <div id="pnlListingPager">
                            Page:
```

```
                        <span id="dpListing">
                                    <span class="activePage">1</span>
                            <a class="previous" data-pager-page="0">previous</a>
                            <a class="next" data-pager-page="0">next</a>
                        </span>
                        of 1
                    </div>
                </li>
            </ul>
        </div>
    </div>
</div>

<div class="grid">
    <div class="item">
        <div class="item-text">
            <div class="item-date">Jul <span>4 - 7</span></div>
            <h3 class="item-title">
                <a href="/calendar/flea-market-under-the-bridge/1647514">Flea Market Unde
            </h3>
            <div class="item-city">Marquette</div>
            <div class="item-venue">
                <span>Venue:</span> <a href="/calendar/flea-market-under-the-bridge/1647
            </div>
        </div>
    </div>
    <div class="item">
        <div class="item-text">
            <div class="item-date">Jul <span>1 - </span>Oct <span>31</spa
            <h3 class="item-title">
                <a href="/calendar/historic-hills-scenic-byway-bale-trail/1643930">Histo
            </h3>
            <div class="item-city">Fairfield</div>
            <div class="item-venue">
                <span>Venue:</span> <a href="/calendar/historic-hills-scenic-byway-bale-
            </div>
        </div>
    </div>
    <div class="end-grid-action">
        <button class="button" id="btnShowMoreEvents">Show More Events</button>
    </div>
</div>
```

```
    <div ID="pnlControls2" style="display:none;">
        <div class="listControls" id="div1">

            <div class="pager">
                <ul>
                    <li>
                        <div id="pnlListingPager">
                            Page:
                            <span id="dpListing">
                                        <span class="activePage">1</span>
                                <a class="previous" data-pager-page="0">previous</a>
                                <a class="next" data-pager-page="0">next</a>
                            </span>
                            of 1
                        </div>
                    </li>
                </ul>
            </div>
        </div>
    </div>



</div>


"""
```

```
import pandas as pd
from bs4 import BeautifulSoup
pd.set_option('display.max_colwidth', None)
import warnings
warnings.filterwarnings('ignore')

# Base URL
base_url = "https://www.traveliowa.com"

# Parse the HTML content
soup = BeautifulSoup(html_doc, 'html.parser')

# Find all divs with class 'item'
items = soup.find_all('div', class_='item')
```

```
# Initialize a list to store the extracted data
data = []

# Extract data from each item
for item in items:
    date = item.find('div', class_='item-date').get_text(strip=True)
    title_tag = item.find('h3', class_='item-title').find('a')
    title = title_tag.get_text(strip=True)
    link = base_url + title_tag['href']
    city = item.find('div', class_='item-city').get_text(strip=True)

    # Check if the venue div and its a tag exist
    venue_div = item.find('div', class_='item-venue')
    if venue_div and venue_div.find('a'):
        venue = venue_div.find('a').get_text(strip=True)
    else:
        venue = 'Venue information not available'

    # Append the extracted data to the list
    data.append({
        'Date': date,
        'Title': title,
        'Link': link,
        'City': city,
        'Venue': venue
    })

# Create a DataFrame from the list of dictionaries
df = pd.DataFrame(data)

df
```

	Date	Title	Link
0	Jul4 - 7	Flea Market Under the Bridge	https://www.traveliowa.com/calendar/flea-market-u
1	Jul1 -Oct31	Historic Hills Scenic Byway Bale Trail	https://www.traveliowa.com/calendar/historic-hills-

Events_per_10k

Measure Description

In our analysis, we focused on the total number of past and future events (2021-2025) for each city. We calculated the total events per 10,000 people to evaluate their distribution across the cities.

Measure Calculation

After extracting the relevant data from the website, we tallied the total number of events for each city. To calculate our measure of interest (Events_per_10k), we used the following steps:

- Extracted city population data from the American Community Survey (ACS).
- Divided the total number of events in each city by the population of the corresponding city.
- Multiplied the results by 10,000 to express the number of events per 10,000 people.

Thus, the final measure Events_per_10k was obtained using this formula:

$$\text{Events\_per\_10k} = \frac{\text{Total Number of Events}}{\text{Population}} \times 10,000$$

US Census Bureau

Brief Description

The 2020 Decennial Census Respondent Rates dataset provides detailed information on the participation rates of households in the 2020 Census, reflecting the level of engagement and response across different regions. This data is crucial for understanding demographic representation and ensuring accurate population counts.

Source: 2020 Census Response Rates

Measures

census_response_rate: Cummulative Self-Response Rate - Overall

This dataset was extracted using the US Census API call, specifically for Iowa using the state code 19. No modifications or additions were made to the data.

Access Decennial Self-Response Rates

US Religion Census

Brief Description

The Dataset provides data about the distribution of different religious traditions.

Religious Diversity

Measure Description

Religious Tradition Diversity (consider non-religious people as a tradition)

Measure Calculation

Simpson Index

$$D = 1 - \left(\frac{\sum n(n-1)}{N(N-1)} \right)$$

Where:

- n = the total number of organisms of a particular species
- N = the total number of organisms of all species

Interpretation: Simpson's Diversity Index is a measure of diversity which takes into account the number of species present, as well as the relative abundance of each species. As species richness and evenness increase, so diversity increases. It is also the probability of picking two random individuals which are not of the same group.

Religious Diversity

From the dataset, we summarise the dataset by the number of adherents across different religious traditions. Calculate the Simpson Index for different traditions (species) using the formula above.

US Department of Agriculture Census

Brief Description

The United States Department of Agriculture is the federal executive department responsible for developing and executing federal laws related to farming, forestry, rural economic development, and food.

prop_crop

Measure Description

Percentage of crop lands in county

Measure Calculation

crop_area: areas of land used for agriculture - crop land, taken from USDA data county_area: areas of the whole counties accross the state of Iowa, calculated by using sf package and passing the geometry

$$\frac{\text{crop area}}{\text{county area}}$$

value_per_acre

Measure Description

Value of every acre of agricultural land - crop land, taken from USDA data

Measure Calculation

Directly extract the value per acre from USDA data.